



# Universitat de les Illes Balears

## GENETIC DIVERSITY AND GEOGRAPHIC PATTERNS OF THE AMPHIPOD GENUS *METACRANGONYX* (CRUSTACEA; GAMMARIDEA)

TESI DOCTORAL

AUTORA: Maria del Mar Bauzà Ribot

### DIRECTORS

Dr. Carlos Juan Clar

Dr. Damià Jaume Llabrés

Dr. Joan Pons Pons

Departament de Biologia / Programa de Doctorat de Biologia

Universitat de les Illes Balears

2013





# ÍNDEX GENERAL

<i>Summary</i>	i
Plantejament i objecte de la tesi	ii
<b><u>PART 1: INTRODUCCIÓ</u></b>	<b>1</b>
<b>1. Característiques de la fauna subterrània</b>	<b>1</b>
1.1 Biodiversitat de la fauna subterrània .....	2
1.2 La qüestió biogeogràfica als organismes subterranis: vicariança o dispersió? .....	4
<b>2. Model d'estudi: <i>Metacrangonyx</i></b>	<b>8</b>
Història evolutiva de la família Metacrangonyctidae .....	11
<b>3. Sistemàtica i inferència filogenètica</b>	<b>15</b>
3.1 Determinant un model de substitució .....	17
3.2 Mètodes de reconstrucció filogenètica .....	19
<b>4. Biogeografia i filogeografia</b>	<b>21</b>
<b>5. Marcadors moleculars</b>	<b>24</b>
Tipus de marcadors moleculars .....	25
<b>6. Relotge molecular i calibratge</b>	<b>31</b>
Models relaxats de rellotges moleculars .....	36
<b>7. Història geològica de les Illes Balears</b>	<b>40</b>
Illes Gimnèsies .....	41
<b>8. Mar de Tetis</b>	<b>44</b>
<b>9. Eustatisme a la Mediterrània Occidental i a nivell global</b>	<b>47</b>
<b><u>PART 2: MATERIALS I MÈTODES</u></b>	<b>51</b>
<b>1. Organismes d'estudi i mostrejors</b>	<b>53</b>

1.1 Filogeografia de l'espècie <i>Metacrangonyx longipes</i> .....	53
1.2 Filogènia i biogeografia del gènere <i>Metacrangonyx</i> .....	55
<b>2. Protocols experimentals</b> .....	<b>55</b>
2.1 Extracció i purificació del DNA genòmic .....	55
2.2 Amplificació i purificació del producte de PCR convencionals .....	56
2.3 Seqüenciació i precipitació .....	58
2.4 Amplificació del mitogenoma sencer (Long-PCR) .....	59
2.5 Seqüenciació dels mitogenomes .....	63
2.5.1. Seqüenciació de genoteques pel mètode de Sanger .....	63
2.5.2. Genoteques per PCR en emulsió i piroseqüenciació .....	66
<b>3. Mètodes analítics</b> .....	<b>69</b>
3.1 Edició de seqüències .....	69
3.1.1. Marcadors mitocondrials i nuclears .....	69
3.1.2. Mitogenomes obtinguts pel mètode de Sanger .....	69
3.1.3. Mitogenomes obtinguts per piroseqüenciació .....	69
3.2 Anotació dels mitogenomes .....	70
3.2.1. Anotació dels PCGs .....	71
3.2.2. RNA ribosomal i de transferència .....	72
3.3 Alineament .....	74
3.3.1. Alineament dels PCGs .....	76
3.3.2. Alineament dels gens ribosomals .....	77
3.4 Anàlisi poblacionals .....	77
3.5 Freqüències nucleotídiques i nombre efectiu de codons .....	80
3.6 Divergència als PCGs dels mitogenomes anotats .....	82
3.7 Anàlisi de reordenació gènica .....	83
3.8 Test de Saturació .....	84
3.9 Metodologia de reconstrucció filogenètica .....	85
3.9.1. Selecció del model evolutiu .....	85
3.9.2. Inferència filogenètica .....	88
3.9.2.1. Anàlisi del set de dades per a la filogeografia de <i>M. longipes</i> .....	91
3.9.2.2. Anàlisi del set de dades dels mitogenomes .....	95
3.10 Estimació del temps de divergència .....	101
3.10.1. Taxa d'evolució i temps de divergència a <i>M. longipes</i> .....	102
3.10.2. Estimació del temps de divergència per a la família Metacrangonyctidae .....	103

<b><u>PART 3: RESULTATS</u></b>	<b>105</b>
<b>Capítol I: Filogeografia de <i>Metacrangonyx longipes</i> a les Illes Balears</b>	<b>107</b>
<i>Introducció</i> .....	107
<i>Resultats</i> .....	109
<i>Discussió</i> .....	122
<i>Conclusions</i> .....	126
<b>Capítol II: Caracterització del DNA mitocondrial <i>Metacrangonyx longipes</i></b>	<b>127</b>
<i>Introducció</i> .....	127
<i>Resultats i Discussió</i> .....	128
<i>Conclusions</i> .....	143
<b>Capítol III: Anàlisi filogenètica i biogeogràfica dels <i>Metacrangonyctidae</i></b>	<b>145</b>
<i>Introducció</i> .....	145
<i>Resultats</i> .....	147
<i>Discussió</i> .....	163
<i>Conclusions</i> .....	166
<b><u>PART 4: DISCUSSIÓ GENERAL</u></b>	<b>169</b>
<b><u>PART 5: CONCLUSIONS GENERALS</u></b>	<b>183</b>
<u>GENERAL CONCLUSIONS</u>	185
<b>Bibliografia</b>	<b>189</b>
<b>Material Addicional</b>	<b>219</b>
<b>Publicacions</b>	<b>230</b>



## SUMMARY

Speciation and biogeographic patterns of biota in extreme isolated environments have fascinated researchers for decades. The subterranean habitat represents one of the last frontiers in the knowledge of extreme environments. Although information on the evolutionary patterns and processes of the hypogean (= subterranean) fauna is scarce, cave ecosystems have been considered true “natural laboratories” to study the effects of temporal and spatial isolation on genetic divergence.

Environmental stability, permanent darkness and oligotrophy are characteristics of the cave habitat for both terrestrial (hypogean) and freshwater (stygobiont) fauna. One interesting habitat is the anchialine cave system, made of flooded brackish or marine water caverns with no connection to the sea. Most of anchialine crustacean species are clearly of marine origin, and are supposed to be the only living representatives of old evolutionary lineages. Many of these species display an extreme disjunct distribution in tropical and subtropical regions. Vicariance rather than dispersal would explain the present distribution patterns of these crustacean species. This assumption is based on the mentioned disjunct distributions, absence of long distance dispersal mechanisms and the existence of enormous geographical barriers for dispersal among congeneric species. In addition, present distributions for some of the taxonomic groups almost perfectly fit with the areas covered by epicontinental seas during the late Mesozoic, suggesting that they could be the result of vicariance by plate tectonics. Presently, the use of molecular phylogenetic techniques allows contrasting dispersal vs. vicariance hypotheses as explanations for the biogeographic patterns displayed by these fauna.

I have examined in this thesis the phylogeny of Metacrangonyctidae, a family of strictly stygobiont (= subterranean groundwater) amphipod crustaceans of marine origin. This amphipod family, composed by 18 species and about 18 other taxa not formally described, display a bizarre disjunct geographic distribution. A single endemic species (*Metacrangonyx longipes*) is found in groundwaters of the Balearic Islands of Mallorca and Menorca, while the remaining species occur in Fuerteventura (Canary Is.), North Africa, Middle East, Elba Island (Italy), and at the other side of the Atlantic in La Hispaniola Island. Because the dispersal capacities of these organisms

---

are very limited, the global distribution of the family has been associated traditionally to vicariance by plate tectonics and to peripatric speciation. The marine to inland groundwater transition would be related to coastline regression episodes in the Tethys Sea.

In the present study, we have examined for this family the tempo and colonisation pattern of subterranean habitats, both at a micro- and macroevolution level. In Chapter I we explore the phylogenetic relationships within *M. longipes* in the Balearic Islands, finding that *M. longipes* display a surprisingly old and highly fragmented population structure, with major events of cladogenesis within the species roughly correlating with some of the major transgression-regression episodes that affected the Mediterranean basin during the last 6 Ma. Eustatic changes, acting as vicariant events, seem to have been the main modulators of the evolutionary relationships found among Balearic population of *M. longipes*. Chapter II provides a description of the first complete nucleotide mitochondrial sequence of an amphipod species (*M. longipes*) and put the methodological basis for obtaining other crustacean mitogenomes. We then sequenced and annotated 18 mitogenomes from species of the family (plus another two mitogenomes from species representatives of other related amphipod families used as outgroups). This vast mitogenomic sequence information was subsequently used in Chapter III to test the alternative vicariance and dispersal hypotheses as explanations of the global disjunct metacrangonyctid distribution throughout phylogenetic analyses. Molecular clock estimation of divergence time and phylogeny patterns obtained for this family are consistent with a scenario in which *Metacrangonyx* is an old thalassoid lineage present in the shore of the western Tethys during the mid-Cretaceous. The separation of North Atlantic landmasses due to continental drift and subsequent sea-regression episodes would explain the present diversification and distribution *Metacrangonyx* patterns.

---

## PLANTEJAMENT I OBJECTE DE LA TESI

Els ambients extrems o aïllats han intrigat els científics durant molt de temps, sent objecte de nombrosos afanys per entendre els processos d'especiació i biogeogràfics que han forjat llur biodiversitat. Una de les fronteres en la recerca de la fauna present en ambients extrems emperò, es troba en el medi subterrani. Malauradament la informació relativa als patrons i processos evolutius que afecten a la fauna subterrània és escassa i sols recentment s'han començat a obtenir dades precises sobre l'adaptació a aquest hàbitat. Culver (1982) ja feia menció a la importància de considerar els ecosistemes subterranis com a vertaders "laboratoris naturals" per a l'estudi dels efectes de l'aïllament espacial i temporal en la divergència genètica. Cada cop són més els casos coneguts en els quals la investigació de la fauna cavernícola ha ajudat a aclarir processos evolutius i factors històrics relacionats amb la biogeografia i especiació talment han fet anteriorment estudis de la fauna d'illes oceàniques (Cooper et al., 2007; Page et al., 2008; Juan et al. 2010).

L'estabilitat ambiental, l'obscuritat permanent i l'oligotròfia són característiques de les coves, tant terrestres com les ocupades per aigües subterrànies. Dins de la classificació de coves hi trobem aquelles anomenades anquihalines, que es defineixen com cavernes inundades per aigua salobre o marina sense connexió directe amb la mar (Holthuis, 1973). La majora d'espècies anquihalines, clarament d'origen marí, són els únics representants vivents d'antics llinatges evolutius i són notables pels patrons de distribució extremadament disjunts i confinats a les latituds tropicals i subtropicals. Aquesta observació, combinada amb altres evidències com l'absència de potencial dispersiu a gran distància i l'existència de barreres immenses per a la dispersió entre les estacions ocupades per espècies congenèriques, apunten més cap a la vicariança que cap a la dispersió a l'hora d'elucidar els processos determinants dels patrons de distribució actuals d'aquestes espècies.

*Metacrangonyx* Chevreux, 1909 és un gènere de crustacis amfípodes estrictament estigobiont (=aquàtic subterrani) del qual no es coneix cap parent epigeu (és a dir, habitant de la superfície terrestre), i al qual s'hi ha atorgant rang de família diferenciada (F. Metacrangonyctidae Boutin & Messouli 1988). La seva posició taxonòmica es troba encara sota debat. Una única espècie endèmica balear, *Metacrangonyx longipes* Chevreux, 1909, es troba àmpliament distribuïda a Mallorca i

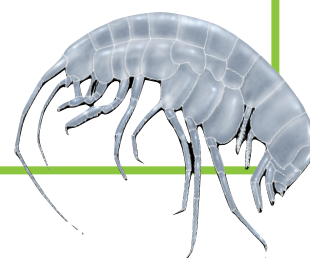
Menorca. La família Metacrangonyctidae compta amb 18 espècies formalment descrites la qual presenta una distribució majorment perimediterrània al nord d'Àfrica (11 espècies, més almenys 18 espècies més encara no descrites), conca de la Mediterrània Occidental (dues espècies), Orient Mitjà (dues espècies), arxipèlag canari (una espècie) i dues espècies a l'altre costat de l'oceà Atlàntic a la illa de la Hispaniola (República Dominicana). Abans de la troballa de les dues espècies a l'altre costat de l'oceà Atlàntic, només es coneixien representants al Vell Món i s'havia proposat que la família Metacrangonyctidae s'hauria originat posteriorment a l'obertura de l'oceà Atlàntic central a partir d'ancestres marins litorals que colonitzaren les aigües subterrànies continentals durant un episodi de regressió marina (Boutin & Messouli, 1988). No obstant això, el descobriment de dues noves espècies de *Metacrangonyx* a l'illa de La Hispaniola més la gran dificultat d'aquests organismes per dispersar-se a grans distàncies suggereix un origen anterior del gènere, com a mínim anterior a l'obertura de l'oceà Atlàntic central (Jaume & Christenson, 2001). Tot plegat semblaria que la distribució d'aquesta família podria respondre a un esdeveniment vicariant per la tectònica de plaques i especiació peripàtrica associada a episodis de regressió de les paleolíniees de costa de la mar primigènia de Tetis.

Atès que *Metacrangonyx* compta amb una elevada diversificació específica en una regió geogràfica relativament petita (conca Mediterrània), àdhuc dues espècies a una mateixa illa del Carib, s'ha escollit a aquest gènere com un model ideal per testar, mitjançant l'ús de tècniques moleculars, la hipòtesi biogeogràfica que retrotreu els patrons de distribució molt disjunts i vicariants de molts altres tàxons d'origen marí a conformacions continentals i marines pretèrites (en aquest cas a la paleogeografia anterior a l'obertura total de l'oceà Atlàntic central). D'altra banda, l'existència de nombroses poblacions de *M. longipes* al territori insular balear, ha permès abordar una filogeografia en detall de com ha transcorregut l'ocupació del medi subterrani continental. En concret, s'ha pogut comprovar el paper que han exercit els diferents moviments eustàtics i la topografia balear en els processos de colonització i distribució de variabilitat genètica en les diferents poblacions balears. A més, l'assoliment d'una filogènia molecular de les espècies del gènere, així com el calibrat de rellotges moleculars, ha resultat de gran ajuda per a la resolució de la història biogeogràfica i evolutiva d'aquest interessant tàxon subterrani.

---



**I N T R O D U C C I Ó**





## 1. CARACTERÍSTIQUES DE LA FAUNA SUBTERRÀNIA

La bioespeologia moderna nasqué al començament del s. XX, amb la publicació de la primera síntesi sobre fauna subterrània i els seus aspectes evolutius i ecològics pel naturalista romanès Emil G. Racovitza (1907). Des del descobriment del primer organisme cavernícola, la salamandra *Proteus anguinus* Laurenti, 1768 (Fig. 1.1), els bioespeleòlegs i biòlegs evolucionistes de tot arreu han quedat fascinats per les peculiaritats de la vida animal als ambients subterranis, on es combinen (1) l'extrema oligotròfia imperant a la majoria de coves, (2) la seva relativa estabilitat ambiental; (3) l'obscuritat permanent; (4) l'elevat nombre d'endemismes; (5) i la seva extremada especialització adoptant estranyes morfologies (Barr, 1967, Poulson & White, 1969, Culver 1970, Juan et al., 2010).



**Figura 1.1.** Individu de *Proteus anguinus*, primer organisme cavernícola descobert.

El procés d'adaptació a condicions hipogees comporta tota una sèrie d'adaptacions morfològiques, fisiològiques i etològiques conegudes en conjunt com a troglomorfoosi (Oromí et al., 1991; Caccone & Sbordoni, 2001). Els caràcters més distintius són la reducció ocular i la pèrdua de pigmentació (Vandel, 1964; Culver, 1982; Caccone &

Sbordoni, 2001). Altres exemples d'adaptacions molt comuns en els organismes hipogeus són els canvis morfològics cap a l'augment i l'exageració, com són l'elongació del cos i apèndixs a artròpodes, o bé l'exacerbació d'altres òrgans sensorials, com a resposta a l'obscuritat permanent. Aquestes presumptes adaptacions apareixen a grups diversos, des de platihelminths, a artròpodes, fins als vertebrats, i s'ha suggerit que són resultat de convergència evolutiva arran de les pressions selectives similars que operen al medi hipogeu. És precisament això el que va colpir més a Darwin arran de l'estudi d'espècies hipogees a Nord-Amèrica i Europa, dedicant tres pàgines a la qüestió a *The Origin of species* (Darwin, 1859),

destacant sobretot la convergència de trets regressius, com la pèrdua ocular o la carència de pigmentació corporal.

### **1.1. Biodiversitat de la fauna subterrània**

Fins el dia d'avui es coneixen més de 7,000 espècies subterrànies aquàtiques (estigobionts) (Botosaneanu, 1986; Gilbert & Deharveng, 2002), mentre que l'estima del nombre de tàxons cavernícoles terrestres (troglobionts) sembla arribar com a mínim a 21,000 espècies (Juberthie & Decu, 1994; Sket, 1999). Malauradament, a causa de la inaccessibilitat (cavitats fins ara inabastables, microcavitats, etc.) i de la crisi taxonòmica (= falta d'especialistes en molts grups taxonòmics; Proudlove & Wood, 2003), aquestes xifres probablement no representen més que una grollera subestimació (Gibert & Deharveng, 2002; Juan et al., 2010). Per aquest motiu, els hàbitats subterranis representen al dia d'avui una de les fronteres en la recerca faunística en ambients extrems. Si a més a més, s'hi afegeix el problema addicional plantejat pel fet que molts trets morfològics són resultat de convergència evolutiva, és lícit pensar que es puguin arribar a confondre les relacions entre espècies i els límits geogràfics entre elles (Buhay & Crandall, 2005).

Un tret característic de la fauna subterrània és que mostra un patró de diversitat diferent a aquell observat al medi epigeu. Mentre certs grups d'organismes s'hi troben àmpliament representats, altres que són comuns al medi epigeu en són absents. Per exemple, tal com resumeixen Gibert i Deharveng (2002) al seu treball, els crustacis són un dels grups que amb més èxit ha colonitzat l'hàbitat aquàtic subterrani. Entre els 12 ordres de crustacis presents a la fauna aquàtica subterrània, sis d'aquests (Mystacocarida, Gelyelloida-Copepoda, Syncarida, Mictacea, Thermosbaenacea, Remipedia) són exclusivament estigobionts (Stoch, 1995). Copepoda, Isopoda i Amphipoda són dels grups d'organismes estigobionts més abundants, més extensament distribuïts, i taxonòmicament diversos trobats al medi hipogeu. Holsinger (1993) va llistar fins a 740 espècies d'Amphipoda Gammaridea estigobionts, distribuïdes entre 138 gèneres i 36 famílies, tot i que la majoria pertanyen solament a 12 famílies.

---

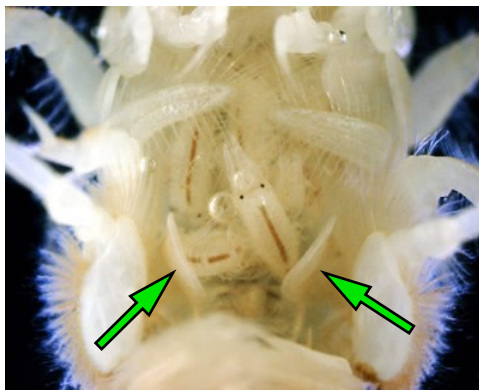
Gràcies a l'ús de tècniques moleculars durant els darrers anys s'ha anat modificant l'idea de la baixa diversitat de la fauna hipogea que proposaren anteriorment Poulson i White (1969) i Botosaneanu i Holsinger (1991), entre d'altres. Botosaneanu i Holsinger (1991) afirmaren que els sistemes subterrànics aquàtics més rics en nombre d'espècies tenen sols unes poques desenes d'espècies, contrastant doncs amb la gran diversitat trobada a la fauna epigea. Això sembla així a certs grups d'organismes, com és el cas dels vertebrats, o per exemple en el cas dels insectes, on el nombre d'espècies és molt superior en el medi epigeu (Stoch, 1995). No obstant això, els crustacis semblen no ajustar-se gens ni mica en aquesta afirmació i de fet el nombre d'espècies hipogees trobades a certs indrets poden igualar o superar el nombre d'espècies trobades a la superfície.

A petita escala, s'ha vist que la proporció d'endemismes trobada a la fauna hipogea d'isòpodes terrestres en el carst ibèric o en amfípodes en el carst Dinàric d'Eslovènia ha arribat fins al 90% (Dalens et al., 2002; Sket, 1996). De fet, en ecosistemes aquàtics epigeus, només es troben ratis similars en els llacs antics, de més de 10,000 anys. Trontelj et al. (2009) resumeixen les possibles causes que explicarien aquesta insòlita riquesa d'endemismes: (1) es creu que molts dels estigobionts actuals d'aigües subterrànies continentals serien descendents de relictos marins que quedaren aïllats per "varament" durant regressions marines (Stock, 1980; Botosaneanu & Hoslinger, 1991; Notenboom, 1991; Humpreys, 2000); (2) la majoria de la fauna estigobiont presenta una baixa capacitat de dispersió (Coineau & Boutin, 1992; Verovnik et al., 2003); (3) les poblacions queden aïllades en un hàbitat fragmentat arran de diversos esdeveniments vicariants (Barr & Holsinger, 1985; Kane et al., 1994); (4) la fragmentació hidrològica de les àrees càrstiques afavoreix l'aïllament i la divergència a llarg termini (Marmonier et al., 1993; Sket, 1997; Gibert et al., 2000; Verovnik et al., 2004). Altres autors recalquen l'estabilitat ambiental com un altre distintiu essencial que defineix el medi aquàtic subterrani (Gibert et al., 1994). Aquesta manca de canvis succeïts durant llargs períodes geològics, hauria fet minvar l'especiació per diversificació ecològica en el medi hipogeu, i conseqüentment, hauria contribuït també a l'acumulació de tàxons relictos, parents d'aquells epigeus àmpliament distribuïts els quals es van extingir arran de grans canvis ambientals (Barr, 1968; Juberthie i Decu, 1994, 1998; Danielopol, 2000; Gilbert i

Deharveng, 2002).

## 1.2. La qüestió biogeogràfica als organismes subterranis: vicariança o dispersió?

La qüestió central que més ha confrontat els biogeògrafs que treballen amb fauna subterrània ha sigut destriar el rol que han exercit la vicariança i la dispersió en els patrons de distribució dels estigobionts (Culver et al., 2007; Porter et al., 2007; Juan et al. 2010). Les espècies subterrànies sovint s'han vingut considerant com a "super especialistes", incapaços de sobreviure fora les condicions ambientals altament estables de llur habitat hipogeu, a més de presentar un baix potencial dispersiu (Holsiner, 1991; Coineau, 1994; Ketmaier et al., 2003; Culver & Pipan, 2009). No



**Figura 1.2.** Marsupi ventral de l'amphípode *Corophium volutator*. Les flexes verdes apunten als oostegits.

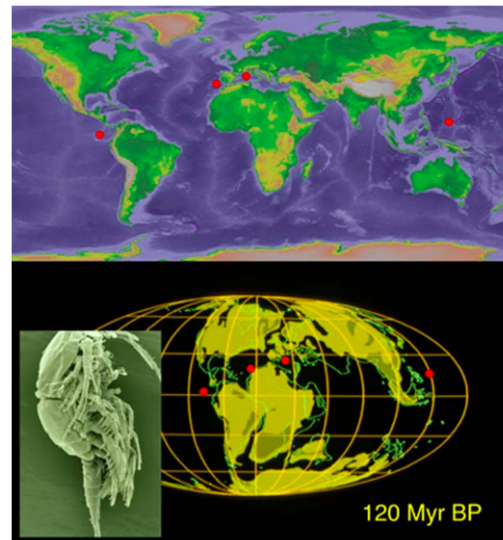
obstant això, diversos estudis han afavorit la preponderància d'un model o un altre en la determinació dels patrons espacials de la gran diversitat present a la fauna càrstica estigobiont. Per exemple, un model de vicariança és difícilment aplicable a tàxons presents a hàbitats que s'han format recentment (Danielopol et al., 1994; Lefébure et al. 2007), o bé a tàxons que es troben àmpliament distribuïts. Pel darrer cas, arran de l'aplicació de tècniques moleculars, s'ha pogut comprovar no obstant que l'àmplia distribució presentada per moltes espècies subterrànies és en realitat un artefacte, i del que es tracta és d'un complex de múltiples llinatges críptics (Lefébure, et al., 2006a; Finston et al., 2007; Guzik, et al., 2008; Trontelj et al., 2009; Zaksez et al., 2009). Per tant, certs esdeveniments vicariants serien responsables de l'àmplia distribució de diversos llinatges a diferents grups de crustacis (Ketmaier et al., 2003; Verovnik et al., 2004; Culver et al., 2007; Page et al., 2008; Finston et al., 2009).

Així i tot, hi ha estudis que mostren que l'àmplia distribució d'alguns tàxons subterranis s'explica més clarament per una combinació de vicariança i dispersió. De

fet, la importància que pugui tenir la dispersió o la vicariança variarà quant a les regions geogràfiques, a l'hàbitat subterrani (terrestre o aquàtic), al tipus de desenvolupament, tipus d'alimentació de les larves, etc. (Christiansen & Culver, 1987; Ketmaier et al., 2003; Verovnik et al., 2004; Holsinger 2005; Buhay & Crandall, 2005; Lefébure et al., 2006; Culver et al., 2007).

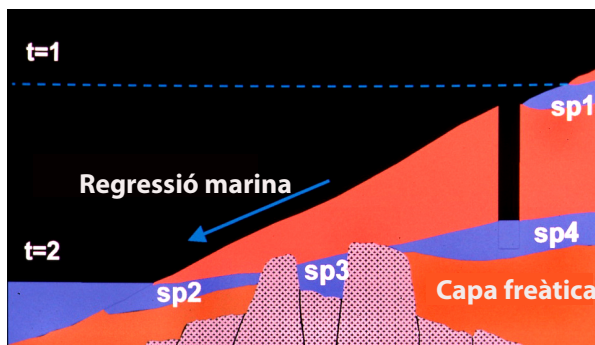
El tipus de desenvolupament que presentin els tàxons és un factor essencial a tenir en compte a l'hora d'inferir qualsevol hipòtesi biogeogràfica. Per exemple, els crustacis peracàrides (que engloben diversos ordres, com amfípodes, isòpodes, misidacis i afins) manquen de larves pelàgiques per ésser dispersats. En aquests, les cries que fan eclosió ho fan ja en un estadi de desenvolupament molt avançat, pràcticament idèntic a l'adult no-nedador. Els ous són transportats per la femella en un marsupi ventral (Fig. 1.2) en comptes de ser alliberats directament al medi i deixats a la mercè dels corrents, la qual cosa redueix encara més llur potencial dispersiu.

Cal dir que alguns grups d'amfípodes tenen més facilitat que d'altres per dispersar-se passivament, com ara els organismes algícoles, dels quals s'espera que sobrevisquin amb més facilitat i per períodes més perllongats en aigües obertes amb associació a masses flotants d'algues (Kingsford & Choat, 1985; McGrath & Myers, 1989). Mentre que aquells amfípodes típics de la infauna (és a dir, els que viuen enterrats en sediments no consolidats), els tubícoles, o bé aquells adaptats a la vida en caveres submarines i medi crevicular associat, presenten un potencial pràcticament nul per a la dispersió passiva, pel que constitueixen *per se* grups idonis per testar teories biogeogràfiques basades en la vicariança.



**Figura 1.3.** Distribució global del gènere *Expansohria*, un copèpode habitant de coves anquialines. Els patrons de distribució extremadament disjunts, aparentment inexplicables de molts llinatges anquialins, prenen sentit quan es traslladen a un mapa de paleolínies de costa cap a finals del Mesozoic. Les localitats actuals tan distants entre elles, en aquella època es trobaven més properes i connectades per mars epicontinental.

Un cas extrem d'aquests darrers el constitueixen diversos grups de amfípodes reclosos en aigües subterrànies continentals, i que són de derivació marina directa, per exemple, pertanyen a gèneres o famílies típicament marins, el que es coneix com a organismes talassoids (és a dir, descendents directes d'avantpassats marins). Curiosament, les àrees de distribució d'aquests amfípodes mostren una marcada correlació amb aquelles ocupades per mars epicontinentals presents cap a finals del Mesozoic i el Terciari (Fig. 1.3). Notenboom (1991) proposà un mecanisme que explicava l'ocupació de les aigües subterrànies d'aquesta fauna marina, suggerint que el varament terra endins de poblacions d'espècies litorals adaptades i restringides a la vida en el medi crevicular costaner (i per tant ja pre-adaptades a la vida en aigües subterrànies) durant alguna de les nombroses regressions experimentades pel nivell de la mar en el transcurs del temps geològic, seria el vertader motor d'aquesta colonització passiva. L'aïllament resultant conduiria una ulterior especiació i diversificació (Fig. 1.4). A més, els patrons anfi-atlàntics de distribució exhibits per alguns d'aquests llinatges no serien el resultat de dispersió, sinó més bé de vicariança per tectònica de plaques (Sterrer, 1973; Iliffe et al., 1983; Stock, 1993).



**Figura 1.4.** Model de Regressió (Notenboom, 1991) per explicar el mecanisme de colonització de les aigües subterrànies continentals per part de la fauna marina. Una espècie marina ancestral (sp1) adaptada a la vida en el medi crevicular costaner dona lloc a tres espècies a diferents aquífers després d'una regressió marina (t=2).

Cert és que les hipòtesis biogeogràfiques són més robustes si es coneixen en detall les relacions filogenètiques de les espècies considerades, a més de la seva distribució. Lamentablement, la inferència filogenètica a grups faunístics subterranis basada exclusivament en caràcters morfològics no sol ésser

concloent atès de l'homoplàsia morfològica inherent a aquest hàbitat tan restrictiu (Proudlove & Wood, 2003; Lefèbure et al., 2006). Dins del grup dels amfípodes, com a la majoria d'artròpodes subterranis, la homogeneïtat morfològica queda reflectida, no només en la regressió de l'aparell ocular i de la pèrdua de pigmentació corporal,



sinó sobretot en l'elongació desmesurada dels apèndixs corporals (antenes, pereopodis, uropodis), així com en la compensació sensorial a la pèrdua de visió com un increment de la dotació de mecano- i quimio-receptors. En contraposició a aquesta convergència morfològica, l'aïllament i la submissió a la pressió selectiva que suposa l'adopció d'una vida subterrània, pot conduir al mateix temps a fer que aquests organismes desenvolupin una elevada divergència genètica (Cooper et al., 2007; Finston et al., 2007; Lefébure et al., 2006).

Els mètodes moleculars actuals han demostrat llur eficiència a l'hora de desvetllar taxons o llinatges críptics que han passat desapercibuts a la taxonomia tradicional basada exclusivament en caràcters morfològics (Proudlove & Wood, 2003). D'aquesta manera, la història biogeogràfica inferida per a una suposada espècie basada en el seu patró de distribució, es demostra com totalment espúria si aquella "espècie" es tracta en realitat d'un conglomerat d'espècies al·lopàtriques críptiques (Lefébure et al., 2007). Per tots aquests motius, l'ús de la sistemàtica molecular en combinació amb l'anàlisi morfològica està especialment indicat en els estudis de fauna cavernícola, de forma que pugui destriar-se clarament la influència exercida per les forces selectives imposades per l'hàbitat subterrani de la de l'ancestre comú. No obstant això, els estudis filogenètics d'amfípodes hipogeus són escassos i basats sobretot en caràcters morfològics, i les anàlisis biogeogràfiques de grups de crustacis peracàrides hipogeus publicats fins al dia d'avui presenten aquesta limitació (Stock, 1980; Notenboom, 1988; Boutin, 1994; Koenemann & Holsinger, 1999; Vonk & Schram, 2003). Són excepcions notables dins d'aquest conjunt les recents contribucions de Cooper et al. (2007), Finston et al. (2007) i Lefébure et al. (2006; 2007) on s'apliquen tècniques de sistemàtica molecular per resoldre la filogènia i filogeografia de diversos grups de amfípodes hipogeus, tot i que a una escala geogràfica reduïda. Nogensmenys, els patrons biogeogràfics primaris que es puguin destriar en absència d'una bona evidència filogenètica poden arribar a ésser instructius (Holsinger, 1991; Stock, 1993).

---

## 2. MODEL D'ESTUDI: *Metacrangonyx*

Com ja s'ha esmentat a l'apartat anterior, els amfípodes són uns dels grups més abundants, més àmpliament distribuïts, i taxonòmicament diversos de la fauna subterrània. Per mor d'aquestes característiques, més llur aïllament geogràfic, la seva dificultat per dispersar-se a grans distàncies i el gran nombre d'endemismes a nivell local, els amfípodes subterranis són excel·lents candidats per abordar estudis biogeogràfics (Holsinger, 1986a; 1991; 1994).

*Metacrangonyx* és un gènere d'amfípodes estrictament estigobiont (és a dir, aquàtic subterrani) del que no es coneix cap parent epigeu, al qual s'hi ha atorgat rang de família diferenciada (F. Metacrangonyctidae) (Boutin & Messolui, 1988a) (Fig. 1.5). Malauradament, la seva posició taxonòmica es troba encara en discussió. Aquest gènere es pot trobar a aigües subterrànies tant anquihalines en el medi costaner, com totalment dolces situades a més de 1,400 m d'alçada a les muntanyes de l'Atlas marroquí.

La seva diversificació és moderadament elevada, i la seva distribució és circumscrita a la conca mediterrània occidental (Illes Balears, Marroc i Illa d'Elba), desert del Sinaí, depressió de la Mar Morta (Israel), sud-oest d'Oman així com a la illa de Fuerteventura (Illes Canàries) i les costes septentrionals i meridionals de la



**Figura 1.5.** *Metacrangonyx dominicanus*, espècie endèmica de les aigües subterrànies costaneres de l'Illa de la Hispaniola. Fotografia: T. Iliffe.

República Dominicana (La Hispaniola), al mar Carib (Fig. 1.6). Generalment, les diferències morfològiques entre les espècies són molt petites, i els caràcters diagnòstics més importants rauen en el gnatopodi 2, les plaques coxals, les bases i ungles dels

pereopodis 5-7, l'uropodi 3 i el telson. La majoria de les espècies d'aquest gènere són al·lopàtriques, sent excepcional trobar-ne dues compartint hàbitat (com és el cas de *M. spinicaudatus* i "*M. paurosexualis*"), i a més a més les seves àrees no mostren

solapament. Totes aquestes observacions suggereixen el limitat potencial dispersiu dels membres d'aquest gènere.

Les diferents espècies es distribueixen de la forma següent: Fuerteventura (1 espècie; Stock et al., 1986); Marroc (11 espècies, més al menys 18 espècies que encara no han estat formalment descrites; Balazuc & Ruffo, 1953; Ruffo, 1954; Karaman & Pesce, 1979; Boutin & Messouli, 1988b; Messoululi et al., 1991; Boutin, 1994; Oulbaz et al., 1998); l'illa d'Elba (Itàlia; 1 espècie; Stoch, 1997); conca de la Mar Morta (Israel; 1 espècie; Karaman, 1989); península del Sinaí (Egipte; 1 espècie; Ruffo, 1982); Oman (1 espècie, Jaume & Vonk, 2012); i dues espècies a l'altre costat de l'oceà Atlàntic, a la República Dominicana (La Hispaniola; Jaume & Christenson, 2001). Una única espècie balear (*Metacrangonyx longipes*) es troba àmpliament distribuïda a Mallorca i Menorca (Chevreux, 1909; Margalef, 1952).

D'altra banda, la distribució anfi-atlànica del gènere suggereix que el seu origen podria retrotraure's a èpoques anteriors a l'obertura de l'oceà Atlàntic central (Jaume & Christenson, 2001). La seva distribució actual seria producte de vicariància per tectònica de plaques i d'especiació peripàtrica associada a episodis de regressió de les paleolíniees de costa del mar de Tetis. D'aquesta manera, la presència de *Metacrangonyx* a illes situades a ambdós costats de la dorsal mesoatlànica (Canàries i Hispaniola) planteja la qüestió de com ha tingut lloc la seva colonització, atès que es troben separades per una conca oceànica àmplia i profunda. Això s'ha explicat recurrent a l'existència de pretèrits trams d'aigües somes associades a la mar de Tetis. A una fase ulterior de la seva història geològica, aquestes illes haurien quedat separades per aigües profundes (Iliffe et al., 1983). No obstant això, l'evidència geològica es troba en discrepància amb aquesta argumentació zoogeogràfica, ja que al menys les Canàries, s'han originat directament per activitat volcànica en el fons oceànic i a una fase més recent que la de la presència d'aigües somes entre ambdós costats de l'Atlàntic (veure Iliffe et al., 1993 i referències allà citades). La datació molecular de les divergències filogenètiques de les espècies insulars atlàntiques podria contrastar-se amb les edats d'expansió oceànica i de formació de les illes per testar aquestes hipòtesis.

---

El model d'especiació de *Metacrangonyx* (recordem: on una espècie ancestral marina d'àmplia distribució i ja adaptada a la vida en el medi intersticial/crevicular costaner deixa simultàniament poblacions perifèriques aïllades en aigües continentals subterrànies després d'una regressió marina) molt probablement obeeix el tipus peripàtric de Mayr (1982). Aquest es caracteritza per la gran disparitat entre la mida de la població parental i la derivada. El model assumeix també que, si les diferents poblacions de l'espècie parental tendeixen a diferenciar-se d'una forma estocàstica en resposta a factors ambientals locals, el flux gènic entre aquelles situades a la part central de l'àrea de distribució és prou important com per prevenir la seva diferenciació. No obstant això, es suposa que aquest flux gènic no és prou intens a la perifèria com per evitar l'establiment de nous fenotips (reconoscibles com a morfo-espècies). Per això mateix, si el flux gènic interdèmic cessa o bé queda reduït considerablement, aquests nous fenotips podrien arribar a fixar-se com espècies noves. El patró de descendència resultant d'aquest procés seria polítom: les neoespècies compartirien entre elles només aquells caràcters també compartits amb la població parental. Com a conseqüència, les espècies originades a partir de poblacions perifèriques que haguessin quedat aïllades arran d'una mateixa regressió marina, però en àrees geogràfiques separades, no tenen necessàriament que presentar apomorfies.

Les regressions marines poden considerar-se com a esdeveniments vicariants a la història evolutiva dels llinatges talassoids subterrànies (Holsinger, 1991). Dintre la Biogeografia, el terme vicariança s'aplica principalment a especiació al·lopàtrica a grans poblacions ocorreguda arran d'una fragmentació de l'espècie ancestral, per exemple per tectònica de plaques. Aquest tipus d'especiació tindria un reflex en un patró de descendència essencialment dicòtom. Però en el cas específic de vicariança per efecte d'una regressió marina s'esperaria que aquest patró fos polítomic (Holsinger, 1991, Crow et al., 1999). En el cas d'existir sinapomorfies (apomorfies compartides) entre neoespècies talassoides, aquestes serien el producte de: (1) que haguessin especiat ulteriorment al seu origen i implantació a les aigües subterrànies continentals; (2) que a l'espècie ancestral marina ja s'hagués donat variació geogràfica; o (3) que l'àrea de distribució d'aquesta ja se trobés fragmentada abans de l'adveniment de la regressió marina. En els darrers dos casos, les poblacions de

---

l'espècie ancestral marina ja compartirien apomorfies. En el cas (1), l'especiació dins aigües continentals subterrànies pot ser el resultat de l'existència de barreres entre conques hidrogràfiques, o de dispersió entre estigohàbitats seguida d'aïllament.

Aquest tàxon subterrani d'origen marí representa un organisme ideal per emprendre un estudi sobre evolució i biogeografia de fauna subterrània atès que: (1) presenta una àmplia distribució geogràfica, amb un nombre elevat d'espècies endèmiques, abastant ambdós costats de l'Atlàntic i la conca Mediterrània fins a l'Orient Mitjà, alhora que està moderadament diversificat -18 espècies formalment descrites-; (2) aquestes morfo-espècies en contades ocasions són simpàtriques, de manera que els seus patrons evolutius i filogeogràfics a petita escala poden ser fàcilment destriables; i per últim (3) la distribució anfi-atlàntica del gènere, al tractar-se d'animals amb baix potencial dispersiu, permet també la reconstrucció de patrons evolutius i biogeogràfics a gran escala, rarament estudiats a la fauna subterrània.

### **Història evolutiva de la família Metacrangonyctidae**

S'ha proposat que el procés de diferenciació en els metacrangonyctids s'assemblaria al que hom ha proposat per altres amfípodes estigobionts d'origen talassoide (= marí): "especiació en dues fases" (Fig. 1.4). Aquest model combina esdeveniments dispersius, als quals els ancestres marins litorals colonitzarien les aigües subterrànies continentals durant episodis de transgressió marina; amb esdeveniments vicariants, aquests últims associats a regressions del nivell de la mar (Stock, 1977, 1980; Boutin & Messouli, 1988b; Boutin & Coineau, 1990; Notenboom, 1991; Coineau & Boutin, 1992).

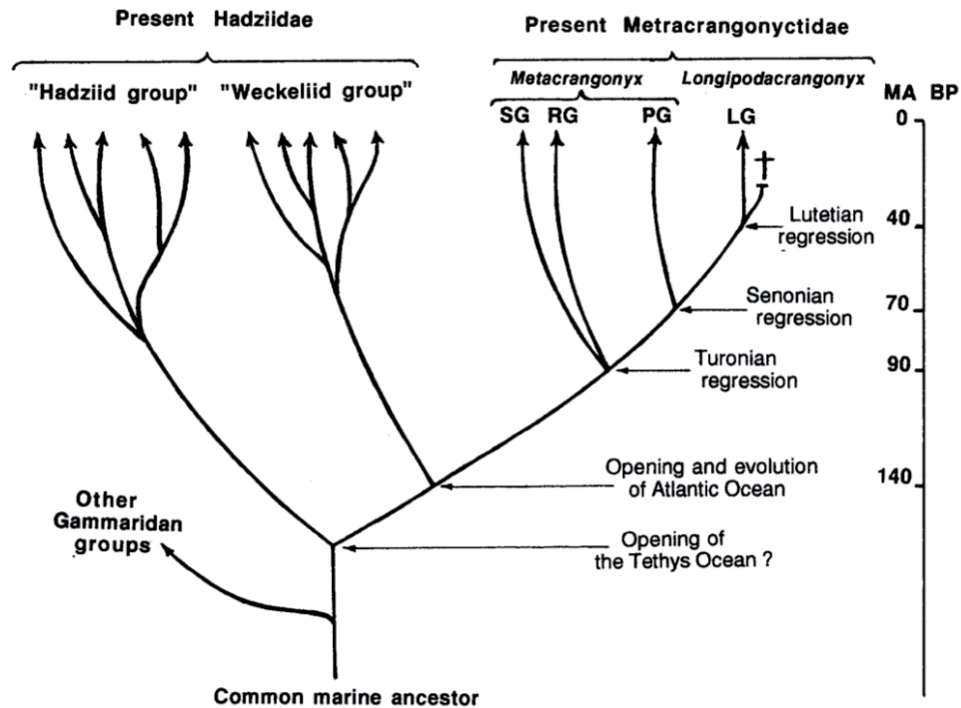
Després de cada procés de "colonització", l'ancestre marí podria haver-se extingit o bé continuar evolucionant independentment a la mar. La reduïda capacitat de dispersió d'aquests amfípodes subterrànies (Coineau, 1984; Wägele, 1985; Boutin & Coineau, 1990; Holsinger, 1991; Boutin, 1994) permetria la correlació de les àrees on aquests organismes són presents a l'actualitat amb els llocs on va ocórrer la colonització de les aigües continentals subterrànies (Boutin, 1994); es a dir, llur distribució ens indicaria fins a on va penetrar la mar en èpoques passades.

---

Boutin (1994) proposà una taxonomia i filogènia detallada de la família on es distingeixen quatre llinatges aparentment monofilètics, els quals denominà grups *panousei*, *remyi* i *spinicaudatus* pel gènere *Metacrangonyx*, al seu torn també distingí un segon gènere, *Longipodacrangonyx* Boutin i Messouli, 1988, de validesa qüestionable (Jaume, comunicació personal). Combinant la informació zoogeogràfica i geològica, Boutin (1994) suggerí una datació aproximada per a la separació d'aquests llinatges, proposant que els grups *spinicaudatus* i *remyi* colonitzaren les aigües continentals subterrànies durant la regressió del Turonià esdevinguda al Cretàcic superior (fa uns 90 Ma) (Fig. 1.7). La següent onada de colonitzadors, que involucraria als membres del grup *panousei*, portaria a l'establiment de poblacions continentals subterrànies durant la regressió del Senonià, en el Cretàcic superior (70 Ma). L'onada colonitzadora més recent hauria sigut protagonitzada per les espècies de *Longipodacrangonyx*, que haurien penetrat en els estigohàbitats durant la regressió de l'Eocè mitjà (40 Ma). Boutin (1994) proposà també datacions per a l'origen i extinció dels presumptes membres marins de la família.

L'absència de metacrangoníctids a àrees del Marroc que van ser inundades per la mar durant èpoques posteriors a l'Eocè s'interpretà com evidència la qual els metacrangoníctids marins ancestrals desaparegueren poc temps després d'aquest període. D'altra banda, la suposada absència de metacrangoníctids a l'altre costat de l'Atlàntic suggerí que l'origen de la família no podia ésser més antic al de l'obertura de l'oceà Atlàntic central. L'aparent absència de metacrangoníctids a certes àrees del Marroc cobertes per la mar durant el Juràssic recolzava també aquesta hipòtesi.

---



**Figura 1.7.** Filograma suggerit per Boutin (1994) on es presenten les relacions entre Hadziidae i Metacrangonyctidae, i els possibles esdeveniments geològics responsables de la separació dels principals llinatges amb la seva corresponent datació (dreta) en milions d'anys (Ma.)

No obstant això, el descobriment de *Metacrangonyx* en aigües subterrànies continentals de les Antilles Majors capgirà l'anterior hipòtesi sobre la datació de l'origen de la família (Jaume & Christenson, 2001). Com a primera conseqüència d'aquesta troballa, aquests autors suggereixen un origen molt més antic del gènere, en qualsevol cas anterior a l'obertura de l'Atlàntic. Tenint en compte que existeixen evidències geològiques que l'establiment d'aigües profundes entre la costa Est i Oest d'aquest oceà no va ocórrer fins fa 110 Ma (Sclater et al., 1977), proposen aquesta edat, com a mínim, per a l'origen del gènere *Metacrangonyx*. La presència de membres d'aquest gènere a ambdós costats de l'Atlàntic assenyalaria que el grup ancestral marí que donà origen a les espècies actuals ja s'hauria diferenciat abans de l'establiment d'aigües profundes entre ambdues riberes oceàniques.

Segons Jaume i Christenson (2001) la pertinença de les dues espècies de *Metacrangonyx* de La Hispaniola a dos dels grups proposats per Boutin (1994) contradiu la datació proposada per aquest darrer autor per a les onades colonitzadores de les aigües subterrànies continentals. La presència d'un membre del

grup *spinicaudatus* (*M. samanensis*) i un altre del grup *panousei* (*M. dominicanus*) a la Hispaniola contradiu la datació inferida pels integrants d'aquests dos llinatges (90 i 70Ma, respectivament) atès que és inferior a l'estimada per l'obertura de l'Atlàntic central (110Ma). Endemés existeixen evidències geològiques que indiquen que a les Antilles Majors no hi ha hagut terres permanentment emergides anteriors a l'Eocè mitjà (Iturralde-Vinent & MacPhee, 1999), pel que els llinatges de *Metacrangonyx* a les Antilles Majors no podrien ésser més antics que l'Eocè mitjà. En realitat, les coves on es troben les espècies de La Hispaniola estan excavades a terrasses de la plataforma coral·lina del Quaternari localitzades molt a prop de la línia de costa actual. Totes aquestes evidències suggereixen un origen molt recent per a aquests tàxons (Jaume & Christenson, 2001).

Aquests autors infereixen a partir de la troballa de les espècies de La Hispaniola que les poblacions marines de metacrangoníctids persistiren a la costa Atlàntica fins al Quaternari o fins i tot fins a l'actualitat, el que contradiu l'escenari proposat per Boutin (1994) per a l'extinció de l'ancestre marí dels Metacrangonyctidae, que ell suggereix ulterior a l'Eocè. A més a més, és important destacar la persistència de poblacions aparentment totalment marines de *M. longipes* Chevreux, 1909, segons Boutin (1994) membre del grup *remyi*, a una cova litoral del sud-est de Mallorca (Cova del Dimoni). Aquesta espècie endèmica balear presenta una ampla distribució a Mallorca i Menorca, amb poblacions establertes des d'aigües salobres properes a la costa fins a aigües totalment dolces situades a 200 m sobre el nivell de la mar.

Finalment, Jaume i Christenson (2001) intenten reconciliar l'escenari evolutiu proposat anteriorment per Boutin (1994) amb el descobriment de les dues espècies presents a l'altre costat de la costa Atlàntica. Assumint en un començament que els diferents grups de *Metacrangonyx* són realment monofilètics, l'única forma de fer-ho seria especular que els ancestres de les dues espècies de La Hispaniola haguessin romàs en estasi morfològica en el litoral marí de les Antilles durant un període prolongat després del qual els membres dels seus respectius llinatges de l'altre costat de l'Atlàntic haguessin colonitzat les aigües subterrànies continentals, i posteriorment s'haguessin extingit a o bé haguessin continuat evolucionant

---



independentment a l'ambient marí. Les espècies de La Hispaniola haurien per tant colonitzat les aigües subterrànies de l'interior de l'illa diacrònicament respecte al que feren els membres dels seus propis llinatges en el Vell Món. Encara que aquesta hipòtesi es debilita després del descobriment de poblacions suposadament marines de metacrangoníctids a la Mediterrània, pertanyents a un dels grups més antics (el grup *remyi*), la gran qüestió a resoldre seria com presumir l'absència de membres del grup *remyi* a territoris marroquins posteriors al Turonià quan membres marins del mateix grup encara persisteixen a l'oest de la Mediterrània. Una explicació seria que l'estoc marí ancestral d'aquest grup no hauria estat involucrat en més episodis de colonització de les aigües continentals subterrànies després de la regressió del Turonià, i en canvi hauria persistit en estasi morfològica a la mar. Les localitats de *Metacrangonyx* situades terra endins a les Balears emperò no recolzen això: corresponen a afloraments del Miocè o bé del Quaternari i no a terrenys cretàtics, com s'esperaria si l'escenari evolutiu proposat per Boutin (1994) fos el correcte.

### 3. SISTEMÀTICA I INFERÈNCIA FILOGENÈTICA

Es coneix com a sistemàtica a la disciplina que estudia la diversitat dels organismes, i que intenta explicar les relacions evolutives entre aquests d'acord als seus trets comuns (Ridley, 1996). La sistemàtica tradicional depenia majoritàriament de l'evidència empírica fornida per les dades morfològiques diagnòstiques que exhibien els organismes actuals i fòssils. L'existència d'un registre fòssil, emperò, era i segueix essent escàs o absent per molts de tàxons i si a més s'hi afegeix el problema ocasionat per la possible ocurrència de convergència de caràcters entre diversos llinatges, és evident la necessitat de l'ús de tècniques d'anàlisi filogenètica. El cladisme, la metodologia aplicada per la sistemàtica filogenètica parteix exclusivament de trets compartits, derivats i homòlegs (sinapomorfies) per reconstruir la història evolutiva de tàxons exclusivament monofilètics (on s'inclouen l'ancestre i tots els seus descendents) (Hennig, 1966). Durant els darrers decennis, els estudis sistemàtics s'han perfeccionat gràcies a l'adquisició i inclusió de dades moleculars, desenvolupant-se així la filogenètica molecular que emprava seqüències nucleotídiques o aminoacídiques com a font d'informació per a la reconstrucció

---

filogenètica. Malgrat que això ha revitalitzat la sistemàtica, aquesta disciplina s'ha vist sobretot afavorida pel perfeccionament en els mètodes d'anàlisi que s'utilitzen per generar i avaluar hipòtesis evolutives. Aquests mètodes constitueixen els procediments d'inferència que permeten relacionar o contrastar les hipòtesis amb l'evidència empírica (Kluge, 1997).

La metodologia filogenètica parteix d'una matriu de caràcters (trets variables i heretables) de la qual s'inferixen les relacions filogenètiques dels tàxons comparats. En aquestes matrius, les files representen individus, espècies o tàxons mentre que les columnes representen el caràcter homòleg (per exemple, posicions nucleotídiques) a cada uns del tàxons. A les seqüències de DNA, els caràcters poden tenir quatre estats (A, C, G, T) i en determinades circumstàncies pot donar-se que l'estat d'una posició no sigui clara pel que es pot codificar amb una de les ambigüitats reconegudes pel codi IUPAC (R, Y, W, S, M, K, H, B, D, V, N, X, ?). Quan hom pretén reconstruir una filogènia amb seqüències nucleotídiques ortòlogues, primer ha d'estimar les identitats entre les diferents posicions nucleotídiques de cada seqüència, és a dir, establir quines són les posicions putativament ortòlogues (hipòtesis primàries d'homologia). Aquest procés es denomina alineament de seqüències i s'aplica tant a DNA, RNA com aminoàcids. Les insercions i delecions (*indels*) de posicions nucleotídiques són inferències del procés d'alineament i es representen com un estat nou anomenat *gap* i és representat pel símbol del guió "-". L'alineament és clau en el procés d'anàlisi ja que seran les dades que empraran els models d'evolució per obtenir les matrius de dissimilitud.

La inclusió de regions altament divergents i de longitud hipervariable a un alineament ha sigut objecte de nombroses controvèrsies (Giribet & Wheeler, 1999; Simmons & Ochotorena, 2000; Lee 2001; Pons & Vogler, 2006). Els *indels* representen processos mutacionals diferents de les substitucions nucleotídiques i altament estocàstics que dificulten l'establiment acurat d'identitats (hipòtesi d'homologia). Això és perquè els *indels* que inclouen un gran nombre de posicions nucleotídiques es podrien explicar tant per una mutació única o bé per tractar-se de múltiples mutacions que inclouen poques posicions nucleotídiques. A més, aquest problema es veu agreujat ja que depenen de la penalització aplicada per l'obertura i extensió dels

---

gaps (*opening & extension gap penalty*) sorgeixen generalment diferents alineaments, és a dir, s'estableixen hipòtesis diferents d'homologia primària (identitats). El manteniment de la pauta de lectura aminoacídica dels gens codificants per a proteïna (PCGs de l'anglès *protein coding genes*) o bé l'estructura secundària dels gens ribosomals, poden ajudar a establir la homologia entre les seqüències (Kjer, 1995). No obstant la dificultat que presenten els esdeveniments d'inserció i deleció per alinear les seqüències, poden contenir un senyal filogenètic important (Simmons et al., 2001; Pons & Vogler, 2006), pel que existeixen diferents mètodes per incorporar-los a l'estudi. La inclusió de la informació continguda als *gaps* es pot solucionar tractant-los com a "cinquè estat" juntament als quatre nucleòtids; o bé com a *missing data* atès que en la seva evolució es troben sotmesos a processos diferents dels que afecten a les substitucions nucleotídiques (Swofford et al., 1996) o no contenen un senyal fiable (Simmons et al., 2001); o per acabar, poden tractar-se com a caràcters addicionals de presència-absència o codificacions "multiestat" més complexes (Simmons & Ochotenera, 2000). Recentment, Simmons & Ochotenera (2000) han posat seriosament en dubte el tractament dels *gaps* com a "cinquè estat" ja que comportaria el tractament de posicions amb *gaps* continus com a múltiples caràcters independents quan el més probable és que siguin el resultat d'un sol esdeveniment evolutiu.

Una vegada que s'ha obtingut l'alineament, és a dir, s'ha formulat la hipòtesi de homologia posicional dels diferents caràcters, cal decidir quin és el mètode de reconstrucció filogenètica més adequat.

### 3.1. Determinant un model de substitució

La reconstrucció filogenètica podria ésser un procés senzill si es parteix de la premissa que el nombre de diferències entre dues seqüències s'incrementa a mesura que ho fa el temps en què aquestes divergeixen des d'un ancestre comú. Si fos aquest el cas, seria suficient comptar el nombre de diferències i agrupar-les per similitud (Holder & Lewis, 2003). Malauradament, la taxa d'evolució d'una seqüència no és constant en el temps i el càlcul de les distàncies observades podria subestimar la vertadera magnitud de canvi evolutiu. La necessitat de l'ús de models probabilístics a la filogènia molecular es va fer evident ja durant la primera anàlisi comparativa

---

entre seqüències de DNA (Brown et al., 1982; Jukes & Cantor, 1969). De fet, la selecció natural o les taxes de mutació durant la història evolutiva d'un llinatge poden fer que diverses regions del DNA canviïn més ràpid de l'esperat o altrament, que diverses regions siguin més semblants del que caldria esperar. Algunes posicions de la seqüència de DNA no són útils per fer reconstruccions; és el cas de llocs que no varien al llarg del temps (*invariant sites*), posicions que varien ràpidament (*multiple hits*), etc. És per aquest motiu, que s'han desenvolupat mètodes de correcció de distància basats en models probabilístics del nombre de substitucions per convertir les distàncies observades en mesures de distància evolutives reals. Tots els models de substitució existents es troben relacionats entre si, tot i que difereixen en els aspectes del procés de fixació de mutacions. Els diversos models evolutius es distingeixen pel seu grau de parametrització. Existeixen tres tipus de paràmetres principals dins aquests models de substitució de seqüències de DNA que poden avaluar-se computacionalment:

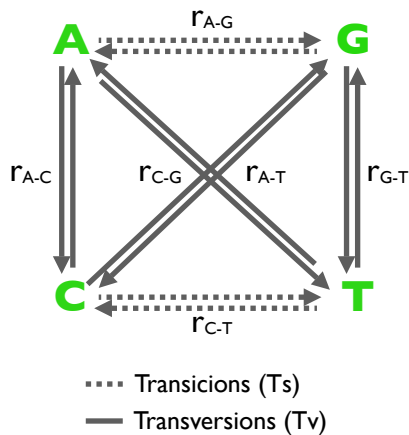
- **Paràmetres de freqüència nucleotídica.** Es distingeixen els models que consideren que les freqüències nucleotídiques són iguals (JC69; K2P; K3P; etc.) i aquells que consideren que la composició nucleotídica es troba esbiaixada (F81; HKY85; GTR; etc.).

- **Paràmetres de taxes d'intercanvi.** Es distingeixen els models que consideren que les taxes de substitució transicionals/transversionals són equiprobables (JC69) d'aquells models que tenen en compte que certes substitucions poden ser més freqüents que d'altres (K2P; GTR...) (Fig. 1.8). De fet, solen ésser més freqüents les substitucions entre nucleòtids que són bioquímicament més similars. En el cas del DNA, els quatre tipus de transicions (de purina a purina o pirimidina a pirimidina: A→G, G→A, T→C, C→T) són generalment més freqüents que els vuit tipus de transversions (de purina a pirimidina o viceversa: A↔T, A↔C, T↔G, G↔C). Aquest fenomen s'incrementa sobretot quan la divergència entre les seqüències a comparar és molt alta. Pel que si aquest biaix no es tingués en compte podria afectar a la divergència estimada entre les seqüències.

- **Paràmetres d'heterogeneïtat de taxes de substitució nucleotídica.** La

---

variació entre taxes de substitució entre diferents llocs d'una seqüència influeix notablement en el procés de reconstrucció filogenètica. Aquests models inclouen el model invariant que estima la proporció de llocs que no varien a la seqüència (paràmetre  $I$ , *proportion of invariant sites*), i en cas de variar calcula la freqüència a la qual ho fa seguint una distribució de tipus gamma ( $G$ ). Aquesta distribució gamma assigna una probabilitat de canvi als diferents llocs de la seqüència, tot i que per simplificar els càlculs s'aplica una distribució discreta la qual emprava quatre categories. El paràmetre alfa (o en anglès també anomenat *shape parameter*) controla la forma d'aquesta distribució. A major valor d'alfa menor és l'heterogeneïtat de taxes entre els diferents llocs de la seqüència (Fig. 1.9). És fàcil veure llavors que  $I$  i  $G$  en realitat es troben relacionats (a major  $I$ , menor serà el valor d'alfa i viceversa) i això fa que en ocasions la seva co-estimació pugui resultar difícil.



<i>Nst</i>	<i>models</i>	<i>Freq. nucl</i>
1	JC69 (Ts=Tv)	equivalents
1	F81 (Ts=Tv)	distintes
2	K2P (Ts=Tv)	equivalents
2	HKY (Ts=Tv)	distintes
3	K3P (2 Ts, 1 Tv)	equivalents
6	GTR (cada canvi una taxa diferent)	distintes

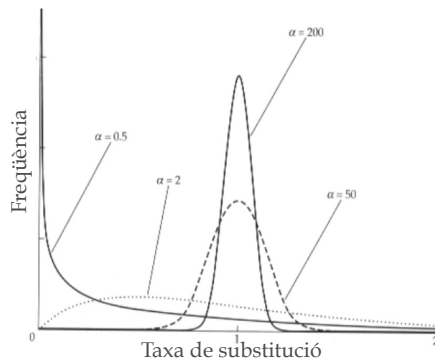
**Figura 1.8.** Esquerra: esquema representatiu dels sis tipus i taxes de substitució diferents. Dreta: principals models de substitució nucleotídica definits quant a les freqüències nucleotídiques (distintes o equivalents) i al nombre de categories de substitució nucleotídica (*Nst*).

### 3.2. Mètodes de reconstrucció filogenètica

Els diferents mètodes d'anàlisi es poden dividir en mètodes basats en distàncies i els mètodes basats en criteris d'optimització. Els mètodes basats en distàncies computen en un únic algorisme les distàncies genètiques entre cada parell de seqüències i agrupen aquelles seqüències en funció de dissimilitud, sense considerar l'evolució dels caràcters de cada posició. Per contra, els mètodes basats en criteris d'optimització solen acoblar dos algorismes, els quals generen noves hipòtesis evolutives (topologies d'arbres) i on s'optimitzen el nombre de canvis (substitucions

en cas de àcids nucleics i proteïnes) segons les dades reals per a cada posició.

(i) *Mètodes basats en distàncies*



**Figura 1.9.** *Diferents valors  $\alpha$  d'una distribució gamma* desavantatges són que no permeten estimar l'evolució de cada un dels caràcters de la matriu i només s'avalua una topologia. El més utilitzat és el de *neighbour-joining* (NJ, Saitou i Nei, 1987) el qual permet que cada branca tingui la seva pròpia taxa de substitució, i és àmpliament utilitzat quan la divergència és molt baixa, o bé per l'estudi d'hipòtesis preliminars. Un altre mètode basat en distàncies genètiques és el de UPGMA (de l'anglès *Unweighted Pair Group Method with Arithmetic Mean*) (Sneath & Sokal, 1973), però a diferència del NJ, aquest construeix arbres ultramètrics ja que considera que totes les branques tenen una taxa constant i proporcional en el temps i entre els diferents llinatges. Aquest de fet, és el principal motiu pel qual l'UPMGA no és considerat un bon algorisme atès la seva manca de realisme.

(ii) *Mètodes basats en criteris d'optimització*

A diferència dels mètodes anteriors, la longitud de les branques de l'arbre (distàncies) no estan fixades sinó que són avaluades a cada topologia. Els principals criteris d'optimització són els de màxima parsimònia (*maximum parsimony*, MP), que troba l'arbre bifurcat que explica amb menor nombre de passes totes les diferències en un alineament múltiple de seqüències (ha estat durant molt de temps el mètode principal usat pel cladisme); el de màxima versemblança (*maximum likelihood*, ML) que troba la topologia i la longitud de les branques que tenen la probabilitat conjunta més alta, i la probabilitat bayesiana (*bayesian inference*, BI). Aquest últim estima la "probabilitat posterior" dels arbres. Aquesta mesura bayesiana es pot interpretar

com a la probabilitat que un arbre concret sigui el que descriu correctament la història del conjunt de tàxons que s'està estudiant (Rannala & Yang, 1996; Huelsenbeck & Ronquist, 2001; Huelsenbeck et al., 2002) (veure més endavant a l'apartat de Materials i Mètodes).

#### **4. BIOGEOGRAFIA I FILOGEOGRAFIA**

La biogeografia estudia la distribució a l'espai i en el temps dels éssers vius i el processos que l'han originat. És una ciència interdisciplinària que presenta un doble caràcter descriptiu (caracterització dels patrons de distribució) i interpretatiu (la recerca de les causes d'aquests patrons). De Candolle (1820) fou el primer en diferenciar entre factors ecològics i històrics en la interpretació biogeogràfica. Segons aquest autor, la principal diferència entre els efectes d'aquests dos factors és l'escala temporal: mentre la interpretació ecològica cerca les causes actuals dels patrons de distribució, la interpretació històrica cerca les causes responsables que ocorregueren fa milions d'anys (Morrone & Crisci, 1995; Crisci, 2001; Posadas et al., 2006), posant l'accent en els grups taxonòmics i esdeveniments històrics biogeogràfics (Crisci et al., 2006). Una de les qüestions més rellevants de la biogeografia històrica moderna és la de reconstruir la història d'un tàxon particular o d'un grup de tàxons basant-se en hipòtesis filogenètiques (Cox & Moore, 2005; Lomolino et al., 2006). Durant els darrers anys, la qüestió temporal ha adquirit una gran rellevància en estudis de sistemàtica. De fet, actualment existeix una sinèrgia entre la datació (registre fòssil, rellotge molecular) i la biogeografia històrica per generar hipòtesis sobre l'origen, evolució i la distribució dels organismes (Posadas et al., 2006).

Es parla de distribucions disjunctes quan espècies/llinatges propers ocupen diferents regions allunyades entre elles i a més a més són absents en àrees intermèdies. Els patrons de distribució disjuncts han estat uns dels problemes que més han intrigat als biogeògrafs durant la història. Aquests patrons reflecteixen esdeveniments passats que poden ésser explicats de dues maneres: o bé l'ancestre comú habitava una de les àrees i posteriorment es dispersà a l'altra àrea, on divergí i va sobreviure fins al dia d'avui; o bé l'ancestre que es trobava àmpliament distribuït per tot arreu quedà posteriorment fragmentat on els seus descendents han

---

sobreviscut fins al dia d'avui. Aquestes hipòtesis es coneixen amb el nom de dispersió i vicariança, respectivament (Nelson & Platnick, 1981). El dispersalisme ha sigut l'explicació dominant a causa del paradigma geològic emergent sobre l'estasi geogràfica arran dels treballs de Wallace (1855, 1876) i Darwin (1859). Ambdós autors consideraven que la dispersió dels organismes és a l'atzar, a partir de centres d'origen i és la causa principal dels patrons disjunts. Alfred Wegener (1915) proposà per primera vegada la idea de la deriva continental amb la primera versió de l'obra de *Die Entstehung der Kontinente und Ozeane* (L'origen dels continents i oceans) (Fig. 1.10). Tot i així, aquesta teoria fou rebutjada per la majoria de geòlegs i no fou fins els anys seixanta del segle XX, amb el desenvolupament de la teoria de la tectònica de plaques, quan s'aconseguí explicar el moviment de continents i l'expansió del fons oceànic. Aquestes teories revolucionaren els paradigmes de la biogeografia. Léon Croizat (1958) fou uns dels primers científics en qüestionar les idees dispersalistes com a única explicació dels patrons de distribució disjunts proposant un nou concepte: la vicariança. El desenvolupament de la sistemàtica filogenètica, la concepció de la deriva continental i la sistemàtica molecular han influenciat notablement la biogeografia històrica (Crisci et al., 2000; 2003). En concret, l'ús de caràcters moleculars a la sistemàtica ha permès incrementar les dades disponibles per realitzar estimacions filogenètiques (Hillis et al., 1996). Així mateix, en molts de casos les filogènies moleculars han possibilitat l'ús de rellotges moleculars que daten els esdeveniments cladogenètics dins una filogènia.

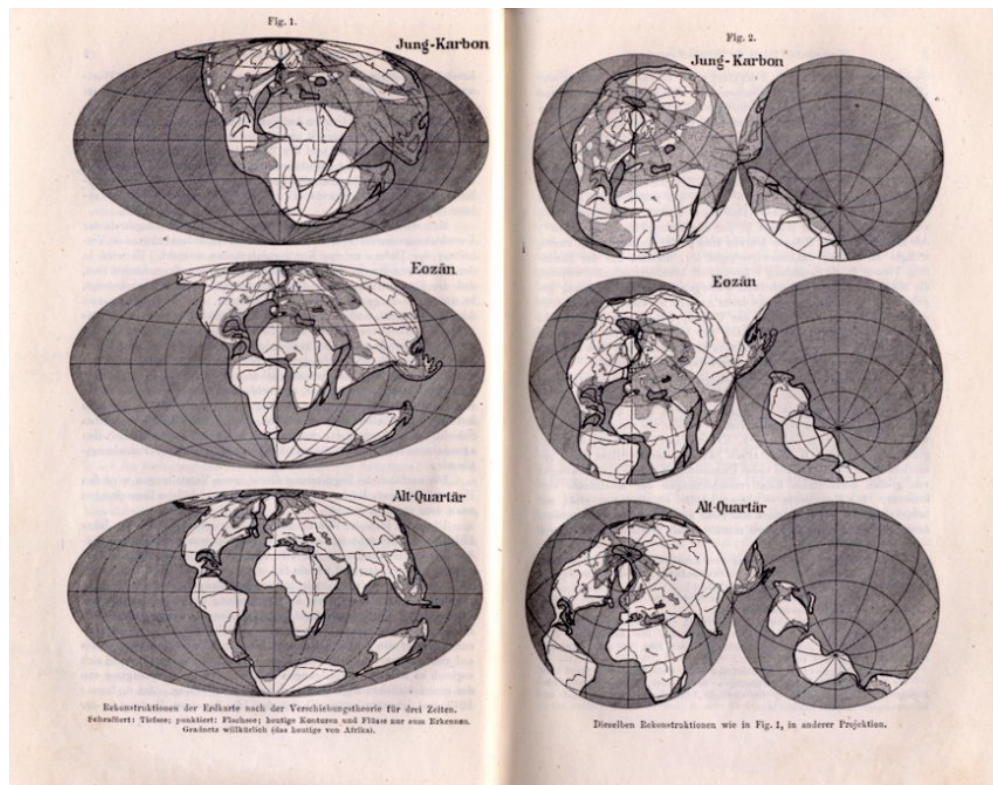
La influència d'aquestes disciplines ha quedat reflectida en l'emergència de nous enfocaments a la biogeografia històrica (resum a Posadas et al., 2006):

- ***Panbiogeografia***: desenvolupada per Léon Croizat (1964), que es basa en la idea que els patrons disjunts comuns a diferents tàxons són relictos. La panbiogeografia considera equiprobables la dispersió, vicariança i extinció com a mecanismes responsables de la distribució dels organismes. "*La Terra i la vida evolucionen plegades*" (Croizat, 1952; 1958; 1964; 1974).
  - ***Biogeografia cladística o de vicariança***: cap a inicis dels setanta neix aquest nou enfocament a partir de l'associació que feren diversos biòlegs (Nelson, 1969; Rosen, 1978; Nelson & Platnick, 1981) entre la panbiogeografia de Croizat amb la
-



metodologia proposada per la sistemàtica filogenètica de Hennig (1966). Aquest enfocament concep normalment els esdeveniments d'especiació com a causa d'aïllament (especiació al·lopàtrica). D'aquesta manera les relacions filogenètiques dels llinatges reflecteixen el temps relatiu de separació entre les biotes o àrees. Nogensmenys també assumeix en casos ambigus la possibilitat d'intervenció de mecanismes dispersius per explicar les distribucions actuals.

- **Filogeografia:** aquest enfocament és un exemple més d'aplicació de filogènies moleculars i es defineix, en sentit estricte, com l'anàlisi espacial dels llinatges gènics (Avise et al., 1987). Aquesta disciplina va sorgir fa poc més de vint anys gràcies al progrés de la biologia evolutiva en dos aspectes: un tecnològic, que va suposar l'adquisició de bases de dades moleculars així com el desenvolupament d'eines bioinformàtiques (ús d'algorismes d'inferència filogenètica); i un altre teòric, que va permetre l'aplicació de la teoria de la coalescència a l'estudi de processos microevolutius.



**Figura 1.10.** Il·lustracions extretes de l'obra *Die Entstehung der Kontinente und Ozeane* d'Alfred Wegener de l'edició de 1929 on es representa la separació dels continents des del Carbonífer Superior al Quaternari Inferior.

Disciplines tradicionalment allunyades com són la microevolució (genètica de poblacions) o macroevolució (filogenètica, sistemàtica) han trobat punts en comú gràcies a la filogeografia (Avise, 2000). La filogeografia aplica l'anàlisi de genealogies gèniques a l'estudi de la genètica de poblacions permetent treure conclusions d'acord a com han sigut els esdeveniments de colonització, diversificació i extinció dels llinatges gènics a determinades àrees. L'estudi comparat dels diferents patrons filogeogràfics inferits a diverses espècies o llinatges codistribuïts permet la formulació d'hipòtesis sobre possibles esdeveniments de vicariança-dispersió i identificar les causes geològiques, ecològiques i etològiques que hagin pogut modelar aquests patrons (Lanteri & Confalonieri, 2003).

La filogeografia es considera com a una subdisciplina de la biogeografia històrica, ja que es basa en els aspectes històrics de la distribució actual dels llinatges. Encara que, a diferència de la biogeografia, l'anàlisi filogeogràfica s'aplica principalment a nivell intraespecífic o de llinatges gènics molt propers. Un dels trets més destacables de la filogeografia és que permet integrar conceptes i tècniques típiques de la demografia, etologia, genètica de poblacions, sistemàtica, filogenètica i paleontologia (Avise 1998, 2000). No obstant això, darrerament, els estudis filogeogràfics estan basats majoritàriament en dades moleculars en lloc de característiques morfològiques o etològiques (Avise, 1998).

## 5. MARCADORS MOLECULARS

Carl Woese (Woese & Fox, 1977) fou un dels primers investigadors en utilitzar les diferències observades entre les seqüències nucleotídiques en un sol gen per investigar les relacions evolutives entre arqueobacteris. Poc temps després, cap al començament dels anys vuitanta, el desenvolupament de noves tècniques moleculars com és el cas de la reeixida reacció en cadena de la polimerasa (coneguda com PCR, les sigles angleses de *Polymerase Chain Reaction*) ha permès que nombrosos estudis poblacionals, filogenètics i taxonòmics abordin la diferenciació existent dins nivells taxonòmics diferents mitjançant l'ús de seqüències de DNA. A la taxonomia moderna, s'ha comprovat que l'ús combinat de dades moleculars a més de les dades

---

morfològiques facilita la delimitació de moltes espècies i el descobriment de molts llinatges i tàxons críptics desapercebuts anteriorment per la taxonomia tradicional (Balakrishnan, 2005; DeSalle et al., 2005; Goetze, 2003, etc.).

Cruickshank (2002) resumeix les diferents propietats que hauria de reunir un marcador molecular ideal: (1) ser un marcador de còpia única a un genoma haploide; (2) fàcilment alineable, és a dir, que fos senzill establir les identitats entre les posicions nucleotídiques de dues o més seqüències com a hipòtesi primària d'homologia; (3) tenir la mateixa taxa de substitució a totes les posicions; (4) tenir una taxa de substitució prou alta per garantir un nombre suficient de llocs informatius; (5) i una taxa de substitució prou baixa per evitar *multiple hits*; (6) presentar una composició nucleotídica no esbiaixada; (7) sense variació en la composició nucleotídica entre els tàxons estudiats; (8) tractar-se de marcadors àmpliament emprats, per disposar de *primers* universals i per poder contribuir a la realització d'altres estudis.

Malauradament, hom pot comprovar que és gairebé impossible que cap marcador molecular reuneixi totes aquestes propietats, i que a més a més, cal tenir en compte el model d'organisme a emprar i el tipus d'estudi que es vol realitzar a l'hora de triar el marcador molecular idoni.

### Tipus de marcadors moleculars

El DNA es pot classificar en dues categories: DNA nuclear i l'organular. El DNA nuclear (nDNA) és de major mida que l'organular i es troba format per seqüències úniques o bé per regions repetitives en les quals una seqüència principal es repeteix un nombre variable de vegades. Aquesta multiplicació pot tenir lloc a diferents escales, des d'unes poques bases, a un gen complet, o a un genoma sencer. Tal com resumeix Fitch (1970) al seu treball, existeixen dos tipus d'homologia entre gens: si l'ancestre comú més recent de dos gens és un esdeveniment de duplicació gènica llavors els gens són paràlegs, per contra es diu que els gens són ortòlegs. L'exemple clàssic és el cas de la família de les globines: totes les globines descendeixen del mateix ancestre comú. El gen de l'hemoglobina  $\alpha$  i el gen de

---

l'hemoglobina  $\beta$  són gens paràlegs mentre que l'hemoglobina  $\alpha$  d'un humà i l'hemoglobina  $\alpha$  d'un ratolí són ortòlegs. Els esdeveniments de duplicació gènica és una de les raons del perquè els arbres dels gens no corresponen amb l'arbre vertader de les espècies (Page, 1994; Martin & Burg, 2002) pel que la identificació dels gens ortòlegs és de vital importància dins el camp de la sistemàtica atès que són aquests els únics capaços de revelar la vertadera història dels llinatges (Finch, 1970).

Un altre exemple de família gènica són els gens ribosomals (rDNA) que codifiquen pel RNA ribosomal (rRNA). El rDNA nuclear d'eucariotes es troba organitzat en tàndem amb nombroses còpies (fins a 5000). Cada repetició consisteix en gens codificants per la subunitat nuclear petita (*SSU* de les sigles en anglès *small subunit* o *rRNA 18s*), la subunitat nuclear gran (*LSU* de les sigles en anglès *large subunit* o *rRNA 28s*), i el rDNA 5.8S, els quals són transcrits en rRNA però no traduïts a proteïnes. Aquestes molècules de rRNA formen part del ribosoma, pel que mutacions a la seva seqüència poden perjudicar greument l'estructura del ribosoma i per tant, afectar negativament a la maquinària de la síntesi proteica. Tot plegat es tradueix en la presència de fragments de rDNA altament conservats, amb una taxa d'evolució molt lenta, els quals han sigut objecte de nombrosos estudis per determinar les relacions filogenètiques entre llinatges distants (Hillis & Dixon, 1991; Hwang & Kim, 1999). Concretament, el gen nuclear *SSU* ha sigut molt emprat com a marcador a la reconstrucció de filogènies profundes (Field et al., 1988). No tots els dominis que conformen la *SSU* es troben altament conservats. En aquests gens s'hi poden trobar segments de ràpida evolució o alta variabilitat, anomenats "segments d'expansió" o "regions variables" responsables de la variabilitat en longitud existents entre tàxons propers (Vogler et al., 1997; Choe et al., 1999). Aquestes regions altament variables han sigut sovint emprades per resoldre qüestions filogenètiques d'espècies o famílies més properes. Nogensmenys, aquestes regions tenen el gran inconvenient que, en llinatges no tant propers, resulten molt problemàtiques d'alinear.

Els genomes d'òrgànuls normalment consisteixen en una molècula senzilla circular i llurs gens es troben en còpia única. A diferència del DNA nuclear (nDNA), els DNA mitocondrial (mtDNA) o cloroplàstic (cpDNA) codifica exclusivament per part de les proteïnes requerides per exercir les seves funcions (respiració i fotosíntesi,

---

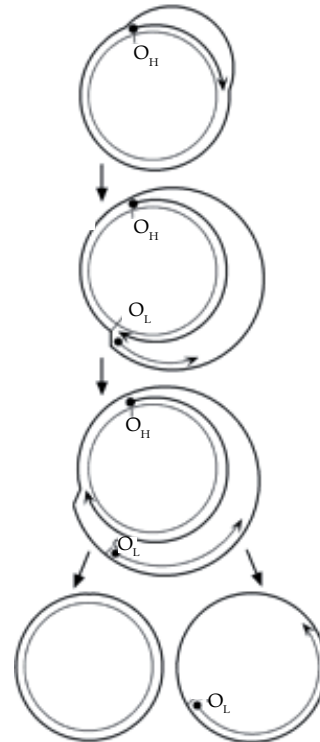
respectivament), i per a la maquinària necessària per expressar aquests gens. De fet, no és gens habitual trobar seqüències de DNA no codificants a mitocondris o cloroplasts (Page & Holmes, 2000).

El genoma mitocondrial (mitogenoma) dels metazous consisteix en una molècula circular tancada covalentment sense extrems de dues cadenes de filaments, anomenats filament pesat o H i filament lleuger o L (de l'anglès *heavy* o *light*, respectivament) (Clayton, 1982; 1996), on a Crustacea també s'utilitza la denominació de - (de *minority strand*) i + (de *majority strand*) en referència al nombre de gens codificats per a cada filament. La seva grandària ateny una longitud de 12 a 20 kb dependentment del grup de Metazoa. Fins i tot s'han trobat diferències en la grandària de mitogenomes dins un mateix grup, tot i que aquesta variació sol ser més infreqüent dins Arthropoda (Gissi et al., 2008). Aquesta variació en la mida del genoma normalment és a causa de la variació en la longitud de la regió control, en la qual sovint es troben zones amb moltes repeticions i on es poden donar freqüents fenòmens de duplicacions-inversions. Una altra diferència important entre els gens mitocondrials i nuclears és l'elevat biaix que existeix en la composició nucleotídica en el mtDNA, especialment en Arthropoda on es freqüent la riquesa en nucleòtids A+T (Simon et al., 1994).

El contingut gènic del mitogenoma en animals és gairebé constant. Habitualment conté 37 gens distribuïts arreu dels filaments + i -. (2rRNA, 22 tRNA i 13 gens que codifiquen per a proteïnes responsables de la fosforilació oxidativa) i normalment manca d'espaiadors intergènics (Hoffmann et al., 1992; Wolstenholme, 1992; Boore 1999). A més hi ha una regió molt rica en A+T que conté els elements reguladors essencials per a la iniciació de la transcripció i replicació anomenada regió control (Zhang & Hewitt, 1997). La replicació d'aquesta molècula bicatenària és un procés semi-conservatiu (Rocha, 2004) i majoritàriament duu a terme de forma asimètrica pels dos filaments, és a dir, comença en dos orígens de replicació diferents i distants dins el mitogenoma ( $O_H$ ,  $O_L$ ) (Clayton, 1982). Primer comença la síntesi del nou filament líder (H) a partir de l'origen de replicació  $O_H$ , situat a la regió control del filament L. Un cop la nova cadena H sintetitzada arriba a uns dos terços de la seva longitud, queda exposat  $O_L$  situat al filament H, permetent l'inici de la síntesi

---

del nou filament L (Clayton, 1982; Brown et al., 2005; *strand-displacement model*; Fig. 1.11). Per mor d'aquesta asimetria direccional pel que fa a la replicació, el filament parental H que donarà lloc al filament nou L (filament retardat) queda durant més temps en estat de cadena simple pel que està sotmès a més processos mutacionals que no pas el filament parental L que donarà lloc al nou filament H (filament conductor) (Hassanin et al., 2005; Wei et al., 2010). Aquestes mutacions sobretot són desaminacions a les adenosines i citocines. La desaminació d'una A dona una hypoxantina (I) que resulta emparellar-se millor amb una C que no pas amb una T, mentre que la desaminació d'una C dona lloc a una base uracil, la qual s'emparella millor amb una A que amb una G (Lindahl, 1993). Aquest fet ha sigut considerat per nombrosos autors com la causa principal de l'existència d'una notable asimetria composicional nucleotídica entre els dos filaments del mitogenoma (*Strand Asymmetry Compositional bias*) (Clayton, 1982; Hassanin et al., 2005; Francino & Ochman, 1997; Brown et al., 2005; Wei et al., 2010; Lin et al., 2012) que normalment es determina pel biaix en AT, expressat com a  $(A-T)/(A+T)$ , i pel biaix GC, expressat com a  $(G-C)/(G+C)$  (en anglès *AT skew* i *GC skew*, respectivament) (Nikolaou & Almirantis, 2006; Albrecht-Buehler, 2006; Perna & Kocher, 1995; Wei et al., 2010). El més freqüent és trobar un biaix positiu quant a GC i un biaix negatiu quant a AT a la cadena H; tot i que s'han trobat casos contraris a diversos grups d'artròpodes com és el cas dels isòpodes (Hassanin et al., 2005; Kilpert & Podsiadlowski, 2006; Hassanin 2006). La composició nucleotídica dels mitogenomes a més, varia notablement entre diferent tàxons d'animals i fins i tot entre espècies congenèriques.



**Figura 1.11.** Model de replicació de "strand-displacement". La replicació de la cadena H comença a l'origen de replicació  $O_H$  i tot seguit es produeix una extensió i desplaçament de la regió control. Aquest procés té lloc fins que l'origen de replicació  $O_L$  queda exposat i s'inicia la síntesi de la nova cadena L en la direcció oposada (Brown et al., 2005).



Un altre tret important a esmentar és que el mitogenoma es trobi en presència d'un ambient altament reactiu, la qual cosa afavoreix les elevades taxes de mutació característiques d'aquesta molècula. Diversos estudis demostren que la taxa de substitució nucleotídica és unes 5-10 vegades major al mtDNA de Metazoa que a les regions codificants de nDNA (Brown, et al., 1979, 1982; Brown, 1983). Una altra característica essencial del mtDNA és que s'hereta habitualment per via materna (Sutovsky et al., 1999). Aquest fet, evita que les noves mutacions que sorgeixen a diferents individus es recombinin durant la reproducció sexual, pel que aquells canvis observats a la seqüència seran deguts exclusivament a mutacions produïdes al llarg de generacions (Avise et al., 1987). De tota manera, s'ha detectat ocasionalment herència mitocondrial paterna a animals (Kvist et al., 2003) que pot donar lloc a l'aparició d'heteroplàsmia (es a dir, la presència de múltiples genotips mitocondrials a una cèl·lula o a un organisme) (veure el resum que fa White et al., 2008). Una altra peculiaritat que cal tenir en compte del mtDNA és la facilitat que suposa la seva obtenció/manipulació en comparació del nDNA per haver-hi normalment nombroses còpies per cèl·lula i ésser circular. És per totes aquestes raons que el mtDNA s'ha convertit en una eina fonamental en estudis microevolutius i filogeogràfics (Avise, 2000).

Històricament, els gens mitocondrials més emprats per establir genealogies han sigut la Citocrom oxidasa I (*cox1*), Citocrom oxidasa II (*cox2*), subunitat ribosomal petita (*rrnS*), subunitat ribosomal grossa (*rrnL*), Citocrom b (*cob*) i la regió control (de l'anglès *control region*, CR) (Caterino et al., 2000; Meyer & Zardoya, 2003). No obstant això, hi ha estudis que demostren que la majoria de les regions del mitogenoma podrien resultar similarment útils (Simon et al., 1994).

Malgrat les seqüències mitocondrials evolucionin més ràpidament que les nuclears, s'ha observat que les seves reordenacions gèniques solen romandre inalterades durant llargs períodes de temps evolutiu, pel que les comparacions d'aquestes reordenacions tenen un gran potencial per resoldre filogènies profundes dins Metazoa (Boore, 1999; Boore et al., 1998; Lavrov & Lang, 2005). De fet, arran de la translocació de *trnL2* (UUR) entre els gens *cox1* i *cox2*, compartida per Hexapoda i Crustacea, Boore i col·laboradors (1998) proposaren unificar aquests dos subfilums

---

en un nou grup anomenat Pancrustacea. Malauradament, s'ha vist que aquest supòsit general que les reordenacions gèniques es troben solament conservades entre les espècies molt properes filogenèticament i al seu torn difereixen entre aquelles espècies distants resulta ésser en certes ocasions inapropiat, i sobretot en el cas de Peracarida on s'ha observat que existeix una notable diversitat quant als ordres gènics (Krebes & Bastrop, 2012), principalment en el cas dels isòpodes (Kilpert & Podsiadlowski, 2012). No obstant això, tot i que aquestes reordenacions es considerin caràcters filogenètics *per se*, en cert casos s'ha detectat la presència d'homoplàsia (Kilpert & Podsiadlowski, 2012). Addicionalment a les dades obtingudes a partir de les seqüències nucleotídiques, i a les reordenacions gèniques, una altra informació útil que pot extreure's dels mitogenomes són els esdeveniments d'inserció gènica i deleció (Rokas & Holland, 2000) o variabilitat en la longitud gènica o dels espais intergènics (Schneider & Ebert, 2004). La utilitat filogenètica dels mitogenomes complets ha sigut estudiada rigorosament als darrers anys, especialment per a insectes i grups propers com els crustacis (Nardi et al., 2003; Cameron et al., 2004, 2007; Carapelli et al., 2007; Kjer & Honneycutt, 2007).

La quantitat de mitogenomes complets i la seva anàlisi corresponent de tots o la majoria dels 13 gens codificants per a proteïnes (de l'anglès *protein coding genes*, PCGs) s'ha vist incrementada notablement durant el passat decenni (Boore et al., 2005). Aquest fet ha permès la recerca de filogènies més profundes i testar diferents hipòtesis de reconstrucció filogenètica (Nardi et al., 2003; Cook et al., 2005; Hassain et al., 2005), així com el desenvolupament de models d'evolució més realistes i l'estimació millor suportada de reconstruccions filogenètiques (Hassanin, 2006). Tanmateix, la proporció de mitogenomes coneguts fins el dia d'avui representa pobrament als diferents grups de Metazoa i es troba altament esbiaixada. Per exemple, abans de la realització d'aquesta tesi, dins Crustacea només es coneixien 37 mitogenomes que representaven 30 de les 800 famílies que conformen aquest grup (Martin & Davis, 2001).

---



## 6. RELLOTGE MOLECULAR I CALIBRATGE

La hipòtesi del rellotge molecular (Zuckerlandl & Pauling, 1965) nasqué arran de la observació que feren Zuckerlandl i Pauling (1962) en comparar la seqüència de l'hemoglobina a diferents espècies de mamífers. Aquests autors constataren que hi havia una relació lineal entre el nombre de substitucions aminoacídiques entre dues seqüències pertanyents a dues espècies determinades i el temps de divergència entre aquestes espècies, calculat a partir de dades fòssils d'un antecessor comú. L'observació els portà a afirmar que les seqüències nucleotídiques i aminoacídiques evolucionen a una taxa relativament constant a través del temps i entre els diferents llinatges. Malgrat més endavant, a mesura que va anar creixent l'evidència molecular, s'observà que la taxa de substitució no era igual a totes les proteïnes. Un exemple n'és el treball de Dickerson (1971), on es demostrà que la taxa de substitució als fibrinopèptids és major que en el cas del citocrom b. Arran de tots aquests estudis Motoo Kimura proposà la Teoria Neutralista de l'Evolució Molecular (1968). Kimura suggerí que una gran part de mutacions noves no tenen cap efecte en l'eficàcia biològica (*fitness*), i per aquest motiu la selecció natural no hauria d'afavorir-les ni desfavorir-les. Aquestes mutacions neutrals podrien arribar-se a fixar en una població, o bé podrien perdre's, en els dos casos per processos purament estocàstics (e.g. deriva genètica). Kimura deduí que la taxa de substitució de les mutacions selectivament neutres és igual a la seva taxa de mutació; així doncs, grans canvis a la taxa de substitució podrien indicar grans canvis en els factors evolutius (Li, 1997). En casos de marcadors no neutrals, emperò, la taxa de substitució d'aquelles mutacions selectivament avantajoses dependria a més d'altres paràmetres, com és el seu avantatge o desavantatge selectiu i la mida poblacional, pel que aquests marcadors no serien viables per estudis de datació, atès a la desigualtat entre la taxa de substitució i la de mutació (Page & Holmes, 1998).

L'ús del rellotge molecular s'ha transformat en una eina indispensable en estudis de biologia evolutiva. Alguns exemples són l'especiació en aus ocorreguda durant Pleistocè arran dels cicles climàtics característics d'aquesta època

---

(Bermingham et al., 1992; Avise & Walker, 1998), la radiació evolutiva en grans grups de plantes i animals (Doolittle et al., 1996; Hedges et al., 1996, 2001), etc. Malgrat aquestes valuoses contribucions de l'ús del rellotge molecular, les datacions estimades a partir de dades moleculars no han estat exemptes de controvèrsies (Swofford et al., 1996). Uns dels debats fonamentals es centra en la constància del rellotge molecular. Fins als anys vuitanta del segle XX la validesa d'un rellotge universal i constant aplicable a tots els organismes fou àmpliament acceptada. Per citar-ne un exemple: el "rellotge universal de mtDNA" proposat per Brown i col·laboradors (1979, 1982) el qual donà una taxa de divergència d'aproximadament del 2% per milió d'anys a tot el mtDNA animal. No obstant això, a mesura que va anar augmentat la quantitat de dades moleculars, s'observà un alt grau d'heterogeneïtat en les taxes de substitució entre diferents llinatges, gens i regions genòmiques. Martin (1999) resumeix els factors més importants implicats en la variació de la taxa de canvi entre llinatges evolutius: el temps de generació als diferents llinatges, a menor temps de generació major és la taxa de substitució per unitat de temps (Kohne, 1970); la quantitat de dany sofert pel DNA a causa del metabolisme oxidatiu, de forma que aquells organismes amb alta taxa metabòlica tindrien una major taxa d'evolució que aquells amb metabolisme més lent (Martin & Palumbi, 1993); i per acabar, l'eficiència del sistema de reparació del DNA en els diferents llinatges (Britten, 1986). A causa d'aquesta variabilitat en la taxa de substitució als diferents llinatges i marcadors, el que es coneix com a heterogeneïtat de les taxes (Britten, 1986; Bromham & Penny, 2003), s'han proposat diferents metodologies d'estimació del temps de divergència per tal de "relaxar" el rellotge molecular. D'una banda l'aplicació d'un rellotge molecular universal estricte s'ha transformat en l'ús de rellotges moleculars "locals" que s'apliquen a marcadors particulars i a diferents parts de l'arbre (Hasegawa & Kishino, 1989; Kishino & Hasegawa, 1990; Swofford et al., 1996; Yoder & Yang, 2000). Aquests "rellotges locals" suggereixen que malgrat l'absència d'una constància d'una taxa universal, els llinatges propers evolutivament, o bé gens amb similitud funcional, comparteixen la mateixa taxa o similar de substitució, pel que en lloc d'aplicar un rellotge global a tot l'arbre convé aplicar taxes específiques als diferents grups de branques similars. La identificació de regions dins l'arbre on hi ha taxes similars requereix d'una gran quantitat d'informació *a priori* de les diferències de taxes entre les branques, per no

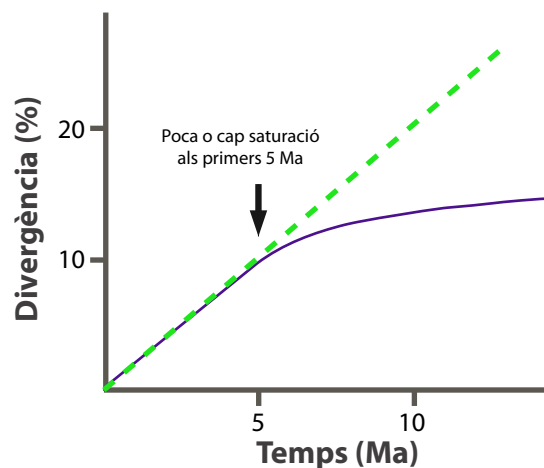
---

parlar de la complexitat computacional associada a l'aplicació de diferents models a cada una d'aquestes regions. Tot i de l'existència de mètodes que identifiquen regions de l'arbre que poden compartir el mateix rellotge molecular (*Heuristic rate smoothing*, Yang, 2004) aquests models serien útils solament en casos molt determinats on es veu clarament que la taxa varia notablement entre els tàxons i aquests a més, es troben evolutivament propers (Drummond et al., 2006). D'altra banda, existeixen mètodes que apliquen models paramètrics de relaxació bayesians com alternativa als rellotges locals (Thorne et al., 1998; Huelsenbeck et al., 2000; Kishino et al., 2001; Drummond et al., 2006). Aquests models ofereixen la possibilitat d'explorar un gran nombre de models alternatius que inclouen com a paràmetres la topologia de l'arbre i la variació de la taxa de substitució al llarg del temps i entre els llinatges (Lepage, et al., 2007) (es parlarà de la seva implementació al subapartat següent).

Malauradament, la diferència existent entre taxes evolutives entre diferents llinatges no és l'únic problema potencial amb relació a l'estimació del temps de divergència. Els punts de calibratge que permeten l'estima d'edats absolutes també són un component fonamental a les datacions moleculars. El calibratge consisteix en la incorporació d'informació cronològica independent (no molecular) en una filogènia donant un interval de temps o bé fixant una data a un node a partir de l'ús del registre fòssil i/o d'esdeveniments geològics prèviament datats (emergència d'illes oceàniques, formació d'una conca hidrogràfica, esdeveniments de transgressions-regressions marines, etc.). D'aquesta manera s'aconsegueixen transformar les edats relatives en absolutes. Aquests punts de calibratge, emperò, sovint han resultat ser responsables d'errors considerables (Hillis et al., 1996), sobretot en els casos on solament s'ha emprat un punt de calibratge (Arbogast et al., 2002). Atès que la incorporació d'aquests punts requereix d'una sèrie d'assumpcions no sempre falsables i tenint en compte el problema afegit de l'heterogeneïtat de les taxes, és recomanable sempre que sigui possible, l'ús de diferents punts de calibratge per evitar possibles errors de datació. Allò idoni seria poder incorporar més d'un mínim o màxim com a punts de calibratge (Ochman & Wilson, 1987; Graur & Martin, 2004; Near et al., 2005; Bell et al., 2010; Moreau & Bell, 2011), o bé tenir un punt a prop de l'arrel de l'arbre i un altre prop d'una de les branques terminals (Renner, 2005).

Tot i així, el fet de tenir més informació no sempre es tradueix en millors resultats ja que en cert casos les dades podrien estar en conflicte per diverses raons: incongruència entre les edats derivades del registre fòssil i els mètodes de datació molecular (Smith & Peterson, 2002; Bromham & Penny, 2003), punts de calibratge erròniament situats a la filogènia (Benton & Ayala, 2003), incongruències entre els múltiples punts de calibratge (Near & Sanderson, 2004; Near et al., 2005), o bé relacions filogenètiques de l'arbre mal inferides provocant que el punt de calibratge es situés erròniament a l'arbre (Moreau & Bell, 2011). La metodologia desenvolupada per Drummond et al. (2006) per a la relaxació del rellotge molecular ofereix una oportunitat per estimar simultàniament la topologia i la longitud de les branques i així incloure tant la incertesa en la topologia (relacions filogenètiques) com quan van ocórrer (temps o longitud de branques). Near i col·laboradors (Near & Sanderson, 2004; Near et al., 2005) proposaren el mètode de "validació creuada" (*fossil cross-validation*) per identificar incongruències entre distints punts de calibratge fòssil o esdeveniments paleogeogràfics i permetre així, la seva exclusió. Cal incidir que la demostració d'incongruència implica tant sols que algun punt del calibratge és erroni i no pas que cap d'ells sigui correcte. De fet, només la congruència de les datacions obtingudes arran de les anàlisis independents de varis punts de calibratge és un bon indicatiu de la robustesa dels resultats.

Un altre possible conflicte a l'hora de calibrar i estimar els processos de divergència és l'existència de saturació (homoplàsia) a les seqüències. La saturació es dona quan el nombre de substitucions observades (estimades) són inferiors a les reals (ocorregudes). Aquest gran nombre de substitucions nucleotídiques a la mateixa posició (*multiple hits*) no són detectades pel fet que la substitució final torna a mostrar el nucleòtid



**Figura 1.12.** Exemple de representació gràfica de la relació  $d/t$  (on  $d$  representa el nombre de canvis i  $t$  el temps transcorregut). En el cas ideal de no haver-hi saturació (línia verda) hom esperaria que la taxa de substitució fos igual a  $d/t$ . En canvi, sota la presència de saturació (línia blava) la relació entre  $d/t$  decreix al llarg del temps, atès  $d$  es troba subestimada.

inicial, el que es coneix com a substitucions homoplàsiques o emmascarades. És a dir, la saturació disminueix el senyal filogenètic contingut a les seqüències. El problema de la saturació és pot representar gràficament de la següent manera: en cas de no haver-hi saturació, s'esperaria que la taxa d'evolució molecular fos igual a la quantitat observada de canvis acumulats entre dos llinatges ( $d$ ) dividit pel temps que aquests han divergit ( $t$ ). A la realitat, emperò, s'observa com el ratio  $d/t$  decreix amb el temps a causa de la saturació (Fig. 1.12). La relació curvilínia entre  $d$  i  $t$  ha sigut àmpliament estudiada (Brown et al., 1979, 1982; Arbogast et al., 2002; Ho et al., 2005). Per aquest motiu, si no es corregeix aquesta saturació (amb l'aplicació de models d'evolució nucleotídics) es pot arribar a subestimar les taxes de substitució, o bé sobreestimar-les si s'assignessin erròniament els polimorfismes a les seqüències (Ho et al., 2005). Un altre factor a considerar a l'hora de calibrar nodes interns és que cal evitar de totes totes el calibratge a branques massa internes per datar processos evolutius recents, ja que es podria sobreestimar les edats reals de divergència a causa d'una subestimació de la taxa intraespecífica (Ho et al., 2008). A més, Emerson (2007) mostra com l'extinció de llinatges, o bé la representació incompleta dels tàxons estudiats, pot conduir a sobreestimacions de les taxes d'evolució quan es calculen edats d'esdeveniments de colonització a partir de dades moleculars. En els darrers anys, un gran nombre d'estudis en biogeografia històrica han treballat amb informació temporal, sent el rellotge molecular una de les eines més emprades per a la datació de nodes (punts de calibratge). Menció apart mereix l'aportació de l'estudi de Heads (2005) on es critica ferventment l'ús de la datació dels nodes a les filogènies com a edats absolutes ja que llurs datacions solament indicarien un mínim d'edat pels esdeveniments filogenètics. Segons aquest autor, solament els esdeveniments geològics que fossin més joves que l'edat dels nodes podrien ser descartats com a responsables de l'estructura filogenètica dels llinatges estudiats, mentre que aquells esdeveniments geològics anteriors a l'edat de cladogènesi del node no podria descartar-se d'haver influït a la història dels tàxons estudiats.

En última instància, quan no es disposa de cap tipus de registre fòssil útil ni d'esdeveniments geològics fiables per dur a terme el calibratge, com és el cas per a molts artròpodes, es pot recórrer a la utilització d'una taxa de substitució estandarditzada (Papadopoulou et al., 2010). La taxa de substitució estandarditzada

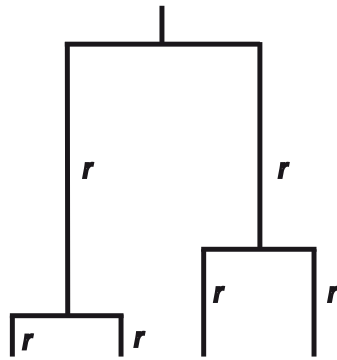
estimada pel mtDNA d'artròpodes més emprada és la proposada per Brower (1994), de 2.3% de divergència genètica per a cada milió d'anys (Ma) (0.0115 mutacions/posició nucleotídica per seqüència per Ma). Cal dir emperò, que altres estimacions de taxes de substitució del mtDNA a insectes basats en vicariança biogeogràfica, edats d'illes, fòssils, etc. han donat taxes més baixes (Sperling et al., 1997; Andersen et al., 2000) o més elevades (Fleischer et al., 1998; Shapiro et al., 2006).

### **Models relaxats de rellotges moleculars**

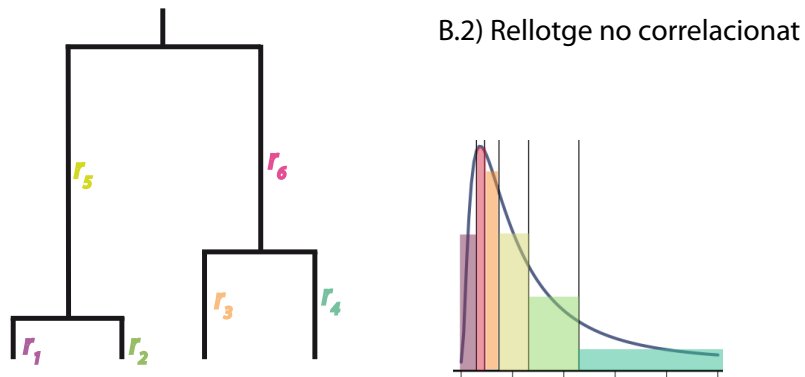
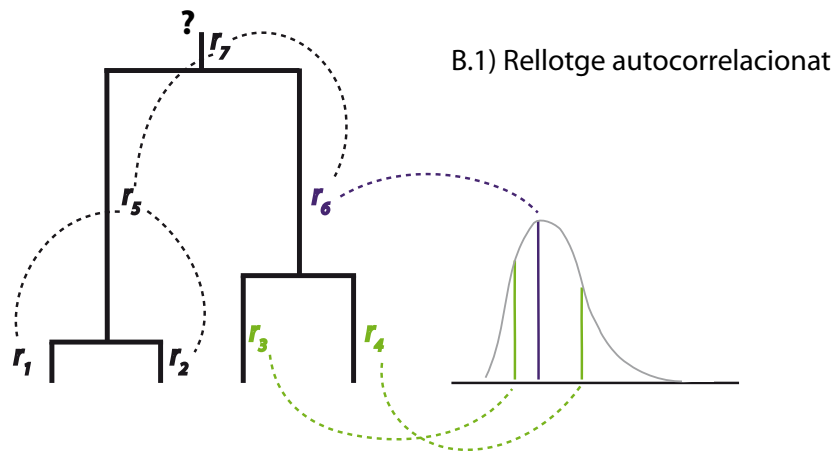
En els primers estudis on s'aplicava la hipòtesi del rellotge molecular, qualsevol fos la informació emprada pel calibratge d'una filogènia (taxa de substitució estandarditzada, ús de esdeveniment paleogeogràfics, registre paleontològic, etc.) s'assumia que totes les branques evolucionaven a la mateixa taxa. I això malgrat que, com ja s'ha esmentat anteriorment, les taxes evolutives depenen de la combinació d'una sèrie de factors (temps generacional, mida poblacional efectiva, taxa metabòlica, eficàcia dels mecanismes de reparació del DNA, etc.) que poden variar segons els llinatges. De fet, al llarg del temps s'ha vist que la diferència de taxes entre llinatges és generalitzada més que excepcional. Per aquest motiu, s'han desenvolupat mètodes estadístics que han obert la possibilitat d'afinar el calibratge del rellotge molecular. Es coneixen com a mètodes de rellotges relaxats per a l'estimació dels temps de divergència d'una filogènia. Un rellotge molecular relaxat permet que les taxes de substitució variïn entre les diferents branques o entre les diferents parts d'una filogènia (Sanderson, 2002; Thorne & Kishino, 2002; Yang, 2004; Drummond et al., 2006). Tot el contrari ocorre amb l'ús d'un rellotge estricte, el qual aplica una taxa única per a tota la filogènia (Fig. 1.13).

---

## A) Rellotge estricte o global



## B) Rellotges relaxats



**Figura 1.13.** Exemples dels diferents rellotges moleculars existents. A) Sota el supòsit d'un relotge estricte totes les taxes són iguals entre llinatges i al llarg del temps. B.1) Exemple de filogènia sota un relotge relaxat autocorrelacionat, on les taxes de substitució de cadascuna de les branques són estimades a partir d'una distribució normal logarítmica centrada en el valor de la taxa de l'ancestre comú (Thorne et al., 1998). B.2) Exemple de filogènia estimada sota un relotge relaxat no autocorrelacionat, on les taxes de substitució per a cadascuna de les branques és estimada independentment i estocàsticament a partir d'una distribució (a l'exemple, una distribució normal logarítmica) (Drummond et al., 2006).

Gillespie (1979, 1984) introduí el concepte d'autocorrelació de les taxes d'evolució entre els diferents llinatges dins una filogènia. Aquest autor suggerí que existia una alta probabilitat que els llinatges propers evolutivament tinguessin taxes de substitució similars vist que els llinatges hereten dels seus ancestres aquells trets que determinen aquestes taxes. Aquest concepte d'autocorrelació temporal de les taxes evolutives fou posteriorment desenvolupat per Sanderson (1997) per determinar automàticament els canvis en taxes evolutives entre diferents llinatges. Com a mètode d'optimització, Sanderson (1997) proposà un mètode no paramètric en el qual es minimitzava el canvi de taxa entre una branca i els seus descendents (*non-parametric rate smoothing*, NPRS) (implementat al programa informàtic *R8s*). Més endavant, Sanderson (2002) desenvolupà un altra metodologia la qual implementava com en el cas anterior la penalització numèrica pels canvis abruptes de taxes entre branques, tot i que fent-se servir d'un model *semi*-paramètric de Màxima Versemblança que optimitzaria les taxes de substitució per a cada branca. Aquest mètode es coneix com a Versemblança Penalitzada (*penalized-likelihood method*, PL) (implementat també a *R8s*).

Paral·lelament a la metodologia desenvolupada per Sanderson aparegueren altres mètodes d'autocorrelació temporal. Diversos autors proposaren un nou mètode paramètric d'implementació bayesiana per estimar la variació de la taxa evolutiva i el temps de divergència entre llinatges (Thorne et al., 1998; Kishino et al., 2001; Thorne & Kishino, 2002) fins i tot a diferents particions. Els diferents paràmetres a modelar en aquest mètode són la taxa de substitució per a cada branca (com hipòtesi inicial o *prior*: usa una distribució normal logarítmica (*log-normal*) al voltant de la taxa d'evolució del node inicial d'una branca), el temps de divergència a cada node, i el paràmetre  $\nu$  que determina la quantitat d'autocorrelació al llarg del temps (aquest és el mètode que s'implementa a *Multidivtime Package*) (Fig. 1.13).

Va ésser llavors quan Drummond i col·laboradors (2006) qüestionaren la validesa d'aquesta metodologia d'autocorrelació tant per filogènies profundes com més recents, i proposaren un nou mètode de relaxació per estimar els temps de divergència dins d'una filogènia el qual emprava també una aproximació bayesiana. La crítica de Drummond et al. (2006) afirmava que quan les distàncies evolutives són

---



grans la probabilitat que les espècies germanes estiguin sotmeses a pressions selectives similars decreix, i per tant, evolucionaran nous mecanismes bioquímics resultant en un canvi a les taxes evolutives. En canvi, en el cas de llinatges més propers, on la distància genètica és menor, els processos estocàstics que modelen l'acumulament de diferències entre les seqüències jugarien un rol més important que no pas els factors hereditaris. Per aquest motiu, els models no correlacionats s'ajustarien millor a les dades que als models autocorrelacionats (Drummond & Rambaut, 2007; Yang, 2007).

Aquest mètode va una passa més enllà dins la metodologia d'estimació filogenètica: permet que la taxa d'evolució variï entre les diferents branques, i a diferència dels anteriors mètodes de relaxació d'autocorrelació, no assumeix l'existència d'una correlació entre les taxes dels llinatges més propers. D'aquesta manera podria dir-se que aquesta metodologia permet al rellotge "relaxar-se" encara més. Un dels avantatges d'aquest mètode és que permet la co-estimació simultània de la topologia de la filogènia, de la variació de les taxes de substitució, i del temps de divergència; tres tasques que prèviament havien de realitzar-se per separat. El paquet informàtic BEAST (Drummond et al., 2006; Drummond & Rambaut, 2007) incorpora aquesta metodologia que a més, permet la implementació de diversos models que s'ajusten a les dades. Si hi ha variació de taxes entre els llinatges, les taxes de les branques descendents es prenen estocàsticament d'una distribució. Les més freqüents serien la distribució *log-normal* (*uncorrelated log-normal clock*) o bé les distribucions normals i exponencials. Un dels principals avantatges dels mètodes d'aproximació bayesiana, tant l'implementat a MultiDivTime -que permet que les taxes variïn sota l'assumpció d'autocorrelació- com l'implementat per BEAST -el qual permet variar les taxes a l'atzar-, proporcionen càlculs per estimar l'error assumit en forma d'interval de credibilitat bayesiana els quals permeten comparar i testar hipòtesis que cauen fora d'aquest interval.

La selecció del model de rellotge molecular òptim (per exemple rellotge estricte vs. relaxat o rellotge relaxat autocorrelacionat vs. no autocorrelacionat) no és una tasca senzilla. De fet, encara no és a l'abast un test estadístic òptim per saber quin model és el que millor s'ajusta a les dades. Fins aleshores han sigut molt pocs els

---

estudis que han analitzat el comportament dels models de rellotge molecular disponibles. Lepage et al. (2007) analitzen el poder estadístic a partir de la computació del rati dels valors marginals de versemblança, el que es coneix com a Factor de Bayes (de l'anglès *Bayes factors* (BFs) (veure apartat 3.9.2 de Materials i Mètodes); aquest test emperò, sembla donar resultats força distints dependentment de la profunditat de la filogènia, de la selecció gènica i del mostreig taxonòmic (Drummond et al., 2006; Lepage et al., 2007; Brown et al., 2008; Ho, 2009). Altres autors han comparat els diferents models amb el *Posterior Bayes factors*, el qual integra la distribució de la probabilitat posterior en lloc de la distribució *a priori*, el que implica un menor temps computacional (Aitkin, 1991, Aris-Brosou & Yang, 2002), tot i que sembla que tendeixen a afavorir els models amb major nombre de paràmetres (Lepage, et al., 2007).

## 7. HISTÒRIA GEOLÒGICA DE LES ILLES BALEARS

L'arxipèlag Balear es situa al bell mig de la Conca Mediterrània Occidental i representa la part emergida del Promontori Balear, que resulta d'una prolongació cap al nord-est de les Serralades Bètiques que s'estenen al sud de la península Ibèrica (Rodríguez-Perea et al., 1987; Fornós et al., 2002). Aquest promontori comprèn 350 km de llarg i 105 km d'amplada, i es troba separat de la costa ibèrica per un solc submarí (solc de València) amb una fondària de 813 m (Fig. 1.14), mentre que limita pel sud-est amb la conca Algeriana. Des del punt de vista geomorfològic consta de dos grans subgrups estructurals, cada un dels quals té la seva plataforma continental: les illes Gimnèsies, formades per Mallorca, Menorca i Cabrera, amb llurs corresponents illots; i les illes Pitiüses, formades per Eivissa i Formentera amb els seus respectius illots. Els dos grups del Promontori Balear es troben separats pel canal de Mallorca amb una fondària de 748 m. El canal de Menorca separa les illes Gimnèsies a l'extrem nord-est del promontori, tot i que comparteixen la mateixa plataforma continental (Acosta et al. 2004).

---

## Illes Gimnèsies

Mallorca és la major de les Illes Balears amb una superfície de 3,640 km<sup>2</sup>. Les dimensions màximes de l'illa atenyen els 95 km des del seu extrem occidental (Sant Elm) fins a l'oriental (Cap de Capdepera), i de 79 km des del punt més septentrional (Cap de Formentor) fins al més meridional (Cap de Ses Salines).

El basament geològic de tot l'arxipèlag balear és de tipus calcari a excepció Menorca, on es troben afloraments rocosos antics d'arenisques silícies i de roques metamòrfiques a tota la meitat nord de la illa. Durant tot el període del Juràssic

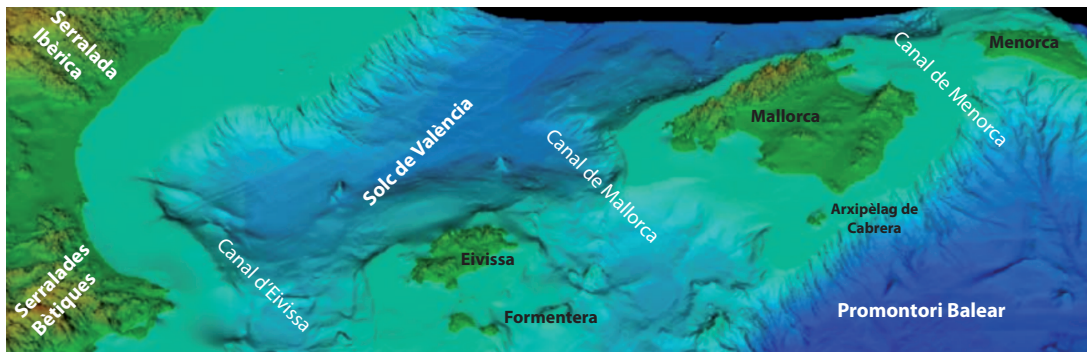


Figura 1.14. Esquema del Promontori Balear i de l'arxipèlag Balear (Acosta, 2002).

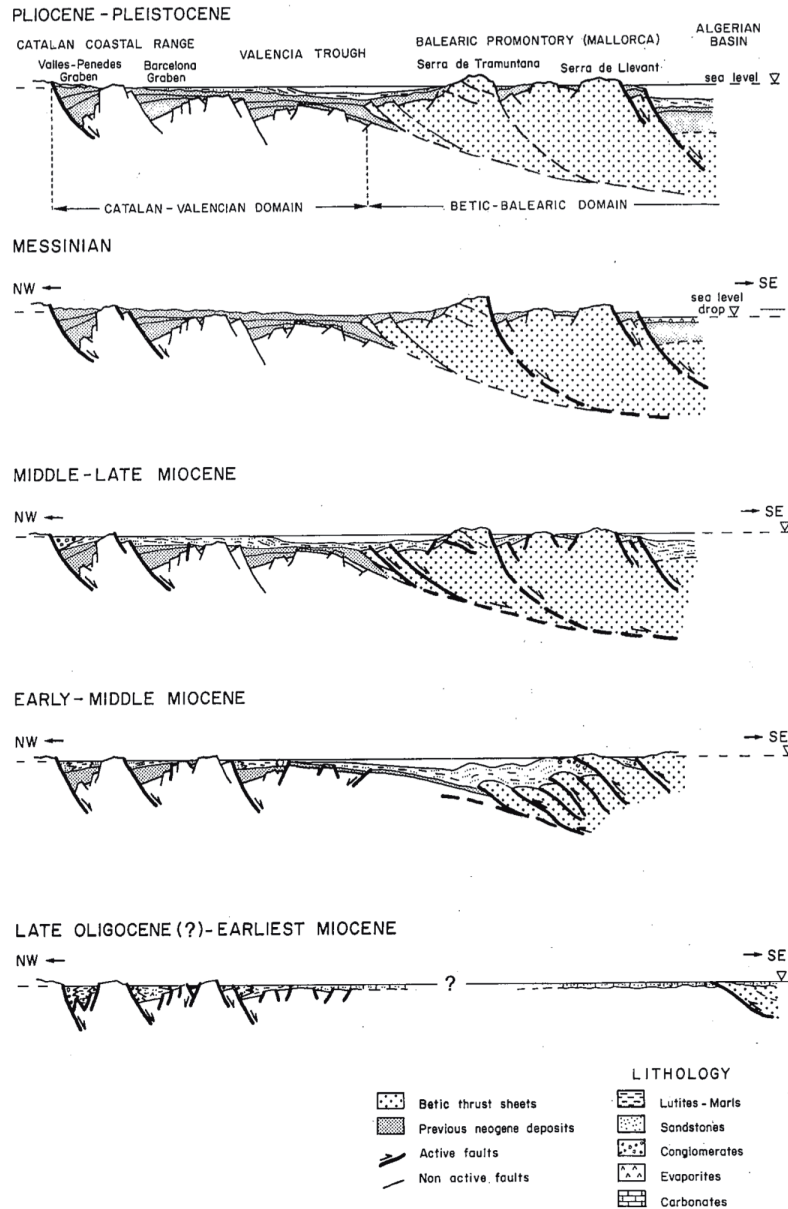
àdhuc part del Cretàcic Inferior les aigües del Mar de Tetis cobrien totalment el que en el futur donaria lloc a les illes Gimnèsies, i es creu que no fou fins a finals del Cretàcic/Paleocè/Eocè Superior que van emergir -vist que no es té registre sedimentari a l'àrea d'aquest període (Pomar, 1979, Pomar et al., 1983)-. Aquesta emersió general que afectà a tota l'àrea de la Mediterrània Occidental (Alps, Pirineus, Serralada Bètica, Nord-Àfrica, etc.) durant els períodes del Mesozoic Superior i Terciari, podria estar relacionada amb una seqüència complexa d'esdeveniments controlats per processos de compressió i d'extensió, originats arran de la col·lisió/subducció entre les dues grans plaques d'Euràsia i Àfrica deguda al gir antihorari d'Àfrica i Aràbia com a resposta a l'obertura de l'Atlàntic sud (Hsü & Bernoulli, 1978; Pomar, 1979; Pomar et al., 1983; Olivet et al., 1984; Banda & Santanach, 1992; Dañobeitia et al., 1992). Aquestes dues fases d'estructuració tectònica tingueren un gran impacte sobre les illes configurant llur aparença geomorfològica actual. La primera fase suposà processos de compressió entre el

Paleogen i el Miocè Mitjà; provocant l'inici de la separació de diverses microplaques continentals que formaven part del cinturó Hercinià de la Península Ibèrica i sud de França (que donarien lloc més endavant a Còrcega, Sardenya i les Illes Balears). La segona fase, fou de distensió i va tenir lloc durant el Miocè Superior, amb la separació completa de les I. Balears de la resta de microplaques continentals cap al Langhià (ca. 16.5 Ma), donant lloc a la formació de sistemes diferenciats de blocs, *horsts* (pilars tectònics) i *grabens* (fosses tectòniques), orientats preferentment NE-SO i limitats per grans falles normals formades a partir del Miocè Superior (Pomar, 1979; Fontboté et al., 1990; Gelabert et al., 1992; Roca, 1994; Tuccimei et al., 2006; Fornós & Gelabert, 2011) (Fig. 1.15). L'estructura de Mallorca correspon fisiogràficament a aquesta sèrie de pilars i fosses tectòniques, no més antics que el Langhià, disposats alternativament i que donaren lloc respectivament a les serres i plans actuals de l'illa. Els pilars tectònics consisteixen en dues serralades orientades en direcció NE-SO (Serra de Tramuntana i Serra de Llevant) i una sèrie de petits pujols disposats al mig d'aquestes dues (Pomar, 1979; Marzo et al., 1983; Palomino et al., 2009). Les fosses tectòniques es corresponen amb els plans, que no són més que els blocs enfonsats a partir de falles normals i que equivalen a les conques amb un farcit de materials del Miocè Mitjà-Superior i del Plio-Quaternari (Tuccimei et al., 2003). La Serra de Tramuntana és una alineació muntanyosa paral·lela a la costa nord-occidental de l'illa, amb una longitud d'uns 90 km i una amplada mitjana de 15 km. El relleu d'aquesta serralada és important, 14 dels seus cims superen els 1000 m (sobre el nivell de la mar), tots ells ubicats en el sector central. Així mateix, s'obren algunes valls (Son Marc, Sóller, etc.). La Serra de Llevant està formada per un conjunt de muntanyes d'altituds més modestes que la Serra de Tramuntana, amb suaus pendents i formes arrodonides.

L'illa de Menorca és la més septentrional i oriental de l'arxipèlag Balear, amb una superfície de 700 km<sup>2</sup>, 53 km de longitud i 19 km d'amplària màxima. Menorca es compon de dues grans unitats geomorfològiques: la regió de la Tramuntana i la regió de Migjorn, separades per un traç orientat SE-SO que s'estén des de Cala Morell fins a Maó. Cada un d'aquests sectors de l'illa són resultat de diferents etapes d'estructuració mesozoica i cenozoica de l'illa. La regió de la Tramuntana, que comprèn la meitat septentrional, és el resultat de l'etapa tectònica compressiva

---

desenvolupada durant l'Oligocè Superior i el Miocè Inferior. Es caracteritza per un relleu accidentat, amb nombrosos pujols de menys de 250 m d'altitud, a excepció de la muntanya d'El Toro de 358 m. L'altra regió, la de Migjorn, situada a la meitat meridional de Menorca, reflecteix bàsicament l'etapa extensiva que tingué lloc durant el Miocè Superior i el Plio-Quaternari. Aquesta regió correspon a una àmplia plataforma aplanada i poc elevada suaument inclinada cap a la mar (Fornós & Gelabert, 2011).



**Figura 1.15.** Esquema sintètic extret de Fontboté et al. (1990) de les principals etapes evolutives d'un perfil hipotètic NO-SE, situat entre Catalunya i Mallorca, des de l'Oligocè Superior fins al dia d'avui. Es creu que la separació completa de les Illes Balears de la resta de les microplaques continentals no seria més antiga que el Langhià (edat del Miocè Mitjà, ca. 16.5 Ma), quan es formaren els sistemes de pilars i fosses tectòniques donant lloc a les actuals serralades i plans de l'illa.

## 8. MAR DE TETIS

El terme Tetis (del grec Τηθύς, Têthýs) va ésser introduït pel geòleg austríac Suess al 1893 referint-se a una vasta mar circumtropical interior d'aigües somes avui extinta que s'estenia per l'Equador entre els supercontinents Lauràsia i Gondwana des del Juràssic Mitjà (175- 146 Ma) fins fa 20 Ma.

Cap a finals del Paleozoic totes les plaques continentals es trobaven agrupades en una única placa, formant un gran continent conegut com a Pangea. Aquest gran continent únic ràpidament començà a fragmentar-se fins a arribar a la disposició actual. Aquesta fragmentació de Pangea tingué lloc en dues fases principals. La primera, esdevingué durant un breu interval de temps durant el Triàssic Superior-Juràssic Inferior (230-200 Ma), en la qual es separaren Lauràsia i el sud-est de Gondwana del bloc format per Àfrica i Sud-Amèrica (Schettino & Turco, 2009; 2011). Aquesta separació fou arran d'una fissura que es donà en direcció ENE-OSO la qual tallà el centre de Pangea des del que equivaldria actualment a la mar Mediterrània fins al Carib (Bortolotti & Principi, 2005), formant-se així la primera escorça oceànica en el proto-Atlàntic. La fissura entre aquests grans blocs anà avançant durant el Juràssic, quedant completament separats al Juràssic Superior (164-145 Ma) per l'obertura de la primigènia Mar de Tetis, que més endavant donaria lloc a l'actual oceà Atlàntic Nord i a la Mar Mediterrània (Bortolotti & Principi, 2005). Posteriorment, en una segona fase, més perllongada, transcorreguda durant tot el Cretàcic (145-65.5 Ma), es fragmenten en blocs més petits Lauràsia i Gondwana. L'obertura de l'oceà Atlàntic nord comença ca. 180 Ma, separant Iberia i Nord-Amèrica, malgrat que al començament es tractava d'un oceà d'aigües somes i no fou fins fa 90-110 Ma quan es creu que es s'establiren condicions d'aigües profundes entre ambdues ribes (Sclater et al., 1979; Jones et al., 1995).

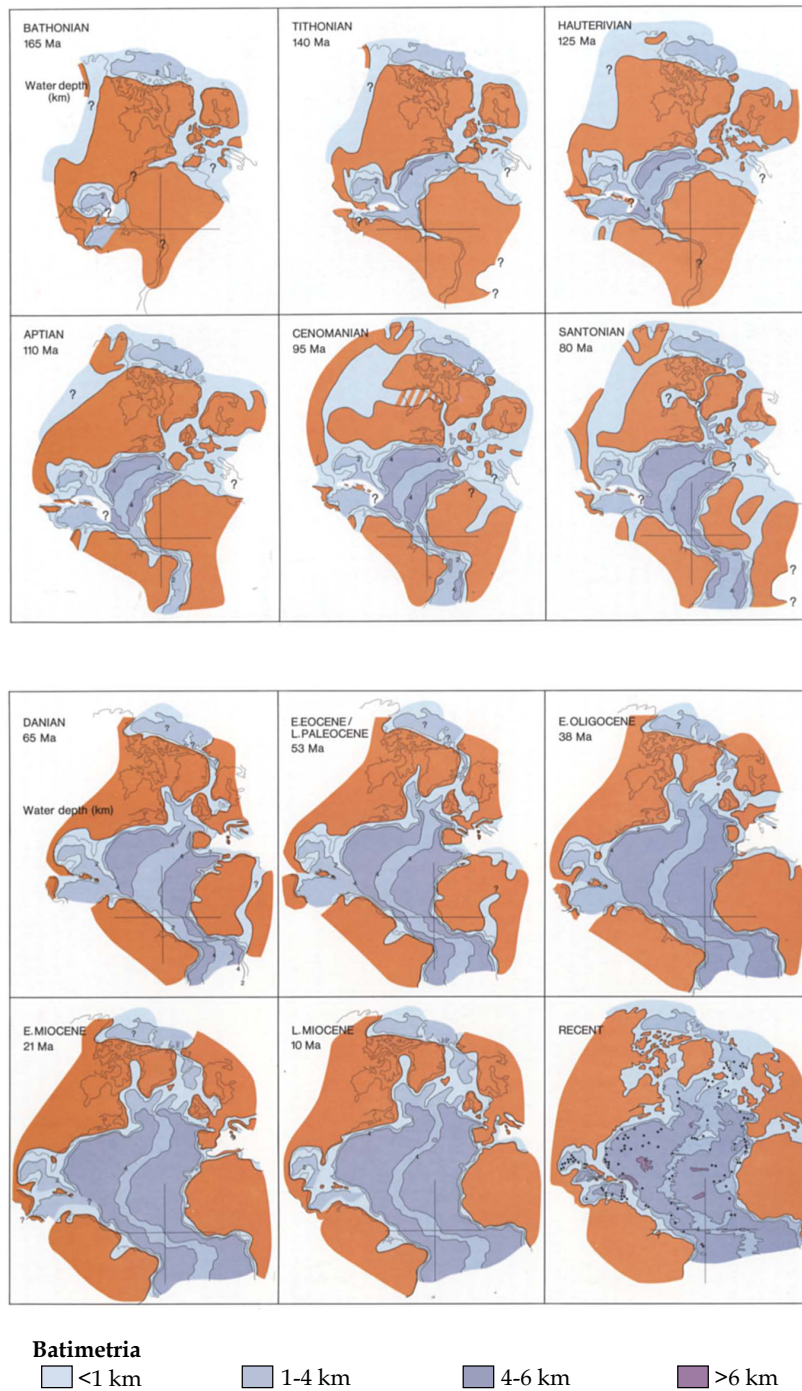
L'era de finals del Mesozoic i tot el Cenozoic és caracteritzada per cicles de transgressions-regressions que feren variar notablement la paleolínia de costa arreu del continents. Durant els períodes de transgressions, la línia costanera es traslladà terra endins, i ocasionalment negà enormes porcions, formant diverses mars epicontinentals poc profundes tal com es pot observar a la Fig. 1.16 (Thiede, 1979).

---

Un dels cicles de transgressions-regressions més importants es creu que fou el que transcorregué durant el Cenomanià-Turonian-Santonian del Cretàcic Tardà (96-83 Ma) i afectà notablement a la major part de la geografia del Marroc (Choubert & Faure-Muret, 1962) (Fig. 1.16). El treball de Boutin (1994) suggereix aquesta transgressió-regressió del Cenomanià-Turonian com a causa principal per la individualització de mols llinatges de *Metacrangonyx* marroquins a aigües continentals subterrànies.

Cap a finals del Cretàcic Superior (80-70 Ma) el proto-Carib, l'Atlàntic oriental i la porció de la Mar de Tetis que ocupava la posició de l'actual Mediterrània Occidental, formaven una ruta marítima contínua on les seves línies costaneres i illes es trobaven molt més properes entre elles del que es troben avui en dia (Smith et al., 1994). Cal esmentar que cap a aquella època, l'illa de La Hispaniola, l'arxipèlag Canari i la conca de la Mediterrània Occidental restaven encara per formar-se. Tanmateix, l'actual configuració caribenya no és anterior a l'Eocè Superior (Iturralde-Vinent, 2006); l'edat de l'illa de Fuerteventura, la més antiga de les Illes Canàries, no data d'abans de 22 Ma (Coello et al., 1992); mentre que la conca de la Mediterrània occidental es formà ca. 20 Ma (Álvarez, 1972). Això no obstant, existeixen nombroses evidències geològiques que suggereixen de la presència d'arxipèlags submergits i monts submarins al centre de l'Atlàntic Oriental des d'almenys 60 Ma (Geldmacher et al. 2001, 2005; Fernández-Palacios et al., 2011), les quals estaven més properes a la Mediterrània Occidental que actualment (Fernández-Palacios et al., 2011). D'altra banda, es creu de l'existència d'illes efímeres al Proto-Carib (illes volcàniques, crestes i bancs poc profunds) que probablement romangueren emergides uns pocs milions d'anys cap al començament del Cretàcic (145 Ma) (Iturralde-Vinent, 2006).





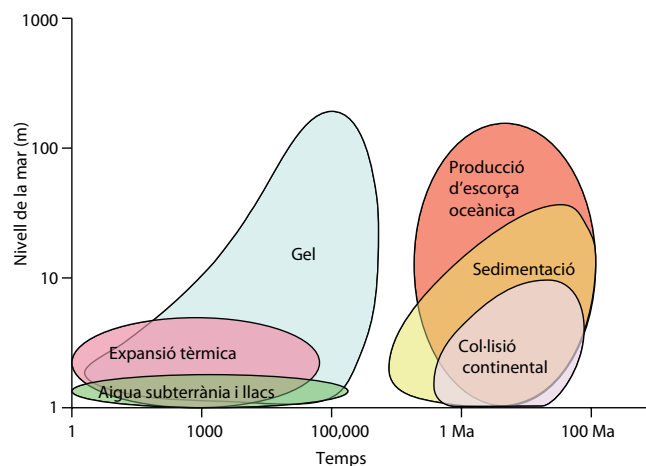
**Figura 1.16.** Paleogeografia i paleobatimetria del Nord Atlàntic i part de la Mar de Tetis Occidental durant l'era Mesozoica i Cenozoica extreta del treball de Thiede (1979). Durant el Juràssic Inferior-Mitjà comença l'obertura de l'oceà Atlàntic nord. Malgrat ja s'estableixen condicions d'aigües profundes fa 90-110 Ma, durant el Cenomanià-Santonian al Cretàc Tardà encara hi ha gran quantitat de mar epicontinental que connecten ambdós costats de l'Atlàntic nord. Concretament es pot observar a les figures corresponents del Cenomanià-Santonian com els cicles de transgressions-regressions afectaren la major part de la geografia del Marroc i del nord



## 9. EUSTATISME A LA MEDITERRÀNIA OCCIDENTAL I A NIVELL GLOBAL

Fou Eduard Suess (1888, 1906) qui proposà per primer cop el concepte de moviment eustàtic per explicar el fenomen conegut amb anterioritat de les transgressions (inundacions de la mar que desplacen la línia costera terra endins) i regressions (retrocés de la línia costanera deguda un descens del nivell de la mar). Es coneix com a eustatisme a les fluctuacions del nivell global de la mar respecte als continents a causa de tres grans fenòmens: els moviments verticals de la terra, la sedimentació i el clima; que amb moltes ocasions, poden donar-se simultàniament. L'eustatisme diastròfic o tecto-eustatisme, és aquell que, sense variar la quantitat global d'aigua, canvia el nivell dels fons oceànic arran de deformacions tectòniques de les conques -tectogènesi- o a l'acumulació de depòsits o sediments -epirogènesi-. Els canvis en el volum del oceà o glacio-eustatisme, és el fenomen que explica l'ascens o descens del nivell de la mar per canvis climàtics (Payton, 1977; Miller et al., 2005; Miller 2011) (Fig. 1.17).

Els canvis cíclics orbitals de la Terra al voltant del Sol, han generat els cicles de refredament global (edats de gel) des de l'era Precàmbrica fins al Quaternari (Hays et al., 1976). El període Quaternari ha sigut dominat per aquestes grans oscil·lacions climàtiques que han acabat determinant la distribució



**Figura 1.17.** Esquema modificat a partir del treball de Miller et al. (2005) sobre els diferents mecanismes responsables dels moviments eustàtics arreu de la superfície terrestre. El tecto-eustatisme (tectogènesi i epirogènesi) seria el més comú a èpoques més remotes, mentre que el glacio-eustatisme seria el que explicaria millor els canvis del nivell de la mar més recents.

de la fauna i flora de tota la Terra (Hewitt, 2004). Uns dels efectes més importants que han tingut les edats de gel ha sigut els canvis eustàtics a nivell global deguts a processos de glaciació i desglaciació ocorreguts durant el Plistocè (Fairbridge, 1961, 1962). Aquests cicles glacials-interglacials ocorreguts durant el Plistocè han deixat constància de transgressions de + 90 m i regressions de fins -135 m a nivell global

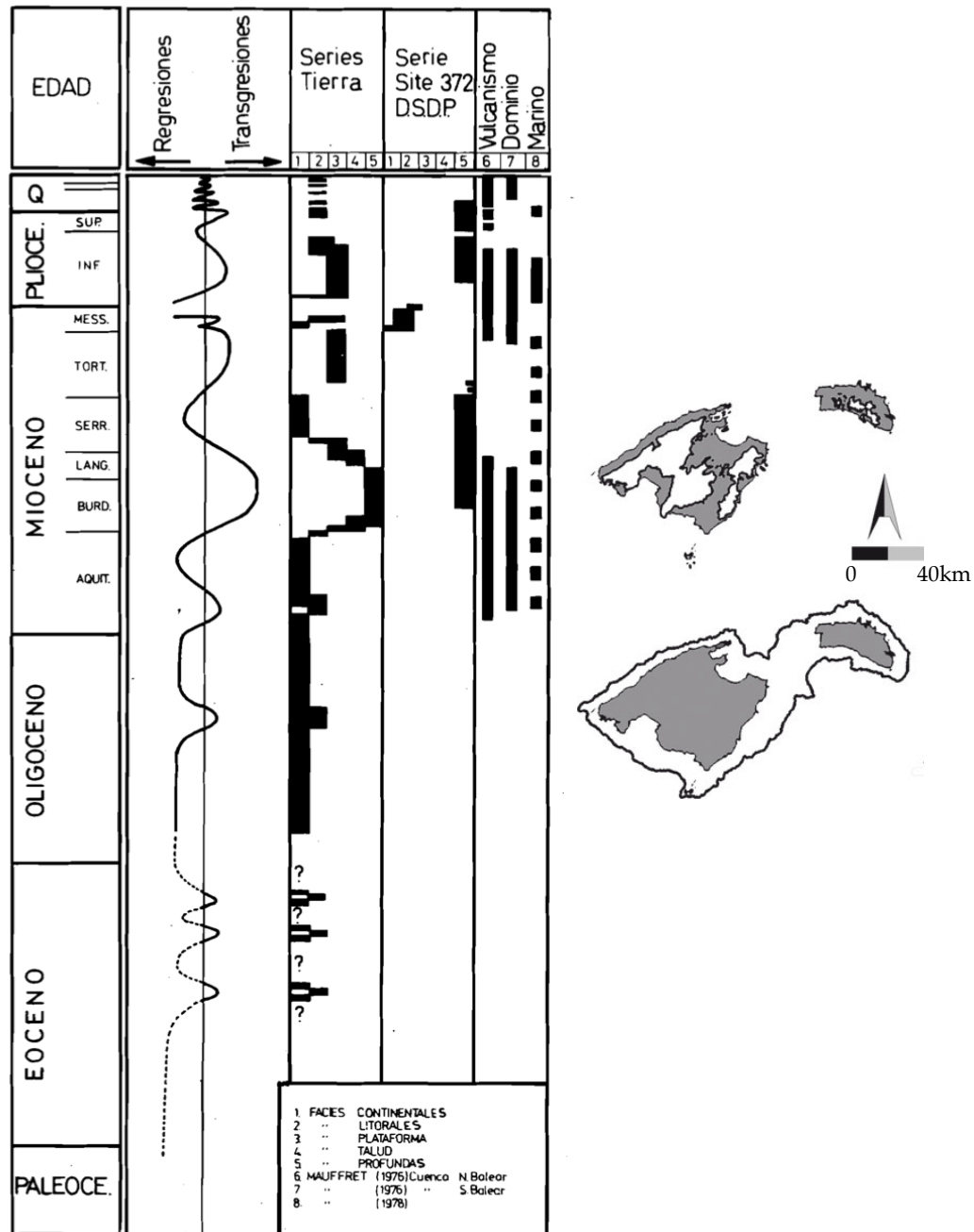
amb variacions a nivell local per efectes tectònics i hidrogràfics (Dillon & Odale, 1968; Yokoyama et al., 2001; Peltier & Fairbanks, 2006) (Fig.1.18). L'impacte produït arran d'aquests moviments eustàtics de la mar poden deduir-se a partir del registre fòssil (Willis & Niklas, 2004) o bé a partir de l'estudi filogeogràfic de determinats llinatges (Kadereit et al., 2004; de Bruyn et al., 2005). Cal esperar que la seva influència sigui més forta a l'hora de forjar la distribució i estructura genètica de llinatges associats a sistemes costaners sobretot per aquells llinatges habitants d'àrees envoltades per mars epicontinents.

Pel que fa a l'estudi de l'eustatisme a la Mediterrània, la cosa es complica perquè es tracta d'una conca restringida que, ocasionalment, ha pogut quedar closa respecte als oceans i que a més ha tingut un dinamisme molt actiu durant tota la segona part del Terciari (Riba, 1981). De les anàlisis sedimentològiques i biostratigràfiques realitzades se n'extreu que a partir del Neogen la Mediterrània sofrí una sèrie de moviments eustàtics que afectaren a tot el Promontori Balear. A la Figura 1.18 es mostra la corba de transgressió-regressió terciària de la Mediterrània Occidental. Una de les primeres transgressions més importants del Miocè esdevingué durant el Burdigalià Superior-Langhià (16.5-15 Ma), en la qual gran part de les Balears quedaren negades, seguida d'una regressió durant el Serraval·lià (13.8 Ma) que concorregué amb l'última fase de compressió balear i col·locació dels mantells de la Serra de Tramuntana (Riba, 1983). Amb el Tortonià (11.3 Ma) arribà una altra transgressió post-tectònica instaurant-se un nou domini marí que cobrí parcialment Mallorca i Menorca, amb aigües d'una mar lliure i de salinitat normal (Riba, 1981, 1983; Jolivet et al., 2006). Aquesta mar epicontinental emperò, no va atènyer nivells prou alts per enaiguar del tot les illes ja que els relleus muntanyosos acabats de formar a Mallorca (Serra de Tramuntana, Serra de Llevant, blocs elevats al centre de Mallorca) o bé la regió de la Tramuntana de Menorca, més elevada, quedaren emergides. Com a resultat d'aquesta transgressió les Illes Gimnèsies esdevingueren un grup reduït de petites illes corresponent a les parts més altes (Pomar, 1991). Just abans de finalitzar el Miocè, durant el Messinià (6.5-5.1 Ma), l'orogènia alpina provocà el tancament de l'estret de Gibraltar. Conseqüentment, la conca Mediterrània sofrí una gran regressió, el que provocà el seu l'assecament total o parcial en un breu espai de temps i deixant grans àrees de la conca emergides (Hsü et

---

al., 1977; Krijgsman et al., 1999; Duggen et al., 2003). Aquest fenomen anomenat "Crisi de salinitat del Messinià" (*Messinian salinity crisis*) ha sigut objecte de nombrosos estudis atès que ha tingut grans repercussions sedimentàries, paleogeogràfiques i paleobiològiques per a diferents grups d'organismes, entre els quals es compten els crustacis estigobionts talassoides (Coineau, 1994). Cap a començaments del Pliocè Inferior (5.5-5.35 Ma), l'estret de Gibraltar s'obrí de nou, suscitant una transgressió marina a la conca Mediterrània. Es creu que el procés d'inundació tardà menys de 20,000 anys a completar-se (Pierre et al., 2006), tot i que el nivell de la mar no arribà al nivell que hi havia durant el Tortonjà. Aquest augment del nivell de la mar deixà grans parts de l'arxipèlag Balear emergides, quedant com a resultat un arxipèlag format per nombroses illes sense connexió amb altres illes del Mediterrani o altres territoris de la Península Ibèrica, nord d'Àfrica o sud de França. Els canvis climàtics succeïts durant el Quaternari produïren grans oscil·lacions glacio-eustàtiques que modificaren notablement la línia de costa i l'àrea emergida de les Illes Balears. Com a casos extrems es podrien mencionar les fortes regressions causants de la unió temporal entre els diferents grups geomorfològics (les I. Gimnèsies d'una banda, i les I. Pitiüses de l'altra) o bé el cas d'acusades transgressions, que produïren un dràstic minvament de l'àrea emergida de les illes reduint-se a gairebé a menys de la meitat. D'aquesta manera la mar entrà per les badies i altres zones baixes inundant grans superfícies de terreny i acabà reduint certes àrees en conjunts de petits illots circumdats per aigües somes (Hillaire-Marcel et al., 1996; Vesica et al., 2000; Gràcia et al. 2001) (Fig. 1.18).

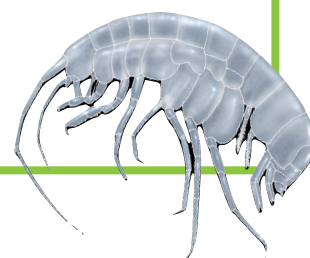
---



**Figura 1.18.** *Esquerra:* esquema sintètic extret de Pomar (1983) de les sèries terciàries a la Mediterrània Occidental. La corba eustàtica ha estat inferida a partir de fluctuacions a nivell de la mar i amb comparació de les dades del domini marí. Al Miocè s'observen diverses transgressions importants: la primera durant l'Aquitanià, una altra de major extensió durant el Burdigalià superior-Langhià seguida d'una fase regressiva durant el Serravallia, i la darrera transgressió durant el Tortonian que fou més lleugera i negà parcialment l'arxipèlag, convertint a Mallorca i Menorca en un grup d'illots a les zones més elevades (Tramuntana i Llevant). A la corba també es pot observar la gran regressió ocorreguda durant el Messinià arran del tancament de l'estret de Gibraltar que es continuà d'una transgressió cap al Pliocè Inferior. El Plistocè és caracteritzat per cicles de glaciació i deglaciació molt seguits que afectaren a la línia de costa nombroses vegades.

*Dreta:* imatge extreta de Gràcia et al. (2001) que mostra la configuració aproximada de les I. Gimnèsies durant una fase de màxima transgressió (a dalt) i una fase de màxima regressió (a sota) durant el Plistocè. En gris la distribució actual de les terres emergides a les illes.

**MATERIALS I MÈTODES**





## 1. ORGANISMES D'ESTUDI I MOSTREJOS

El material zoològic sobre el qual versa aquesta tesi es capturà a diverses coves i pous d'aigua dolça i/o salobrosa durant el període 2007-2010. Els animals es capturaren utilitzant una xarxa de plàncton confeccionada amb *Nytral* de 100 micres de llum de malla i 20 cm de diàmetre de boca acoblada a un mànec telescòpic extensible fins a 3 m, el que facilitava els seu transport dins les coves. Pels pous s'utilitzà una xarxa *Cvetkov* modificada (Cvetkov, 1968) i confeccionada amb el mateix material, de 25 cm de diàmetre de boca. Aquesta xarxa portava al seu extrem un gobelet proveït d'una vàlvula que impedia als amfípodes eventualment capturats escapar nedant quan la xarxa es submergia de nou per iniciar una altra captura. La xarxa estava llastrada de manera que s'enfonsés amb facilitat i es resuspengués el sediment del fons del pou per facilitar-ne les captures. L'objectiu era col·lectar, quan fos possible, 10 individus per població.

Els exemplars foren fixats immediatament en etanol absolut i conservats a -20°C per preservar el DNA en condicions òptimes fins a la seva posterior identificació i extracció en el laboratori.

A cada capítol es detallen les localitats de la col·lecta, les seves coordenades geogràfiques, els individus estudiats, juntament a mapes als quals es representen els punts de col·lecta.

### 1.1. Filogeografia de l'espècie *Metacrangonyx longipes* (Capítol I)

Aquesta espècie endèmica de les illes Gimnèsies (Fig. 2.1) fou descrita com a una nova espècie i gènere a 1909 pel cèlebre naturalista francès Chevreux gràcies al material col·lectat pel biòleg marí Emil G. Racovitza l'any 1905 a les Coves del Drac de Porto Cristo (Mallorca). Es distingeix amb facilitat i a primera vista de qualsevol altre amfípode present a les illes, sigui subterrani o epigeu, gràcies als seus tercers uropodis extremadament reduïts.

Ulteriorment, l'espècie s'ha trobat en nombroses localitats de l'illa de Mallorca inclosos als plans de l'interior. A la resta d'illes de l'arxipèlag balear només es té

---

constància de l'existència de dues poblacions localitzades a Menorca mentre que es desconeix de cap altra població a les Pitiüses. Aquestes dues poblacions menorquines habiten dues coves anquihalines ubicades a extrems oposats (est i oest) de l'illa (sempre a substrat calcari del Miocè). No s'ha trobat mai a més de 125 m d'altitud, i



**Figura 2.1.** Imatge d'un exemplar de *Metacrangonyx longipes* obtinguda per microscopia electrònica de rastreig. (Foto: Ferràn Hierro, UIB).

s'ha observat que ocupa un rang ben ampli de salinitats: des d'aigües marcadament salobres del litoral costaner (aquífers salinitzats, coves anquihalines) a totalment dolces de pous de la zona del pla de llevant de Mallorca o bé a rius subterranis actius, com la cova dels

Estudiants a Sóller. Com a curiositat cal esmentar que a una cavitat litoral de Porto Cristo (Cova del Dimoni), connectada directament a mar oberta per un ample túnel submarí, es capturaren en una ocasió diversos exemplars (Jaume & Christenson, 2001). Malauradament, existeix el dubte de què un aiguaneix desguassi directament a la mar a l'interior de la cavitat, pel que els espècimens capturats podrien en realitat haver vingut d'aigües dolces subterrànies interiors, pel que no es tindria constància de parents marins.

Poc es coneix de l'ecologia de l'espècie, tret que no li atreuen els esquers (s'ha provat amb carn, peix, formatge...) i que es captura amb més freqüència a prop de fustes submergides, amb el que es probable que s'alimenti de fongs o bacteris que creixen sobre la fusta en descomposició.

En conjunt es mostrejaren 162 espècimens de *M. longipes*. S'utilitzaren com a localitats de referència les presents a la bibliografia disponible (set coves anquihalines i una d'aigua dolça) i es prospectaren noves localitats a 23 pous d'aigua dolça a l'interior de l'illa de Mallorca. Addicionalment es cercaren més localitats fora de la distribució geogràfica descrita fins aleshores, com diverses zones properes a la Badia de Palma (zona de Calvià-Andratx), encara que no se'n trobaren exemplars. Les localitats mallorquines



i menorquines considerades a l'estudi per la filogeografia de *M. longipes* es mostren a la Figura 3.1 (Capítol I) mentre que llurs coordenades geogràfiques i el nombre d'individus analitzats per a cada marcador molecular es poden veure a la Taula 3.1 (Cap. I). Com a grups externs s'utilitzaren espècies congenèriques originàries del Marroc: *M. remyi* Balazuc & Ruffo, 1953 col·lectada a una alçada de 1,280 m a l'Alt Atlas (Marroc), "*M. nicoleae tamri*" col·lectada a Tamri (costa d'Agadir, Marroc), i *M. ilvanus* Stock 1997 provinent de l'illa d'Elba (arxipèlag toscà, Itàlia).

## 1.2. Filogènia i biogeografia del gènere *Metacrangonyx* (Capítol III)

S'han analitzat 214 espècimens del gènere *Metacrangonyx* corresponents a 17 espècies i dues espècies del gènere *Longipodacrangonyx* provinents de pous d'aigua dolça i coves anquihalines arreu de tota la seva distribució geogràfica: República Dominicana (illa de La Hispaniola), Fuerteventura (Arxipèlag Canari), Marroc, Mallorca i Menorca, i l'illa d'Elba (Itàlia), a excepció de les espècies originàries de l'Orient Mitjà, de les quals no es van poder aconseguir exemplars (Fig. 3.24, Capítol III). Alguns dels tàxons originaris del Marroc inclosos en aquest estudi no han estat encara descrits formalment, tot i que es troben identificats i citats amb la seva nomenclatura binomial en els treballs de Messouli (1988, 1994). Per aquesta raó, en aquesta tesi s'han fet servir els noms emprats al treball de Messouli (1988, 1994) tot i que entre cometes i no en lletra itàlica per poder distingir-los d'aquells que ja es troben formalment descrits. A més s'han inclòs altres dues espècies d'amfípodes que s'han emprat com a grups externs: *Bahadzia jaraguensis* (República Dominicana) i *Pseudoniphargus daviui* (Cabrera, Illes Balears), pertanyents a les famílies Hadziidae i Pseudoniphargidae, respectivament. La localització dels mostrejos amb les coordenades geogràfiques corresponents i el nombre d'individus estudiats es troben a la Taula 3.7 (Cap. III).

## 2. PROTOCOLS EXPERIMENTALS

### 2.1. Extracció i purificació de DNA genòmic

Prèviament a l'extracció del DNA genòmic, els espècimens es van deixar eixugar a temperatura ambient fins a s'evaporés tot l'etanol.

---

Les extraccions de DNA es realitzaren a partir de l'espècimen sencer amb *DNAeasy™ Tissue kit* (Qiagen, Hilden, Alemanya), seguint les recomanacions del fabricant tot i que amb les següents modificacions: incubació amb proteïnasa K durant tota la nit, adició d'un volum de tampó d'elució de 50µl i incubació de 10 minuts a temperatura ambient abans de l'elució. Aquest kit tal com fa la tècnica estàndard empra enzims proteases i detergents per extreure el DNA però per a la purificació substitueix el tradicional protocol de fenol/cloroform per matrius de silici, les quals a determinades condicions de força iònica i pH, presenten una alta afinitat pel RNA i DNA però no per la resta de molècules.

Per comprovar l'èxit de l'extracció de DNA es quantificà el total de DNA genòmic extret a partir d'un espectrofotòmetre *NanoDrop© ND-1000*. En certes ocasions, la integritat del DNA es va estimar amb electroforesi de gel d'agarosa al 1%, TAE 1% i bromur d'etidi 50 µg/mL.

## **2.2. Amplificació i purificació del producte de PCR convencionals**

L'estudi del patró espai-temporal de la invasió dels hàbitats subterranis per a l'espècie *M. longipes* a poblacions balears es realitzà a partir de l'anàlisi de les seqüències de DNA del gens mitocondrials que codifiquen per citocrom oxidasa I, citocrom b, i rRNA 16s (*cox1*, *cob*, *rrnL*, respectivament) i les seqüències del marcador nuclear Histona H3A (*H3*).

Per a la reconstrucció de la filogènia del gènere es va fer prèviament un estudi per conèixer quins eren aquells llinatges més representatius i divergents per amplificar-ne el seu mitogenoma. Per esbrinar-ho, s'amplificaren els marcadors *cox1* i *rrnL* de 214 espècimens de la família *Metacrangonyctidae* (detall a la Taula 3.7, Cap. III). Un cop obtingut aquest arbre preliminar, es van identificar els llinatges més representatius per a *Metacrangonyx* i *Longipodacrangonyx* (Figura 3.20, Cap. III). En alguns dels casos, s'observà que individus de la mateixa morfo-espècie presentaven una inesperada divergència genètica suggerint la presència d'espècies críptiques i per tant es seleccionaren representants de cada un d'ells. Com a resultat, es seleccionaren 21 llinatges per obtenir-ne posteriorment llurs mitogenomes complets així com el gen nuclear que codifica per a la subunitat petita ribosomal 18S (*SSU*).

---

Tots aquests fragments (*cox1*, *cob*, *rrnL*, *H3*, *SSU*) han sigut amplificats per tècniques convencionals de PCR (*Polymerase Chain Reaction*) (Mullis & Faloona, 1987; Saiki et al., 1988). Per a l'amplificació dels fragments s'utilitzaren els encebadors descrits anteriorment a Folmer et al. (1994), Barraclough et al. (1999), Xiong & Kocher (1991); i Colgan et al. (1998). Per alguns dels individus, aquests encebadors "universals" no donaren amplificacions positives pel que es van dissenyar encebadors específics a partir de les seqüències obtingudes a altres individus. Totes les seqüències dels encebadors emprats, així com el nombre d'individus amplificats es presenten a la Taula 2.1. El volum final de cada reacció de PCR fou de 25  $\mu$ L, i s'utilitzà aproximadament uns 15 ng de DNA genòmic, amb les següents concentracions finals de: una unitat de Taq DNA polimerasa (EcoTaq, Ecogen), 1,8 mM de  $MgCl_2$ , 0,25  $\mu$ M de cada encebador i 0,5 mM de dNTPs. La barreja de reactius, exceptuant el DNA motlle, es manipulà a una zona específica del laboratori lliure de DNA per evitar possibles contaminacions, i es treballà en tot moment sobre gel per evitar l'activació prematura de la reacció. Les condicions d'amplificació de la PCR es presenten a la Taula 2.2. L'eficiència de l'amplificació i l'absència de bandes secundàries inespecífiques es comprovaren per electroforesi en gels d'agarosa a l'1% tenyits amb Bromur d'Etidi (50  $\mu$ g/mL). Un cop comprovada l'amplificació baix llum ultraviolada, aquelles mostres que presentaven una única banda d'amplificació i la mida esperada, foren purificades mitjançant el kit comercial de columnes *MSB® spin PCRapace* (Invitex GMBH, Berlin, Alemanya) que empra matrius de silici per purificar el DNA, seguint les recomanacions del fabricant. Els purificats foren després quantificats per un espectrofotòmetre *NanoDrop® ND-1000* per normalitzar-ne llur concentració.

A *Metacrangonyx*, les amplificacions del gen *SSU* amb els encebadors "universals" donaren dos amplicons de diferent grandària i pes molecular. Ambdues bandes es tallaren del gel d'agarosa i es purificaren independentment amb les columnes de silici del kit *QIAquick™ Gel Extraction Kit* (Qiagen, Hilden, Alemanya). Ambdós amplicons es seqüenciaren i s'observà a partir d'una cerca a les bases de dades genòmiques mitjançant l'algoritme BLASTn (de l'anglès *Basic Local Alignment Search Tool*, Altschul et al., 1997; disponible a l'adreça <http://www.ncbi.nlm.nih.gov/blast>) que un d'aquests corresponia a una contaminació per ciliats, mentre que l'altre corresponia al gen de *SSU*. A partir d'aquesta seqüència es dissenyaren dos encebadors específics per a *Metacrangonyctidae*, més tres encebadors interns per seqüenciar el fragment sencer (Taula 2.1).

Taula 2.1. Oligonucleòtids emprats a l'amplificació i seqüenciació dels fragments de DNA estudiats.

Gen	Nom encebador	Sentit	Mida esperada (pb)	Seqüència 5'-3'	Referència
Cox1	LCO1490	F	649	GGTCAACAAATCATAAAGATATTGG	Folmer et al., 1994
	HCO2198	R	649	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al., 1994
	MetaCoxF2	F	620	GAACCTTAGATACCCWGGTAATTTGATYG G	Aquest estudi
	MetaCoxR2	R	620	TCAGTTAATAAYATAGTAATAGCYCC	Aquest estudi
	M14	F	490	CGCCTGTTTAACAAAAACAT	Xiong & Kocher, 1991
rrnL	M74	R	490	CTCCGGTTTGAACCTCAGATCA	Xiong & Kocher, 1991
	16SmetaF	F	400	RGTATTTTGACCGTGCTAAGG	Aquest estudi
	16SmetaR	R	400	TGTAAAAATTAARGTTGAACAAAC	Aquest estudi
Cob	CB3	F	360	GAGGAGCAACTGTAATTACTAA	Barraclough et al., 1999
	CB4	R	360	AAAAGAAARTATCATTGAGTTGAAT	Barraclough et al., 1999
H3	H3A	F	325	ATGGCTCGTACCAAGCAGACVGC	Colgan et al., 1998
	H3B	R	325	ATATCCTTRGGCATRATRGTGAC	Colgan et al., 1998
SSU	SSU-F	F	2300	CCTAYCTGGTTGATCCTGCCAGT	Englich et al., 2003
	SSU-R	R	2300	TAATGATCCTTCCGCAGGTT	Englich et al., 2003
	18s-F123	F	770	CGGCTCATTAAATCAGTCTTGG	Aquest estudi
	18s-F651	F	960*	AGAACCTACTGAAGGGCAAG	Aquest estudi
	18s-F1228	F	900*	GAGTAAATCAGAGTGCTCAAAG	Aquest estudi
	18s-F1553	F	780*	TGCCAACCATCAATCCGC	Aquest estudi
	18s-R2605	R	480	CGGCTCATTAAATCAGTCTTGG	Aquest estudi

\* Fragments amplificats interns un cop amplificat el gen SSU amb els primers SSU-F/R o 18s-F123/R2605.

Taula 2.2. Condicions d'amplificació per a cadascun dels fragments.

Fragment	Desnaturalització inicial		Desnatur.		Annealing		Extensió			Extensió final	
	T(°C)	temps	T(°C)	temps	T(°C)	temps	T(°C)	temps	cicles	T(°C)	temps
Cox1	94	4'	94	30"	47	30"	72	45"	40	72	10'
Cox1*	94	4'	94	30"	50	30"	72	45"	40	72	10'
rrnL	94	4'	94	30"	48	30"	72	30"	40	72	10'
rrnL*	94	4'	94	30"	46	30"	72	30"	40	72	10'
SSU	94	4'	94	30"	51	30"	72	2'30"	35	72	10'
Long-mit <sup>#</sup>	95	1'	94	20"	58-64	20"	62-64	8'	35	68	10'
Short-mit <sup>#</sup>	95	1'	94	15"	58-64	15"	62-64	5'	35	68	10'

\* Fragment amplificat amb encebadors específics de Metacrangonyx.

# Fragments amplificats per la tècnica de Long-PCR. El Long-mit correspon al fragment de 10 kb i el Short-mit a 4.5kb aprox. Les temperatures d'annealing en aquest cas depèn de cada parella d'encebadors emprats.

### 2.3. Seqüenciació i precipitació

Es seqüenciaren ambdues cadenes dels fragments de PCR purificats mitjançant el mètode de seqüenciació cíclica automàtica ABI Prism Big Dye™ Terminator Cycle Sequencing Ready Reaction kit versió 2.0. (Applied Biosystems, Foster City, CA, USA). Aquest mètode combina la terminació de la reacció de polimerització per la

incorporació dels di-desoxinucleòtids (Sanger et al., 1977), la utilització d'enzims termostables i l'ús de fluorocroms diferents per identificar cada un dels quatre di-desoxinucleòtids diferents.

La barreja de la reacció de seqüenciació es realitzà en un volum total de 10 µL: 0.6 µL *Big Dye*<sup>TM</sup>, 1,2 µL 5xbuffer, 0,2 µL d'encebador [10 µM], 15-20 ng de DNA i completant amb aigua miliQ esterilitzada fins als 10 µL. El programa utilitzat per a la reacció de seqüenciació fou el següent: una desnaturalització inicial de 95 °C durant 3 minuts, seguida de 35 cicles amb una fase de desnaturalització de 10 segons a 96 °C, una fase d'hibridació de 5 segons a 50 °C i una fase d'extensió durant 4 minuts a 60 °C. Els productes obtinguts foren precipitats afegint a la reacció 50 µL d'etanol absolut i 2 µL d'acetat sòdic (3 M a pH 5,2). Després d'una incubació a temperatura ambient durant 15 minuts, el precipitat es concentrà per centrifugació a 14.000 rpm durant 30 minuts. Posteriorment, aquest precipitat es rentà amb 250 µL d'etanol al 70% centrifugant-se a 14.000 rpm durant 15 minuts. El precipitats de les reaccions una vedada assecats a temperatura ambient s'enviaren al Servei de Seqüenciació del Parc de Recerca Biomèdica de Barcelona on s'analitzaren en el seqüenciador automàtic ABI 3100 (Applied Biosystems, Foster City, CA, USA).

#### 2.4. Amplificació del mitogenoma sencer (Long-PCR) (Capítols II i III)

La PCR "llarga" o més coneguda com a Long-PCR (*Long-range polymerase chain reaction*), és una variant de la tècnica de la PCR que permet l'obtenció de fragments més extensos que la PCR convencional (entre 5 Kb i 30 Kb). Aquesta amplificació de fragments tant llargs és possible gràcies a l'ús d'enzims polimerases modificats que mostren una termostabilitat superior gràcies a la incorporació per enginyeria genètica de cofactors específics que tenen una funció correctora i reparadora a causa de la seva eficient activitat exonucleasa 3'-5' (per exemple la *Pfu* DNA polimerasa). En aquest estudi s'utilitzà l'enzim *Herculase*<sup>TM</sup> II *Fusion DNA polimerase* (Agilent, Sta Clara, CA, USA).

Els genomes mitocondrials de 21 llinatges de *Metacrangonyx* i dels grups externs *Bahadzia jaraguensis* i *Pseudoniphargus daviui* es varen amplificar en dos fragments solapants de 4.5 i 10 Kb mitjançant PCR "llarga". Per a cada mitogenoma, es

dissenyaren encebadors específics d'uns 30 pb i al voltant d'un 40% de G+C, basant-se en les seqüències nucleotídiques obtingudes dels gens de *cox1* i *rrnL* d'un individu representant de cada llinatge. És a dir, el fragment llarg cobria des del gen de *cox1* fins al *rrnL*, i el curt en la cadena complementària des del gen *rrnL* fins al *cox1*. No obstant això, el mitogenoma quasi complet de cinc de les espècies fou possible amplificar-los amb una parella d'encebadors degenerats basats en les comparacions de les seves seqüències dels marcadors de *rrnL* i *rrnS* (Taula 2.3).

A l'hora de dissenyar els encebadors per *Long-PCR* s'han de tenir en compte diversos requisits per assegurar-se'n l'èxit: òbviament cal que no contenguin bases complementàries entre elles o amb l'altre encebador, però sobretot s'ha d'evitar la complementarietat a l'extrem 3', per evitar així l'aparició de *primer-dimer* o *cross-dimer*; és recomanable dissenyar també encebadors rics en contingut G+C (superior al 40%) vist que augmenta la temperatura de hibridació i per tant la seva especificitat; i finalment, cal evitar dissenyar els encebadors a regions d'estructures secundàries o a regions repetitives per evitar l'aparició d'estructures secundàries internes (*hairpins*, *loops*). Pel disseny dels nous encebadors s'utilitzà la versió online del programa *Netprimer* (Premier Biosoft International, USA, <http://premierbiosoft.com/servlet/com>) que permet valorar diferents paràmetres fonamentals (contingut en G+C, temperatura d'hibridació, formació de *primer-dimer* o *cross-dimer*, estabilitat de l'extrem 3', etc.) i així destriar, entre els encebadors possibles, aquell que sigui l'òptim per a la *Long-PCR*. A totes les amplificacions es va dur a terme la tècnica de la *Hot-start PCR*. Aquesta tècnica consisteix en orbar a la reacció de l'enzim polimerasa fins que s'hagi assolit una temperatura superior de 80 °C. D'aquesta manera, s'evita la hibridació de l'encebador a zones de similitud parcial a la diana, pel que s'augmenta considerablement l'especificitat de l'amplificació. La mescla de la reacció de la *Long-PCR* fou: 1-2 µL de DNA genòmic (concentració aproximada de 100 ng/µL), 5 µL de 10 x PCR buffer, 8 µL de mescla de dNTP (2.5 mM cada un), 2.5 µL de cada un dels primers (10 µM), 1 µL d'*Herculase*<sup>TM</sup> (5 U/µl) i fins 50 µL d'aigua destil·lada MiliQ. Les temperatures òptimes d'hibridació del encebador es van determinar empíricament, realitzant una PCR amb gradient de temperatura, la qual anava d'un rang de 5° C menys de la temperatura teòrica estimada amb el programa *Netprimer* fins a 2°C graus més. A la Taula 2.2 es mostren les temperatures resultants dels cicles dutes a terme per l'amplificació dels mitogenomes.

---

L'eficiència de la reacció d'amplificació es comprovà per electroforesi en gel d'agarosa a l'1%, TAE 1x i tinció amb Bromur d'Etidi (50 µg/mL). Aquelles mostres que presentaven una única banda foren purificades mitjançant el kit comercial de columnes *MSB® spin PCRapace* (Invitak GmbH, Berlin, Alemanya). Vist que amb una reacció no s'obtenia prou producte, es repetiren les reaccions de *Long-PCR* fins aconseguir la quantitat necessària. Per aquelles mostres on a l'amplificació s'hi obtingueren algunes inespecificitats (i fou impossible eliminar-les variant les condicions d'amplificació) es procedí a purificar les bandes de la mida esperada directament del gel tenyit amb cristall violeta (1.6 mg/mL) i fer-ne una re-amplificació del producte. El cristall violeta s'uneix a les càrregues negatives de la molècula de DNA de forma no permanent (pel que no és mutagènic) i són visibles a la llum natural. Això, estalvia tenyir el gel amb Bromur d'Etidi i així evita la ruptura i la degradació de la molècula de DNA al sotmetre el gel baix llum ultraviolada (Rand, 1996). Un cop purificats tots els productes de PCR, s'ajuntaren els fragments del mateix pes molecular i del mateix individu i es precipitaren segons el protocol de Etanol/Acetatsòdic descrit a l'apartat anterior 2.3.

---



**Taula 2.3.** Llista d'encebadors emprats per l'amplificació dels mitogenomes de *Metacrangonyctidae* i grups externs.

Espècie	Codi	Fragment	Nom encebador	Seqüència
<i>M. dominicanus</i>	DOM	<i>cox1-rrnL</i>	MetDom_cox1F4	GAGAATAGTTGAGAGGGGGGTAGGG
			MetDom_16s_R3	TGAAAACITGGAATGAAGGGTCTAACAA
		<i>rrnL-cox1</i>	MetDom_16s_F3	TTGTTATACCCCTTCATTCCAGTTTCA
<i>M. samanensis</i>	SAM		MetDom_cox1_R4	CCTACCCCTCTCAACTATITCTCC
		<i>cox1-rrnL</i>	Sam_cox1_F1	CTGGGCTAGGATAGTGGGTACTGCTA
			Sam_16s_R6	AAATAAAAAGGAAAGGATTAAGTTACTCTAGGG
		<i>rrnL-cox1</i>	Sam_16s_F7	TATCCAGAACTATTACGCTGTTATCCCTAGAG
<i>M. repens</i>	REP		Sam_cox1_R4	GGTCCATTTTTATCCCACTAACTCG
		<i>cox1-rrnL</i>	Repens_cox1_F8	GGTCATAGTGGTTGTTCTGTAGATTTGG
			Repens_16s_R12	AATGGGATGTAAGAAAGTTACTCTAGGGATA
		<i>nad1-cox1</i>	Repens_nad1_F5	CCCAGATATAAATCCACACCCCTTC
			Repens_cox1_R9	CAACCAGTTCCTACTCCACTCTCCAC
<i>M. ilvanus</i>	ILV		Ilv_cox1_F5	GAGGTATGGTAGAAAAGAGGAGTTGGTACAGG
		<i>cox1-rrnL</i>	Ilv_16s_R2	GGATAACAGCGTAATAATTTTGATAGCCC
			Ilv_16s_F2	GGCTATCCAAAATTATTACGCTGTTATCCC
		<i>rrnL-cox1</i>	Ilv_cox1_R2	CAGTCCATCCTGTACCAACTCCTCTTTTACC
"M. nicoleae tamri"	TAM	<i>cox1-rrnL</i>	Tam_cox1_F1	TAGTATAGTGGGACGGCAATAAGGG
			Tam_16s_R2	AGTTACTCTAGGATAACAGCGTAATAATTTGG
		<i>rrnL-cox1</i>	Tam_16s_F1	AAGGTCTAACCAAAATTATTACGCTGTTATCCC
<i>M. spinicaudatus</i>	SPI		Tam_cox1_R1	IGTTCATCCTGTCCCTACTCCTCTCTCCAC
		<i>cox1-rrnL</i>	SpinSK_cox1_F2	ACTGATAAGAAGAATAGTAGAAAGAGCGGTAGG
			SpinSK_16s_R3	TTGGAAACTGGAATGAAAGGTTAAACAAT
		<i>rrnL-cox1</i>	SpinSK_16s_F1	GGTCTATCCAAAATTATACGCTGTTATCCC
"M. boveei"	VOM		SpinSK_cox1_R2	CAACCTGTACCTACGCCCTTTTCTACTATCTC
		<i>cox1-rrnL</i>	Bov6_cox1_F1	GAGAATAGTAGAAAGAGGGGTAGGGACGGG
			Bov6_16s_R1	ATTTGGGGACTAGAATGAAAGGTTTACAACTATG
		<i>rrnL-cox1</i>	Bov6_16s_F3	ATTCATAAGTGTGAAAACCTTTCATCTACTCCCC
<i>M. remyi</i>	REM		Bov6_cox1_R2	CCGTCCCTACCCCTCTTTCTACTATTTCTCC
		<i>cox1-rrnL</i>	Rem_cox1_F3	GACTGTTTACCCTCCTTTATCCAGAGTTATTCG
			Rem_16s_R1	CGATGTGGAATAAAATATCCCTGTAGAGTAGAAAC
		<i>rrnL-cox1</i>	Rem_16sF3	TATAAGCCGTTGGGACCCCTTCTACTAGTC
<i>M. longicaudus</i>	LML		Rem_cox1R3	TCCAACCAGTTCCTACCCTCTTTCTACTATT
		<i>cox1-rrnL</i>	IMLM_cox1_F2	ATAAGAAGAATAGTAGAGAGAGGAGTAGGGACTGG
			IMLM_16s_R1	TAATTGGGAACITGGAATGAAGGGTTGAAC
		<i>rrnL-cox1</i>	IMLM_16s_F1	GATCTATCCAAAATTATCGTGTCTGTTATCCC
<i>M. panousei</i>	AGM		IMLM_cox1_R2	TGAAAAAATAGCCAAATCCACCGAAGC
		<i>cox1-rrnL</i>	AGTS_cox1_F1	GCCAGAATGGTAGGCCTGCTATAAGAG
			AGTS_16s_R1	ATCTGAGGTTTGAATAACTTTGATTTGGGG
		<i>rrnL-cox1</i>	AGTS_16s_F1	CCTATCCAAAATTATTACGCTGTTATCCCTAG
<i>M. goulmimensis</i>	OUM		AGTS_cox1_R4	TCAACCTGTTCCTACTCCTCTTTCACTATTCT
		<i>cox1-rrnL</i>	OUM_cox1_F1	ATTTGGAGGGTTTGGTAATGACTATTGGCC
			OUM_16s_R1	TTAATTGGAGACTGGAATGAAAGGTTGAAC
		<i>rrnL-cox1</i>	OUM_16s_F2	TIGTTCAACCTTTTCTCCAGTCTCC
<i>M. goulmimensis</i>	LMG		OUM_cox1_R1	GCAATAGTCAATTACCAAAACCCCTCC
		<i>cox1-rrnL</i>	L10_cox1F3	GTAGAGAGAGGTGTAGGGACGGGG
			L10_16sR1	ATTTGGAACTGGAATGAAGGGTTG
		<i>rrnL-cox1</i>	L10_16sF2	AAAAATTGTTCAACCCCTTCTATCC
<i>M. goulmimensis</i>	ZOM		L10_cox1R3	CACCCCGTCCCTACACCTCTCTC
		<i>cox1-rrnL</i>	ZOM_cox1F3	AAGAAGAATAGTAGAAAGGGGAGCCG
			ZOM_16sR1	TAGGGATAACAGCGTAATAATTTGGATAG
		<i>rrnL-cox1</i>	ZOM_16sF1	TATCCAAAATTATTACGCTGTTATCCC
"Longipodacrangonyx stocki"	HOM		ZOM_cox1R1	CGGCTCCCTTTCTACTATTCTTC
		<i>rrnS-rrnL</i>	Uni_Meta_12sF	TATGTAACTAATCTAATTTTAARTATATCTGCACCT
			Uni_Meta_16sR	ATTTTTTACATGATTTGAGTTCAGACCCG
		<i>rrnS-rrnL</i>	Uni_Meta_12sF	TATGTAACTAATCTAATTTTAARTATATCTGCACCT
"M. paurosexualis"	SKP		Uni_Meta_16sR	ATTTTTTACATGATTTGAGTTCAGACCCG
		<i>rrnS-rrnL</i>	Uni_Meta_12sF	TATGTAACTAATCTAATTTTAARTATATCTGCACCT
			Uni_Meta_16sR	ATTTTTTACATGATTTGAGTTCAGACCCG
		<i>rrnS-rrnL</i>	Uni_Meta_12sF	TATGTAACTAATCTAATTTTAARTATATCTGCACCT
<i>M. boutini boutini</i>	BBM		Uni_Meta_16sR	ATTTTTTACATGATTTGAGTTCAGACCCG
		<i>rrnS-rrnL</i>	Uni_Meta_12sF	TATGTAACTAATCTAATTTTAARTATATCTGCACCT
			Uni_Meta_16sR	ATTTTTTACATGATTTGAGTTCAGACCCG
		<i>rrnS-rrnL</i>	Uni_Meta_12sF	TATGTAACTAATCTAATTTTAARTATATCTGCACCT
<i>M. nicoleae tamri</i>	AKM		Uni_Meta_16sR	ATTTTTTACATGATTTGAGTTCAGACCCG
		<i>cox1-nad1</i>	Met_nd1F1	CAGTATTAATCTGACACAAGCTCTGAACCTCTCTC
			Met_16s_R1	TGAAAAATAGAAAGTATAGCTGCCC
		<i>rrnL-cox1</i>	Met_16sF2	CTATCCAAAATTATTACGCTGTTATCCC
<i>M. longipes</i>	SFM		Met_cox1R2	GAAGAAGTCAGTTACCGAACCCCTCC
		<i>cox1-rrnL</i>	MAM_cox1F3	ATGGTAGAAAGAGGGGTTGGTACAGGATG
			MAM_16sR1	ATTTGTTTTTAATTGAGAAGTGAAGG
		<i>rrnL-cox1</i>	MAM_16sF1	CCTATCCAAAATTATTACGCTGTTATCCC
"M. notenboomi"	MAM		MAM_cox1R5	ACCAACCCCTTCTTCTACCATTTCTAC
		<i>cox1-rrnL</i>	Bah_cox1_F3	GGACAGTTTATCTCCATTAGCAGCAGC
			Bah_16s_R3	GGAGAGATCATATCTATAAAATGATTGCGACC
		<i>rrnL-cox1</i>	Bah_16s_F1	GAGGTGCAATCAATTTTATAGATATGATCTCTCC
<i>Bahadzia jaraguensis</i>	BAH		Bah_cox1_R2	GCTGTAATGAAGACTGACCAAGCAAATAAAGG
		<i>cox1-rrnL</i>	Pni_cox1_F2	CAGACATAGCCCTCCCTCGCATAAAATAAC
			Bah_16s_F1	AAATACTATAATAAAATAAAGTGACGATAAGACCCTATAAGC
		<i>rrnL-cox1</i>	Pni_16s_R1	ATCTTATGCTCATCCATCAATGTAAGTAACTCAC
<i>Pseudoniphargus daviui</i>	PSN	<i>cox1-rrnL</i>	Pni_16s_F1	GTTATTTATGCGAGGGAAGGCTATGTCTG
		<i>rrnL-cox1</i>	Pni_cox1_R1	



## 2.5. Seqüenciació dels mitogenomes (Capítols II i III)

Per a la seqüenciació dels 23 mitogenomes obtinguts en aquesta tesi es seguiren diferents metodologies.

### 2.5.1. Seqüenciació de genoteques pel mètode de Sanger

#### 2.5.1.1. Construcció de genoteques a partir de fragments de DNA amplificats per long-PCR i digerits amb endonucleases (Capítol II)

Els fragments llargs del mitogenoma de *M. longipes* (aprox. 10 kb i 5 kb) es digeriren independentment amb tres endonucleases de restricció (*DraI*, *RsaI* i *TaqI*) segons les especificacions del fabricant. Les quantitats utilitzades per a la digestió foren de 20 µL de DNA purificat (aprox. 1 µg), 2,4 µL de tampó i 1,6 µL (10.5 unitats) d'enzim de restricció. Les digestions dels dos primers enzims s'incubaren tota la nit a 37 °C i a 65 °C per *TaqI*. Després, les digestions s'incubaren a 96 °C durant 10 minuts per inactivar la reacció enzimàtica. Els fragments resultants es comprovaren en gels d'agarosa al 2% i s'observà que produïren un rang de fragments de DNA de 150 pb a 1,5 kb. La resta de les digestions de les tres endonucleases es mesclaren a un sol tub i es aquest es purificà amb el *MiniElute PCR Purification Kit* (Qiagen, Hilden, Alemanya). Tot seguit es lligaren els fragments de DNA purificat dins el vector de clonatge *pJET1.2* (*CloneJET™ PCR Cloning kit*, Fermentas, Glen Burnie, MD, USA) segons les recomanacions del fabricant. Aquest vector confereix a la cèl·lula portadora resistència per a l'antibiòtic ampicil·lina i addicionalment, conté un gel letal situat al lloc de clonatge, el qual es veu interromput en el moment que el producte de PCR s'insereix en el lloc de clonatge. La transformació de les cèl·lules competents amb CaCl<sub>2</sub> de *Escherichia coli* (*One-shot® TOP10 E. Coli cells*, Invitrogen, Madison, WI, USA) es va fer mitjançant la tècnica de xoc tèrmic a 42 °C. Les cèl·lules transformades es seleccionaren en plaques amb medi de cultiu agar-LB que contenia ampicil·lina (100 µg/mL) i s'incubaren tota la nit a 37°C.

Es realitzà un triatge de 96 colònies per PCR per identificar aquelles que haguessin inserit fragments superiors a 300 pb. Aquesta tècnica consisteix en prendre

les colònies i resuspendre-les individualment en 50 µL de tampó TRIS 10 mM i després desnaturalitzar-les a una temperatura de 95 °C durant 10 minuts. Un cop desnaturalitzades, es centrifugaren durant 60 segons a 12.000 rpm per precipitar les restes cel·lulars fragmentades. Seguidament, s'empraren 1 µL de sobrenedant com a DNA motlle per fer una PCR estàndard en la qual s'utilitzaren els encebadors universals pel vector *pJET*. La mida de cada un dels productes amplificats es comprovà en gels d'agarosa al 1,5% i es triaren aquells que presentaven inserts de més de 400 bp. En total s'obtingueren 63 clons positius, els quals foren purificats i seqüenciats en ambdues direccions emprant els encebadors universals pel vector *pJET*. Per visualitzar la cobertura inicial dels 63 clons, es feu servir com a bastida la seqüència parcial del mitogenoma de l'amfípode *Parhyale hawaiiensis* (Cook et al., 2005). Vist que les seqüències obtingudes no cobrien la totalitat del mitogenoma, es dissenyaren encebadors addicionals per cobrir els forats existents i s'empraren els fragments llargs de 10 Kb i 5 kb com a DNA motlle per a les noves reaccions de seqüenciació. La seqüència dels oligonucleòtids dissenyats per completar el mitogenoma de *M. longipes* s'inclouen a la Taula Addicional 1. Totes les reaccions de seqüenciació es feren segons el protocol desenvolupat a l'apartat 2.3.

#### 2.5.1.2. Construcció de genoteques a partir de la fragmentació mecànica (Hydroshear®) de fragments de DNA obtinguts per long-PCR (Capítol III)

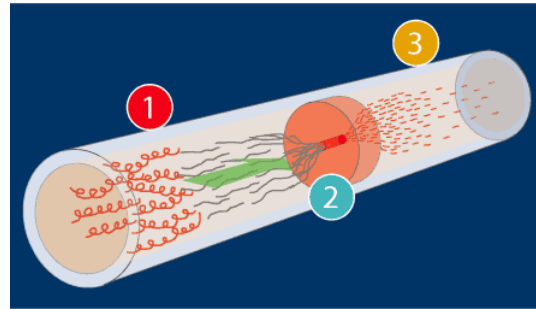
En el cas de les espècies *M. repens*, *M. dominicanus* i *M. samanensis* els fragments amplificats es fragmentaren mecànicament amb el sistema de *Hydroshear*® (GeneMachines® Corporation, San Carlos, CA, USA). Aquest aparell fou dissenyat per fragmentar el DNA a partir de forces hidrodinàmiques i així crear un conjunt de fragments a l'atzar de DNA, independentment de la concentració de DNA, de la longitud del DNA i del volum de la mostra. El DNA passa a través d'una xeringa, la qual disposa d'un orifici estret que presenta un punt de contracció (Fig. 2.2). Depenent de la pressió i de la velocitat de bombeig a la qual es sotmeti el DNA, aquest quedarà més o menys fragmentat. Aquest sistema és per ara el millor ja que permet optimitzar la llargària de les lectures (*reads*) i, al mateix temps, el nombre de clons necessaris per a la construcció de llibreries genòmiques i així facilitar l'obtenció d'un *contig* (=conjunt ordenat de fragments on les seves seqüències solapen parcialment) sense forats i amb una alta cobertura. El principal avantatge és que aquesta tècnica no està esbiaixada tal

---

com succeeix en la digestió enzimàtica, la qual depèn de la localització dels llocs de reconeixement i del nombre de dianes de restricció. Amb aquesta metodologia hom pot triar la mida dels fragments tallats aplicant la pressió adequada. Malauradament, presenta una sèrie de desavantatges: la limitació de la mida dels fragments que pot generar, el tractament laboriós que necessita la preparació de la mostra abans de ser introduïda a l'aparell i la reparació dels extrems dels fragments un cop tallats per poder-ne ser posteriorment clonats.

Per a la fragmentació es partí d'una concentració de 0.1 µg/mL en un volum de 110 µL, és a dir, 10 µg en 100 µL. S'ajustaren els paràmetres per a què els fragments obtinguts fossin al voltant de 1,5 kb (25 cicles a una velocitat de nivell 5).

Després de la fragmentació es dugué a terme una reacció de polimerització per reparar i restaurar els grups fosfats 5' dels extrems dels fragments. Al producte obtingut rere la fragmentació s'hi afegí la següent barreja de reacció de polimerització: 13 µL de NEbuffer 2 10x (New England BioLabs ); 10 µL de dNTPs (2.5mM); 6 µL de la polimerasa de DNA Klenow I (5,000U/mL) (New England BioLabs); 2 µL de la polimerasa T4 (7.9 U/µL, Promega ). La reacció s'incubà a temperatura ambient durant 1h. Per poder visualitzar i recuperar els fragments de mida òptima es realitzà un gel d'agarosa a l'1% i es carregà tot el producte fragmentat, fent-se córrer a 100 mV durant 20 min. Es retallà la taca difusa (*smear*) obtingut al voltant de 1,5 kb, i es purificà amb el kit comercial de columnes d'extracció de banda de Wizard® SV Gel & PCR Clean-Up System (Promega, Madison, USA). Tot seguit es quantificaren els purificats amb un espectrofotòmetre NanoDrop® ND-1000. Per a la reacció de lligació s'utilitzà el kit comercial Fast-Link™ DNA Ligation Kit (Epicentre® Biotechnologies, Madison, USA). La reacció és dugué a terme amb els següents reactius: 1.5 µL de Fast-Link Buffer 10X, 0.75 µL d'ATP (10 mM), 2 µL d'enzim Fast-Link Lligasa (2 U/ µL), 100 µg vector de clonatge, 450 ng de producte tallat purificat i completant amb aigua MiliQ esterilitzada fins arribar a un volum final de 15 µL. Atès que la lligació entre extrems roms és més ineficient que per aquells extrems



**Figura 2.2.** Representació del procés de tall produït dins la xeringa de l'Hydroshear®. (1) DNA abans de fragmentar; (2) el DNA passa a través d'un orifici estret pel punt de contracció que produirà la fragmentació a l'atzar del DNA; (3) diferents fragments de DNA de diferents llargàries. (imatge presa de [www.jpjgi.doe.gov](http://www.jpjgi.doe.gov))

cohesius es deixà que la reacció tingués lloc durant tota la nit a 16° C. Per a la transformació de les cèl·lules competents es va seguir la tècnica d'electroporació. Aquesta tècnica consisteix en aplicar un xoc elèctric a les cèl·lules competents per tal d'augmentar la conductivitat elèctrica i permeabilitat permetent-ne així l'entrada del vector amb l'insert per la membrana temporalment afeblida. Per a la transformació s'empraren les cèl·lules competents de *Escherichia coli* (*One-shot*® TOP10 *E. Coli* cells, Invitrogen, Madison, WI, USA) amb 2 µL de producte lligat i es va fer servir l'electroporador de *BioRad-Micropulser* amb un voltatge de 1,8 kV. Posteriorment, les cèl·lules transformades es sembraren en un medi de agar-LB amb ampicil·lina (100 µg/mL) i s'incubaren tota la nit a 37°C.

L'endemà es realitzà un triatge a l'atzar de 96 de les colònies obtingudes per PCR per identificar-ne aquelles que haguessin inserit fragments de més de 300 bp. Es seqüenciaren els inserts en una sola direcció fins a aconseguir una cobertura del *contig* com a mínim d'un 5x (Carapelli et al., 2008) (Taula 3.8, Cap. III). Malauradament, els clons seqüenciats no van ésser suficients per cobrir la totalitat del mitogenoma pel que es van haver de dissenyar encebadors específics per amplificar i seqüenciar les regions amb forats o de baixa cobertura. Totes les reaccions de seqüenciació es feren segons el protocol desenvolupat a l'apartat 2.4. Les seqüències dels oligonucleòtids dissenyats per complementar els mitogenomes de *M. repens*, *M. samanensis* i *M. dominicanus* es troben a la Taula Addicional 2.

### **2.5.2. Genoteques per PCR en emulsió i piroseqüenciació** (Capítol III)

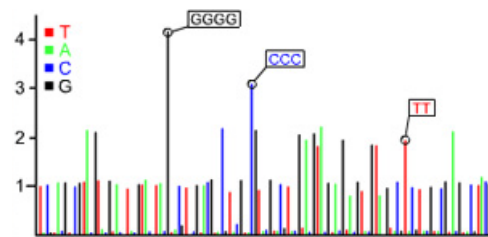
Al començament del segle XXI, en la recerca d'abaratir els costos oferts per la seqüenciació capil·lar del DNA, es desenvoluparen metodologies alternatives basades en la nanotecnologia sorgint així els primers seqüenciadors de DNA en paral·lel d'alt rendiment (*high-throughput*). Entre les diferents tecnologies de seqüenciació de nova generació, la plataforma Roche (454) GS FLX *Sequencer* és sens dubte la que permet llegir els fragments més llargs, fins a 500 bp (Meyer et al. 2008) i des de mitjans de 2011 permet la lectura de fins a 1 kb (<http://my454.com/index.asp>). Aquesta segona generació de seqüenciadors són capaços de generar cents de milers de reaccions de seqüències en paral·lel gràcies a la creació d'una llibreria de DNA a partir d'una modificació de la PCR, coneguda com PCR d'emulsió, i a la immobilització dels

---

fragments amplificats en una superfície sòlida (microesferes d'afinitat específica).

Primer de tot es duu a terme la fragmentació mecànica del DNA en un procés conegut com a nebulització, on el DNA és fragmentat a l'atzar. De forma similar a l'*Hydroshear*<sup>®</sup>, la mida dels fragments obtinguts depèn de la pressió del gas del nebulitzador que condicionarà la velocitat a la qual passa la solució de DNA a través d'un petit orifici. La nebulització és un mètode senzill, ràpid, econòmic i requereix de poca quantitat de DNA. Un cop fragmentat el DNA es reparen els extrems dels fragments i es lliguen a dos adaptadors que es faran servir més endavant com a encebadors per l'amplificació i seqüenciació. Un dels adaptadors conté un marcatge de biotina a l'extrem 5' que permetrà als fragments de DNA unir-se a les microesferes d'esptreptavidina. Un cop s'ha reparat el DNA, els fragments són desnaturalitzats, alliberant-se la cadena sense marcar, quedant d'aquesta manera un fragment de DNA monocatenari, aïllat i immobilitzat a la microesfera. Finalment les diferents microesferes són emulsionades juntament amb els reactius d'amplificació necessaris per què es dugui a terme la PCR

d'emulsió a cada una de les esferes. La piroseqüenciació té lloc sobre una placa especial (*PicoTiter*<sup>™</sup> Plate) que presenta uns diminuts pouets, de 44 micres de diàmetre, als quals sols hi entra una microesfera amb milions de còpies que han estat amplificades i a



**Figura 2.3.** Exemple de pirograma on es veu que el senyal lumínic emès és proporcional al nombre de nucleòtids incorporats.

on s'introdueixen els enzims i els productes per a realitzar la seqüenciació. Aquesta reacció química de seqüenciació consisteix en una tècnica de determinació del DNA a gran escala no fluorescent que mesura l'alliberament de pirofosfat a una reacció de polimerització mitjançant una sèrie de reaccions enzimàtiques acoblades que alliberen llum cada cop que s'incorpora un nucleòtid. El procés de piroseqüenciació es fa de forma paral·lela a cada un dels pous de la placa pel que cents de senyals són llegits ensems. Finalment, aquesta quimioluminescència produeix una imatge que es captada i interpretada per l'ordinador el qual proporciona la seqüència final de nucleòtids. A diferència dels cromatogrames obtinguts per una reacció de seqüenciació convencional on cada pic correspon a un nucleòtid, en el cas dels pirogrames produïts per piroseqüenciació el senyal lumínic emès per la cadena motlle és proporcional al nombre de nucleòtids d'una mateixa base incorporats a la cadena (Figura 2.3).

La resta de mitogenomes s'obtingueren per piroseqüenciació de llibreries a partir dels sistemes de *Roche* FLX/454 o el *GS Junior*. Tots els amplicons de PCR foren quantificats per un espectrofotòmetre *NanoDrop*® ND-1000 i es barrejaren en concentracions equimolars per a cada mostra a una concentració final de 300 ng/μl per fer-se servir com a substrat per a la preparació de la llibreria FLX (454 Life Sciences, Roche). Set dels 19 mitogenomes restants s'obtingueren a partir de la realització d'una llibreria de FLX (etiquetant prèviament per separat cadascuna de les set mostres). Després de la reacció de piroseqüenciació per l'aparell *Roche* FLX/454 s'obtingueren les lectures de 200.000 seqüències corresponents a un octau de la placa *PicoTiterPlate*. El protocol per a la preparació del marcatge i de la llibreria seguit es pot consultar a Meyer et al. (2007). El programa *Newbler Assembler* (Margulies et al., 2005) classificà les seqüències obtingudes prèviament segons els seu etiquetatge (és a dir, per espècimen) i a continuació s'assemblaren les seqüències de cada espècimen per aconseguir el mitogenoma complet. Aquest programa és considerat el millor programa d'assemblatge de seqüències/genomes *de novo* (Kumar & Blaxter, 2010) i està dissenyat específicament per a dades generades per la tecnologia de *Roche* FLX/454.

Per a les 12 espècies restants es va explorar la possibilitat de realitzar una piroseqüenciació *multiplex* sense la necessitat de l'etiquetatge individual a cadascuna de les mostres, abaratint així el cost final (Timmermans et al., 2010). Primer de tot, es comprovà si s'assemblaven correctament les 200.000 lectures obtingudes de les set espècies al procés anterior (amb l'eliminació prèvia de l'etiquetatge específic). Aquesta exploració es va fer amb el programa *CodonCode Aligner v 3.7*. (CodonCode Corporation, Denham, MA, USA) al qual s'hi aplicà un mínim d'identitat (*minimum match*) del 95%. D'aquesta forma s'observà que la divergència trobada als mitogenomes de les diferents espècies era prou alta per reconstruir el mitogenoma sencer a cada una de les espècies sense la formació de quimeres. Aquestes 12 espècies es feren en dues tandes de 12 amplicons (un fragment de 10 kb i un altre de 4.5 kb per mitogenoma) purificats, quantificats i de concentracions equimolars a una concentració final de 100 ng/μL per mitogenoma. Per a cada sis mitogenomes es va construir una llibreria única que es seqüencià a un analitzador *Roche* *GS Junior* el qual donà un total aproximat de 100.000 lectures. Les seqüències parcials de *cox1*, *cob* i *rnl* prèviament amplificades i seqüenciades per la metodologia de Sanger (apartat 2.4.) i encebadors "universals", foren inclosos al conjunt de lectures obtingudes per poder confirmar d'aquesta manera

---

el correcte assemblatge de cada uns dels mitogenomes i saber a quina espècie corresponia a cada *contig*/mitogenoma obtingut (similarment a les “seqüències de pesca” descrites a Timmermans et al., 2010). L’assemblatge es realitzà en el programa *CodonCode Aligner* v 3.7. Tant els assemblatges realitzats amb la plataforma FLX com la GLS es feren basant-se en un mínim de cobertura del 59x (Taula 3.8, Cap. III).

### 3. MÈTODES ANALÍTICS

#### 3.1. Edició de seqüències

##### 3.1.1. Marcadors mitocondrials (*cox1*, *rrnL* i *cob*) i nuclears (*H3* i *SSU*) (Capítol I i III)

Els cromatogrames d’ambdues cadenes obtingudes s’examinaren i editaren emprant el programa *Codon Code Aligner* v.3.7. Ambdós sentits de les cadenes es feren servir per a l’obtenció d’una seqüència consens. L’ortologia de les seqüències es confirmà a partir de cerques a les bases de dades genòmiques mitjançant l’algorisme BLAST. L’absència de codons d’aturada a seqüències codificants de proteïna es confirmaren traduint-les amb el codi genètic apropiat (codi genètic mitocondrial per a invertebrats) i les seqüències amb substitucions no sinònimes es confirmaren rellegant els cromatogrames.

##### 3.1.2. Mitogenomes obtinguts pel mètode de Sanger (Capítol II i III)

Els cromatogrames de les seqüències dels clons i les seqüències obtingudes per PCR per cobrir els forats existents entre els clons es visualitzaren i editaren amb el programa *Codon Code Aligner* v.3.7.

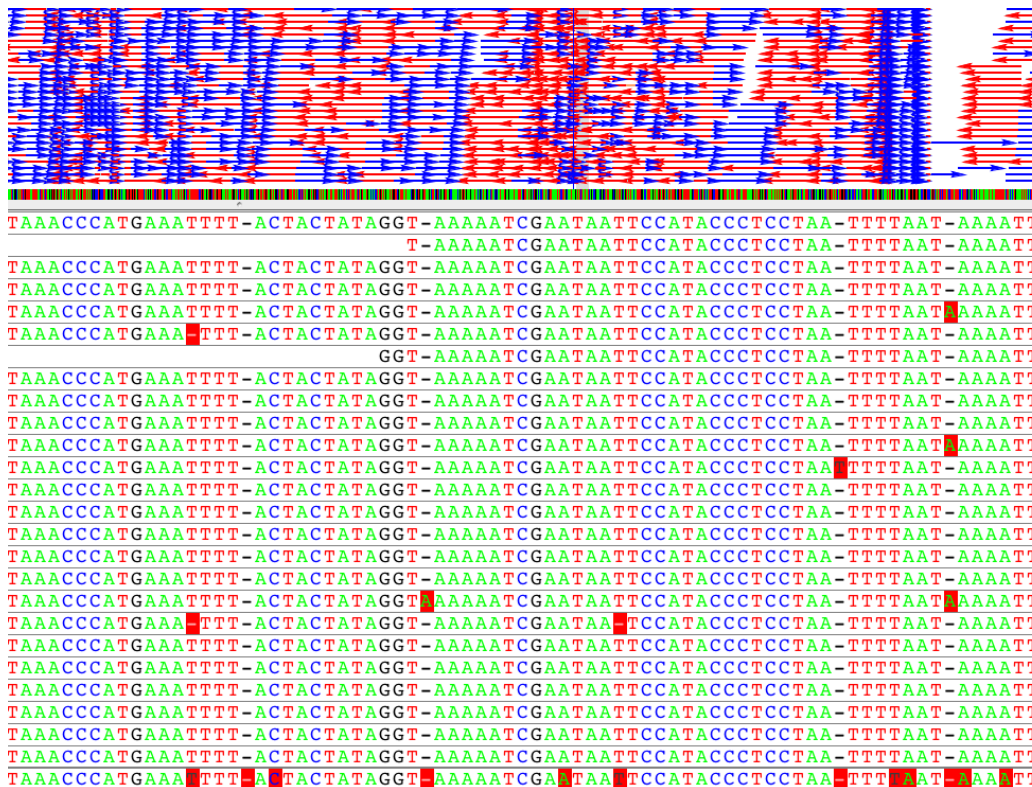
##### 3.1.3. Mitogenomes obtinguts per piroseqüenciació (Capítol III)

Com ja s’ha esmentat anteriorment, per l’assemblatge s’utilitzà el programa *Codon Code Aligner* v.3.7. De les lectures obtingudes per piroseqüenciació solament s’empraren les de més alta qualitat (score > 400) per formar el *contig* complet o gairebé complet de cadascun dels mitogenomes. S’inspeccionaren acuradament aquelles regions que incloïen regions homopolimèriques i/o amb insercions/delecions. En els

---



casos de regions homopolimèriques es decidí el nombre d'A o T d'acord amb la seva freqüència tot i que en alguns casos el nombre d'A o T es decidí segons el marc de lectura pels gens codificants per a proteïnes (PCGs) o segons fos el cas, en virtut de l'estructura secundària dels gens ribosomals de transferència (tRNAs) (Figura 2.4). Els detalls de la longitud mitjana de les seqüències, nombre i qualitat de les lectures de les seqüències i cobertura per a cadascuna de les espècies es troba indicat a la Taula 3.8, Cap. III.



**Figura 2.4.** Diferents reads analitzats pel mitogenoma de *M. goulmimensis* amb el programa Codon Code Aligner. Observi's com per cadascuna de les posicions ambigües (ombrejat en vermell) de la seqüència consens (última fila) es pot esbrinar el nucleòtid correcte d'acord amb la freqüència en la qual apareix als reads.

### 3.2. Anotació dels mitogenomes (Capítol II i III)

La pre-identificació i pre-anotació dels gens mitocondrials com els PCGs, els gens codificants per RNA ribosomal (rRNA) i els gens codificants per a RNA de transferència (tRNA) es realitzà mitjançant l'eina DOGMA (de l'anglès *Dual Organellar GenoMe Annotator*; <http://dogma.ccbb.utexas.edu/>) (Wyman et al., 2004). DOGMA és la primera eina informàtica que permet l'anotació automàtica dels genomes mitocondrials i cloroplàstics, proporcionant endemés, una interfície gràfica per a la visualització i l'edició de les anotacions. Pels PCGs, el mitogenoma es traduït als sis marcs de lectura diferents i es realitza una cerca BLASTX (cerca de seqüències



aminoacídiques) amb la base de dades aminoacídica disponible per a mitogenomes. Aquests gens són identificats en base a la seva similitud (*e values*) i a la presència de llocs altament conservats. Així mateix, aquest programa també localitza estructures secundàries altament conservades de rRNAs i tRNAs en animals, mitjançant algorismes BLASTn i de covariació (Eddy & Durbin, 1994) respectivament. Per últim, DOGMA construeix una llista de gens possibles a partir dels resultats de les tres diferents cerques. Un cop identificats tots els gens mitocondrials amb aquest sistema es procedí a fer-ne la seva anotació d'una manera més acurada segons la naturalesa dels gens. Les seqüències completes dels mitogenomes foren depositades a la base de dades de l'EMBL-Bank (Taula 3.8, Cap. III).

### 3.2.1. Anotació de PCGs (Capítols II i III)

La identificació dels extrems dels PCGs no és una tasca trivial. No tots els gens es troben delimitats per un codó d'inici Metionina (ATG/ATA) (Smith & Marcker, 1968) o Isoleucina (ATT/ATC) a l'extrem 5' i un codó d'aturada (TAA/TAG) a l'extrem 3'. A la literatura existeixen nombrosos exemples de casos on els PCGs comencen per inusuals codons d'inici - com Leucina (TTG/TTA/CTT/CTA), Valina (GTT/GTG), etc.- o bé acaben amb codons d'aturada truncats (Taa o TAA). De fet s'han proposat altres codons d'inici alternatius com ATN, GTG, TTG pel mitogenoma a invertebrats (Wolstenholme, 1992). Totes aquestes variacions dificulten molt la comesa de la delimitació dels extrems i el fet de tenir només una seqüència si escau, ho complica encara més. Per aquest motiu, normalment és requereix de la comparació d'espècies congenèriques o properes per fer més senzill la detecció dels extrems dels gens.

Normalment s'assumeix que el codó d'inici correcte és el més proper a aquell "consens" en comparació de la resta de seqüències homòlogues d'espècies properes i evitant, sempre que sigui possible, el solapament a l'extrem 5' amb altres gens (Boore et al., 2005).

Per a gens amb codons d'aturada complets (TAA o TAG) que no solapin amb els gens situats a l'extrem 3' i amb una longitud similar a altres gens homòlegs, no hi ha dubte de la seva delimitació. En molts de casos emperò, s'hi troben al seu lloc codons d'aturada incomplets o truncats (Taa o TAA) que són enzimàticament processats i completats per poliadenilació post-transcripcionalment (Ojala et al., 1980).

---

La detecció d'aquests codons d'aturada no resulta una tasca senzilla i en aquestes ocasions és crucial poder comparar diferents seqüències homòlogues ja anotades prèviament. D'aquesta manera és més senzill esbrinar si és més versemblant allargar la proteïna fins a trobar un codó d'aturada (que a més solapa amb altres gens adjacents situats en direcció 3') o en suposar un codó d'aturada truncat compatible amb la longitud dels gens homòlegs comparats d'espècies congenèriques.

En el cas de *M. longipes*, que resultà ésser el primer mitogenoma seqüenciat i anotat en aquesta tesi, els extrems 5' i 3' dels PCGs se detectaren manualment per comparació als PCGs alineats, anotats i publicats d'altres crustacis i hexàpodes (Taula Addicional 3). Per a la resta de mitogenomes analitzats en aquesta tesi, els diferents extrems 5' i 3' dels PCGs s'identificaren manualment per comparació als alineaments de PCG anotats a altres espècies de crustacis peracàrides, les quals foren descarregades de *MitoZoa* (<http://mi.caspur.it/mitozoa/>) (Taula addicional 4). *MitoZoa* és una base de dades especialitzada, acurada i actualitzada que posseeix gairebé tots els mitogenomes complets de Metazoa fins aleshores publicats.

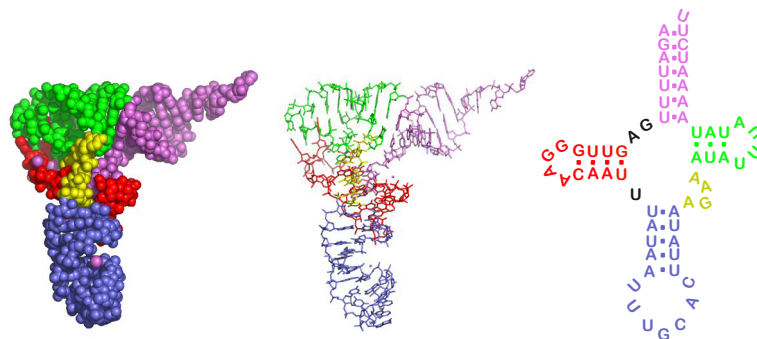
Quant als PCGs es va comprovar la total absència de *numts* (de l'anglès, *nuclear mitochondrial DNA*; són gens que han sigut transferits al genoma nuclear) o codons d'aturada (*stop codons*) al mitogenoma.

### 3.2.2. RNA ribosomal (rRNA) i de transferència (tRNA) (Capítols II i III)

Els gens de tRNA i rRNA són gens essencials pels processos biològics considerant que són les molècules claus a la traducció del RNA missatger a proteïna. Els nucleòtids constituents de la seqüència de RNA solen formar estructures secundàries, on una part de la cadena d'RNA s'aparella amb altres nucleòtids de la mateixa molècula, causant el plec del RNA amb si mateix. Aquests aparellaments vénen a formar estructures de cadena doble (*stems* o braços), els quals solen estar més conservats que les seqüències de cadena senzilla (*loops* o bucles) que són altament variables. El joc d'aquests aparellaments es coneix amb el nom d'estructura secundària. A més, els *stems* presenten mutacions compensatòries com a resposta a una mutació apareguda al nucleòtid parella que permeten restablir les unions i per tant, aquestes estructures secundàries.

---

El RNA de transferència es compon d'uns 60-95 nucleòtids i és ubic a tots els organismes. Tots els tRNAs es caracteritzen per presentar zones de seqüència nucleotídica de complementarietat intracatenària (braços) altament conservades, les quals adopten una estructura secundària en forma de fulla de trèvol característica a quasi tots els tRNAs (Figura 2.5). Una vegada es produeixen els plecs necessaris en la conformació de l'estructura secundària aquesta molècula passa a tenir la forma de "L", en moltes ocasions representada en orientació "Γ". Aquesta configuració és el resultat de l'amuntegament del braç acceptor amb el braç TΨC i del braç D amb el braç anticodó (Dirheimer et al. 1995).



**Figura 2.5.** Estructura secundària adoptada per la majoria de tRNAs -en aquest cas del tRNA-Ala de *M. longipes*-. Els distints colors indiquen els diferents braços-bucles característics d'aquesta molècula: braç acceptor (lila), braç anticodó (blau), braç extra de longitud variable (groc), braç D (vermell) i braç TΨC (verd). Imatge modificada de <http://chemistry.umeche.maine.edu>.

Gran part del coneixement que es té actualment sobre l'evolució mitocondrial és gràcies a l'anàlisi de mitogenomes complets per reconstruir les relacions filogenètiques de molts de grups d'organismes (Masta, 2000). S'ha observat que el codi genètic als mitocondris de metazous es troba altament modificat, el que es tradueix en un insòlit ús de codons d'inici (Wolstenholme, 1992). Tal i com s'ha fet menció anteriorment, el DNA mitocondrial és altament susceptible d'acumular mutacions, i els tRNAs mitocondrials no en són cap excepció (Lynch, 1996). De fet, és freqüent trobar als tRNAs mitocondrials de metazous estructures secundàries atípiques que destorben greument el seu alineament posterior. S'ha observat que certs transcrits de tRNAs a diversos grups de metazous no tenen la típica estructura secundària canònica en forma de fulla de trèvol, mancants fins i tot el braç TΨC (Wolstenholme et al., 1987; Dirheimer et al., 1995; Keddie et al., 1998; Masta, 2000). Avui dia encara es van descobrir nous mecanismes que permeten l'acumulació d'aquestes mutacions aparentment deletèries

als tRNAs i que encara conserven la seva funcionalitat. S'ha suggerit que un possible mecanisme seria que els transcrits de tRNA siguin posteriorment "editats", tant és així que els aparellaments erronis (*mismatches*) que es donen a regions dels braços podrien ésser postranscripcionalment corregits (Börner et al., 1997). Per totes aquestes raons no és gens recomanable confiar exclusivament en la detecció i alineament de tRNAs d'acord amb la seqüència primària consens. Durant els darrers anys s'han desenvolupat diverses eines bioinformàtiques que tenen en compte els elements conservats de les estructures secundàries d'aquestes molècules per la seva detecció i també pel seu alineament posterior.

Per a la identificació dels tRNAs en aquesta tesi, es realitzà una primera cerca de tRNAs putatius amb el programa *tRNAscan-SE* 1.21 (Fichant & Burks, 1991; Lowe & Eddy, 1997; Schattner et al., 2005; <http://lowelab.ucsc.edu/tRNAscan-SE/>). Aquest programa utilitza models probabilístics de covariació (*covariance models*) (Eddy & Durbin, 1994; Lowe & Eddy, 1997) que tenen en compte tant l'estructura primària com les restriccions de l'estructura secundària per a la detecció de gens homòlegs. Aquests models de covariància es basen en alineaments múltiples de seqüències i poden capturar la informació consens d'estructures primàries i secundàries disponible a la base de dades. Malauradament, alguns dels gens de tRNA no van ésser detectats amb *tRNAscan-SE* 1.21, pel que van haver de ser localitzats manualment a partir de la cerca de aparellaments de bases que poguessin indicar estructures secundàries típiques dels tRNA i de possibles anti-codons (Kumazawa & Nishida, 1993).

Els gens ribosomals (rRNA) foren localitzats amb *DOGMA* mitjançant un algoritme BLASTn, mentre que els extrems dels gens s'identificaren tenint en compte l'homologia de la seqüència amb altres rRNAs de crustacis peracàrides –disponibles a la base de dades de *MitoZoa* (Taula Addicional 4)-, l'estructura secundària proposada per altres crustacis (Crease, 1999), i els extrems dels gens adjacents. En tots els mitogenomes analitzats, a excepció dels grups externs, els gens *Val* i *Leu1* ajudaren a l'anotació dels extrems 5' i 3' del gen *rrnL*; mentre que els gens *Val* i l'inici de la regió control (inici de regió poliT i poliA) facilitaren l'anotació del gen *rrnS*.

### 3.3. Alineament

Per a l'alineament múltiple es trià la versió online del programa *MAFFT* 6 (de

---

l'anglès *Multiple Alignment with Fast Fourier Transform*) (Kato et al., 2002, 2005, 2009; Kato & Toh, 2010; <http://align.bmr.kyushuu.ac.jp/jp/mafft/online/server/>) atès que és un dels mètodes d'alineament múltiple (o MSA de l'anglès *Multiple Sequence Alignment*) més ràpids i precisos dels que existeixen actualment (Grasso & Lee, 2004; Golubchik et al., 2007). Aquest mètode incorpora una sèrie d'algorismes basats en FFT (de l'anglès *fast Fourier transform*), el qual permet una ràpida detecció dels segments homòlegs entre les diferents seqüències. Per a l'alineament de tots els marcadors a excepció dels ribosomals (*rrnL* i *SSU*) s'utilitzà el procés d'alineament progressiu (Feng & Doolittle, 1987) incorporat a MAFFT. Aquest mètode a MAFFT comprèn tres etapes (Kato et al., 2005):

- i. *Alineament progressiu-1*: MAFFT construeix primer de tot un alineament a parells i alhora elabora un arbre guia a partir de la matriu de distàncies mitjançant el mètode UPGMA (de l'anglès *Unweighted Pair Group Method with Arithmetic Mean*) (Sneath & Sokal, 1973). Les seqüències es van alinear progressivament (Feng & Doolittle, 1987) seguint l'ordre establert pel patró de ramificació de l'arbre guia. Aquest procés es refereix com a FFT-NS-1.
- ii. *Alineament progressiu-2*: es calcula per segona vegada una matriu de distàncies basada en la informació obtinguda del primer alineament FFT-NS-1 (Edgar, 2004), i es torna a construir un alineament progressiu en base al nou arbre guia generat a partir de la nova matriu de distàncies. Aquest mètode es refereix com a FFT-NS-2.
- iii. *Refinament interactiu*: l'alineament FFT-NS-2 és millorat pel mètode de refinament interactiu (Gotoh, 1996) que maximitza el valor WSP (de l'anglès *weighted sum of pairs*). Aquest procés es refereix com a FFT-NS-i.

No obstant això, aquest mètode de refinament interactiu no té en compte la possibilitat que les seqüències no siguin globalment alineables, com succeeix amb els gens no codificants. L'ús de RNAs funcionals no codificants (ncRNAs) ha portat al desenvolupament de nous algorismes de refinament que sí tenen en compte l'estructura secundària a l'hora de construir un MSA.

---

A *MAFFT v. 6* (Kato et al., 2008; 2009) s'inclou una nova estratègia d'alineament anomenada Q-INS-i (<http://mafft.cbrc.jp/alignment/software/source66.html>; Kato & Toh, 2008; Letsch et al., 2010), una eina que incorpora la informació estructural de les seqüències en forma de probabilitat d'aparellament de bases calculada per l'algorisme de McCaskill o el de *CONTRAFold* (McCaskill, 1990 & Do, et al., 2006). A diversos estudis de marcadors ribosomals s'ha comprovat que la implementació d'aquests algorismes produeixen millors resultats que no pas els basats exclusivament amb mètodes de seqüències, i a més, presenten un bon rati d'exactitud/temps computacional (Edgar & Batzoglou, 2006; Letsch et al., 2010; Kornobis et al., 2011; veure resultats presentats a <http://mafft.cbrc.jp/alignment/software/eval/rna.html>). Cal dir emperò, que tot i augmentar l'eficàcia i fiabilitat dels alineaments, l'ús d'aquests algorismes no és suficient en casos d'alineament entre gens estructurals altament divergents. Vist que tant els gens estructurals com els ribosomals presenten zones altament conservades (*stems*) i per tant fàcilment alineables, mentre que d'altres són molt divergents i es troben replets d'*indels* i *multiple hits (loops)*, s'ha vist que és recomanable eliminar les regions extremadament divergents de l'alineament (Swofford et al., 1996; Lee, 2001; Castresana, 2000). En altres estudis s'ha demostrat que l'eliminació d'aquestes regions problemàtiques permet millorar els arbres filogenètics tot i la pèrdua d'informació ja que augmenta notablement el senyal filogenètic (Talavera & Castresana, 2007).

### 3.3.1. Alineament dels PCGs

La similitud o identitat (hipòtesi primària d'homologia) entre PCGs s'identifica millor a nivell aminoacídic atès que aquestes evolucionen més lentament perquè la majoria de substitucions nucleotídiques són sinònimes. És a dir, no es produeix canvi aminoacídic ja que la majoria de substitucions tenen lloc a les terceres posicions del triplets (codons), i com el codi genètic és degenerat diferents triplets codifiquen pel mateix aminoàcid. Per aquest motiu, els alineaments que tenen en compte la seqüència aminoacídica dels PCGs permeten comparar seqüències més divergents. A més a més, aminoàcids diferents tot i que amb característiques fisicoquímiques similars, són tractats com a quasi-identitats. Un altre gran avantatge d'aquests alineaments d'aminoàcids (per triplets) és que eviten la introducció d'insercions i delecions dins un codó, respectant així la pauta de lectura. Cadascun dels PCGs dels mitogenomes

---

(MPCGs) més el gen nuclear *H3* van alinear-se d'acord amb les seves seqüències aminoacídiques a partir del programa *TranslatorX* (Abascal et al., 2010; disponible a <http://translatorx.co.uk>; Katoh & Toh, 2008). Aquest programa realitza un retroalineament segons sigui la traducció aminoacídica. *TranslatorX* permet l'ús de diversos programes d'alineament múltiple entre els quals es va triar MAFFT 6 per la realització d'aquesta anàlisi.

Finalment es revisaren manualment tots els alineaments amb el programa *MESQUITE*® v. 2.5. (Madison & Madison, 2008 <http://mesquiteproject.org>.) per descartar la presència de pseudogens, codons d'aturada, substitucions no sinònimes poc freqüents i altres errades en l'edició.

Els 13 MPCGs corresponents als mitogenomes sencers es van concatenar emprant el programa *Phyutility* v.2.2. (Smith & Dunn, 2008) fins a obtenir un alineament únic de 11,142 caràcters per a la matriu que contenia els grups externs i 11,076 caràcters per a la matriu amb els 21 mitogenomes sense els grups externs, equivalent aproximadament a un 77% del mitogenoma.

### 3.3.2. Alineament dels gens ribosomals (*rrnL* i *SSU*)

Per a l'alineament del gen mitocondrial *rrnL* i del gen nuclear *SSU* s'utilitzà l'estratègia Q-INS-i que considera les estructures secundàries presents als ncRNA implementada a *MAFFT 6* (Katoh & Toh, 2008). Un cop obtingut l'alineament, les regions hipervariables i blocs ambiguament alineats es varen eliminar amb el programa *Gblocks* v. 0.91b (Castresana, 2000) per estimar el seu impacte sobre la filogènia. Aquest programa elimina aquelles posicions altament divergents i per tant, probablement mal alineades. S'empraren els paràmetres més relaxats i com a resultat s'obtingueren 49 blocs que contenien un màxim de 2,159 caràcters (abans 2,882) per seqüència.

### 3.4. Anàlisis poblacionals (Capítol I)

Totes les anàlisis poblacionals de *M. longipes* es realitzaren amb el set de dades complet per a *cox1* (Taula 3.2, Cap I).

Per comprovar l'existència d'aïllament causada per distància geogràfica es

---



realitzà un test de Mantel (Mantel, 1967; Mantel & Valand, 1970) que es troba implementat al programa *ZT* (Bonnet & Van der Peer, 2002). Aquesta prova s'utilitza per calcular la correlació entre dues matrius (en aquest cas distància genètica vs. distància geogràfica). Atès que les distàncies no són totalment independents unes de les altres, no és factible avaluar sols el coeficient de correlació existent entre les dues matrius i la seva significació estadística. El test de Mantel ha resolt aquest problema a partir d'un test de permutació on la hipòtesi nul·la diu que els valors de la matriu *A* són independents dels valors de la matriu *B* pels mateixos objectes. Per testar-la es realitza un test de permutació on es compara el valor original de l'estadístic amb la distribució trobada un cop es canvia l'ordre a l'atzar (permutació) dels objectes que componen una de les matrius.

Els nivells de diversitat intrapoblacional s'han determinat d'acord a la diversitat nucleotídica ( $\pi$ ) i a la diversitat haplotípica ( $H_d$ ). En primer lloc,  $\pi$  correspon al número mig de diferències entre parells de bases per lloc entre dues seqüències. És un paràmetre que indica el grau mitjà de divergència nucleotídica entre els individus presents a la població i es pot estimar com (Nei, 1987):

$$\pi = \frac{\sum_{i=1} p_i p_j d_{ij}}{L}$$

On  $d_{ij}$  és una estimació del nombre de substitucions nucleotídiques entre els haplotips  $i$  i  $j$ ,  $p_i$  és la freqüència de l'haplotip  $i$  i  $L$  és la longitud en parell de bases de la seqüència analitzada. El valor de  $\pi$  va de 0 a 1.

$H_d$  és el paràmetre que indica la probabilitat de què dos haplotips agafats a l'atzar dins una població siguin diferents entre ells. Es calcula de la següent forma:

$$H_d = \frac{(1 - \sum x_i^2)n}{n - 1}$$

on  $x_i$  és la freqüència d'un haplotip i  $n$  la mida mostral (Nei, 1987). Es seu valor va de 0 a 1. Per exemple, a una població monomòrfica el seu valor seria 0, mentre que a una població on el nombre d'haplotips diferents se correspon al nombre d'individus el valor de  $H_d$  seria 1.

---



Els índex de diversitat poblacional, el nombre d'haplotips,  $H_d$  i  $\pi$ , i les distàncies  $F_{ST}$  i llur significació estadística basada en 10.000 permutacions s'obtingueren amb el programa *ARLEQUIN* v. 3.01 (Schneider et al., 1997; Excoffier et al., 2005). S'excloueren de les anàlisis aquelles poblacions que només eren representades per un individu.

L'anàlisi espacial de variància molecular es realitzà amb el programa *SAMOVA* v. 1.0. (*Spatial Analysis of MOlecular Variance*, Dupanloup et al., 2002) amb el qual s'identificaren les agrupacions geogràficament homogènies que maximitzaven la variància genètica entre els grups de poblacions ( $F_{ST}$ ). Aquest programa implementa el mètode d'AMOVA (*Analysis of MOlecular Variance*; Excoffier, et al., 1992) que es fa servir per calcular la variació molecular dins d'una espècie o llinatge, basant-se en un model jeràrquic o niat. A diferència del mètode d'ANOVA tradicional, aquí la hipòtesi nul·la utilitza mètodes de permutacions que no requereixen la suposició d'una distribució normal. Per calcular la variació molecular dins d'un llinatge, aquest mètode calcula els estadístics de  $F$  que són paràmetres d'estructura gènica (Wright, 1951, Excoffier et al., 1992) -variació entre poblacions dins de grups ( $F_{SC}$ ), variació dins de poblacions ( $F_{ST}$ ) i variació entre grups ( $F_{CT}$ )- i identifica el nombre òptim de grups de poblacions per a un set de poblacions mostrejades segons una distribució geogràfica específica. Els estadístics  $F$  permeten l'anàlisi de l'estructura a poblacions subdividides en grups. A més, també poden emprar-se per mesurar la distància genètica entre els diferents grups, un concepte que es fonamenta en la idea que aquells grups que no presentin aparellament entre ells tindran freqüències al·leliques diferents a les de la població total. Per a aquesta anàlisi s'utilitzaren 100 processos de simulació anellats per a cada valor de  $K$ , des de  $K = 2$  a  $K = 20$ . Els test de neutralitat permeten saber si un marcador particular evoluciona baix el cànon de la teoria neutral. És a dir, segons les prediccions del model neutral, en absència de selecció, el polimorfisme d'un marcador dins d'una espècie i la seva divergència entres espècies són proporcionals a la taxa de mutació neutra. La possible associació no atzarosa entre els llocs polimòrfics dins del *cox1* es va avaluar amb dos estadístics:  $F_S$  de Fu (Fu, 1997) que fou calculat amb *ARLEQUIN* v. 3.01, i el paràmetre  $R_2$  (Ramos-Onsins & Rozas, 2002), calculat amb el programa *DnaSP* v. 5.10.1 (Librado & Rozas, 2009) realitzant 10,000 replicats i assumint que no hi havia recombinació. L'estadístic  $F_S$  empra la informació obtinguda de la distribució dels haplotips, detectant un excés en el nombre d'al·lels estranys o nous a les poblacions a causa d'una expansió poblacional. L'estadístic  $R_2$  es basa en la

diferència existent entre el nombre de mutacions úniques (*singleton mutations*) i la mitjana de les diferències nucleotídiques. Estudis de simulació han demostrat que els estadístics  $R_2$  i  $F_S$  són els més adients a l'hora de detectar creixement poblacional en comparació d'altres tests (sempre que no hi hagi recombinació); concretament  $R_2$ , resultaria millor per a sets de dades petits mentre que  $F_S$  seria millor per a sets de dades més grans (Fu, 1997; Ramos-Onsins & Rozas, 2002).

### 3.5. Freqüències nucleotídiques i nombre efectiu de codons (Capítol II)

Tal com ja s'ha esmentat a l'apartat 6.1 de la Introducció, la composició nucleotídica dels mitogenomes varien notablement entre diferents tàxons d'animals. A més, dins el mateix mitogenoma existeixen també diferències en la composició nucleotídica entre els dos filaments que conformen aquesta molècula, el que es coneix com a "asimetria composicional nucleotídica entre els diferents filaments" (*Strand Asymmetry Compositional bias*). Existeixen diversos estadístics que s'empren per descriure i caracteritzar els patrons de distribució de composició nucleotídica als mitogenomes, tal com els estadístics de biaix en GC, biaix en AT, contingut en G+C, i per tant i complementari al darrer, contingut en A+T (Perna & Kocher, 1995). Les primeres i segones posicions dels codons dels PCGs es troben subjectes a selecció a nivell de seqüència aminoacídica a la proteïna resultant, i per aquest motiu aquestes posicions no són bones candidates com a estimadores per estudiar els patrons composicionals (Perna & Kocher, 1995). Tot i que les terceres posicions (posició degenerada) serien en un principi lliures de les pressions selectives a nivell d'especificació aminoacídica, fóra bo esmentar aquí que la composició d'aquestes posicions també es troba afectada per la composició global del mitogenoma i que moltes terceres posicions no són totalment lliures per alternar entre tots els nucleòtids per mor de l'existència d'aminoàcids que es troben solament representats per dos i no quatre codons diferents (Wei et al., 2010). A més a més, els dos tipus de parells de bases, A-T per una banda i G-C per l'altra, poden presentar comportaments diferents i no estar necessàriament relacionats directament. Hassanin i col·laboradors (2005) observaren que valors absoluts en el valor de biaix en GC eren sempre més alts que aquells valors de biaix en AT a totes les posicions dels codons, pel que suggeriren que l'asimetria en la composició nucleotídica als diferents filaments quedava millor reflectida en els valors de biaix en GC. Aquest fet ha sigut comprovat també

---

recentment pel treball de Kilpert i col·laboradors (2012), els quals observen que els valors en biaix en GC són millors indicadors del biaix existent entre les freqüències nucleotídiques que no pas els valors de biaix en AT.

Les freqüències aminoacídiques i nucleotídiques dels MPCGs per a totes les posicions i de forma independent per a cada una de les tres posicions del codó es calcularen amb el programa *DAMBE* v.5.2.28 (*Data Analysis and Molecular Biology and Evolution*, Xia & Xie, 2001). Aquest paquet conté una sèrie de programes per a l'anàlisi descriptiva i comparada de dades moleculars, incloent seqüències nucleotídiques i aminoacídiques, freqüències al·lèliques o matrius de distàncies entre d'altres. *BioEdit* v.7.0.8. (Hall, 1999) s'utilitzà a més pel càlcul dels valors en biaix en AT i GC.

Des de que Wright (1990) ideà el concepte de nombre efectiu de codons (d'hora endavant ENC, de l'anglès *Effective Number of Codons*) com a mesura del nombre de codons emprats (*codon usage*) en un gen o espècie, aquest valor ha estat emprat com estimador del biaix de codons sinònims (*synonymous codon usage bias*), o sia, com a quantificador del grau de desviació de la igualtat d'ús de codons sinònims dins gens o genomes. Aquesta mesura s'ha denominat també  $N_c$  i s'ha observat que varia considerablement segons l'organisme (Grantham et al., 1980). El valor d'ENC pot anar des del més baix (20) que indicaria que solament s'empra un codó per a cadascun dels vint aminoàcids possibles assenyalant un fort biaix en l'ús de codons, fins al més alt (61) que indicaria que s'empren tots els codons possibles per a codificar els aminoàcids. S'ha observat que el biaix en l'ús de codons és deu a diversos factors biològics: abundància de tRNAs determinats, biaix mutacional específic de cada cadena del mitogenoma, nivell d'expressió gènica, longitud gènica, composició aminoacídica, contingut en G+C, etc. (Wan et al., 2004; Sueoka & Kawanishi, 1999; Blake et al., 2003). En moltes espècies, aquesta mesura ha contribuït a una "selecció translacional" ja que els gens que es troben altament expressats tendeixen a emprar aquells codons que corresponen als tRNAs més abundants (Ikemura, 1981, 1985; Gouy & Gautier, 1982). Per aquesta raó, es diu que en aquest tipus d'espècies, on l'ús de codons es troba esbiaixat, els mètodes per calcular o analitzar el biaix en l'ús de condons sinònims són eines de gran importància en biologia evolutiva (Fuglsang, 2008). Els patrons d'ús de codons també varia entre gens del mateix genoma, i s'ha observat que la relació del contingut en G+C a les terceres posicions sinònimes dels codons (GC3) amb l'ENC és una mesura molt efectiva per explorar aquesta heterogeneïtat (Wright, 1990). El valor

---

d'ENC i el valor de GC3 (només calculant el contingut en GC als codons que són *four-fold*, és a dir, que tenen les quatre possibles degeneracions a la tercera posició) es determinà mitjançant el programa *INCA* v.1.20 (INteractive Codon usage Analysis) (Supek & Vlahovicek, 2004). El patró d'ús de codons a les posicions sinònimes dels MPCGs s'examinaren a partir de la gràfica ENC-plot (Wright, 1990), la qual compara la relació d'ENC vs. GC3.

### 3.6. Divergència als PCGs del mitogenomes anotats (Capítol II)

Les divergències nucleotídiques dels MPCGs individuals per a *M. longipes* s'estimaren a partir de comparacions a parells entre els mitogenomes complets de crustacis publicats fins aleshores i posteriorment es compararen amb els valors obtinguts per a 35 espècies que representaven la majoria d'ordres d'Hexapoda dels quals hi havia informació disponible (Taula Addicional 3). S'emprà el programa *MEGA* v. 4.0.2 (*Molecular Evolutionary Genetics Analysis*) (Tamura et al., 2007) pel càlcul de les distàncies corregides a partir del model MCL (*Maximum Composite Likelihood*) (Tamura et al., 2004) i una taxa de variació entre llocs a partir de la distribució gamma amb el paràmetre *G* (*shape parameter*) de 0.40 estimat en *RAxML* v. 7.2.4. (de l'anglès *Randomized Accelerated Maximum Likelihood*) (Stamatakis et al., 2005). Les posicions *gap* no foren considerades a la comparació a parells. Els valors de les mesures de divergència van ser normalitzades dividint el valor obtingut per a cada gen pel valor del gen amb la taxa més alta.

La nova versió del software *MEGA* v. 4.0.2. inclou el mètode MCL (*Maximum Composite Likelihood model*) per estimar les distàncies evolutives ( $d_{ij}$ ) entre múltiples seqüències de DNA. Amb aquesta aproximació, el logaritme compost de versemblança (CL, *Composite Log Likelihood*) obtingut a partir de la suma del logaritme de versemblança per a tots els parells de totes les seqüències en un alineament, és maximitzat gràcies a l'ajust dels paràmetres comuns per a cada patró de substitució nucleotídica ( $\theta$ ) per a cada parell de seqüències ( $i,j$ ):

$$CL = \sum_{ij} \ln l(\theta, d_{ij})$$

Aquest mètode difereix de la resta de mètodes que estimen distàncies

evolutives en què cada distància evolutiva no és estimada independentment de les altres (*independent estimation approach*, IE) (Tamura et al., 2007). El mètode MLC ofereix una sèrie d'avantatges front als de tipus IE. Per començar, per estimar distàncies evolutives per a cada parell de seqüències els mètodes IE solen produir grans errors a no ser que s'utilitzin amb seqüències molt llargues. L'ús de models MLC redueix aquests errors considerablement, ja que cada sèrie de paràmetres estimat a partir de tots els parells de totes les seqüències és aplicat a l'estimació de cada distància. Quan les distàncies són estimades amb errors més baixos, s'espera que els mètodes basats en distàncies per inferir filogènies resultin ésser més precisos. Tot i que les topologies obtingudes pels diferents mètodes coincideixin, l'ús de distàncies obtingudes amb models MCL solen donar valors de *bootstrap* més elevats en comparació dels que han emprat models IE (Tamura et al., 2007). Un altre avantatge d'aquest model és que els models IE no sempre poden aplicar-se quan les seqüències són evolutivament molt distants, o quan el set de dades és molt gran perquè el patró de substitució es converteix cada cop en més complicat (Tamura et al., 2004). L'ús del mètode MCL evita aquest problema de forma efectiva i permet l'ús de models sofisticats per a la inferència de filogènies a partir d'un gran nombre de seqüències àdhuc per aquelles que són evolutivament molt distants.

La implementació d'aquest model MLC permet considerar la taxa de variació de substitució de lloc a lloc, seguint una distribució gamma, així com la incorporació de l'heterogeneïtat de la composició nucleotídica en diferents espècies/seqüències (Tamura et al., 2007).

### 3.7. Anàlisi de reordenació gènica (Capítol II)

S'emprà la versió online del programa CREx (*Common interval Rearrangement Explorer*, <http://pacosy.informatik.uni-leipzig.de/crex>) (Bernt et al., 2007) per deduir els possibles escenaris de reordenació gènica al mitogenoma de *M. longipes* (el primer obtingut) vs. Pancrustacea i Peracarida, segons la detecció de "arbres amb intervals comuns d'ordre gènic" (*strong interval trees*, SITs) (Bérard et al., 2007), que s'obtenen a partir de l'anàlisi de l'estructura o l'ordre que presenten els diferents gens en el genoma estudiat. L'ordre gènic per a tots els mitogenomes de *Metacrangonyx* resultà el mateix, pel que no es repetí aquesta anàlisi un cop obtingudes la resta de seqüències.

Concretament, un interval comú (*common interval*) és un subconjunt de gens que apareixen consecutivament a dos (o més) dels ordres gènics comparats (Bérard et al., 2007); per exemple el cas del bloc *cox1-L2-cox1* conservat a Pancrustacea. Es diu que dos intervals comuns A i B commuten si es dóna qualsevol de les següents circumstàncies:  $A \subset B$ , o  $A \supset B$  o  $A \cap B = \emptyset$ . Un interval comú es converteix en un interval fort (*strong interval*) si commuta amb cada un dels intervals comuns. Un SIT per a dos ordres gènics diferents, és un arbre arrelat on els intervals comuns forts (SCIs, de l'anglès *strong common intervals*) poden representar-se com a nodes interns de l'arbre generat i que té exactament una terminació per a cada gen. És a dir, un SIT representaria gens que apareixen amb el mateix ordre gènic a diferents genomes de l'*input* donat. Existeixen dos tipus de nodes interns: els creixents i els decreixents (en anglès *linear increasing* i *linear decreasing*). Els descendents d'un node vendran representats simplement pels SCIs que incloguin. Si aquests descendents apareixen amb el mateix ordre en ambdues reordenacions gèniques, el node s'anomenarà *linear increasing*, que s'identifica amb el símbol (+); si els descendents d'un node apareixen exactament en l'ordre oposat, s'anomenarà *linear decreasing*, representat per (-); altrament el node és anomenat *prime* (Bernt, et al., 2007).

CREx determina heurísticament els escenaris de reordenació gènica succeïts entre un ordre gènic i un altre que presenti un patró de reordenació diferent. Amb l'ajuda d'una matriu de distàncies, poden identificar-se a partir de comparacions a parells aquells ordres gènics que siguin similars. Posteriorment, gràcies al SIT generat, l'usuari pot identificar regions de similitud entre dos ordres gènics. El mètode heurístic és capaç de determinar els escenaris de reordenació gènica gràcies a la detecció de transposicions, transposicions inverses, inversions i esdeveniments de pèrdua de còpies duplicades en tàndem (TDRL, *tandem-duplication-random-loss*). La identificació dels TDRL ha suposat un gran avanç a l'estudi de les reordenacions gèniques, atès que poden aportar informació molt valuosa sobre la direcció evolutiva, a diferència de la resta d'esdeveniments de reordenacions (Chaudhuri et al., 2006; Perseke et al., 2007).

### 3.8. Test de Saturació als PCGs

En el cas de l'estudi filogenètic de la família Metacrangonyctidae, es va comprovar si hi havia saturació a l'alineament obtingut. Per aquest motiu es realitzà el

---

test de saturació de Xia i Lemey (2009) que es troba implementat a *DAMBE* v. 5.2.28. Aquest test comprova si l'entropia observada a les seqüències és significativament menor a l'entropia en el cas de haver-hi una "saturació completa". El rati de l'entropia observada amb l'entropia en "saturació completa" es coneix com a índex de saturació (*index of substitution saturation, I<sub>ss</sub>*). Quan aquest no és significativament menor al valor crític de *I<sub>ss</sub>* (*I<sub>ssc</sub>*) -valor al qual la informació continguda a les seqüències no seria fiable per obtenir l'arbre correcte- es conclou que les seqüències estudiades es troben severament saturades (Xia & Lemey, 2009).

### 3.9. Metodologia de reconstrucció filogenètica

#### 3.9.1. Selecció del model evolutiu

Els models evolutius consideren una sèrie d'assumpcions sobre el procés de substitució nucleotídica-aminoacídica que permeten el càlcul de probabilitats de canvi entre nucleòtids-aminoàcids a través de les branques d'un arbre filogenètic (Posada & Crandall, 2001; Posada, 2008). Possibles violacions en aquestes assumpcions poden induir a errors significants a l'anàlisi filogenètica (Buckley & Cunningham, 2002; Lemmon & Moriarty, 2004), pel que és de gran importància realitzar una verificació curosa del compliment de les assumpcions preses a cada model particular. A més complexitat del model (major nombre de paràmetres), millor és l'ajust a les dades (major versemblança), malgrat que també augmentarà la variància (incertesa o error) dels paràmetres estimats. De fet, models amb millor versemblança no són necessàriament els que estimaran millor les longituds de branques d'un arbre ni els que obtinguin la millor estimació de la topologia (Sullivan & Joyce, 2005), i àdhuc podrien provocar que les dades fossin intractables computacionalment (Holder & Lewis, 2003). Contràriament, si el model ignora paràmetres que siguin particularment importants, la magnitud de canvi evolutiu serà subestimada, el que conduiria a una estimació filogenètica inconsistent (Gaut & Lewis, 1995; Sullivan & Swofford, 1997, 2001). Les estratègies de selecció de models tracten de trobar el nivell de complexitat apropiat segons les dades disponibles, és a dir comparen els diferents models alternatius per tal de escollir aquell que sigui més adient. Hi ha diversos procediments per destriar entre els diferents models possibles, entre els quals es troben:

- Test de raó de versemblança o hLRT (de l'anglès *hierarchical likelihood*)



*ratio test*): és un dels mètodes estadístics més emprats per a la selecció de models filogenètics. Consisteix en fer comparacions per parells entre models niats (la hipòtesi nul·la és un cas especial de la hipòtesi alternativa) fins que el mètode convergeix en el model més simple que no pot refutar-se per a un nivell de significança determinat (Ripplinger & Sullivan, 2008).

$$\delta = 2(\ln L_1 - L_0)$$

On  $L_0$  és el valor de versemblança de la hipòtesi nul·la (model més simple), i  $L_1$  és la versemblança de la hipòtesi alternativa (model complex). La significació del estadístic  $\delta$  s'avalua normalment sota el supòsit que es distribueixi asimptòticament amb una  $\chi^2$  amb  $q$  graus de llibertat, on  $q$  és la diferència del nombre de paràmetres lliures entre els dos models (Yang et al., 1995; Frati et al., 1997; Huelsenbeck & Crandall, 1997; Posada & Crandall, 2001). Els desavantatges principals d'aquest mètode estadístic és que es requereix partir d'una topologia inicial que es estimada per parsimònia o el mètode de NJ (Sanderson & Kim, 2000) i no permet la comparació entre models no niats, alhora que tampoc penalitza als models amb més paràmetres (Felsenstein, 2004; Sullivan & Joyce, 2005) i a més els seus resultats són sensibles als protocols d'implementació d'aquest test (*top-down vs. bottom-up*; veure Pol, 2004).

- AIC (de l'anglès *Akaike information criterion*): és un estimador no esbiaixat del paràmetre de contingut d'informació de Kullback-Leibler (Kullback & Leibler, 1951), que mesura la informació que es perd quan es fa servir un model per aproximar-se a la realitat (Akaike, 1973). El seus avantatges són que permet la comparació entre models no niats i ordenar les hipòtesis segons sigui el valor d'AIC. D'aquesta manera, es pot quantificar la incertesa associada a la selecció d'un determinat model respecte a la resta a partir de les diferències entre els seus valors AIC i amb el càlcul dels *Akaike weights*. D'acord amb això, els millors models seran aquells que acumulin un *Akaike weight* superior a 0.95 (Burnham & Anderson, 2002). Existeix un factor de correcció per AIC, conegut com a AICc (Hurvich & Tsai, 1989), el qual s'empra quan el rati entre la mida de la mostra i el nombre de paràmetres lliures és petit (Ripplinger & Sullivan, 2008). Ambdós es computen de la següent manera:

---



$$AIC_i = -2 \ln L_i + 2k_i$$

$$AIC_c = AIC + \frac{2k_i(k_i + 1)}{n - k - 1}$$

On  $L_i$  és el valor maximitzat de la funció de versemblança pel model estimat  $i$ ,  $k_i$  és el nombre de paràmetres lliures dins el model estadístic i  $n$  és la mida de la mostra (normalment és pren per  $n$  a la longitud de la seqüència). Emperò, atès els valors d'AIC i  $AIC_c$  convergeixen a mesura que  $n$  augmenta, s'ha suggerit que s'empri  $AIC_c$  per a tots els sets de dades (Burnham & Anderson, 2002, 2004; Posada & Buckley, 2004; Ripplinger & Sullivan, 2008).

- BIC (de l'anglès *Bayesian Information Criterion*) (Schwarz, 1978): com en el cas anterior, consisteix en un criteri d'informació que indica com s'ajusten el model i les dades. Es defineix de la següent manera:

$$BIC_i = -2 \ln L_i + k_i \ln n$$

On  $n$  es la mida de la mostra (normalment es pren per  $n$  a la longitud de la seqüència). A diferència de l'anterior, emperò, sempre té en compte la mida de la mostra, pel que BIC ja no tendeix a seleccionar el model més complex a mesura que  $n$  s'incrementa tal com feia AIC. Atès que la majoria de les vegades el  $\log$  de  $n$  és més gran de 2, BIC sol penalitzar la sobre-parametrització més durament que AIC, pel que tendeix a triar models més simples que AIC (Posada & Crandall, 2001; Abdo et al., 2005).

Finalment, es trià BIC com a mètode de selecció del model adient per a cada marcador i cada partició, vist que diversos estudis com el de Luo et al. (2010), suggereixen que BIC es trobaria entre els millors criteris d'informació per a la selecció de models evolutius gràcies a la seva precisió i exactitud. Per a la selecció del model es trià el programa *jModelTest* v.0.1.1. (Posada, 2008), a excepció dels alineaments de MPCGs pels que es va fer servir l'script de Perl *MrAIC* (Nylander, 2004; <http://www.abc.se/~nylander/mraic/mraic.html>). *jModelTest* permet discernir entre 56 hipòtesis alternatives de canvi nucleotídic, mentre que *MrAIC* permet comparar

entre 24 models evolutius. Per a la selecció del millor model de substitució pel set de dades de tots els PCGs traduïts a proteïnes s'emprà el programa *Prottest* v. 1.4. (Abascal et al., 2005).

### 3.9.2. Inferència filogenètica

Per a la inferència filogenètica s'utilitzaren metodologies amb diferents criteris d'optimització. Es realitzaren anàlisis de Màxima Versemblança (d'hora endavant: ML, de l'anglès *Maximum Likelihood*) (Felsenstein, 1981) i d'inferència bayesiana (d'hora endavant: BI de l'anglès *Bayesian Inference*) (Ronquist & Huelsenbeck, 2003). A diferència de les anàlisis de distàncies que arriben a una solució o arbre únic, les d'optimització comparen múltiples hipòtesis filogenètiques alternatives (arbres) i avaluen quines són les més versemblants. Malauradament, quan augmenta el nombre de seqüències analitzades també augmenta de forma exponencial el nombre d'hipòtesis filogenètiques possibles (Felsenstein, 1978). A partir de 20 unitats taxonòmiques el temps de computació necessari per la cerca exhaustiva dels millors arbres és inviable pel que aquesta és substituïda per una cerca heurística. Una aproximació heurística per un problema d'optimització és un algorisme eficient que explora una fracció del total de les hipòtesis possibles, trobant una solució factible "suficientment bona". El gran desavantatge d'aquestes cerques és que no garanteixen l'obtenció de l'arbre òptim (Sanderson, 1995). Per aquest motiu, es requereix del càlcul del nivell de confiança de les hipòtesis evolutives plantejades que s'indiquen com a estimes de suport estadístic per a cada un dels nodes dels arbres. Els índex de confiança més comuns són els de *bootstrap*, el suport de Bremer i els valors de probabilitat posterior.

La metodologia de Màxima Versemblança (ML) s'ha desenvolupat específicament dins del camp de la sistemàtica molecular. La primera vegada que es va emprar fou amb dades de freqüències gèniques (Cavalli-Sforza & Edwards, 1964) i posteriorment, Felsenstein (1973, 1981) adaptà aquesta metodologia a l'estudi de seqüències aminoacídiques i nucleotídiques. El criteri d'optimització es fonamenta en la cerca de l'arbre que presenta la probabilitat més alta de donar lloc a les dades observades (matriu de seqüències) baix un model evolutiu prèviament definit (Swofford et al., 1996). ML és proporcional a les probabilitats reals: és a dir, entre tots els arbres explorats, la hipòtesi evolutiva més probable serà la que presenti major versemblança global. De forma matemàtica la ML es defineix:

---

$$L(H|D) = \Pr(D|H) = \Pr(D|\tau, \nu, \theta)$$

On  $H$  són els arbres/hipòtesis a testar;  $D$  són les dades;  $\tau$  es defineix com la topologia de l'arbre;  $\nu$  és la longitud de les branques i  $\theta$  és el model evolutiu emprat. Es calcula per a cada una de les posicions de l'alineament i es multipliquen els valors obtinguts per obtenir una versemblança global de l'arbre. Com la versemblança global s'ha de maximitzar per a cada hipòtesi i aquesta no depèn solament de la topologia, sinó també de les longituds de branques i dels valors de cada paràmetre del model evolutiu, els problemes de còmput a la reconstrucció per ML són particularment onerosos malgrat es recorren a aproximacions heurístiques.

Per a la realització d'aquesta tesi s'han emprat els programes *RAxML* v. 7.2.4 i/o *Garli* v. 2.0. ([https://www.nescent.org/wg\\_garli/Main\\_Page](https://www.nescent.org/wg_garli/Main_Page); Ziwckl, 2006) per a les anàlisis de ML. Aquest darrer té l'avantatge que permet particionar les dades i aplicar un model de substitució diferent a cada una de les particions incloent el model del “*nucl model*” per codons de Goldman i Yang (1994).

Mentre que la ML mesura la probabilitat de les dades d'un arbre donat, la inferència bayesiana calcula les “probabilitats posteriors” ( $pp$ ) de les hipòtesis filogenètiques. Aquesta anàlisi produeix una estima de la probabilitat per la qual un arbre sigui el vertader donat un determinat conjunt de dades, així doncs la hipòtesi òptima serà la que maximitzi la  $pp$  (Huelsenbeck & Ronquist, 2001). És a dir, la  $pp$  d'un arbre pot interpretar-se com a la probabilitat condicional que aquell arbre sigui el correcte, i per tant, l'arbre amb major  $pp$  pot ser elegit com la millor estima de la filogènia (Rannala & Yang, 1996; Huelsenbeck & Ronquist, 2001; Huelsenbeck et al., 2002). L'enfocament Bayesià dona com a resultat l'arbre (o el conjunt d'arbres) que són més probables de ser explicats per les dades, o bé la probabilitat que la hipòtesi  $H$  sigui correcte donat les dades observades.

El teorema de Bayes (1763)\* s'utilitza per combinar la probabilitat *a priori* de la filogènia  $P(H)$ , és a dir, la que es dona a un arbre abans de fer qualsevol observació (normalment tots els arbres són, *a priori*, igual de probables) amb la seva versemblança  $P(D|H)$  i produir una distribució de  $pp$  en els arbres  $P(H|D)$ .

$$* P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$

Els mètodes bayesians de reconstrucció filogenètica són bastant similars als de màxima versemblança ja que també necessiten partir d'un model explícit de substitució i els millors arbres són aquells que són més consistents tant amb el model com amb les dades observades. Per una altra banda, existeixen diverses diferències entre ML i BI: mentre ML cerca l'arbre que maximitzi la probabilitat d'observar les dades donat un model de substitució i l'arbre, BI considera que la millor estimació de la filogènia és l'arbre que tingui la màxima probabilitat de ser observat donat un model de substitució i les dades (alineament). Una altra diferència important entre aquests dos mètodes és que ML no tracta a les variables del model com a variables aleatòries mentre que a l'anàlisi bayesiana tant les dades com els paràmetres són considerats com a variables aleatòries; el que significa que inclou incertesa en l'ús dels paràmetres descrita a la distribució de probabilitat posterior (Ronquist et al., 2005).

Malgrat la *pp* és fàcil de formular, calcular la *pp* d'una filogènia es computacionalment enrevessat ja que requereix fer sumes de tots els arbres possibles i, a més a més, per a cada arbre s'han d'integrar totes les possibles combinacions de longituds de branques i tots els possibles valors dels paràmetres dels models de substitució. Afortunadament existeixen diversos mètodes d'optimització numèrica per obtenir una aproximació de la probabilitat posterior d'un arbre. Aquests algorismes tracten de trobar mitjançant un procés d'assaig-error, els arbres òptims. El mètode més emprat d'optimització numèrica en BI és el mètode de Monte Carlo amb cadenes de Markov (d'hora endavant: MCMC, en anglès *Metropolis Coupling Monte Carlo*). Aquest algorisme consisteix en "proposar" a cada passa (o generació) de l'algorisme MCMC un nou arbre que inclou modificacions a l'atzar en la topologia, longitud de branques o bé en els paràmetres del model. Si el nou arbre té una major *pp* que l'anterior, llavors s'agafa com a punt de partida per a la següent interacció (Holder & Lewis, 2003); si en canvi, el nou arbre té una menor *pp* que l'anterior, solament és acceptat si el quocient entre la seva probabilitat i la de l'arbre actual és major que un nombre pres a l'atzar a l'interval 0,1. Si es descarten els primers desplaçaments a l'espai dels arbres (*burn-in*), la distribució de les regions visitades resulta ésser una bona aproximació de les regions amb arbres amb major *pp*. Una qüestió primordial és saber si l'algorisme ha calculat el suficient nombre de generacions. El programa *MrBayes 3.1.2*. (Ronquist & Huelsenbeck,

---

2003) guarda un registre dels arbres visitats i així es pot saber quines regions de l'espai dels arbres s'han explorat més intensament, el que indica quins són els arbres que estan més ben suportats estadísticament; és a dir el MCMC visita els arbres en proporció a les seves *pp*.

### 3.9.2.1. Anàlisi del set de dades per a la filogeografia de *M. longipes* (Capítol I)

S'avaluaren diferents esquemes de particions pel set de dades d'espècimens de *M. longipes*. Les diferents estratègies de partició i anàlisis foren les següents:

- Tots els marcadors (*cox1*, *cob*, *rrnL* i *H3*):
  - Cap partició: totes les posicions analitzades com una sola partició.
  - Dues particions: gens mitocondrials (*cox1* + *cob* + *rrnL*) vs. nuclears (*H3*).
  - Tres particions: gens mitocondrials codificants per proteïna (*cox1* + *cob*) vs. ribosomal mitocondrial *rrnL* vs. nuclears (*H3*).
  - Quatre particions: per gen (*cob/cox1/rrnL/H3*).
  - Quatre particions: primeres + segones posicions dels codons dels gens mitocondrials codificants per proteïna (*cox1+cob*) vs. terceres posicions dels codons d'aquest gens (*cox1+cob*) vs. mitocondrial ribosomal *rrnL* vs. *H3*.
  - Cinc particions: primeres posicions dels codons dels gens mitocondrials codificants per proteïna vs. segones posicions codons vs. terceres posicions codons vs. *rrnL* vs. *H3*.
  
- El set de dades complet pels espècimens de *M. longipes* per a *cox1* (162 + 3 grups externs):
  - Cap partició: totes les posicions analitzades com una sola.
  - Dues particions: primeres + segones posicions dels codons vs. terceres posicions dels codons.
  - Tres particions per posició de codó: primeres vs. segones vs. terceres.

Les diferents estratègies de particionament de les dades es comparen segons el valor de BIC a partir dels valors de versemblança obtinguts amb l'anàlisi de ML,

---

realitzades amb el programa de *RAxML v.7.2.4*. Aquest programa implementa un algorisme ràpid dissenyat per cercar l'arbre amb el millor valor de versemblança en el qual es va fer una anàlisi de *bootstrap* (anàlisi de remostreig) (Felsenstein, 1985) no paramètrica amb 1,000 rèpliques. Addicionalment, a partir dels valors obtingut amb *MrBayes v. 3.1.2*, també es calcularen els valors de factor de Bayes (*Bayes Factors*, d'hora endavant BFs) (Kass & Raftery, 1995) amb el programa *Tracer v. 1.4* (Rambaut & Drummond, 2007; <http://tree.bio.ed.ac.uk/software/tracer/>) per comprovar quin era el millor esquema de particionament.

El BF quantifica i compara el suport relatiu de dues hipòtesis o models ( $M_0$  i  $M_1$ ) segons les dades observades ( $Y$ ), i per a models no niats es calcula com la ràtio de les versemblances marginals (*marginal likelihood*) de les hipòtesis  $f(Y | M_i)$ :

$$B_{10} = \frac{\int(Y|M_1)}{\int(Y|M_2)}$$

El BF és l'anàleg bayesià al LRT (veure apartat 3.8.1.), tot i que té l'avantatge que no necessita que els models siguin niats i la versemblança marginal penalitza la sobre-parametrització. A diferència de BIC i de la resta de metodologies tradicionals de selecció de models com AIC i LRT, aquesta mesura té en compte els priors que s'han emprat a l'anàlisi bayesiana (Xie et al., 2010). Kass & Raftery (1995) introdueixen diferents graduacions per avaluar el valor de BF com a evidència per refutar  $M_0$ , on es veu que a valors superiors a 20 el suport per refutar  $M_0$  davant  $M_1$  és molt elevat (Taula 2.4). La versemblança marginal d'un model és la mitjana ponderada de la versemblança de tots els valors dels paràmetres; és a dir és la mesura de l'ajust mig del model a les dades. El càlcul d'aquesta versemblança marginal és computacionalment inabordable, pel que es fa una aproximació emprant cadenes de Markov Monte Carlo (*mcmc*). Existeixen diverses maneres de calcular les versemblances marginals, entre les quals s'hi troba la que fins ara ha sigut la més emprada, el mètode de la mitjana harmònica (*harmonic mean method*, HM) (Newton & Raftery, 1994). Això no obstant, el HM ha sigut durament criticat els darrers anys i fins i tot s'ha comprovat que pot ésser inapropiat atès que sobreestima les versemblances marginals, no penalitza prou els models amb més paràmetres que no incrementen notablement l'ajust de les dades al model, i presenta una variància extremadament alta o fins i tot infinita (Lartillot & Philippe, 2006; Lewis et al., 2010; Fan et al., 2011; Xie et al., 2011; Baele et al., 2012).

---

Malauradament, ara per ara el HM es l'únic mètode pel càlcul de les versemblances marginals implementat al programa *Tracer* v. 1.4. i v.1.5., que són les que s'han emprat per aquesta tesi.

Per a les anàlisis de IB es feren dues anàlisis independents per a cada cerca bayesiana amb els valors dels priors per defecte a *MrBayes* v.3.1.2. S'utilitzaren quatre cadenes MCMC (tres cadenes calentes i una freda) per a cinc milions de generacions, i es mostrejà a intervals de 1000 generacions. Per a cada partició s'implementà el model evolutiu més adequat segons BIC calculat a *jModelTest* v.0.1.1. Tots els paràmetres es deslligaren entre particions (*unlinked*) i es deixà que cada partició variés amb la seva mateixa taxa. El nombre de generacions a descartar abans d'haver-hi convergència, el que es coneix com el valor del paràmetre de *burn-in* (Huelsenbeck & Ronquist, 2001) s'estimà amb l'ordre *sump* a *MrBayes*, el qual construeix una gràfica del valor de

**Taula 2.4.** Interpretació dels valors de factors de Bayes (BF)

Log (BF <sub>10</sub> )**	2 Ln (BF <sub>10</sub> )*	BF <sub>10</sub> *	Evidència en contra de M <sub>0</sub>
0 a 0.5	0 a 2	1 a 3	no val més que una simple menció
0.5 a 1	2 a 6	3 a 20	positiva
1-2	6 a 10	20 a 150	forta
>2	>10	>150	molt forta

\* Kass & Raftery, 1995.

\*\* Jeffreys, 1935.

versemblança respecte al nombre de generacions. Per estimar la convergència de les anàlisis independents i la mida efectiva de la mostra (ESS, de l'anglès *effective sample size*) s'emprà el programa *Tracer* v. 1.4. D'aquesta manera es comprovà que el valor de ESS fos major de 200 per a totes les anàlisis -Drummond i Rambaut (2007) suggereixen un mínim de 100 en valor de ESS i un òptim per damunt 200- i s'eliminaren aproximadament el 10% de les generacions mostrejades per falta de convergència. La convergència i estacionalitat de la topologia obtinguda a les dues anàlisis independents es testà a partir del programa *AWTY* (de l'anglès: *Are We There Yet?*) (Nylander et al., 2008) amb les ordres *slide*, *cumulative* i *compare*. *AWTY* pot estimar la convergència i estacionalitat de les biparticions al llarg d'una anàlisi (*slide & cumulative command*) o bé pot comparar dues anàlisis independents (*compare command*). L'ordre *slide* testa si les sub-mostres preses de la cadena de mostreig corresponen a arbres en proporció a llurs probabilitats posteriors. És a dir, si es divideix la cadena en X sub-mostres/segments i es calculen les freqüències dels clades-nodes, aquestes haurien de ser similars si s'hagués arribat a l'estacionalitat dels paràmetres. L'ordre *cumulative* fa el mateix



però calcula les freqüències acumulatives de les divisions (*splits*) per a cada una de les simulacions. És a dir, és l'equivalent a aturar periòdicament la cadena de MCMC i calcular les freqüències dels *splits* en aquell punt. Tot i així l'ús de solament l'ordre *slide* i *cummulative* no són suficients per descartar falta de convergència ja que es podria observar un resultat positiu en el cas de que l'anàlisi hagués explorat malament l'espai d'arbres hi hagués caigut en un òptim local. Per aquest motiu es va fer servir també l'ordre *compare* que permet comparar les freqüències de les probabilitats posteriors de les bifurcacions entre les diferents anàlisis; on en el cas de haver-hi convergència s'hi hauria d'observar una alta correlació. Es van prendre les opcions per defecte del programa *AWTY* per calcular la convergència i estacionalitat de les diferents anàlisis: en el cas de l'ordre *slide* es calcularen 10 segments-finestres de sub-mostres, fent un *burn-in* previ del 10% i es mostraren els 20 *splits* més variables. En el cas de l'ordre *cummulative* es calcularen 10 vegades (= increments de 10) els 20 *splits* més variables. Un cop realitzat el *burn-in*, els arbres obtinguts a partir de les dues anàlisis es combinaren per obtenir en una topologia única consens (per la regla de la majoria) emprant l'ordre *sumt* de *MrBayes*, i es prengueren les freqüències de les biparticions (= nodes) de l'arbre consens com a valors de suport de probabilitat posterior (Huelsenbeck & Ronquist, 2001).

Finalment, es dugué a terme la reconstrucció de l'estat dels caràcters quant a habitat dels nodes interns (aigües dolces vs. aigües salobres). Uns dels mètodes més emprats per a la reconstrucció d'estats de caràcters ancestrals mitjançant inferència filogenètica ha sigut el de màxima parsimònia (Maddison & Maddison, 1989). Per a la reconstrucció s'emprà el programa *Mesquite* v. 2.74 (Maddison & Maddison, 2009). La proposta filogenètica sobre la que es realitzà la reconstrucció està presa de l'arbre bayesià de *MrBayes* v. 3.1.2. amb el set de dades complet per *cox1* i la distribució de l'habitat observat entre les poblacions mostrejades per minimitzar el nombre de passes de canvi d'hàbitat.

#### **3.9.2.1.1. Congruència entre les diferents particions**

La congruència entre els diferents marcadors emprats s'estimà mitjançant el test de *Partitioned Bremer Support* (PBS) (Baker et al., 1998) implementat al programa *Treerot* v. 3.0 (Sorenson & Franzosa, 2007) dins *PAUP* v. 4b10 (*Phylogenetic analysis using Parsimony\**, Sunderland, MA, USA) (Swofford, 2002). S'usaren els paràmetres per

---



defecte, excepte que la cerca heurística es realitzà amb 200 rèpliques, guardant 25 arbres per rèplica, per garantir que es trobava l'arbre més curt.

L'índex de decaïment o el valor de suport de Bremer (Bremer, 1988) s'empra per mesurar el suport donat a un node concret dins un arbre filogenètic. Aquest valor es calcula en base a la diferència en la longitud entre els arbres més curts que inclouen un grup i els arbres més curts que no l'inclouen (és a dir, el nombre de passes extra que es requereixen per col·lapsar un grup), mentre que el suport total seria la suma de tots els índex de decaïment. El suport de Bremer particionat (*Partitioned Bremer Support*, PBS) (Baker et al., 1998) calcula la quantitat de suport aportat per a cadascuna de les particions/marcadors a l'índex de decaïment per a cada node de la filogènia resultant de l'anàlisi combinada. Si la partició suporta la relació representada per un node a un arbre combinat, el valor de PBS resultarà ser positiu. Si ocorre el contrari, si la partició suporta una relació alternativa, el valor de PBS serà negatiu. El particionament de les dades és una manera altament escaient d'explorar la incongruència de senyal entre caràcters de diferents marcadors (Brower, 2006).

### 3.9.2.2. Anàlisi del set de dades dels mitogenomes (Capítol III)

Per resoldre la filogènia de la família Metacrangonyctidae s'empraren els MPCGs de tots els 23 mitogenomes amplificats (dos grups externs i 21 metacrangonyctids). S'exploraren diferents esquemes de partició per Inferència Bayesiana i Màxima Versemblança per tal de poder destriar la millor manera de particionar les dades. Aquests esquemes foren els següents:

- Tots els 13 PCGs:
    - o Totes les posicions dels codons analitzades com una sola partició;
    - o Dues particions: primeres + segones vs. terceres posicions dels codons;
    - o Tres particions: primeres vs. segones vs. terceres posicions dels codons;
    - o Segons el model per codons (Goldman & Yang, 1994);
    - o 13 particions: per gen.
-

- Tots els 13 MPCGs més el gen nuclear *SSU*:
  - o Quatre particions: primeres vs. segones vs. terceres posicions dels codons vs. *SSU*;
  - o Segons el model per codons (Goldman & Yang, 1994) vs. *SSU*.

Per a les anàlisis bayesianes s'emprà *MrBayes* v. 3.1.2. implementat en una targeta gràfica de Tesla c2050 que feia servir la llibreria *Beagle* (Suchard & Rambaut, 2009), que redueix notablement el temps computacional, especialment en el cas del model de codons (fins a 30-50 vegades més ràpid). Es corregueren dues anàlisis independents per a cada cerca bayesiana començant amb els valors dels priors per defecte i arbres a l'atzar. S'utilitzaren quatre cadenes de MCMC (tres cadenes calentes i una freda) durant cinc milions de generacions i es mostrejà a intervals de 1000 generacions a partir d'arbres a l'atzar. Per a cada partició de seqüències s'implementà el model evolutiu més adequat segons els valors de BIC calculats amb l'script de perl *MrAIC* (Burnham & Anderson, 2002; Nylander, 2004).

Tots els paràmetres es deslligaren entre particions i també es deixaren variar les taxes de substitució nucleotídica lliurament entre aquestes. El valor del paràmetre de *burn-in*, la convergència i estacionalitat dels paràmetres i diferents topologies es determinaren amb els programes *Tracer* 1.5 i *AWTY* (veure apartat 3.8.2.1). Un cop realitzat el *burn-in*, els arbres restants obtinguts a les dues anàlisis es combinaren per obtenir una sol arbre consens de regla de la majoria derivat dels arbres visitats i es prengueren les freqüències dels nodes com a valors de suport de probabilitat posterior (*pp*) (Huelsenbeck & Ronquist, 2001). Les anàlisis que es realitzaren pel model de partició per codons de Goldman i Yang (1994) es corregueren de 20 a 30 milions de generacions per assegurar-se tenir un valor suficientment alt de ESS un cop haguessin convergit tots els paràmetres. En aquest cas es va aplicar un *burn-in* del 75%.

Pel set de dades dels 13 MPCGs es realitzà una segona anàlisi bayesiana amb el programa *PhyloBayes* v.3.3 (Lartillot et al., 2009) amb el set de dades dels MPGCs traduïts a proteïnes. El principal avantatge d'aquest programa és que les particions no s'han de definir *a priori*. Es diferencia principalment per tenir el seu propi model

---

probabilístic de treball, anomenat CAT i per resultar particularment útil per alineaments grans de múltiples gens (>1000 posicions) (Lartillot & Philippe, 2004). Talavera i Vila (2011) han estudiat com s'ajusta aquest model CAT en el cas d'emprar mitogenomes complets per reconstruir la filogènia a insectes. Aquests autors suggereixen que CAT és el model que millor s'ajusta a alineaments amb gens que presenten elevades taxes de substitució que cerquen esclarir la filogènia amb esdeveniments de divergència molt profunds, i a més, evita satisfactòriament el problema de l'atracció de les branques llargues o LBA (de l'anglès "*long-branch attraction*"). De fet, CAT és una barreja de models que agrupa en infinites categories ( $K$ ) les diferents posicions (nucleotídiques o aminoacídiques) que tenen un taxa i composició semblant. En aquest estudi les taxes d'intercanvi global s'inferiren a partir de les dades, tal com es recomana fer en sets de dades superiors a 1000 posicions (Lartillot et al., 2009). Es realitzaren tres anàlisis independents i es deixaren córrer fins a la seva convergència; o sia, fins que el valor de la freqüència màxima de separació (*maximum split frequency*) fou menor de 0.1 i un valor d' ESS major de 100 (Lartillot et al., 2009).

Per a les anàlisis de Màxima Versemblança s'utilitzà el programa *Garli* v. 2.0. que com ja s'ha esmentat permet aplicar un model de substitució diferent a cada una de les particions incloent el model del "*nucl model*" per codons de Goldman i Yang (1994). Pel set de dades dels 13 MPCG i el gen nuclear *SSU* es realitzà una anàlisi amb *RAxML* v. 7.0.4. en el qual es va realitzar una anàlisi ràpida de *bootstrap* (Felsenstein, 1985) no paramètrica amb 1,000 rèpliques.

Un cop obtinguts els resultats s'observà que els grups externs utilitzats per aquest estudi es trobaven molt allunyats de la família Metacrangonyctidae i vist que aquests grup és monofilètic, es repetiren les anàlisis sense aquests dos mitogenomes per comprovar que la presència dels grups externs no afectava a les relacions filogenètiques de l'*ingroup*. Per finalitzar es trià quin era el millor model de particionament segons el millor BFs calculat per *Tracer* v. 1.5. a partir dels resultats de *MrBayes* v. 3.1.2.; o segons el valor de BIC calculat a partir dels valors de versemblança obtinguts amb *Garli* v. 2.0. (veure apartat 3.8.1.).

La congruència entre els diferents marcadors emprats (MPCGs vs. *SSU* i cadascun dels MPCGs per separat) s'estimà mitjançant el test de PBS (*Partitioned Bremer*

Support) (apartat 3.8.2.1.1.).

### 3.9.2.2.1. Comparació entre topologies alternatives

Les filogènies que són inferides a partir de particions o sets de dades distints en ocasions poden mostrar topologies diferents tot i tractar-se del mateix grup d'organismes (Rodrigo et al., 1993). La majoria de sistemàtics consideren que les diferents particions de dades poden combinar-se solament quan es demostra que no són significativament incongruents entre elles. Per aquest motiu es requereixen de certs tests estadístics per avaluar la incongruència taxonòmica entre les diferents particions (Kishino & Hasegawa, 1989; Hipp et al., 2004). Per determinar si topologies alternatives obtingudes amb diferents sets de dades (només els MPCGs versus MPCGs-SSU) eren significativament diferents es realitzaren anàlisis preliminars amb *RAxML* 7.2.4. (Stamatakis, 2006) i es testaren les dues topologies alternatives amb els tests de Shimodaira-Hasegawa (1999) (d'hora endavant, test de S-H) i el test dels "pesos de versemblança esperats" (d'hora endavant ELW, de l'anglès *Expected Likelihood Weights*) (Strimmer & Rambaut, 2002). El test de S-H ha substituït a l'anterior test de Kishino-Hasegawa (1989) ja que s'ha observat que aquest pot induir a errors al rebutjar falsament la hipòtesi nul·la (Shimodaira & Hasegawa, 1999; Goldman, 2000). El test de S-H emprava un re-mostreig tipus *bootstrap* no paramètric (recerca *RELL: Resampling Estimated Log Likelihoods*), i corregeix aquells valors crítics realitzant múltiples comparacions el que permet saber si es pot rebutjar la hipòtesi nul·la (tots els arbres considerats s'ajusten igualment a les dades), i ha sigut dissenyat per emprar-se baix el criteri d'optimització de ML. L'estadístic del test de S-H ( $\delta$ ) és la diferència entre el millor arbre amb màxima versemblança i qualsevol altre que es compara:

$$\delta_T = \ln L_{ML} - \ln L_T$$

Les hipòtesis que es volen testar són:

$H_0$  = tots els arbres considerats s'ajusten igualment a les dades

$H_1$  = qualcun dels arbres o tots ells no s'ajusten prou bé a les dades

Això no obstant, Strimmer i Rambaut (2002) demostren que el test de S-H pot resultar en ocasions massa conservatiu, pel que suggereixen un nou test estadístic, el test ELW, per testar diferents topologies i construir intervals de confiança per

---

topologies obtingudes a partir d'anàlisi de ML. Aquesta prova és la proporció de la versemblança de l'arbre a comparar sobre la suma de les versemblances de cadascun dels arbres candidats. Els arbres testats són llavors ordenats segons el seu "pes" i s'inclouen dins l'interval de confiança fins que la suma dels seus "pesos" sigui igual o lleugerament major que el nivell de confiança. Aquells arbres que presentin un major "pes" són el més propensos per a ser els vertaders i per tant es trobaran a l'interval de confiança.

#### 3.9.2.2.2. Presència de politomies a les filogènies

Són moltes les ocasions en què estudis dedicats a la reconstrucció filogenètica de diferents tàxons són incapaços de resoldre les relacions filogenètiques satisfactòriament per topar-se a la topologia de l'arbre amb el que es coneix vulgarment com a "espina de peix". No obstant això, no tots els casos de relacions politòmiques són resultats de inferències incongruents o deficientes, el que es coneix com a politomies falses (*soft polytomies*) (Maddison, 1989). S'ha observat que la resolució de nodes precedits per branques curtes (nodes interns) pot ésser un procés ardu fins i tot tenint gran quantitat de dades (Tajima, 1983; Wiens et al., 2008). Teòricament, la presència de politomies veritables (= *hard polytomies*) a un arbre filogenètic pot ser resultat d'un esdeveniment de ràpida especiació-radiació on branques internes extremadament curtes són seguides de branques molt llargues (Walsh, et al., 1999; Braby et al., 2005; Shavit et al., 2007). Aquest patró de politomia veritable s'observa quan una diversificació tan ràpida no permet la fixació de substitucions nucleotídiques informatives suficients (branques curtes) i cal diferenciar-les de les politomies falses, que provenen de la incongruència filogenètica. A més, es podria esperar un resultat similar en el cas d'esdeveniments vicariants on queden aïllats al mateix temps diversos llinatges.

Hi ha diversos estudis que suggereixen que les probabilitats posteriors inferides a les anàlisis bayesianes poden veure's artificialment augmentades, produint valors de confiança "inflats" i alts valors incorrectes de *pp* a nodes incerts o inestables, especialment quan l'arbre vertader presenta longitud de branques internes iguals o properes a zero. El problema sembla residir en el fet que *MrBayes* considera sempre els arbres com a binaris, és a dir, força als nodes a la dicotomia. Diversos estudis de simulació que han emprat arbres politòmics de quatre espècies (tres possibles arbres no

arrelats) demostraren que els valors de *pp* obtinguts per *MrBayes* eren extremadament alts (Suzuki et al., 2002; Cummings et al., 2003; Yang & Rannala, 2005). El problema de no obtenir uns valors de *pp* pels tres arbres binaris possibles iguals a ( $\frac{1}{3}$ ,  $\frac{1}{3}$ ,  $\frac{1}{3}$ ), es coneix amb el nom de *star-tree paradox* (Lewis et al., 2005). El problema principal, emperò, resideix no en com s'allunyen els valors de *pp* inferits del que serien els valors reals sinó més bé, en la resolució de nodes inexistents quant de fet es manca d'informació o senyal filogenètica suficient per ésser resolts (Yang, 2007). Lewis i col·laboradors (2005) suggeriren una modificació a l'algorisme de MCMC, implementada a *Phycas* v.1.2.0. (<http://www.phycas.org> ; Lewis et al., 2005) per tal de solucionar en aquest problema. D'aquesta manera, el nou algorisme (*reversible-jump* MCMC; Green 1995) permet que topologies menys resoltes puguin competir amb hipòtesis on els arbres són totalment resolts o bé són politomies completes (*star tree*).

Per aquest motiu es testà amb *Phycas* v. 1.2.0. la presència de politomies veritables ja que aquest programa explora la probabilitat de trobar politomies. Com ja s'ha esmentat anteriorment, existeixen diverses metodologies per estimar les versemblances marginals (veure apartat 3.9.2.1). *Phycas* v.1.2.0. ofereix diferents metodologies alternatives al tant criticat HM (Lartillot & Philippe, 2006), entre les quals s'hi troba el mètode de l'*stepping stone* (SS) (Xie et al., 2011; Fan et al., 2011), que té una sèrie d'avantatges en comparació del mètode de la mitjana harmònica: dóna millors estimadors de les versemblances marginals i ofereix una major precisió a l'hora de seleccionar un model (Baele et al., 2012). El mètode de SS estima directament les versemblances marginals a diferència del mètode de HM, que estima el rati d'aquestes (Xie et al., 2011). *Phycas* permet la implementació d'un "*prior* de politomia" (de l'anglès *politomy prior*) que permet la relaxació de la condició de dicotomia generalment assumida a les topologies dels arbres. Per aquesta anàlisi es testaren els priors de politomia  $e^1$  o  $e^2$ , el que implica que a un arbre totalment resolt ha de tenir una versemblança d'una o dues unitats de versemblança més per ser afavorida sobre un arbre politòmic (per arbre politòmic s'entén aquell que té al manco un node no resolt). Es corregueren dues anàlisis independents entre elles de 50,000 cicles (corresponent aprox. a cinc milions de generacions a *MrBayes*) amb una cadena freda i una calenta i mostrejant cada cinc cicles. Es va comparar emprant BFs (estimat pel mètode de "*stepping stone method & not marginal likelihood*) la millor *fitness* d'un model sense politomies amb la d'un model que sí les permetés.

Addicionalment, s'emprà el programa *DensiTree* v.2.0.1. (Bouckaert, 2010) per

---

visualitzar els múltiples arbres generats per *MrBayes*. Aquest programa permet dibuixar tots els arbres mostrejats a l'anàlisi bayesiana. Com a resultat, aquelles "àrees" on molts dels arbres coincideixen pel que fa a la topologia (topologia consens) i/o longitud de branques es mostren com a zones d'intensa coloració, mentre que les àrees amb poc consens apareixen borroses i tènues de coloració.

### 3.10. Estimació del temps de divergència

Abans d'aplicar un rellotge molecular cal testat si les seqüències a analitzar mostren una taxa d'evolució constant, es a dir, si segueixen un rellotge molecular estricte o global o si més bé la seva taxa de substitució canvia al llarg del temps o segons els llinatges (Thorne et al., 1998; Sanderson, 1997). En la darrera circumstància s'apliquen rellotges moleculars relaxats que empren algorismes basats en versemblança penalitzada (*penalised likelihood*) (Sanderson, 2002) o taxes logarítmiques no correlacionades (Drummond et al., 2006) per obtenir taxes constants.

Si una filogènia respon a un rellotge molecular estricte, (el que es coneix com a *clock-like*) o per contra és més plausible que segueixi un model de rellotge relaxat, és pot saber calculant els BFs obtinguts a partir de dues anàlisis de MCMC, una sota un model de rellotge estricte i l'altra que segueixi un model de rellotge relaxat. Segons la taula de Kass & Raftery (1995) un valor de  $2x \ln(BF)$  proper a 0 indicarà que no hi ha prou evidència a favor d'uns dels models, pel que es tria el model o la hipòtesi més simple, en aquest cas, un model de rellotge estricte. En canvi, un valor superior a 10 indicaria un suport molt elevat pel model de rellotge relaxat, que s'ajustaria més adequadament a les dades observades. Una altra possibilitat per testar si una filogènia segueix un rellotge estricte és la d'emprar un rellotge no-correlacionat amb una distribució *log-normal*, i observar el valor que es refereix a la desviació estàndard que ha estat estimada, el que equival al valor de *uclsd.stdev* al programa *Tracer*. Si aquest valor és proper a 0 llavors la filogènia és bastant *clock-like*; si per contra el valor a estimar s'allunya de 0 llavors s'assumeix que les seqüències analitzades exhibeixen una heterogeneïtat de taxes prou elevada entre els llinatges, el que fa recomanable aplicar models de rellotge relaxats. Tant pel cas de la divergència dins *M. longipes* com la divergència estudiada a tota la filogènia de Metacrangonyctidae, es testà l'homogeneïtat i la "constància" de la taxa d'evolució en el temps i entre els diferents llinatges a partir dels valors de BFs i els de *uclsd.stdev* obtinguts a *Tracer* arran



d'analitzar els arxius *.log* obtinguts mitjançant *BEAST v.1.7.2*. (*Bayesian Evolutionary Analysis Sampling Trees*) (Drummond & Rambaut, 2007).

### 3.10.1. Taxa d'evolució i temps de divergència a *M. longipes* (Capítol I)

Per estimar el temps de divergència de *M. longipes* es van explorar dues estratègies diferents. La primera fou fixar la taxa estàndard de substitució a 2.3% de divergència per parell de bases cada milió d'anys aplicada pel rellotge molecular mitocondrial a artròpodes (= 0.0115 substitucions per any i per llinatge; Brower, 1994), i per últim s'implementà a la taxa de substitució un rang de 1.4 a 2.6% substitucions per milió d'anys, prèviament estimada a decàpodes marins i aplicada també a altres crustacis (Finston et al., 2007; Knowlton & Weigt, 1998, Kornobis et al., 2010). En ambdós procediments s'emprà el set de dades complet per a *cox1* amb un rellotge molecular relaxat per inferir les dates de divergència dels diferents clades a l'arbre filogenètic emprant mètodes coalescents ja que s'empren dades poblacionals. Es va preferir aquest model ja que s'ha suggerit que els models de rellotge molecular relaxat no correlacionats resulten ésser més adequats i realistes que els rellotges estrictes pel fet que permeten que les taxes de mutació variïn entre els diferents tàxons dins un certs paràmetres (Drummond et al., 2006). Per a les anàlisis d'inferència bayesiana de rellotge molecular s'utilitzà el programa *BEAST v. 1.7.2* que implementa models de rellotge molecular *log-normal* no correlacionat (*uncorrelated log-normal clock*). Es corregueren dues anàlisis independents i s'utilitzaren els diferents models d'evolució i els esquemes de particions descrits anteriorment per a les anàlisis de *MrBayes*. Les anàlisis es començaren a córrer amb un arbre a l'atzar amb un prior de coalescència de població de mida constant. La resta de paràmetres (freqüències nucleotídiques i models de substitució entre les particions) i els models de heterogeneïtat de taxes es "deslligaren" i s'estimaren a partir de les dades. Es corregueren 20 milions de simulacions en MCMC, mostrejant cada 1,000 generacions. El prior pel rang del rellotge mitocondrial de crustacis fou implementat com a una distribució normal amb una mitja de 0.01 substitucions per anys i llinatge, i amb uns valors constrets a un màxim i mínim de 0.013 i 0.007, respectivament. Posteriorment es visualitzaren les dues anàlisis independents amb el programa *Tracer*, on s'observà la convergència i el comportament dels paràmetres estimats. En tots els casos s'observà que era suficient un *burn-in* del 10% de les generacions per estabilitzar les estimacions. El programa *FigTree v.1.3.1*. (disponible a <http://tree.bio.ed.ac.uk/software/figtree/>) serví per a la

---



visualització dels arbres obtinguts.

Per finalitzar es realitzà una altra anàlisi amb el set de dades reduït (la matriu combinada dels marcadors *cox1*, *cob* i *rrnL* per a 34 individus) on es representaven els llinatges majoritaris de *M. longipes*, fixant la taxa estàndard de substitució pel rellotge mitocondrial a artròpodes de 2.3%, però en aquest cas assumint un prior de l'arbre no coalescent sinó un que seguís el procés d'especiació de Yule (només naixement de llinatges). Tal com als models de coalescència de població de mida constant, el model d'especiació que segueix un procés de Yule consta d'un sol paràmetre, aquest darrer emprò, assumeix una taxa constant d'especiació per llinatge i s'ha suggerit ésser més adequat per filogènies a nivell d'espècie (Aldous, 2001; Drummond & Rambaut, 2007).

### 3.10.2. Estimació del temps de divergència per a la família *Metacrangonyctidae* (Capítol III)

Es va fer servir la topologia obtinguda a *MrBayes* de l'anàlisi combinada de MPCGs + *SSU* a *BEAST v. 1.7.2.* (Drummond & Rambaut, 2007) per tal d'optimitzar la longitud de branques basant-se exclusivament en el set de dades de MPCGs, atès l'excessiva dificultat per modelitzar la presència d'*indels* als alineaments de *SSU*. Les anàlisis es realitzaren amb el millor model evolutiu seleccionat segons BIC i el millor esquema de particionament de les dades segons els BFs i BIC (GTR+I+ $\Gamma$  per a les primeres/segones posicions i HKY+I+  $\Gamma$  per a les terceres posicions), assumint dos models distints: un model de diversificació de Yule (només especiació) on un segon paràmetre té en compte l'extinció, o bé el model de *Birth-Death* (creixement/mortalitat) amb un tercer paràmetre que té en compte les espècies no mostrejades. Per aquest darrer model s'empraren els priors per defecte per estimar-ne els seus paràmetres: els valors de la taxa mitjana de creixement i de mortalitat relativa s'obtingueren a partir de distribucions uniformes amb valors limitats de 0 a infinit. Addicionalment, s'assumiren tres taxes de substitució independents implementades com a tres rellotges independents (és a dir, un rellotge independent per a cada una de les tres posicions del codó) per tenir en compte els diferents nivells de saturació en cadascuna de les posicions. Per a les anàlisis es corregueren 100 milions de generacions en Cadenes de Markov Monte Carlo i es mostrejà cada 1,000 generacions. El programa *Tracer v. 1.5.* (Rambaut & Drummond, 2007) es va fer servir per resumir i visualitzar l'efecte de les

simulacions i el *burn-in*. Es van descartar els primers 10 milions de generacions (10% de *burn-in*), i es comprovà la convergència de cadascun dels paràmetres a partir del valor de Mida Efectiva de la Mostra (>200) (EES, *Effective Sample Size*). El valor de ESS indica el nombre de mostres independents per a poder arribar a una mida efectiva de mostreig adient, és a dir, indica si un paràmetre ha sigut mostrejat adequadament.

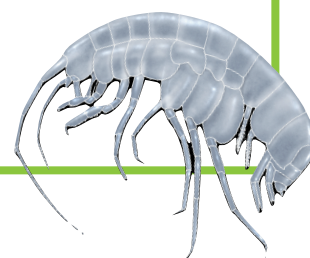
Pel calibratge s'empraren dos grans esdeveniments paleogeogràfics que influïren a la diversificació de dos llinatges de *Metacrangonyx*: el llinatge de *M. longipes* pertanyents a les Illes Balears (16-5.5 Ma) (Pomar, 1991; Krijgsman et al., 1999; Cuerda et al., 1969) i els llinatges de *M. boveei* i *M. remyi* pertanyents a la serralada de l'Alt Atlas al Marroc (37.2-25 Ma) (Frizon et al., 2008).

Es testaren alhora per *Bayes Factors* (BFs) dues hipòtesis diferents: d'una banda l'assumpció d'un esquema de rellotge estricte i d'altra banda l'assumpció d'un esquema de rellotge relaxat no correlacionat amb una distribució *log-normal* (tal com recomana Drummond & Rambaut, 2007). El resultat del test indicà que es descartava la hipòtesi del rellotge molecular estricte per inferir les dades de separació dels diferents clades a l'arbre filogenètic (BF= 167.93,  $p < 0.001$ ; valor de *ucl.d.sdv* per 1eres, 2ones i 3eres posicions= 0.228; 0.281; 0.178, respectivament). El programa *FigTree v.1.3.1.* serví per a la visualització dels arbres obtinguts.

Finalment, es realitzà una validació creuada per comprovar la congruència dels dos punts de calibratge emprats a l'anàlisi, metodologia proposada anteriorment per Near i col·laboradors (2005). Aquest mètode pren l'arbre filogenètic prèviament inferit i fixa cadascun dels punts de calibratge (dos punts, en el nostre estudi); a continuació es calcula la diferència entre les estimacions realitzades (arran de les edats obtingudes per dades moleculars i d'esdeveniments paleogeogràfics) i les edats fixades als nodes com a punts de calibratge.

---

**R E S U L T A T S**





## Filogeografia de *Metacrangonyx longipes* a les Illes Balears

### Introducció

S'estudia en aquest capítol la filogeografia de *Metacrangonyx longipes* Chevreux, 1909, un crustaci amfípode aquàtic subterrani endèmic de Mallorca i Menorca. A Mallorca, ocupa diversos tipus d'hàbitat, des de coves costaneres anquialines de salinitat quasi marina, fins a aigües dolces d'altres coves, fonts i pous situats terra endins. A Menorca, l'espècie es troba aparentment restringida a les coves i pous costaners anquialins, i és absent a les aigües subterrànies dolces de l'interior. En ambdues illes, sols es troba a les terres baixes, i falta a hàbitats aparentment adequats ubicats a altures superiors als 125 m sobre el nivell de la mar. Els Metacrangonyctidae són una família estrictament d'aigües continentals subterrànies sense parents propers coneguts, emperò diverses evidències suggereixen el seu origen marí: [1] els seus membres es troben només a regions continentals que foren cobertes per antics mars epicontinentals (Boutin, 1994; Jaume & Christenson, 2001), i [2] diverses espècies encara mantenen vincles amb el medi marí (és a dir, viuen en pous anquialins i coves a zones costaneres; Jaume & Christenson, 2001).

En aquest estudi hem utilitzat seqüències de tres gens mitocondrials i un nuclear de *M. longipes* i de tres espècies congènériques per realitzar una anàlisi filogenètica i estimar el temps de divergència de les poblacions d'aquesta espècie balear. D'altra banda, s'ha fet servir un set de dades corresponent a un major nombre d'individus, dels que s'ha seqüenciat un fragment de gen mitocondrial de la subunitat citocrom oxidasa 1 (*cox1*) per realitzar l'anàlisi filogeogràfica i examinar l'estructura poblacional d'aquest taxó. Atès la manifesta eurihalinitat de *M. longipes* i l'absència de qualsevol diferència morfològica apreciable entre les poblacions de les dues illes, la nostra predicció inicial fou que l'espècie podria haver-se dispersat arreu de les aigües subterrànies de les illes utilitzant les zones anquialines costaneres, havent colonitzat des d'enllà i de forma recurrent els hàbitats d'aigua dolça de l'interior. Si aquest fos el cas, esperaríem trobar un patró de flux genètic considerable i divergències genètiques poc importants dins de cada illa, on les poblacions de l'interior serien derivades de les

costaneres. No obstant això, el nostre estudi va revelar que aquest amfípode mostra una marcada i antiga estructura poblacional, amb episodis cladogenètics que podrien estar relacionats amb els principals canvis del nivell del mar que van afectar les illes durant els últims sis milions d'anys.

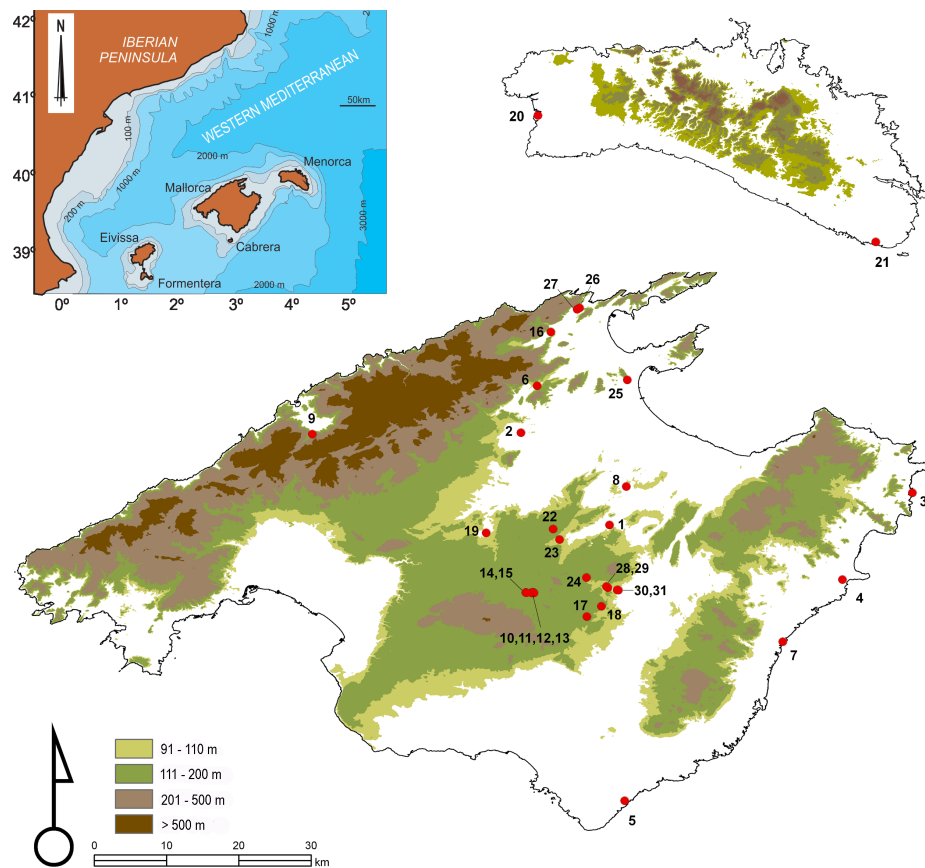
**Taula 3.1.** Localitats dels diferents mostrejos de *M. longipes* a les Illes Balears i llurs coordenades.

Etiqueta	Localitat	Illa	Latitud	Longitud	*N individus
1	Ariany (pou)	Mallorca	39° 38' 38.11"	03° 04' 51.39"	2
2	Búger (pou)	Mallorca	39° 45' 30.92"	02° 58' 14.93"	2
3	Capdepera (Cova de na Barxa)	Mallorca	39° 41' 01.98"	03° 27' 26.70"	2
4	Sant Llorenç des Cardassar (Cova de S'Abisament)	Mallorca	39° 34' 34.20"	03° 22' 13.98"	2
5	Santanyí (Cova des Dracs des Rafal des Porcs)	Mallorca	39° 18' 04.77"	03° 05' 59.31"	1
6	Campanet (pou)	Mallorca	39° 49' 00.70"	02° 59' 27.70"	1
7	Manacor (Cova de Cala Varques)	Mallorca	39° 29' 56.07"	03° 17' 46.87"	1
8	Santa Margalida (Pou d'Hero)	Mallorca	39° 41' 30.35"	03° 06' 06.96"	1
9	Sóller (Cova dels Estudiants)	Mallorca	39° 45' 24.19"	02° 42' 41.92"	2
10	Montuïri (pou A)	Mallorca	39° 33' 36.66"	02° 59' 06.00"	2
11	Montuïri (pou B)	Mallorca	39° 33' 34.38"	02° 59' 03.24"	--
12	Montuïri (pou C)	Mallorca	39° 33' 35.64"	02° 59' 09.96"	--
13	Montuïri (pou D)	Mallorca	39° 33' 34.02"	02° 59' 13.08"	--
14	Montuïri (pou E)	Mallorca	39° 33' 35.52"	02° 58' 36.90"	--
15	Montuïri (pou F)	Mallorca	39° 33' 34.56"	02° 58' 39.12"	--
16	Pollença (Pou Ca'n Colet)	Mallorca	39° 53' 00.86"	03° 00' 29.22"	1
17	Porreres (pou A)	Mallorca	39° 31' 48.48"	03° 03' 09.90"	1
18	Porreres (pou B, Pou de Son Gall)	Mallorca	39° 32' 34.44"	03° 04' 15.60"	1
19	Ruberts (pou)	Mallorca	39° 38' 03.06"	02° 55' 39.57"	2
20	Ciutadella (Cova de S'Aigua)	Menorca	39° 58' 21.48"	03° 50' 08.35"	2
21	St. Lluís (Cova de Sa Figuera)	Menorca	39° 48' 59.40"	04° 15' 20.56"	2
22	Sineu (pou A)	Mallorca	39° 38' 20.34"	03° 00' 39.00"	1
23	Sineu (pou B)	Mallorca	39° 37' 32.76"	03° 01' 07.56"	--
24	Sant Joan (pou)	Mallorca	39° 34' 42.99"	03° 03' 08.63"	2
25	Alcúdia, (Cova de Son Sant Martí)	Mallorca	39° 49' 26.59"	03° 06' 11.31"	3
26	Pollença, Cala Sant Vicenç (pou A)	Mallorca	39° 54' 48.55"	03° 02' 37.32"	1
27	Pollença, Cala Sant Vicenç (pou B)	Mallorca	39° 54' 42.72"	03° 02' 26.79"	--
28	Vilafranca de Bonany (pou A)	Mallorca	39° 34' 02.88"	03° 04' 37.45"	--
29	Vilafranca de Bonany (pou B)	Mallorca	39° 33' 57.72"	03° 04' 43.86"	--
30	Vilafranca de Bonany (pou C)	Mallorca	39° 33' 48.06"	03° 05' 25.83"	--
31	Vilafranca de Bonany (pou D)	Mallorca	39° 33' 46.86"	03° 05' 31.49"	1

\*Nombre d'individus seqüenciats pels marcadors *cob*, *cox1*, *rrnL* i *H3* emprats al set de dades combinat.

## Resultats

A un total de 34 espècimens de *Metacrangonyx longipes* i tres espècies congenèriques que s'empraren com a grups externs (*M. ilvanus*, *M. remyi* i "*M. nicoleae tamri*"), s'hi seqüenciaren quatre marcadors: tres mitocondrials (*cox1*, *cob*, *rrnL*) i un nuclear (*H3*), els quals han atès una longitud total d'1.7 Kb (la informació sobre la procedència així com el nombre d'individus seqüenciats per a tots els marcadors es troba a la Taula 3.1 i a la Figura 3.1). A més a més, es seqüencià un fragment del gen de *cox1* (aprox. 650 pb) a 162 espècimens de *M. longipes*. A la Taula 3.2 es detalla el nombre d'individus seqüenciats per a cadascuna de les localitats.



**Figura 3.1.** Mapa de les Illes Balears (Mediterrània Occidental) amb Mallorca i Menorca, on es mostra el relleu actual i la ubicació dels llocs de mostreig de *M. longipes*. Les corbes de nivell superiors a 90 m esbossarien aproximadament la paleogeografia durant el Plio-Quaternari, on tingueren lloc els principals esdeveniments de transgressió marina, en el supòsit d'escassos o nuls aixecaments o enfonsaments geològics.

Als MPCGs no s'hi observà la presència de codons d'aturada o mutacions que afectessin al marc de lectura ni pics dobles als cromatogrames, descartant així la possibilitat de tractar amb pseudogens nuclears. Les anàlisis realitzades per a cada

marcador han portat a arbres amb topologies congruents, fet reflectit pels valors positius de *Partitioned Bremer Support* (PBS) assolits per la majoria dels nodes (Figura 3.2). Alguns d'ells han mostrat un valor de PBS propers a 0, el que suggereix que el seu baix senyal filogenètic en realitat no es deuria a la incongruència entre marcadors. La majoria de la variació genètica s'ha trobat als gens mitocondrials *cox1*, *cob* i *rrnL*, que presenten 120, 78 i 43 posicions informatives sota parsimònia, respectivament. Les seqüències pel gen de Histona *H3A* representaven únicament cinc haplotips diferents, amb sis posicions informatives sota parsimònia i dues substitucions fixades a *M. longipes* en comparació de les espècies emprades com a grups externs.

**Taula 3.2.** Resum dels estadístics poblacionals pel marcador mitocondrial *cox1*.

Etiqueta	Localitat mostreig	Nombre individus	Nombre haplotips	$H_d \pm SD$	$\pi \times 10^2 \pm SD$
1	Ariany	6	H17 (3), H18 (2), H20 (1)	0.733±0.155	0.480±0.334
2	Búger	4	H11 (1), H12 (2), H13 (1)	0.833±0.222	0.235±0.210
3	Capdepera	19	H2 (15), H3 (4)	0.351±0.111	0.055±0.063
4	Sa Coma (Sant Llorenç des Cardassar)	4	H43 (1), H44 (1), H45 (2)	0.833±0.222	0.157±0.155
5	Santanyí	6	H40 (6)	0.000±0.000	0.000±0.000
6	Campanet	2	H6 (1), H9 (1)	1.000±0.500	0.313±0.383
7	Manacor	8	H41 (7), H42 (1)	0.250±0.180	0.039±0.056
8	Santa Margalida	12	H25 (10), H26 (2)	0.303±0.148	0.047±0.059
9	Sóller	14	H37 (8), H38 (4), H39 (2)	0.615±0.102	0.110±0.099
10	Montuïri	3	H27 (1), H29 (1), H33 (1)	1.000±0.272	0.417±0.375
11	Montuïri	3	H27 (2), H32 (1)	0.667±0.314	0.313±0.295
12	Montuïri	1	H30 (1)	NC	NC
13	Montuïri	1	H49 (1)	NC	NC
14	Montuïri	3	H27 (1), H29 (1), H50 (1)	1.000±0.272	0.313±0.295
15	Montuïri	1	H31 (1)	NC	NC
16	Pollença	1	H1 (1)	NC	NC
17	Porreres	1	H34 (1)	NC	NC
18	Porreres	1	H35 (1)	NC	NC
19	Ruberts	7	H27 (2), H28 (1), H29 (2), H48 (2)	0.857±0.102	0.283±0.212
20	Ciutadella	5	H14 (4), H15 (1)	0.400±0.237	0.063±0.080
21	Maó	13	H16 (13)	0.000±0.000	0.000±0.000
22	Sineu	2	H19 (2)	0.000±0.000	0.000±0.000
23	Sineu	2	H19 (2)	0.000±0.000	0.000±0.000
24	Sant Joan	5	H21 (5)	0.000±0.000	0.000±0.000
25	Alcúdia	23	H4 (1), H5 (20), H8 (1), H46 (1)	0.249±0.117	0.245±0.169
26	Sant Vicenç (Pollença)	4	H7 (3), H10 (1)	0.500±0.265	0.391±0.315
27	Sant Vicenç (Pollença)	1	H47 (1)	NC	NC
28	Vilafranca de Bonany	6	H22 (4), H23 (1), H36 (1)	0.600±0.215	1.701±1.043
29	Vilafranca de Bonany	1	H22 (1)	NC	NC
30	Vilafranca de Bonany	1	H24 (1)	NC	NC
31	Vilafranca de Bonany	2	H22 (2)	0.000±0.000	0.000±0.000

$H_d$ : Diversitat haplotípica

$\pi$ : Diversitat nucleotídica



i. Anàlisis filogenètiques basades en les seqüències de *cox1*, *cob*, *rrnL* i *H3*

Tant els valors de BIC (*Bayesian Information Criterion*) com els de BFs (*Bayes Factors*) indicaren que el millor esquema de particionament en el set de dades combinat era el de quatre particions: (1a) primeres i segones posicions de *cox1* i *cob*, (2a) terceres posicions de *cox1* i *cob*, (3a) *rrnL* i (4a) *H3A*; mentre que en el cas del set de dades per a *cox1* sortí afavorit el de dues particions: (1a) primeres i segones posicions més (2a) terceres posicions (Taula 3.3).

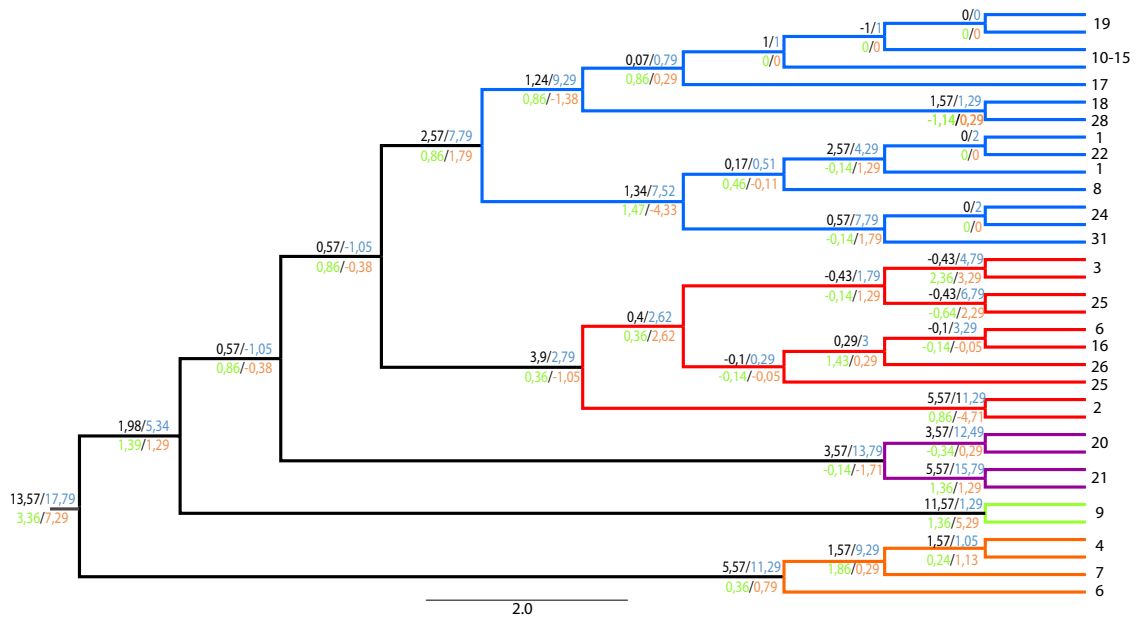


Figura 3.2. Valors obtinguts de PBS per a cadascun dels marcadors emprats del set de dades combinat: *cob* (negre), *cox1* (blau), *H3* (verd) i *rrnL* (taronja).

Segons els resultats obtinguts a Modeltest, els models que s'ajustaren millor a les dades segons el valor de BIC calculats van ésser: HKY+I per a les primeres i segones posicions dels MPCGs *cox1* i *cob*, GTR+G per a les terceres posicions dels MPCGs, HKY+I per a *rrnL*, i F81+I per a *H3*.

Taula 3.3. Valors de BIC i BF pels diferents esquemes de particions emprats a les anàlisis.

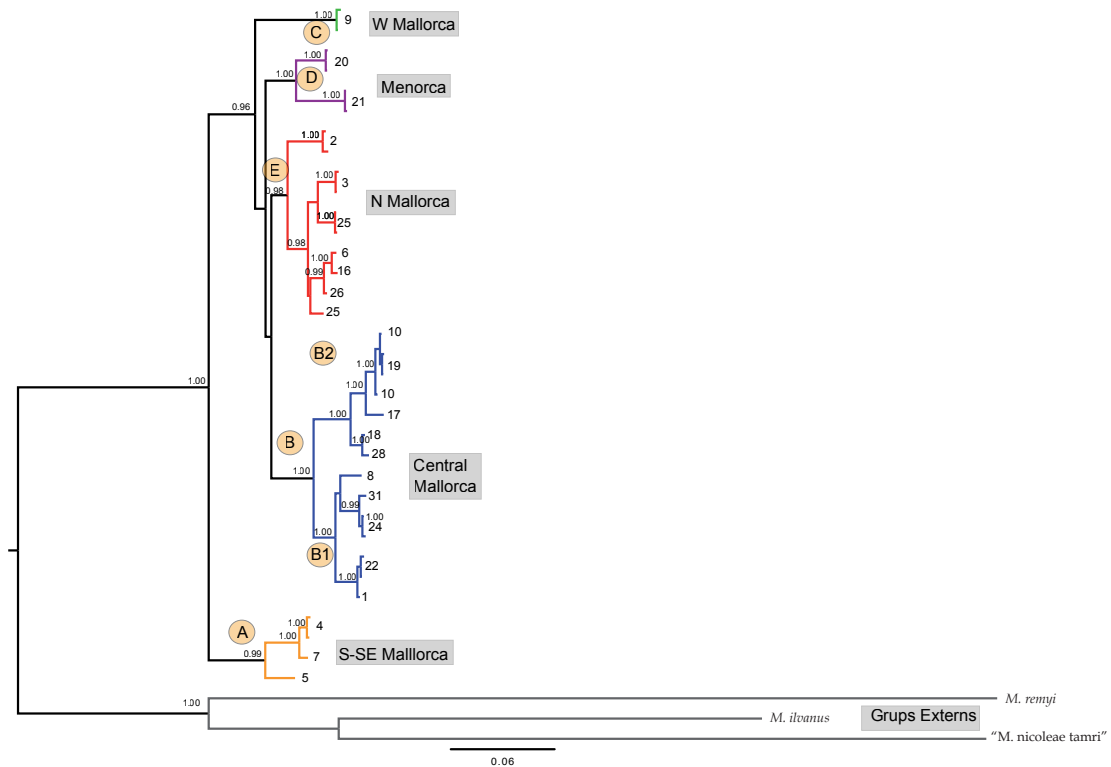
Nombre de particions	Esquema de particions	Ln*	BIC	LnP**	2 LnBF
Set de dades combinat					
0		-7809,32	15631,88	-7892,68	-344,794
2	<i>cob+cox1+rrnS/H3</i>	-7666,03	15355,29	-7783,36	-249,842
3	<i>cob+cox1/rrnS/H3</i>	-7630,10	15293,43	-7749,93	-220,806
4	1+2 mtPCGs/3 mtPCGs/ <i>rrnS/H3</i>	<b>-7336,08</b>	14715,39	<b>-7495,72</b>	n.a.
4	<i>cob+cox1+rrnS+H3</i>	-7606,15	15255,54	-7739,73	-211,95
Set de dades per a <i>cox1</i>					
0	<i>cox1</i>	-3341,80	6698,40	-3500,63	-269,94
2	1+2 <i>cox1</i> vs. 3 <i>cox1</i>	<b>-3117,74</b>	6262,28	<b>-3264,59</b>	n.a.

\* Valor obtingut amb RAxML.

\*\* Valor de LnP (Model | Dades) obtingut amb MrBayes.

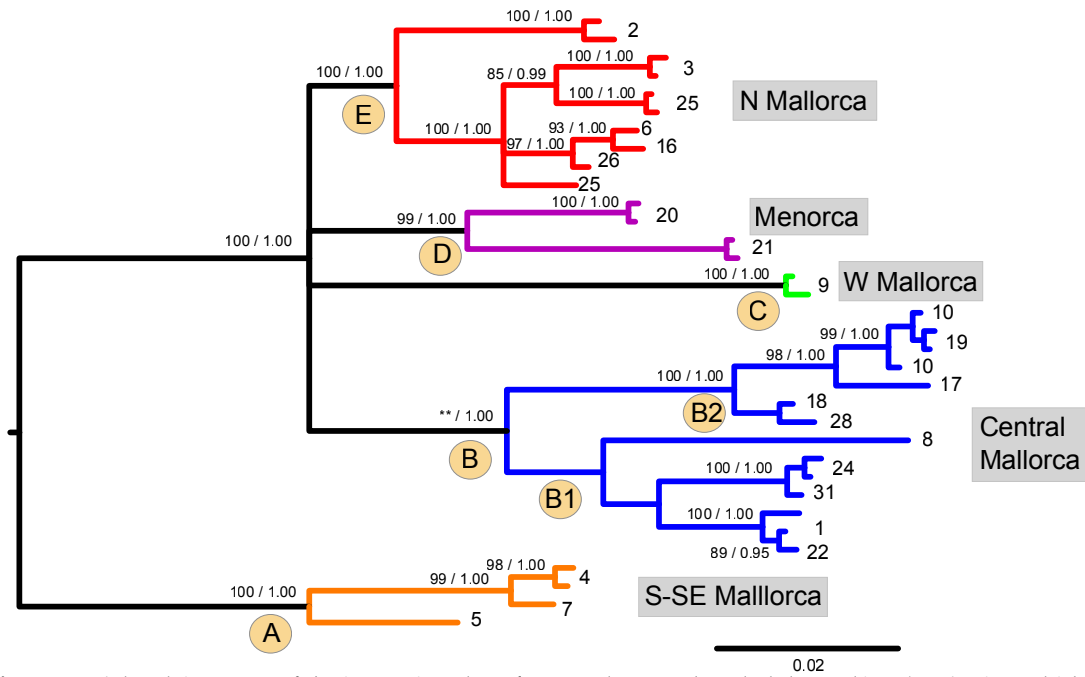
En negreta els valors que tenen el valor més alt de versemblança i que segons BIC i BF són els millors models.

La topologia obtinguda a les anàlisis assenyala la monofília del llinatge *M. longipes* (Fig. 3.3). Vist que els grups externs emprats són molt distants del llinatge de *M. longipes*, s'han repetit totes les anàlisis sense incloure aquests atès que s'ha observat que l'ús de grups externs molt distants pot conduir a topologies errònies o bé a l'obtenció de valors de suport inferiors (Wheeler, 1990; Smith, 1994; Gatesy et al., 2007; Schneider & Cannarozzi, 2009; Rosenfeld et al., 2012).



**Figura 3.3.** Arbre obtingut per inferència bayesiana de *M. longipes* basat en el set de dades combinat (DNA mitocondrial + nuclear) amb els grups externs. Els valors dels nodes corresponen als valors de probabilitat posterior (sols representats >0.95).

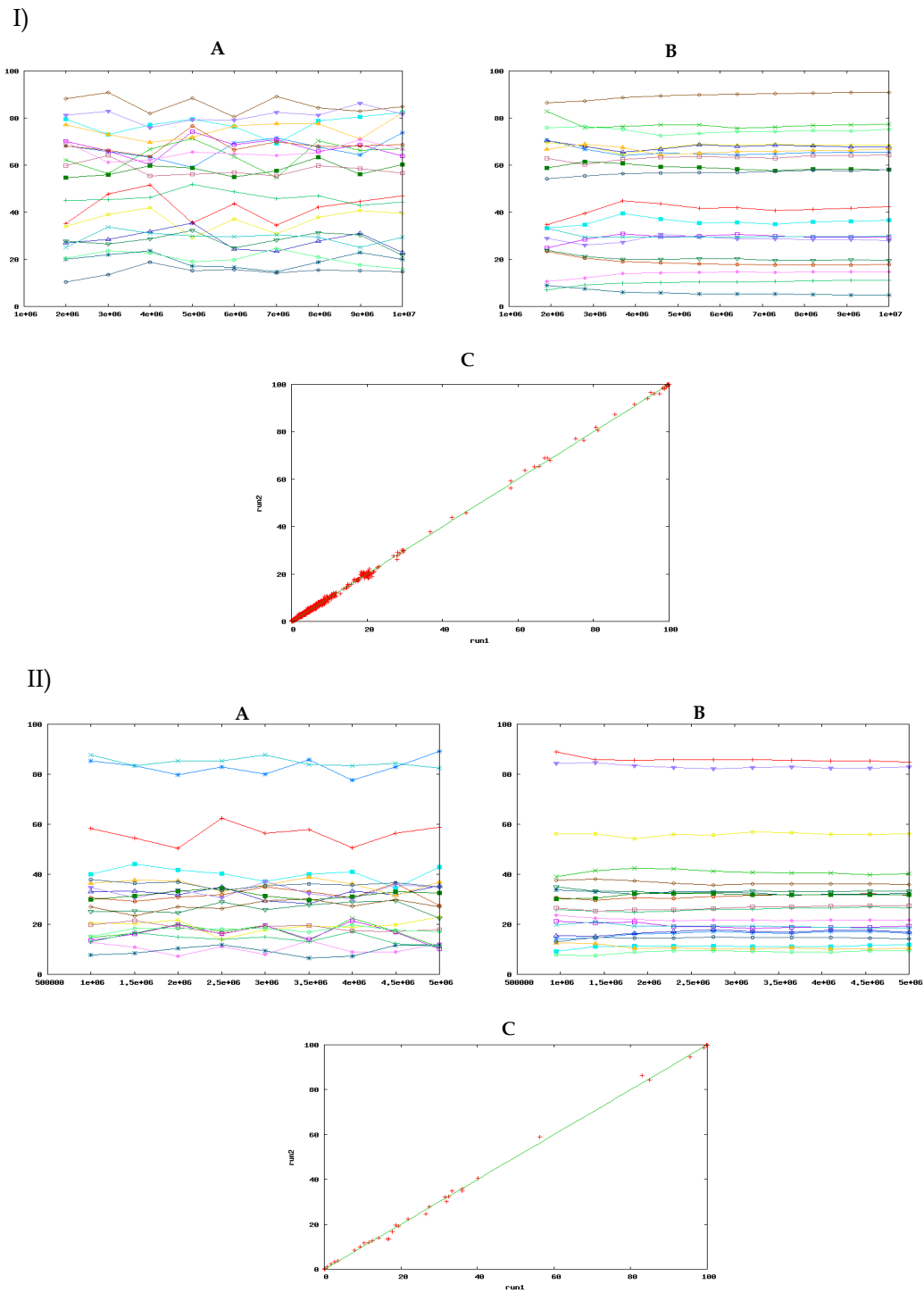
Tant les anàlisis realitzades per Inferència Bayesiana (BI) com aquelles efectuades amb la metodologia de Màxima Versemblança (ML) pel set de dades combinat i els set de dades complet per a *cox1*, han aportat topologies molt semblants, on s'hi han detectat cinc llinatges monofilètics altament divergents els quals no presenten cap tipus de solapament geogràfic (veure la Figura 3.1. per als corresponents llocs de mostreig, i la Figura 3.4). El clade A agrupa les poblacions habitants de les coves anquihalines de la part més meridional de Mallorca (clade A: localitats 4, 5 i 7), que al seu torn són les més divergents a dins del llinatge de *M. longipes*, i representen el grup basal de la filogènia amb un valor màxim de suport estadístic (100%-1.00 com a valors de *bootstrap* a l'anàlisi ML i valors de probabilitat posterior, *-pp-* a l'anàlisi de BI, respectivament).



**Figura 3.4.** Arbre obtingut per Inferència Bayesiana de *M. longipes* basat en el set de dades combinat (DNA mitocondrial + nuclear). Els valors dels nodes representats corresponen als valors de bootstrap (sols representats els valors >85) obtinguts a les anàlisis de Màxima Versemblança (primer número) i als valors de probabilitat posterior obtinguts per Inferència Bayesiana (sols representats els valors > 0.95) (segon número).

La resta de poblacions mostrejades formen quatre clades en gran manera suportats, malgrat les principals relacions evolutives entre aquests no s'han aconseguit resoldre del tot (Fig. 3.4): el clade D (conformat per les dues úniques poblacions menorquines de les coves anquialines 20 i 21; de Ciutadella i Maó, respectivament); dos altres clades altament divergents, el clade C, localitzat a la part occidental de Mallorca (corresponent a la cova d'aigua dolça número 9 a Sóller), i el clade E, conformat per les poblacions més septentrionals de l'illa de Mallorca (corresponents a coves anquialines i pous d'aigua dolça de les localitats 2, 3, 6, 16 i 25-27; localitzades a Campanet, Pollença i Cala Sant Vicenç); i per últim el clade B, el qual agruparia a tots aquells pous d'aigua dolça situats al centre de Mallorca lluny de la línia de costa, i que es subdividiria en dos grups gènics diferents: el clade B1 (localitats 1, 8, 22, 24 i 31; d'Ariany, Santa Margalida, Sineu, Sant Joan i Vilafranca -pou D-, respectivament) i el clade B2 (localitats 10, 17-19 i 28; de Montuïri, Porreres, Ruberts i el pou A de Vilafranca) deixant entreveure una línia de segregació geogràfica amb orientació NE-SO.

A la Figura 3.5 es mostren els resultats obtinguts de *AWTY* per a ambdues anàlisis bayesianes (set de dades complet per a *cox1* i el set de dades combinat per a tots els marcadors), reflectint-se que s'ha arribat a l'estacionalitat i a la convergència de les biparticions realitzades.

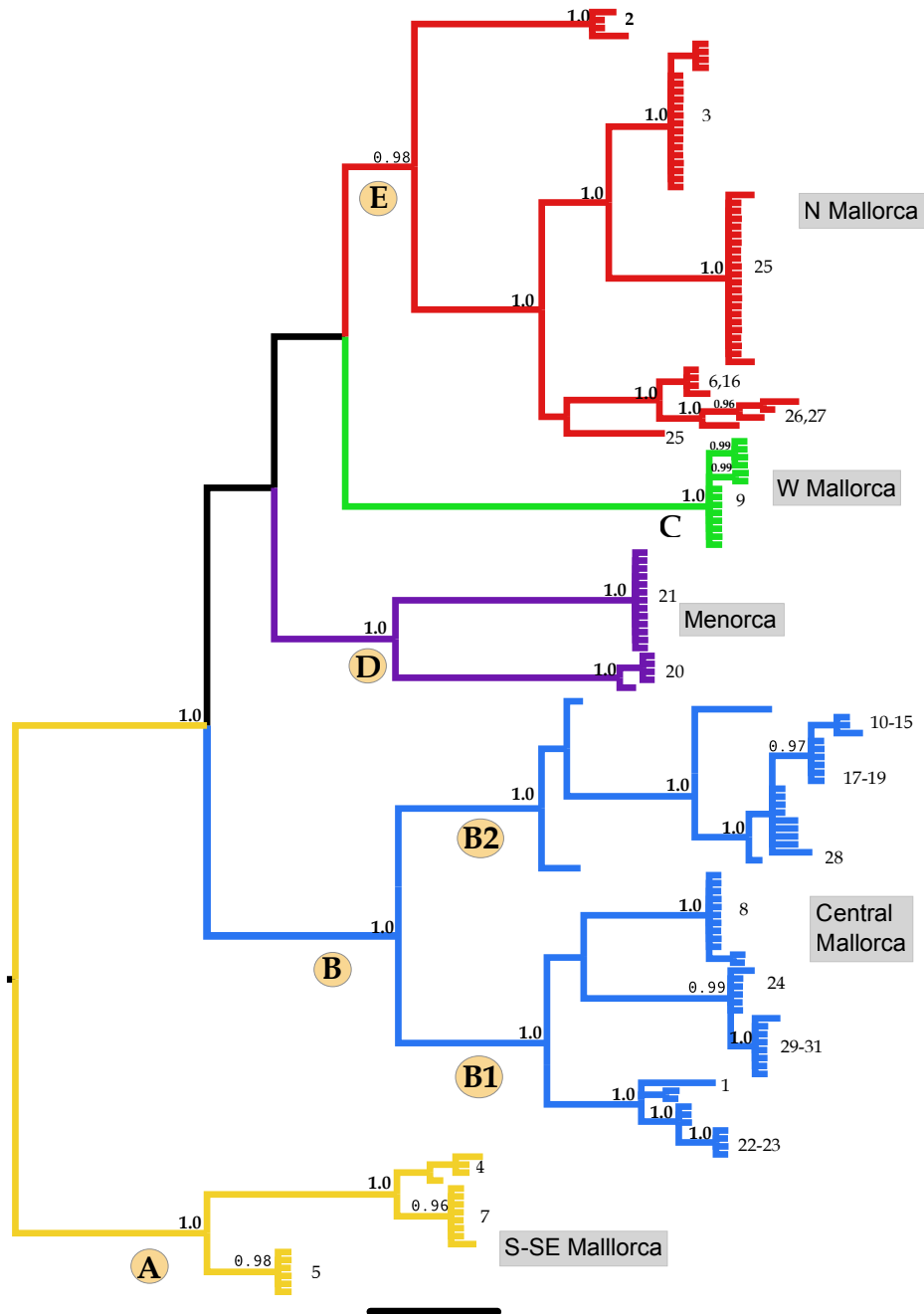


**Figura 3.5.** Estacionalitat i convergència de la topologia obtinguda a MrBayes pel set de dades per a *cox1* (I) i per al combinat (II) a *M. longipes* calculades per AWTY (Nylander et al., 2008), amb els ordres slide (A), cumulative (B) i compare (C).

## ii. Anàlisi filogenètiques i distàncies genètiques basades en *cox1*

Un altre set de dades elaborat a partir d'un mostreig faunístic més extens, i constituït pel fragment del gen per a *cox1* de 162 espècimens, ha estat emprat per a l'estudi de l'estructura poblacional (Taula 3.2). Ambdues anàlisis, realitzades amb les metodologies de BI i ML, han aportat arbres filogenètics compatibles amb l'obtingut a

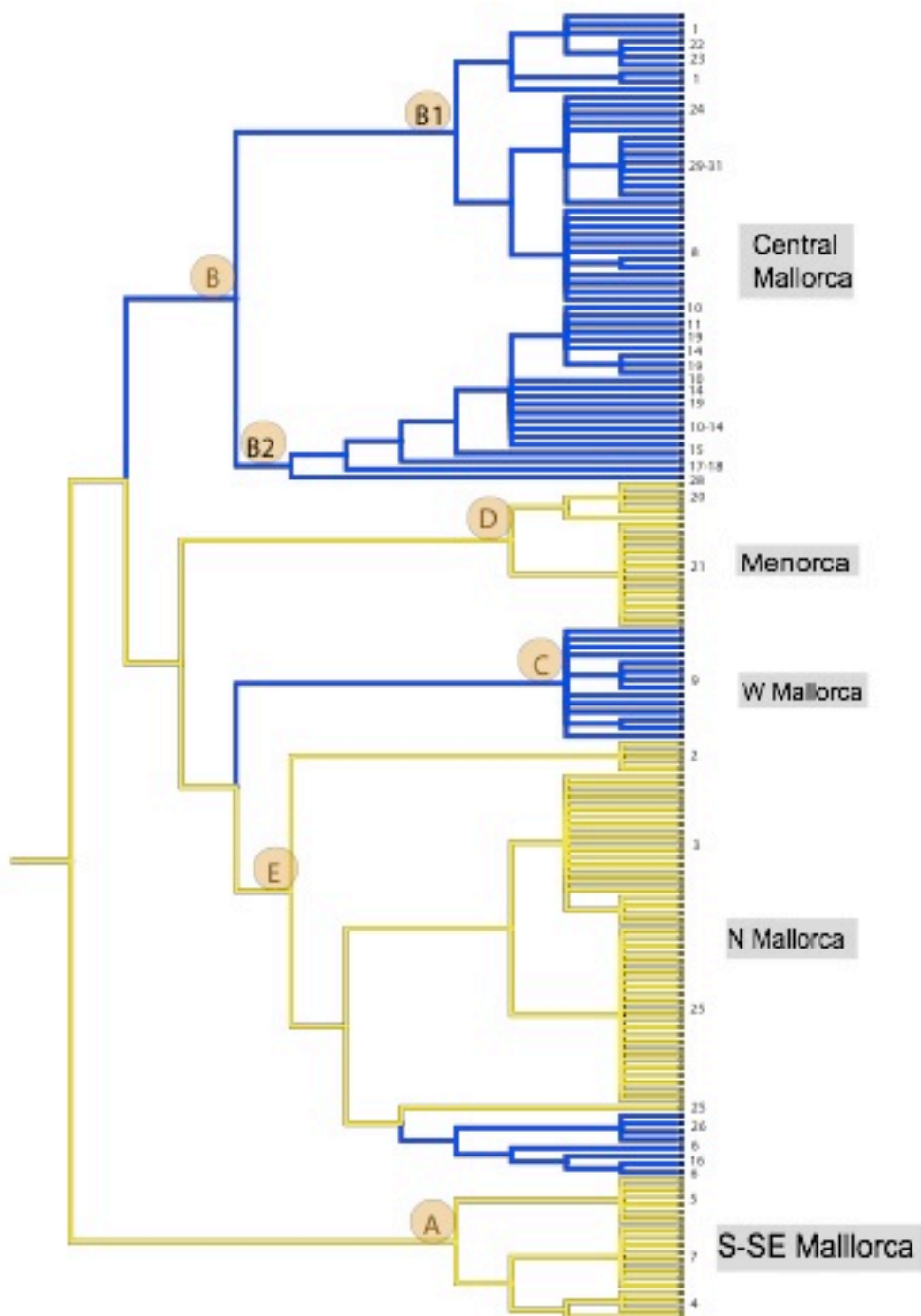
partir de l'anàlisi combinada esmentada al paràgraf anterior (Figura 3.6). No obstant això, encara romanen sense resoldre satisfactòriament (valors de suport baixos als nodes) les relacions filogenètiques entre els clades B-C-D-E, com el cas de les relacions entre les poblacions de Menorca i Mallorca. Els suports corresponents a les relacions filogenètiques entre els clades E i C han resultat ser un poc més elevats.



**Figura 3.6.** Arbre filogenètic obtingut per Inferència Bayesiana de *M. longipes* basat en el set de dades per a *cox1*. Els valors dels nodes representats corresponen als valors de probabilitat posterior obtinguts per Inferència Bayesiana (sols representats els valors > 0.95).

La reconstrucció parsimònica de l'estat dels caràcters pel que fa a l'habitat dels nodes interns (aigües dolces vs. aigües salobres) i basat en la matriu per a *cox1*, apunta

a que han ocorregut com a mínim tres transicions cap a hàbitats d'aigües dolces des d'hàbitats anquialins salobres (Figura 3.7). S'observa també que l'hàbitat ancestral, que és el corresponent a l'ancestre comú més recent de *M. longipes*, és el salobre hipogeu, que també era el corresponent a la majoria de nodes interns.



**Figura 3.7.** Reconstrucció parsimònica de l'estat dels caràcters quant a l'hàbitat de les poblacions de *M. longipes* (aigües dolces -blau- vs. aigües salobres -groc-), basada en el set de dades per a *cox1*.

S'han trobat 78 posicions nucleotídiques variables en els 650 pb de *cox1* analitzats (17.72%). El 86.66% de les posicions informatives corresponen a terceres posicions del

codó, un 2.66% a les segones posicions i el 10.66% a les primeres.

Les distàncies no corregides per a la matriu per a *cox1* entre localitats de mostreig ha donat un rang mínim de 0.5% entre aquelles poblacions situades a llocs molt propers (per exemple, les localitats 18 i 28, separades tan sols per 3.5 km) fins a un màxim de 8.9% (la distància corregida emprant un model GTR donaria 9.8%) per aquelles poblacions situades a les dues illes diferents (localitats 4 i 20, separades per 68 km), o bé entre algunes de les poblacions mallorquines. Segons els resultats obtinguts a les anàlisis filogenètiques, el clade A seria el llinatge més divergent, amb un valor de 7.8-8% de distància genètica (no corregida) amb respecte a la resta de clades, mentre que la distància entre els altres clades seria d'un 6.3-7.5%. La distància entre els subclades B1 i B2 ha resultat ésser d'un 5.5%.

### *iii. Estructura poblacional i diversitat genètica*

S'han aconseguit identificar un total de 50 haplotips diferents per a *cox1* entre els espècimens de les 31 poblacions analitzades (nombre d'accés EMBL: FR729731-FR729892) (Taula 3.2). Quatre d'aquests haplotips són compartits per poblacions veïnades: els haplotips H27 i el H29, que apareixen plegats a diverses localitats de Mallorca (Montuïri/Ruberts; localitats 10, 11, 14 i 19), l'haplotip H19, present a dos pous propers de Sineu (localitats 22 i 23), i l'haplotip H22, que es troba a varis pous de Vilafranca (localitats 28,29 i 31). Els 46 haplotips restants són únics per a cadascuna de les localitats. La Taula 3.2 resumeix la diversitat estàndard intrapoblacional estimada per a *cox1*. Sis poblacions inclouen tant sols un haplotip ( $H_d=0$ ), tot i que cal esmentar que a tres d'aquestes poblacions solament s'hi han pogut col·lectar i analitzar dos espècimens. El contrari ha succeït a tres poblacions de Mallorca, les quals han mostrat un índex de diversitat màxima pel fet que cada individu portava un haplotip diferent ( $H_d=1$ ). La resta de poblacions han mostrat valors de  $H_d$  de baixos a moderats, en un rang de 0.25 a 0.86. El grau de diversificació trobada a les dues poblacions menorquines és molt menor (tres haplotips per a 28 individus) que a les poblacions de Mallorca (47 haplotips per a 144 individus). La diversitat nucleotídica ( $\pi$ ) és bastant baixa a la majoria de les localitats ( $\pi < 0.5\%$ ); la població 28 per si sola ja exhibeix un valor de  $\pi > 1\%$  a causa de la presència d'un individu portador d'un haplotip altament divergent.



Els tests de Neutralitat han resultat ésser no significatius per a tots els casos amb l'excepció de les poblacions 7 ( $F_s$  de Fu = -0.182,  $P < 0.05$ ) i 8 (Ramos-Onsins & Rozas  $R_2 = 0.51$ ,  $P < 0.01$ ).

Adicionalment s'ha realitzat un test de Mantel el qual ha detectat una baixa correlació entre la distància genètica i la geogràfica ( $r = 0.467$ ,  $P < 0.001$ ) a les poblacions balears. Els valors de  $F_{ST}$  per a *cox1* són generalment alts ( $F_{ST} = 0.5-0.99$ ) i altament significatius, amb l'excepció d'aquelles poblacions molt properes geogràficament (la localitat 19 amb les 10-15, i la 24 vs. 28). Les anàlisis de SAMOVA indiquen que el nombre òptim de grups de poblacions necessari per maximitzar el valor de  $F_{CT}$  és  $K = 19-20$  ( $F_{CT} = 0.961-0.963$ ), mentre que els valors corresponents a  $F_{SC}$  han estat dels més baixos testats de la sèrie  $K$ , tal com era d'esperar (Dupanloup et al., 2002). Aquesta anàlisi ha agrupat els pous 10-15 amb el 19 com una mateixa població, mentre que els pous de Campanet (6) i els de Pollença (16) per una part, i els pous de Vilafranca (28-31) i de Cala Sant Vicenç (26 i 27) per l'altra, formarien tres poblacions diferents. La majoria de les localitats que conformarien una mateixa població es troben separades per menys de 2 km, amb l'excepció de les localitats 10 i 19, les quals disten uns 10 km una de l'altra. La resta de localitats representen per elles soles un grup poblacional distint, fet que assenyala una alta subdivisió poblacional, amb el poc flux gènic detectat ocurrent només entre pous que disten a menys de 10 km l'un de l'altre. Adicionalment, amb l'anàlisi d'AMOVA, que permet el reconeixement de grups poblacionals amb qualsevol nombre de  $K$  -a diferència de SAMOVA on el nombre limitant de  $K$  és 20-, es va tractar de detectar una estructura poblacional alternativa amb alts valors de  $K$ . Aquesta anàlisi ha revelat una estructura poblacional similar a l'anterior, tot i que ha identificat els pous 16, 19 i 26 com a poblacions diferenciades; ha presentat uns valors de  $F_{CT}$  que s'han atansat a un màxim de 0.96 a  $K = 19-27$ , amb el valor màxim assolit a  $K = 23$ . Tant és així que les úniques localitats que quedarien agrupades en la mateixa població serien les de Montuïri d'una banda, i les de Vilafranca de Bonany per l'altra; mentre que la resta de localitats correspondrien a poblacions diferenciades.

- iv. Estimació de les edats mitjançant models de coalescència amb població constant i de diversificació sota un model només de naixement de llinatges (Yule)*
-

Primer de tot es descartà l'aplicació d'un rellotge molecular estricte tant al set de *cox1* de 162 individus com al set de dades combinat (és a dir, *cox1-cob-rrnL*) de 34 individus. Pel set de dades de *cox1*, els valors de  $2xlnBF$  (Kass & Raftery, 1995; Nylander et al., 2004) obtinguts amb *Tracer* arran de comparar els diferents models de rellotge relaxat vs. els models de rellotge estricte caigueren dins un rang de 50-52.5, mentre que el valor d'*uclid.std* fou superior a 0.3 a tots els models de rellotge relaxats testats. Pel set de dades combinat, els valors de  $2xlnBF$  de diferents models relaxats vs. models de rellotge estricte caigueren en el rang de 21.2-22.5, i el valor d'*uclid.std* fou superior a 0.39 a tots els models relaxats.

**Taula 3.4.** Estimació del temps de calescència.

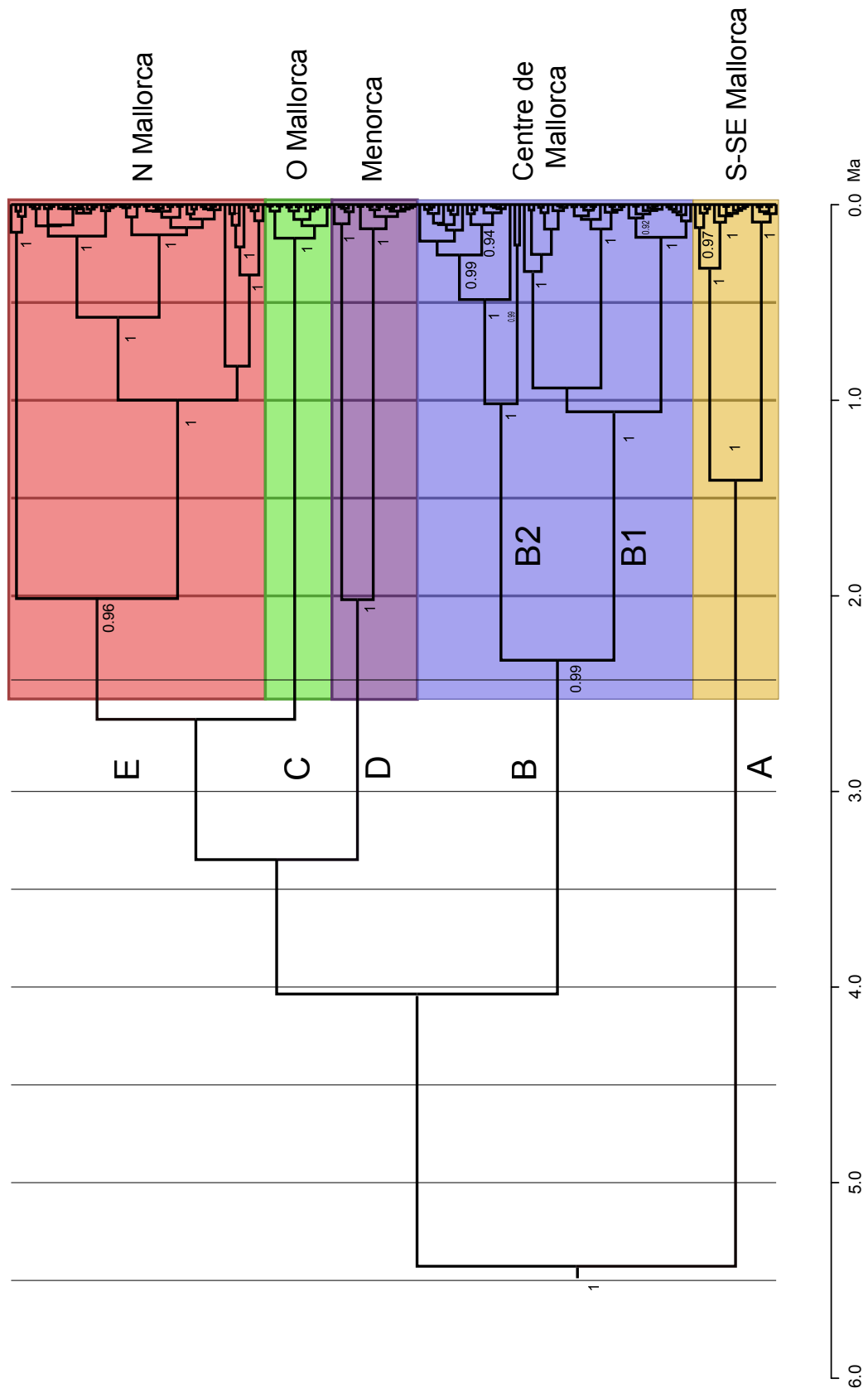
Clade	Model de Coalescència <i>cox1</i>		Model de Yule <i>cox1+ cob</i>
	Rang per a Crustacea 1.4-2.6%	Taxa fixada a 2.3% per Arthropoda	
<b>Node A</b>	1.57 (0.76-2.48)	1.37 (0.64-2.18)	1.83 (1.24-2.48)
<b>Node B</b>	2.66 (1.68-3.70)	2.33 (1.44-3.19)	2.31 (1.78-2.83)
<b>Node C</b>	0.19 (0.05-0.36)	0.17 (0.05-0.32)	0.11 (0.02-0.22)
<b>Node D</b>	2.35 (1.14-3.49)	2.07 (1.04-3.08)	2.11 (1.47-2.77)
<b>Node E</b>	2.36 (1.46-3.27)	2.04 (1.24-2.83)	2.26 (1.68-2.88)
<b>Node B1</b>	1.30 (0.76-1.84)	1.14 (0.68-1.62)	1.17 (0.83-1.54)
<b>Node B2</b>	1.21 (0.68-1.22)	1.05 (0.59-1.56)	1.07 (0.74-1.41)
<b>Arrel</b>	6.22 (4.10-8.70)	5.38 (3.45-7.46)	5.83 (4.46-7.09)

*Edat estimada per als nodes principals de la figura 3.8, a partir d'un rellotge molecular bayesià relaxat no correlacionat assumint un model d'especiació de Yule, o bé un model de coalescència amb població de mida constant i taxes de calibratge alternatives.*

La coalescència de les seqüències mitocondrials de *M. longipes* s'estimà per inferència bayesiana emprant el set de dades complet de 162 individus per a *cox1*, i implementant un model de rellotge molecular relaxat sota un model de coalescència. Es realitzaren dues anàlisis amb diferents taxes; per una banda es fixà una taxa de substitució de 0.0115 per any i per llinatge (Brower, 1994), i per altra banda, s'assumí una taxa amb un rang de 0.007-0.013 –rang estimat anteriorment a crustacis (Knowlton & Weigt, 1998)-. Les edats de l'arrel de l'arbre han resultat caure en un rang de 5.4 a 6.2 Ma segons la taxa assumida, mentre que l'edat dels altres nodes ha sigut notablement similar a les dues anàlisis; no obstant això, el rang per a la taxa mitocondrial per a crustacis ha donat edats lleugerament més antigues amb els rangs d'interval de confiança més amplis (Taula 3.4 i Fig. 3.8). Les estimacions realitzades amb el set de dades combinat (*cox1 + cob*) sota un model d'especiació de Yule i fixant la taxa de substitució a 2.3% són compatibles amb les estimacions obtingudes amb el set de dades

per a *cox1* (Taula 3.4 i Figura Addicional 1). Pel que fa al model de coalescència, la divergència del clade B mallorquí (format per les poblacions del centre de l'illa) data de 2.3-2.7 Ma, mentre que el clade E (format per les poblacions del Nord i Nord-est de l'illa) sembla haver divergit fa uns 2.0-2.4 Ma. En ambdós casos, el 95% del valor de densitats de probabilitat més alta (HPDs, de l'anglès *highest posterior densities*) han caigut en un rang de 3.7-1.2 Ma (Taula 3.4 i Fig. 3.8). El node corresponent al clade D (Menorca) data de 2.1-2.3 Ma, mentre que el clade A (poblacions localitzades al sud i sud-est de Mallorca) data de 1.4-1.6 Ma (amb un valor de HPD del 95% a 3.5-0.6 Ma en ambdós casos). Els nodes corresponents als dos subclades germans del centre de Mallorca, el clade B1 i B2, s'han datat en 1.1-1.2 i 1.0-1.3 Ma, respectivament (HPD del 95%: 1.8-0.6 Ma). En canvi, la coalescència de les seqüències monofilètiques de les coves i pous de Mallorca rep datacions molt més recents, amb estimacions al voltant de 0.1-0.2 Ma.

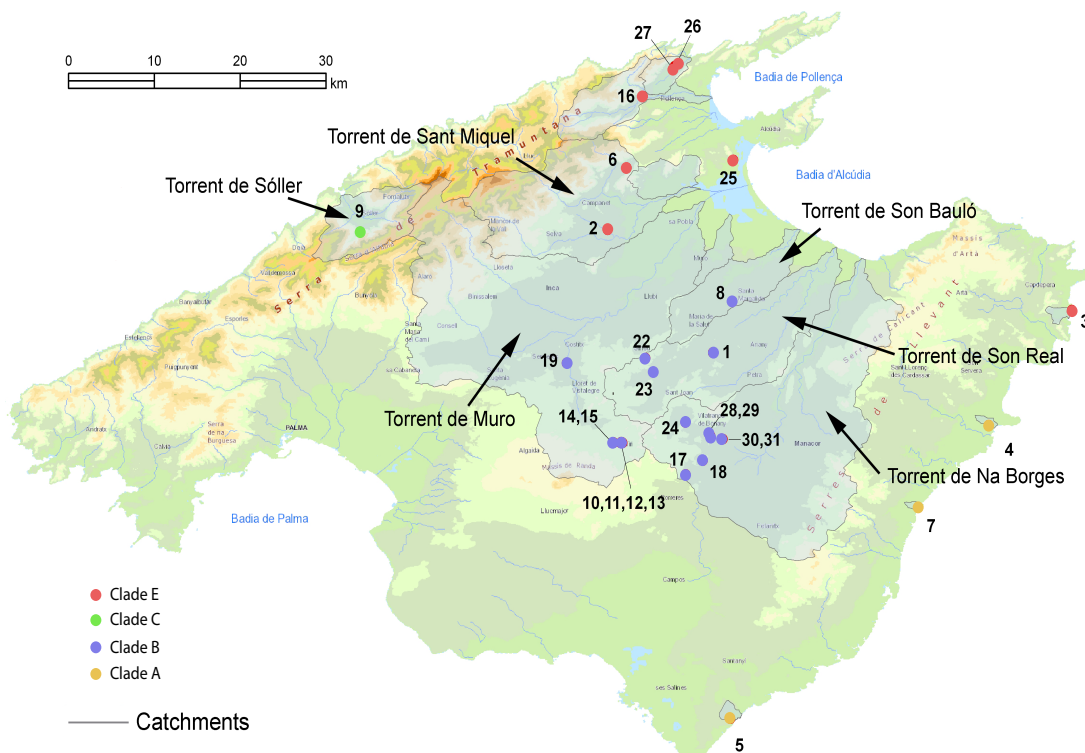
---



**Figura 3.8.** Cronograma basat en l'arbre ultramètric bayesià amb el set de dades per *cox1* de *Metacrangonyx longipes* obtingut a partir d'un rellotge relaxat no correlacionat amb distribució log-normal i baix un model de coalescència amb població de mida constant. La datació dels clades principals es realitzà baix l'assumpció d'una taxa fixa de substitució del 2.3% per parell de bases per milió d'anys.

## Discussió

La condició d'organisme talassoide (= d'origen marí) atorgada prèviament a *Metacrangonyx longipes* ha quedat provada en aquesta tesi; així com s'ha pogut deduir que al menys tres episodis independents de colonització del medi dulciaqüícola subterrani a partir d'ancestres anquialins, i no viceversa, han tingut lloc a dins del llinatge (Fig. 3.7). És difícil datar amb precisió els dos episodis de colonització d'aigües dolces per part de les poblacions del centre i la regió més occidental de Mallorca atès que podrien haver-se donat en un espai de temps molt curt o sincrònicament, tal com queda reflectit amb la politomia vertadera obtinguda per aquests nodes (Fig. 3.2). No obstant això, el darrer i més recent, per part de les poblacions originàries del nord de Mallorca (localitats 6, 16 i 26 de Pollença, Campanet i Cala Sant Vicenç, respectivament) dataria de fa poc més d'un milió d'anys, quan aquests llinatges quedaren aïllats de la resta de poblacions del nord de l'illa habitants d'ambients salobres.



**Figura 3.9.** Mapa de l'illa de Mallorca on es mostra la correspondència entre les conques hidrogràfiques i la distribució dels principals clades registrats de *M. longipes* segons la informació filogenètica extreta del set de dades mitocondrial. No es mostra el mapa de les conques hidrogràfiques de Menorca atès que les dues úniques localitats de mostreig provenen de coves anquialines. La línia discontinua grisa marca el límit de les conques.

Tot i que les diferents poblacions de *M. longipes* analitzades manquen de cap

diferenciació morfològica rellevant, gràcies a l'ús de marcadors moleculars ha estat possible la seva diferenciació observant-se una elevada estructura poblacional, reflectida en cinc grans llinatges notablement divergents. Sorprenentment, s'ha observat que el flux gènic entre les poblacions d'aquesta espècie sembla no excedir els 10 km, i normalment es limita a ocórrer entre aquelles poblacions que no disten més de 2 km l'una de l'altra. Per tant, aquests resultats no sostenen la hipòtesi d'una colonització activa dels hàbitats subterranis d'aigua dolça a partir de l'expansió crevicular/intersticial d'ancestres marins litorals; no obstant això, no es pot descartar per complet l'existència d'episodis esporàdics dispersius que hagin pogut donar-se arran de l'establiment de condicions favorables al passat (Danielopol, 1980; Rouch & Danielopol, 1987; Danielopol & Bonaduce, 1990). De tota manera, si aquest fos el cas, s'hauria d'haver trobat alguna evidència de connectivitat entre les diferents poblacions de *M. longipes* habitants d'aigües dolces a l'interior de l'illa de Mallorca i aquelles habitants del medi anquialí costaner, i també entre les poblacions anquialines entre elles.

Un altre tret important a esmentar és que alguns dels clades de *M. longipes* es troben lligats a conques hidrogràfiques particulars i no s'han trobat a cap altre lloc (Figura 3.9). De fet, els membres del clade C es trobaren exclusivament a la conca del Torrent de Sóller, mentre que els del clade B2 únicament a la capçalera del Torrent de Muro (localitats 10-15 i 19) i en algunes estacions properes de la conca del Torrent de Na Borges (localitats 17-18 i 28). Així mateix, el clade B1 (localitats 1, 8, 22-24 i 28-31) es troba localitzat a la capçalera de tres conques hidrogràfiques diferents: el Torrent de Na Borges, Son Bauló i Son Real; de tota manera, durant els últims períodes glacials quan el nivell de la mar era més baix i la plataforma somera entre Mallorca i Menorca quedà emergida total o parcialment, aquests tres torrents confluïen recurrentment fins a formar un únic sistema de paleodrenatge. Tampoc es descarta que actualment es poguessin donar certs esdeveniments aïllats de dispersió a través de l'interfluvi de conques hidrogràfiques adjacents, vist que les planes al·luvials on es troba l'interfluvi del Torrent de Muro i el de Na Borges, alberguen petits aquífers poc profunds que probablement, formen un continu durant l'hivern, quan l'àrea queda parcialment inundada. Aquests resultats suggereixen per tant que la fragmentació-divisió d'aquestes conques i el desplaçament al llarg del medi hiporreic (aigües dels llits de rius i planes al·luvials) associat a aquests cursos d'aigua, és clau per explicar l'estructura poblacional observada a *M. longipes*.

Fora bo esmentar aquí que el treball realitzat per Cooper et al. (2007) vers els amfípodes estigobionts hiàlids i crangoníctids habitants dels petits aqüífers formats sobre calcretes (depòsits endurits de carbonat càlcic) de la regió àrida de l'Àustràlia occidental. Aquests aqüífers formen un mosaic d'hàbitats aquàtics subterranis seguint antics cursos de drenatge fluvial, avui ja totalment inactius. Cooper et al. (2007) trobaren que els principals llinatges mitocondrials per a *cox1* quedaren confinats a una única i aïllada calcreta, i que la major part de la variació genètica es donà entre calcretes diferents. Encara que algunes de les poblacions properes localitzades al mateix sistema de paleodrenatge eren genèticament semblants –el que suggereix de l'existència de flux gènic al passat–, les poblacions no s'agruparen necessàriament d'acord amb el seu sistema de paleodrenatge. Aquest patró podria ésser resultat de l'existència de flux gènic o d'un rang d'expansió al passat entre els diferents paleodrenatges, abans que les poblacions quedessin aïllades a cadascuna de les calcretes. L'aïllament ulterior de les poblacions podria estar associat amb un important període d'aridificació que afectà la regió entre fa 10 i 4 Ma (Cooper et al., 2007).

Els principals esdeveniments cladogenètics dins del llinatge de *M. longipes* podrien estar associats a la successió de canvis en el nivell de la mar ocorreguts en el passat a tota la Mediterrània (tenint en compte les limitacions i l'error associat a les estimacions del rellotge molecular). Durant el Tortonian (11.3 Ma), Mallorca i Menorca foren envaïdes per una mar epicontinental que reduí a Mallorca en un grup de petites illetes (corresponents a les parts més elevades actualment), mentre que la meitat sud de Menorca fou probablement submergida completament (Fig. 3.1) (Pomar, 1991). Aquests resultats suggereixen que una sola població de *M. longipes* es va distribuir arreu de la plataforma continental de arxipèlag i sobrevisqué a la fase de deposició d'evaporites de l'anomenada "Crisi de Salinitat del Messinià" (*Messinian Salinity Crisis*), que tingué lloc del 5.96 al 5.59 Ma (Krijgsman et al., 1999), coetània a un esdeveniment de regressió generalitzada de la mar que podria haver assecat completament o parcialment la Mediterrània en aquella època. Aquesta gran regressió va ésser probablement la darrera causa que impulsà la divisió de l'espècie en dos llinatges principals: el primer, corresponent al clade A, representat per les poblacions que es mantingueren associades al litoral i que consegüentment, llur distribució es veié afectada rere la retirada de la línia de costa en direcció sud-est; i la resta de clades, que presumiblement derivarien de la part de la població que seguí la línia de costa en

---

direcció nord (Fig. 3.1). En aquestes anàlisis, l'edat de l'ancestre comú més recent del clade A i del seu grup germà (la resta de poblacions) s'ha estimat al voltant de 5.4-6.2 Ma, d'acord amb l'ús un rellotge molecular relaxat basat en seqüències de *cox1* i un model de coalescència. En algun moment entre 4.2 i 2.7 Ma (Pliocè mitjà-superior; Fig. 3.8), les poblacions de Menorca (clade D), la zona central de Mallorca (clade B), el nord de Mallorca (clade E) i l'oest de Mallorca (node C) es van separar. Els esdeveniments cladogenètics corresponents podrien estar vinculats a un únic gran cicle de transgressió-regressió marina, com la provocada pel rebliment del Pleistocè inferior de les conques deprimides de l'àrea de la Mediterrània occidental. La transgressió ocorreguda durant el Pliocè inferior, la qual tingué lloc tot d'una després de la Crisi de Salinitat del Messinià, arribà molt probablement a superar els 100 m per sobre el nivell actual de la mar a la zona balear (Cuerda et al., 1969). Aquest esdeveniment possiblement hagués permès a l'espècie assolir la part central de Mallorca (Fig. 3.1). Aquests resultats assenyalen que més endavant, cap a l'inici del Pliocè superior, els clades B, D i E experimentaren una divisió seguida per una diferenciació de les poblacions, on les principals bifurcacions secundàries tingueren lloc fa 2-0.5 Ma. La incertesa i els errors de càlcul comesos per mor de processos estocàstics associats a les estimacions del rellotge molecular impossibiliten la correlació exacte de les edats dels nodes amb les datacions de determinats episodis transitoris geològics i climàtics. No obstant això, és notable que la topologia de l'arbre obtingut seria compatible amb la successió cronològica documentada dels canvis ocorreguts al nivell del mar durant el Pliocè superior fins al Plistocè mitjà a l'hemisferi nord (Shackleton & Opdyke, 1973; Shackleton et al., 1984; Sosdian & Rosenthal, 2009). Els resultats suggereixen que els esdeveniments cladogenètics més recents al llinatge de *M. longipes* estarien vinculats als dos episodis principals de refredament, és a dir, baixades en el nivell de la mar, que daten de 3 fins 2.5 i de 1.2 fins a 0.85 Ma, respectivament.

---



## Conclusions

Els resultats aquí presentats suggereixen que els cicles de transgressió-regressió (canvis eustàtics) del nivell marí ocorreguts al passat podrien haver induït una sèrie de processos reiterats d'expansió, contracció i fragmentació de les poblacions de *M. longipes*, les quals es troben actualment estructurades en diversos llinatges aïllats i genèticament divergents i adaptades a un ampli rang de condicions de salinitat. Aquest escenari podria explicar el perquè han sorgit tants entrebancs a l'hora de resoldre les relacions filogenètiques entre els diferents llinatges d'aquest amfípode, fossin quins fossin els set de dades o metodologia emprada per aquesta comesa: el sobtat aïllament i la diversificació gairebé sincrònica de poblacions perifèriques d'un mateix ancestre als aqüífers de l'interior podria haver portat a aquesta situació. Aquesta hipòtesi s'ha proposat anteriorment per explicar el patró de distribució observat en alguns tàxons hipogeus habitants d'aigües dolces subterrànies i anquialines a majors escales geogràfiques (Léfebure et al., 2007). Aquest estudi ha posat èmfasi en la importància dels moviments eustàtics com a principals responsables de la gran divergència genètica intra-específica observada a amfípodes subterranis talassoids, un patró que aparentment no vendria acompanyat per cap diferenciació morfològica destacable.

---

## Caracterització del DNA mitocondrial de *Metacrangonyx longipes*

### Introducció

Els Crustacea són un Subfilum d'artròpodes que comprèn més de 52,000 espècies descrites fins aleshores, i presenten una diversitat de plans corporals no assolida a cap altre grup de Metazoa. Es distingeixen actualment sis classes de crustacis: Branchipoda, Cephalocarida, Malacostraca, Maxillopoda, Ostracoda i Remipedia (Martin & Davis, 2001). Dels mitogenomes obtinguts fins ara, 15 corresponen a un sol ordre de malacostracis: els Decapoda (crancs, gambes, etc.) (Carapelli et al., 2007; Yang & Yang, 2008). Dins l'ordre Amphipoda (Peracarida: Malacostraca), es coneixien només a les bases de dades, la seqüència parcial del mitogenoma de *Parhyale hawaiiensis* (Dana, 1853), el qual és incompleta ja que manca una regió d'uns 3 Kb, incloent els gens *rRNA 12s*, parts del *rRNA 16s*, la NADH deshidrogenasa II (*nad2*), i la regió control. D'altres grups de Peracarida, es coneixien sols un mitogenoma complet (*Ligia oceanica*) i dos mitogenomes parcials (*Armadillidium vulgare* i *Idotea baltica*) de l'ordre Isopoda (Kilpert & Podsiadlowski, 2006; Podsiadlowski & Bartolomaeus, 2006; Marcadé et al., 2007).

En aquest capítol presentem la primera seqüència completa d'un genoma mitocondrial d'un amfípode. Hem fet servir el mitogenoma de *Metacrangonyx longipes* per comparar el seu ordre gènic amb el d'altres crustacis, així com la seva composició nucleotídica i l'estructura dels tRNAs. Concretament ens centrem en les comparacions d'altres mitogenomes de peracàrids ja descrits.

---

## Resultats i Discussió

### *i. Organització genòmica a M. longipes*

La seqüència del genoma mitocondrial de *M. longipes* obtinguda ateny 14.113 pb de longitud total (núm. d'accés EMBL: AM944817) i presenta la típica organització circular trobada a la majoria de metazous (Fig. 3.10). Es tracta del mitogenoma més petit descrit fins aleshores dins del grup dels crustacis, similar en llargària al de *Tigriopus californicus* (Copepoda, Harpacticoida, Maxillopoda; amb una longitud de 14.546 pb) (Machida et al., 2002). De l'anotació gènica s'han obtingut els 37 gens habituals d'un mitogenoma animal: 13 gens codificants per a proteïna (MPCGs), dues de les subunitats ribosòmiques dels mitogenomes de metazous (Taula 3.5) i 22 gens codificants per a tRNAs. No obstant això, d'ençà la publicació d'aquest treball a l'any 2009 s'han realitzat algunes modificacions pel que fa als inicis i finals d'alguns dels gens (Taula 3.5). Això ha estat possible gràcies a la informació aportada per l'anotació i alineament de tots els MPCGs i tRNAs de Metacrangonyctidae (veure capítol III), gens òbviament més propers evolutivament que no pas els disponibles aleshores de Crustacea i Hexapoda, emprats a la primera anotació. Un dels canvis més notables és la troballa del gen per a tRNA-Ser<sub>UCN</sub>, el qual havia sigut impossible de localitzar anteriorment. Sorprenentment aquest gen s'ha detectat entre la regió control i el gen per a *cob*. A més a més, s'han acurçat els gens *nad5*, *nad6*, *rrnL*, *trnS<sub>AGN</sub>*, *trnR* i *trnV* mentre que s'han allargat els gens *rrnS*, *nad4*, *nad4L* i *trnK*.

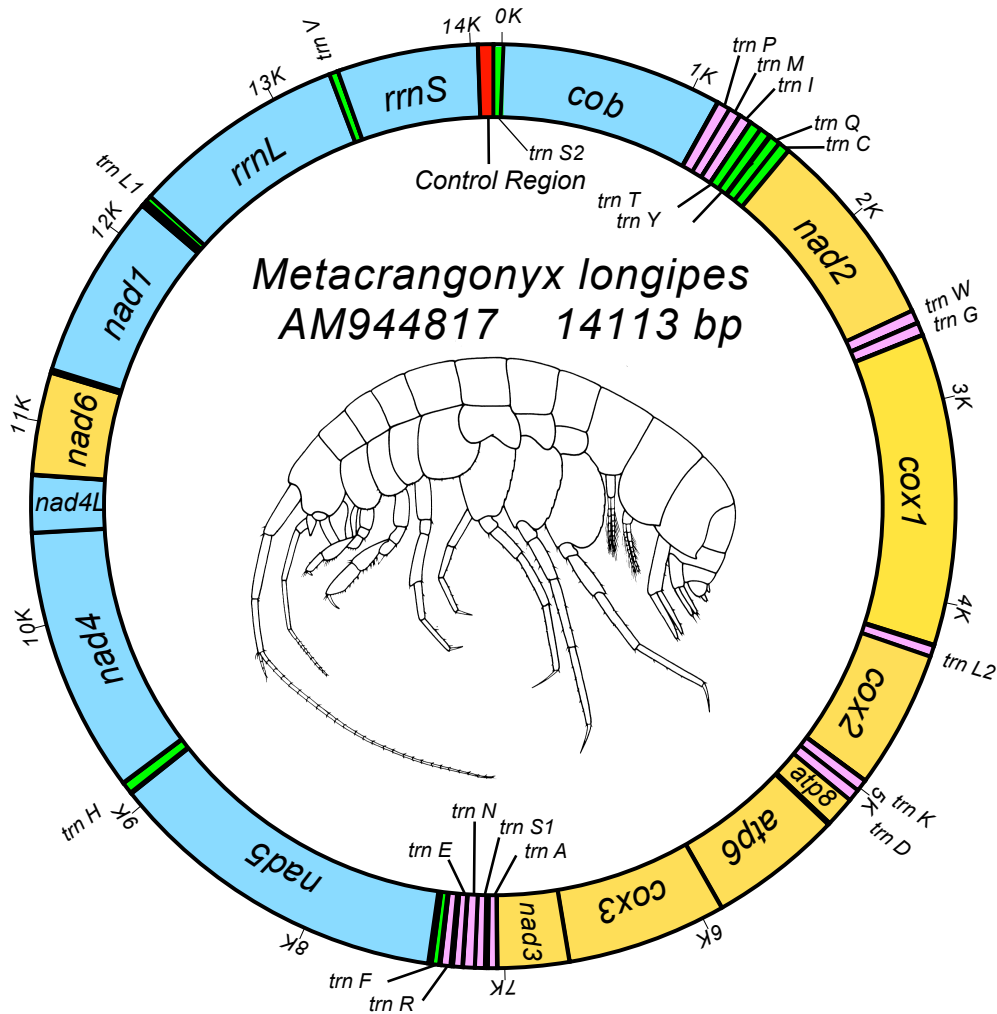
S'ha observat que la gran compactació d'aquest mitogenoma és a causa de la freqüent presència de solapaments entre gens. De fet, s'ha descobert que més de 20 gens solapen entre ells a llurs extrems. Aquestes regions solapades varien en llargària, en un rang que va des d'un sol parell de bases (el més freqüent) fins a un màxim de 9 pb (en el cas del gen codificant per a *atp8*, que solaparia amb 9 pb a l'extrem 5' del *atp6*). Curiosament, sols s'han detectat set espaiadors intergènics no codificants, amb un rang d' 1 a 17 pb; Taula 3.5).

---

Taula 3.5. Contingut gènic al genoma mitocondrial de *M. longipes*.

Gen	Posició	Cadena	Mida	Codó inici	Codó aturada
<i>cob</i>	1-1137	-	1137	ATG	TAA
<b>Espaiador intergènic 1</b>	1138-1140		3	n/a	n/a
<i>trnP</i>	1141-1204	+	64	n/a	n/a
<i>trnM</i>	1205-1265	+	61	n/a	n/a
<i>trnI</i>	1264-1326	+	63	n/a	n/a
<i>trnT</i>	1325-1383	-	59	n/a	n/a
<i>trnY</i>	1384-1443	-	59	n/a	n/a
<i>trnQ</i>	1439-1498	-	60	n/a	n/a
<i>trnC</i>	1499-1556	-	58	n/a	n/a
<i>nad2</i>	1556-2546	+	991	ATT	T
<i>trnW</i>	2547-2610	+	64	n/a	n/a
<b>Espaiador intergènic 2*</b>	2611-2612		2	n/a	n/a
<i>trnG</i>	2613-2674	+	62	n/a	n/a
<b>Espaiador intergènic 3*</b>	2675-2676		2	n/a	n/a
<i>cox1</i>	2677-4218	+	1542	ATT	TAA
<i>trnL1 (UUR)</i>	4214-4273	+	60	n/a	n/a
<i>cox2</i>	4274-4958	+	685	ATT	T
<i>trnK*</i>	4956-5017	+	62	n/a	n/a
<i>trnD</i>	5017-5077	+	61	n/a	n/a
<i>atp8</i>	5078-5236	+	159	ATC	TAA
<i>atp6</i>	5227-5895	+	669	ATG	TAA
<i>cox3</i>	5895-6683	+	789	ATG	TAA
<b>Espaiador intergènic 4</b>	6684-6684	+	1	n/a	n/a
<i>nad3</i>	6685-7038	+	354	ATT	TAG
<i>trnA</i>	7037-7097	+	61	n/a	n/a
<b>Espaiador intergènic 5</b>	7098-7098		1	n/a	n/a
<i>trnS1 (AGN)*</i>	7099-7146	+	48	n/a	n/a
<i>trnN</i>	7147-7207	+	61	n/a	n/a
<i>trnE</i>	7205-7265	+	61	n/a	n/a
<i>trnR*</i>	7265-7320	+	52	n/a	n/a
<i>trnF</i>	7315-7374	-	60	n/a	n/a
<i>nad5*</i>	7374-9076	-	1703	ATT	TA
<b>Espaiador intergènic 6</b>	9077-9082	+	6	n/a	n/a
<i>trnH</i>	9083-9142	-	60	n/a	n/a
<i>nad4*</i>	9140-10460	-	1333	ATG	T
<i>nad4L*</i>	10454-10744	-	291	ATG	TAA
<i>nad6*</i>	10737-11243	+	501	ATT	TAA
<b>Espaiador intergènic 7</b>	11244-11260	+	17	n/a	n/a
<i>nad1</i>	11261-12181	-	921	ATA	TAA
<i>trnL2 (CUN)</i>	12176-12235	-	60	n/a	n/a
<i>rrnL*</i>	12236-13298	-	1063	n/a	n/a
<i>trnV*</i>	13299-13349	-	51	n/a	n/a
<i>rrnS*</i>	13350-14037	-	688	n/a	n/a
<b>Regió Control*</b>	14038-14077	+	40	n/a	n/a
<i>trnS2 (UCN)*</i>	14065-2	-	51	n/a	n/a
<b>Tots els PCGs</b>	n/a	n/a	11073	n/a	n/a
<b>Tots els tRNAs</b>	n/a	n/a	1287	n/a	n/a
<b>Genoma complet</b>	1-14113	+	14113	n/a	n/a

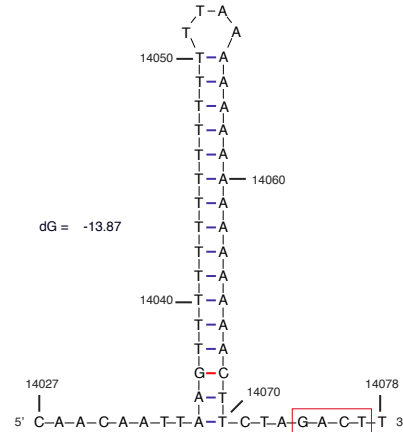
\* Anotació modificada d'ençà la publicació de Bauzà-Ribot i col·laboradors (2009).



**Figura 3.10.** Mapa del genoma mitocondrial seqüenciat a *Metacrangonyx longipes*. El color taronja (PCGs) i el rosa (tRNAs) indiquen els gens que es troben codificats al filament +; mentre que el color blau (PCGs, rrnL i rrnS) i el verd (tRNAs) indiquen els gens codificats al filament -.

Inicialment, tal com es va publicar en aquest treball, la regió codificant de 76 pb localitzada entre els gens *rrnS* i *cob* amb un alt contingut en A+T (84.22%) es creia que corresponia a la regió control atès que contenia l'origen de replicació del mtDNA (Figura 3.11). De fet es detectà en aquesta regió una possible estructura secundària que es plegava en si mateixa en una estructura d'agulla de ganxo (*hairpin structure*) amb un braç de 15 pb, més un llaç (*loop*) de quatre nucleòtids. Similarment, aquesta estructura també s'ha descrit a la regió control del mitogenoma a diversos insectes (Zhang et al., 1995), a la qual se li ha assignat la funció d'origen de replicació del mtDNA. L'extrem 3' de la seqüència flanquejant al voltant de l'estructura d'agulla de ganxo mostra un motiu conservat amb la seqüència GACT, present també a l'isòpode *Ligia oceanica* i a l'hoplocàrid (Malacostraca) *Squilla mantis* (Kilpert & Podsiadlowski, 2006). No obstant això, la regió 5' flanquejant manca de l'element conservat TATA present a molts dels hexàpodes analitzats fins ara. Per contra, presenta el motiu AATT,

la qual cosa suggereix que podria acomplir la mateixa funció que el motiu conservat TATA. Tal com ja s'ha esmentat, s'ha observat que aquesta regió de 76 pb alberga a més a més, el que correspondria al tRNA-Ser<sub>UCN</sub>, deixant sorprenentment una regió control de tan sols 40 pb, el que suposaria ésser la regió control més petita anotada fins aleshores dins Crustacea i d'entre les més petites a Arthropoda. El baix nombre de seqüències no codificants trobat al mitogenoma de *M. longipes* (<1%) i la presència de freqüents solapaments gènics són indicatius d'un mitogenoma extremadament compactat.



**Figura 3.11.** Estructura secundària putativa de la regió control del mitogenoma de *M. longipes*. El quadre assenyala el motiu conservat GACT, present també a l'isòpode *L. oceanica* i a l'hoplocàrid malacostraci *S. mantis*.

**Taula 3.6.** Nombre d'aminoàcids a cadascun dels MPCGs, freqüències nucleotídiques, contingut en A+T i valors de biaix per a AT i GC als gens del mitogenoma de *M. longipes*

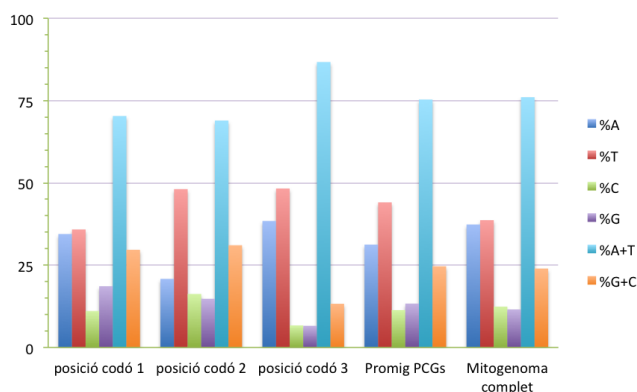
Gen	Cadena	Mida (aa)	%A	%C	%G	%T	% A+T	AT biaix	GC biaix
<i>cob</i>	-	378	30.17	15.83	10.73	43.27	73.44	-0.18	-0.19
<i>nad2</i>	+	330	31.58	8.38	12.71	47.33	78.91	-0.20	0.21
<i>cox1</i>	+	513	28.34	13.81	16.15	41.7	70.04	-0.19	0.08
<i>cox2</i>	+	228	31.83	12.55	13.72	41.9	73.73	-0.14	0.04
<i>atp8</i>	+	52	42.77	8.8	5.66	42.77	85.54	0.00	-0.22
<i>atp6</i>	+	222	30.64	11.81	12.26	45.29	75.93	-0.19	0.02
<i>cox3</i>	+	262	26.87	13.94	16.22	42.97	69.84	-0.23	0.08
<i>nad3</i>	+	117	29.66	9.32	14.69	46.33	75.99	-0.22	0.22
<i>nad5</i>	-	572	34.03	10.3	12.97	42.7	76.73	-0.11	0.11
<i>nad4</i>	-	438	32.24	8.97	13.76	45.03	77.27	-0.17	0.21
<i>nad4L</i>	-	94	36.84	5.61	12.28	45.27	82.11	-0.10	0.37
<i>nad6</i>	+	168	29.39	11.83	7.89	50.89	80.28	-0.27	-0.20
<i>nad1</i>	-	306	32.14	9.45	14.66	43.76	75.9	-0.15	0.22
<i>rrnL</i>	-	n/a	39.93	7.92	13.81	38.34	78.27	0.02	0.27
<i>rrnS</i>	-	n/a	35.83	7.63	12.66	43.88	79.71	-0.10	0.25
Regió control	+	n/a	40.79	7.89	7.89	43.43	84.22	-0.03	0.00
Tots els PCGs	n/a	n/a	31.25	11.34	13.33	44.08	75.33	-0.17	0.08
Tots els tRNAs	n/a	n/a	40.95	9.01	12.98	37.06	78.01	0.05	0.18
Genoma complet	+	n/a	37.36	12.41	11.56	38.67	76.03	-0.02	-0.04

ii. *Gens que codifiquen per a proteïnes: composició nucleotídica i ús de codons (codon usage)*

El contingut en A+T dels gens codificants per a proteïnes de *M. longipes* és del

75.33% (A: 31.25%; C: 11.34%; G: 13.33% i T:44.08%) (Taula 3.6), mentre que el percentatge al mitogenoma complet (cadena +) és del 76.03%; el que representa un dels percentatges més alts descrits el dia d'avui a crustacis, i molt similar als dels mitogenomes a Hexapoda. Fins ara, el mitogenoma d'*Argulus americanus* (Branchiura, Maxillopoda) és el que presenta el contingut en A+T més elevat a crustacis amb un percentatge del 77.80% (Machida, et al., 2002), mentre que el mitogenoma gairebé complet de l'amfípode *Paryphale hawaiiensis* comprèn un 73.66% (Cook et al., 2005; aquest estudi, 2009).

A les freqüències nucleotídiques del mitogenoma de *M. longipes* s'hi observa un cert biaix, amb valors semblants a ambdós filaments (Taula 3.6). Cal tenir en compte emperò que a la majoria d'artròpodes sol donar-se un biaix notable en les freqüències nucleotídiques (Hassanin et al., 2005; Hassanin 2006). Concretament, a insectes i



**Figura 3.12.** Valors mitjans de biaix en composició nucleotídica als PCGs i al mitogenoma complet de *M. longipes*.

crustacis malacostracis (tret d'alguns isòpodes) el més freqüent és trobar valors positius de biaix en AT (és a dir, més A que T) i valors negatius en biaix en GC (més C que G) als gens codificats al filament +, mentre que són freqüents valors negatius en biaix en AT i positius en biaix

en GC als gens codificants al filament - (Hassanin, et al., 2005; Hassanin 2006; Kilpert & Podsiadlowski, 2006; Wei et al., 2010). Emperò, segons Perna i Kocher (1995) i Hassanin i col·laboradors (2005) el biaix composicional entre els gens mitocondrials es reflecteix millor en el biaix en GC i no pas al d'AT. El biaix detectat en GC segons la fórmula de Perna i Kocher (1995) tant pels MPCGs com pel conjunt del mitogenoma al filament + és molt baix, amb valors propers a zero en ambdós casos (-0.035 i 0.08, respectivament). De fet, aquest valor de biaix en GC pel que fa al mitogenoma sencer (-0.035) és el més baix detectat fins aleshores a Amphipoda i un dels més baixos dins Peracarida, on solament l'isòpode *Janira maculosa* presentaria un valor de biaix en GC al filament + encara més baix (-0.030) (Taula Addicional 4) (Kilpert & Podsiadlowski, 2012). Aquests resultats divergeixen notablement dels mostrats pels mitogenomes més propers als dels amfípodes analitzats, els dels isòpodes, els quals presenten valors positius alts de

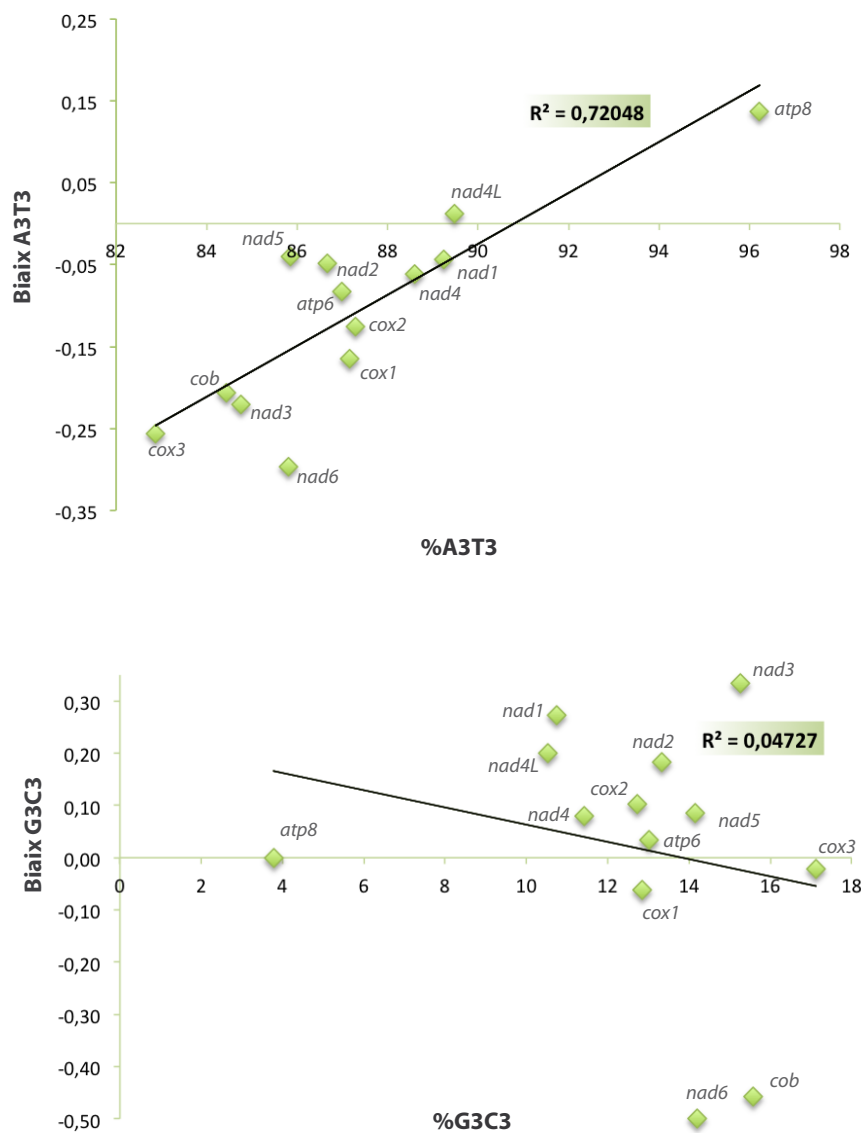
biaix en GC al filament + (d'entre 0.13 a 0.25) a excepció de *J. maculosa* i *Glyptonotus cf. antarcticus*, amb valors de biaix propers a zero i *Armadillidium vulgare* amb valors negatius (Kilpert & Podsiadlowski, 2006; 2012) (Taula Addicional 4). Aquest fet s'ha relacionat amb la presència d'inversions en la regió control que contenen l'origen de replicació a *Ligia oceanica* i *Idothea baltica*, atès que la majoria de crustacis mostren valors de biaix en GC de lleugerament a molt negatius a la cadena + (Hassanin et al., 2005; Hassanin, 2006; Kilpert & Podsiadlowski, 2006; 2012).

En el mitogenoma complet la composició en A i en T és molt similar (37.36% i 38.67%, respectivament) (Fig. 3.12), el que es tradueix en un biaix en AT moderadament negatiu, proper a zero (-0.02), mentre que en el cas dels MPCGs en conjunt hi ha un biaix negatiu en AT més important (-0.17) (43.43% contingut en A i 31.25% contingut en T) (Taula 3.6 i Fig. 3.12). El valor més destacable quant als gens per separat seria el que correspon al gen *nad4L*, amb un valor alt de biaix positiu en GC (0.34). Sorprenentment els gens per *nad2* i *nad3* que es troben codificats al filament + exhibeixen uns valors alts i positius en biaix en GC (0.21 i 0.22, respectivament), el contrari del que seria d'esperar en comparació de la resta de gens codificats al filament + per a la majoria d'artròpodes. Els resultats obtinguts pels MPCGs per a totes les posicions dels codons mostren que la majoria dels gens codificants al filament + com al - presenten valors en biaix lleugerament negatius en AT i lleugerament positius en GC. És important esmentar que l'efecte de l'asimetria composicional nucleotídica als MPCGs entre els diferents filaments difereix segons la posició dels codons (Reyes, et al 1998). Si s'observa la composició nucleotídica per posició dels codons (Fig. 3.12 i Taula Addicional 5) es veu que les primeres posicions presenten un biaix positiu en GC (0.25), mentre que a les segones i terceres posicions aquest valor és proper a zero, indicant un biaix gairebé inexistent en el conjunt dels gens (-0.05 i -0.01, respectivament). A les segones i terceres posicions els biaixos més significatius es troben als valors de biaix en AT (-0.40 i 0.11, respectivament). En aquest estudi s'observa que les terceres posicions són aquelles que tenen un menor contingut en G+C, tan sols un 13.28%, al contrari de les primeres i segones que presenten uns valors en contingut G+C molt superiors (29.68% i 31.06%). Curiosament, el gen que codifica per a *atp8* té el contingut G+C més baix observat a les terceres posicions (3.78%), mentre que el gen de *cox3* és el que mostra major contingut en G+C en aquesta mateixa posició (17.12%). Atès a la degeneració del codi genètic a les terceres posicions, la conseqüència acumulativa de la pressió de mutació direccional pot observar-se



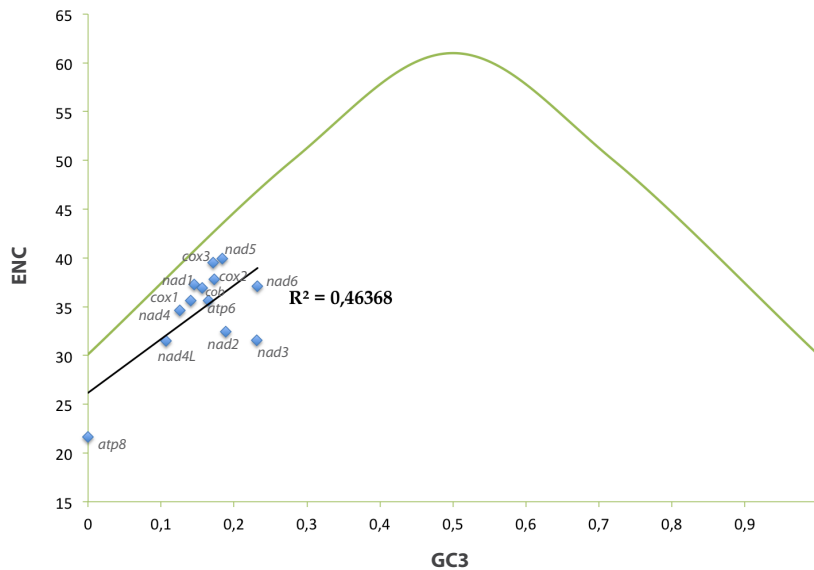
fàcilment a partir de l'anàlisi de composició nucleotídica a les terceres posicions dels codons (Lin et al., 2012).

S'ha estudiat la relació entre el contingut A+T vs. el valor de biaix en AT i també la relació entre el contingut G+C vs. el biaix en GC a les terceres posicions i s'ha observat l'existència d'una correlació positiva ( $r^2 = 0.72048$ ,  $p < 0.005$ ) en el primer cas, sense trobar-se cap relació destacable al segon cas ( $r^2 = 0.04727$ ) (Fig. 3.13). És a dir, aquells gens amb menor contingut en A+T tendrien a presentar valors més negatius en biaix en AT. Per exemple, en el cas de *cox3*, *cob*, *nad3* i *nad6* amb valors de contingut en A+T més baixos que la resta (de 82.2% a 88.6%), presentarien valors en biaix en AT força negatius (de -0.30 a -0.21); *atp8* en canvi, amb el valor més alt en contingut en A+T (96.22%) tindria el valor positiu més alt en biaix en AT (0.14).



**Figura 3.13.** Relació lineal entre el biaix en AT i GC vs. el contingut en A+T i G+C a les terceres posicions dels codons dels MPCGs, respectivament.

S'ha realitzat també el càlcul del nombre efectiu de codons (ENC, de l'anglès *Effective number of codons*) per a tots els MPCGs. ENC representa una mesura senzilla per saber l'ús de codons (*codon usage*), el rang del qual comprèn des de 20 -quan un sol únic codó s'empra per a cada aminoàcid- a 61 (o 62 pel que fa el codi genètic mitocondrial a invertebrats, ja que UGA codifica per a triptòfan) -quan tots els codons sinònims són emprats indistintament- (Wright, 1990). Quant al mitogenoma de *M. longipes*, els MPCGs presenten un valor d'ENC baix ( $35.38 \pm 2.84$ ), indicant que en



**Figura 3.14.** Gràfica que mostra la relació entre ENC i GC3 (%) a cadascun dels gens codificants per a proteïnes de *M. longipes*. La corba verda indica el patró ENC-GC3 esperat sota el supòsit d'absència de selecció en l'ús de codons; és a dir en el cas que el biaix en ús de codons depengui exclusivament del contingut en GC3 (Wright, 1990). La regressió lineal de les dades observades es mostra com a una línia negra.

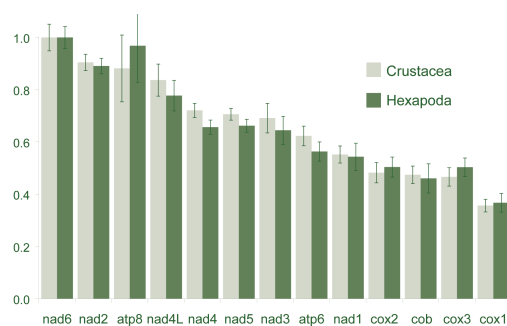
realitat fa servir al voltant de la meitat del tots els codons possibles. Talment com Wright proposà al seu estudi (1990), s'ha estudiat la relació entre ENC vs. contingut en G+C a les terceres posicions (GC3) per tal d'esbrinar els patrons d'ús de codons, amb la gràfica d'ENC-plot. La corba de la gràfica d'ENC-plot mostra la posició esperada dels gens que es troben influenciats exclusivament per la variació en el contingut en GC3. És a dir, si un gen particular es troba subjecte solament a restriccions composicionals o el que és el mateix, a restriccions per biaix mutacionals en el contingut en GC3, aquest es situarà sobre o just per davall d'aquesta corba; per contra, si el gen es troba subjecte a pressió selectiva a nivell traduccional, llavors el valor caurà molt per davall de la corba esperada. Com es pot observar a la Figura 3.14, tots els gens tret d'*atp8* cauen molt pròxims i per davall de la corba esperada, assenyalant que el biaix en ús de codons d'aquests MPCGs es deu principalment a restriccions composicionals originades per pressions mutacionals diferencials, tret comú a molts de gens mitocondrials a animals (Jia & Higgs, 2008). En el cas del gen *atp8*, cal tenir en compte que solament té 53 codons i podria ésser que fossin massa pocs per obtenir un valor prou fiable. A més, es pot veure que existeix una correlació

positiva tot i que baixa entre ENC i GC3 ( $r^2 = 0.46368$ ;  $P < 0.01$ ), com també descriuen per a isòpodes al seu treball Kilpert i Podsiadlowski (2006). No obstant això, els gens *nad2*, *nad3* i *nad6* (més allunyats de la corba esperada) presenten uns valors d'ENC més baixos del que hom s'esperaria d'acord amb l'alt contingut en GC que presenten les seves terceres posicions (18% en el cas de *nad2*, i un 23% per a *nad3* i *nad6*). Si es comparen aquests resultats amb aquells obtinguts a isòpodes, *M. longipes* presenta valors considerablement més baixos d'ENC (pel que també un baix contingut en GC3), mostrant uns valors més semblants a aquells trobats a l'amfípode *Parhyale hawaiiensis* (Cook et al., 2005) i al mitogenoma ric en A+T d'*Argulus americanus* (Branchiura, Maxillopoda) (Machida et al., 2002).

Sis dels tretze MPCGs del mitogenoma de *M. longipes* presenten codons d'inici poc habituals al mtDNA d'artròpodes. El codó ATT es troba a cinc dels gens (Taula 3.5), inclòs el *cox1*, mentre que a la resta de malacostracis començarien amb ACG (Kilpert & Podsiadlowski, 2006). Al seu torn, tres dels MPCGs presenten codons d'aturada truncats (Taula 3.5), com és el cas dels gens de *nad2*, *nad4* i *cox2* els quals acaben en una sola T. Tal com ha quedat reflectit en altres estudis (Ojala et al., 1981), és probable que aquests codons d'aturada truncats siguin completats post-transcripcionalment per poliadenilació, i conseqüentment els transcrits finals acabin presentant codons d'aturada UAA funcionals.

### iii. Divergència genètica als gens codificants per a proteïna

Per avaluar la divergència relativa dels MPCGs individuals, s'ha emprat el conjunt complet de dades disponibles de seqüències de mtDNA de crustacis més una representació de la resta dels principals ordres d'hexàpodes (35 taxons; Taula Addicional 3). Els gens que han mostrat divergències més baixes dins crustacis i hexàpodes són *cox1*, *cox2*, *cox3*, i *cob*, mentre que *nad6*, *nad2*, *atp8* i *nad4L* presenten fins a quasi el doble dels valors mitjans de divergència (Fig. 3.15).

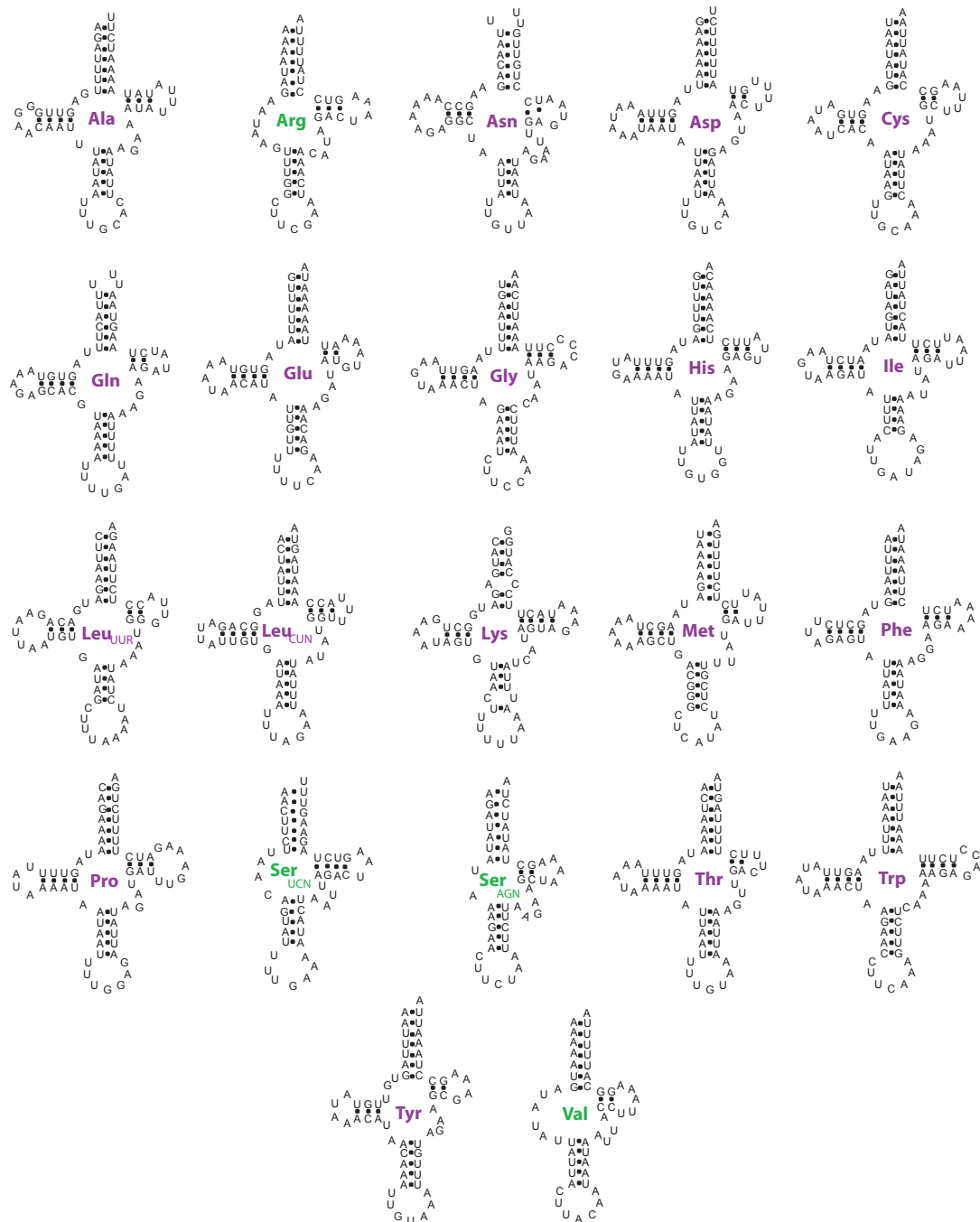


**Figura 3.15.** Mitjana dels valors relatius de divergència genètica corregida dels MPCGs de crustacis i hexàpodes. Nota: les divergències nucleotídiques dels gens individuals s'ha estimat a partir de les comparacions per parells entre els mitogenomes complets de crustacis més 35 espècies que representen els ordres principals d'hexàpodes (Taula Addicional 3).

Sembla que existiria una associació entre la variació gènica i la longitud atès que els gens més curts, com és el cas dels gens per *atp8*, *nad4L* i *nad6*, serien els més divergents. El que no acompliria aquest patró seria el *nad2*, el qual ateny una longitud de 991 pb i presenta una alta taxa de substitució. Com s'ha vist anteriorment a altres estudis (Cameron & Whiting, 2007; Salvato et al., 2008), tant les anàlisis de variabilitat com d'ús de codons dels MPCGs per separat a hexàpodes (insectes) dels ordres Isoptera i Lepidoptera evidencien que alguns dels gens més emprats al camp de la sistemàtica molecular, com seria el cas de *cox1* i *cox2*, presenten la menor variabilitat, mentre que els gens postergats com ara *nad2*, *nad3*, *nad4* i *nad6* podrien ésser del tot adients per a la sistemàtica per mor de l'alta informació continguda i la seva gran variabilitat. Els resultats aquí exposats assenyalen que això podria estendre's als crustacis, els quals presenten un patró de substitució dels MPCGs similar als hexàpodes.

#### *iv. Gens de RNA de transferència*

Com a una primera aproximació s'ha utilitzat el programa tRNAscan-SE per a la identificació dels tRNAs al mitogenoma de *M. longipes*. 17 d'ells s'han localitzat emprant els paràmetres estàndards del programa, mentre que quatre tRNAs (*trnS<sub>AGN</sub>*, *trnN*, *trnF* i *trnV*) han estat inferits a partir de cerques específiques menys restringents a zones no codificants (valor de *COVE cut-off* de -20). En el moment de la publicació del mitogenoma de *M. longipes* no s'havia trobat enlloc el gen per a *trnS<sub>UCN</sub>*. La hipòtesi inicial fou que solapés totalment amb els gens *trnG* i *trnW* (valors de *COVE* de +0.30 i -3.54, respectivament). El gen *trnS<sub>UCN</sub>* sol presentar característiques poc usuals a la majoria dels artròpodes, com és la manca del braç DHU (Kilpert & Podsiadlowski, 2006) i aquest fet podria dificultar la seva cerca amb mètodes estandarditzats com és el cas del programa tRNAscan-SE. Sorprenentment, un cop analitzats la resta de mitogenomes de diverses espècies del gènere *Metacrangonyx* (Cap. III), aquest gen ha sigut localitzat entre la regió control i el gen per a *cob*. Com ha resultat, s'ha observat que aquest tRNA-*Ser<sub>UCN</sub>* ateny una longitud total de 50 pb solapant en 48 pb amb el que anteriorment s'havia anotat com a regió control, on s'inclouen les 4 pb del motiu conservat GACT. Arran de l'anotació de les resta de tRNAs de la família *Metacrangonyctidae*, s'han fet algunes modificacions en els tRNAs de *M. longipes* reflectides a la Figura 3.16. Entre aquests ajustaments s'hi troben els gens que codifiquen per a tRNA-Arg, tRNA-*Ser<sub>AGN</sub>* i tRNA-Val, els quals, igual que el tRNA-*S<sub>UCN</sub>*, manquen del braç sencer DHU. El tRNA-Thr i tRNA-Gln en canvi podrien

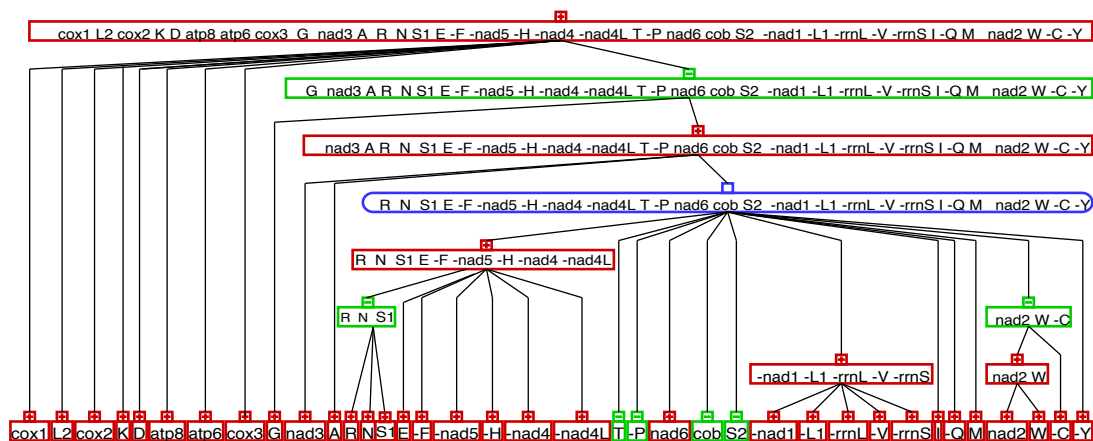


**Figura 3.16.** Estructures secundàries putatives dels tRNAs mitocondrials de *M. longipes*. El color verd indica els tRNAs que s'han modificat d'ençà de la publicació de Bauzà-Ribot i col·laboradors (2009).

mancar del braç sencer de TΨC, o bé presentar un braç de TΨC reduït format per dos enllaços i un llaç anòmal de sols tres nucleòtids. A més a més, s'han evidenciat desparellaments de nucleòtids en el braç acceptor per a tRNA-Asn, tRNA-Gln i tRNA-Lys, i també al braç anticodó de tRNA-Lys, tRNA-Asn, tRNA-Ile i tRNA-Leu<sub>UUR</sub> (Fig. 3.16). Anteriorment a altres estudis, s'han descrit nombrosos casos de desparellaments en els braços dels tRNAs mitocondrials i es creu que podrien ésser modificats per processos posteriors d'edició de RNA (Ojala et al., 1981; Xiong & Kocher, 1991; Yokobori & Pääbo, 1995; Kilpert & Podsiadlowski, 2006). Els tRNAs a *M. longipes* s'han



de reordenacions gèniques que s'han donat entre aquests mitogenomes (transposicions, transposicions inverses, inversions, TDRs, etc.) (veure Materials i Mètodes) (Bérard et al., 2007). En funció dels intervals comuns trobats, la reordenació gènica al mitogenoma de *M. longipes* pot explicar-se per dos mecanismes alternatius diferents. Si més no, tres transposicions que impliquen els gens *trnG*, *trnC* i el segment *trnR-N-S1*, dos canvis de localització de cadena (inversions) –un que implica el segment *cob-trnS2* i un altre el segment– que inclou els gens *trnP* i *trnT*–, i tres repeticions complexes en tàndem amb posteriors pèrdues aleatòries són necessàries per explicar el patró observat a *Metacrangonyx* vs. el patró ancestral a Pancrustacea. Una altra possibilitat (no representada) per explicar aquest patró seria la concurrència d'una única inversió al segment que inclou als gens *cob-nad6-trnP-trnT* al patró ancestral de Pancrustacea, seguida per una nova inversió del gen *nad6* des de la cadena – a la cadena +, més tres duplicacions en tàndem amb posteriors pèrdues d'aquests gens. L'ordre gènic de *M. longipes* també difereix de l'únic amfípode analitzat fins aleshores, *P. hawaiiensis* (Cook et al., 2005). En aquest treball també s'ha completat gairebé la totalitat del mitogenoma d'aquesta espècie, tret d'una petita part del gen *rrnS* i de la regió control (número d'accés a EMBL: FM957525 i FM957526). A més, segons els resultats obtinguts amb el programa de tRNAscan per a *P. hawaiiensis*, han sigut reanotats els gens de tRNA com *trnW* i *trnG* prèviament anotats com a *trnC* i *trnW*, respectivament (Cook et al., 2005) (número d'accés: AY639937). En aquesta espècie, almenys 10 dels gens per a tRNAs presenten canvis posicionals respecte al patró de Pancrustacea.



**Figura 3.18.** Àrbre SIT deduït a partir dels esdeveniments de reordenació gènica necessaris per explicar l'ordre gènic observat en el mitogenoma de *M. longipes*, en comparació amb el patró ancestral per a Pancrustacea.

Després de la publicació d'aquest treball s'han publicat altres mitogenomes d'amfípodes, arribant fins a set mitogenomes complets (Ito et al., 2010; Ki et al., 2010; Kilpert & Podsiadlowski, 2010; Shin et al., 2012; Krebs & Bastrop, 2012) i a tots ells



s'ha pogut observar la translocació del gen *trnG* a la mateixa posició (entre el gen *trnW* i *cox1*, el que podria identificar-se com a una apomorfia comuna a Amphipoda (Fig. 3.19). Kilpert i Podsiadlowski (2010) recalquen l'alta similitud de la seqüència del gen per a *trnG* amb la del seu adjacent *trnW* (74.2% a l'estudi d'aquests autors i 76.6% en el cas de *M. longipes*) pel que suggereixen que *trnG* podria haver emergit arran d'un esdeveniment de duplicació del gen *trnW* seguit d'un canvi al braç anticodó i a l'acceptor fins a acomplir la funció de *trnG*. Aquesta observació ha sigut corroborada més recentment per Krebes i Bastrop (2012) amb l'anotació del mitogenoma de *Gammarus duebeni*. El fet de què siguin presents posicions derivades dels gens *trnR*, *trnS1*, *trnY* i *trnG* a tots els mitogenomes d'amfípodes analitzats fins el dia d'avui (Cook et al., 2005; Shin et al., 2012) respecte al patró de reordenació ancestral, suggereix que podrien haver-se donat en el patró ancestral comú d'amfípodes, tot i que per explicar la translocació dels gens *trnR*, *trnS1* i *trnY* a diferents posicions a *P. hawaiiensis* en comparació de *M. longipes* i a la resta d'amfípodes, es requereix de canvis addicionals per explicar-ne aquesta situació (Kilpert & Podsiadlowski, 2010). No obstant això, tal com assenyalen Krebes i Bastrop (2012) al seu treball, sis dels set mitogenomes coneguts a amfípodes en aquests moments (exceptuant el de *P. hawaiiensis*) comparteixen el mateix ordre gènic pels cinc gens per a tRNAs: els *trnA-trnS1(AGN)-trnN-trnE-trnR* (Fig. 3.19) per la qual cosa suggereixen que aquesta reordenació es donà ja al patró ancestral per a amfípodes i constituiria *per se* un tret apomòrfic tot i que *P. hawaiiensis* presenti un status derivat per la posició d'aquests cinc gens. Els esdeveniments de reordenació més significants a *M. longipes* no obstant són els ocorreguts entre la regió control i el gen *nad2*, on el gen per a *cob* ha sofert una translocació i una inversió passant de la cadena + a la -; inversió comú al mitogenoma d'isòpodes fins aleshores publicats (Kilpert et al., 2012). Els altres mitogenomes de peracàrides coneguts en el moment de la publicació d'aquest treball, els dels isòpodes *Ligia oceanica* i els incomplets *Idothea baltica* i *Armadillidium vulgare*, mostraven algunes translocacions diferents respecte a l'ordre gènic del patró ancestral assumit per a Pancrustacea (Fig. 3.17) (Kilpert & Podsiadlowski, 2006). No obstant això, el treball recentment publicat per Kilpert i col·laboradors (2012) on analitzen tots els mitogenomes d'isòpodes disponibles fins aleshores (nou en total) infereixen un patró ancestral d'ordre gènic, lleugerament modificat als isòpodes moderns (*nad1-trnL1-rrnS-Regió Control -trnS1-cob-trnT-nad5-trnF*). També assenyalen la falta d'evidència d'esdeveniments de translocacions comuns dins Peracarida (Fig. 3.19), pel que el patró ancestral de Pancrustacea sembla ésser l'ordre gènic més plausible per l'ancestre comú



## d'amfípodes i isòpodes.



**Figura 3.19.** Ordre gènic dels mitogenomes de Peracarida en comparació del patró ancestral de Pancrustacea. El color gris indica els canvis succeïts en comparació amb el patró ancestral de Pancrustacea. Esquema extret de Krebs i Bastrop (2012).

## Conclusions

Amb aquest treball s'ha aconseguit la primera seqüència del genoma mitocondrial complet d'un amfípode, el de *Metacrangonyx longipes*, i segona per a peracàrides, superordre de Crustacea, escassament representat fins aleshores a la base de dades de mitogenomes mitocondrials.

Es tracta d'un mitogenoma altament compactat, amb la regió control extremadament reduïda, de només 40 pb, representant el mitogenoma i regió control més curts descrits a crustacis en aquests moments. Exhibeix uns dels percentatges en contingut AT més alts descrits a crustacis (76.03%) i molt similar a l'assolit a mitogenomes d'hexàpodes. Contràriament al que s'observa en els mitogenomes de Pancrustacea, el mitogenoma de *M. longipes* exhibeix un biaix composicional molt moderat, amb valors lleugerament negatius de biaix en GC a la cadena positiva. S'ha descartat que el baix valor en ENC sigui perquè els gens es troben sota pressió selectiva a nivell traduccional, sinó més aviat és perquè existeix una restricció en composició en G+C a les terceres posicions. Del gran nombre de diferències quant a l'ordre gènic en comparació del patró per a isòpodes, se n'extreu que aquest no es troba conservat dins Peracarida, essent el patró ancestral de Pancrustacea l'ordre gènic més probable per l'ancestre comú d'aquest grup (Kilpert & Podsiadlowski, 2006; Kilpert et al., 2012).

La presència d'una transposició comú en el gen de *trnG* tant a *M. longipes* i *P. hawaiiensis*, com a la resta d'amfípodes analitzats, suggereix que aquesta ja es trobava present a l'ancestre comú dels amfípodes, pel que podria tractar-se d'una fidedigne apomorfia d'aquest ordre. Una altra diferència destacable en comparació de la resta de peracàrides estudiats fins aleshores, és que a *M. longipes* no s'hi ha trobat cap inversió destacable en el signe dels valors de biaix en GC als gens que es troben a la cadena positiva, tal com és el cas a isòpodes (Kilpert & Podsiadlowski, 2006, 2010).

Les dades presentades en aquest estudi han ampliat la base de dades dels genomes mitocondrials de crustacis, que en el moment d'aquesta publicació era esbiaixada i exigua. A més, han obert el camí a la seqüenciació d'altres espècies congènriques de diferents àrees geogràfiques, fent possible l'escomesa de l'estudi de la seva filogènia i biogeografia històrica d'aquesta família enigmàtica d'habitants de les aigües subterrànies.

---



# Anàlisi filogenètica i biogeogràfica dels Metacrangonyctidae

## Introducció

Tant la vicariança com la dispersió són sempre inherents als models d'especiació al·lopàtrica, i generalment s'han considerat com a hipòtesis alternatives excloents (Crisp et al., 2011). L'ús de la datació molecular a estudis biogeogràfics ha permès testar, en un marc temporal, la congruència entre els esdeveniments cladogenètics i els vicariants, i reconstruir l'evolució espai-temporal de llinatges ancestrals (Crist et al., 2011; Riddle et al., 2008; Antonelli & Sanmartín, 2011).

Talment s'ha descrit abans a la Introducció general, els Metacrangonyctidae es troben restringits a aigües subterrànies continentals de la regió del Carib, la Mediterrània, el Marroc i Orient Mitjà, i a Fuerteventura (Balazuc & Ruffo, 1953; Ruffo, 1954; Karaman & Pesce, 1979; Ruffo, 1982; Stock et al., 1986; Boutin & Messouli, 1988b; Karaman, 1989; Messoululi et al., 1991; Boutin, 1994; Stoch, 1997; Oulbaz et al., 1998; Jaume & Christenson, 2001; Jaume & Vonk, 2012). Anteriorment, s'ha suggerit que aquesta distribució podria ésser resultat d'esdeveniments vicariants ocorreguts per l'oclusió seqüencial de l'oceà circumtropical Tethys, existent des del Juràssic mitjà fins fa 20 milions d'anys (Stock, 1977, 1980; Holsinger, 1986, 1988). Malgrat tot, els estudis que han intentat abordar aquestes qüestions biogeogràfiques s'han basat en anàlisis morfològiques o dades de seqüències parcials d'ADN, el que ha fet difícil testar aquesta hipòtesi (Neiber et al., 2011; Page et al., 2008; Baratti et al., 2010). En aquesta tesi s'han obtingut les seqüències senceres de tots els gens mitocondrials codificants per proteïnes (MPCGs) dels principals llinatges de *Metacrangonyx* -obtingudes majoritàriament per mètodes de seqüenciació de darrera generació- i a més, s'ha seqüenciat el gen nuclear ribosomal SSU. Amb tota aquesta informació molecular, s'ha intentat resoldre la filogènia i datar els principals esdeveniments de diversificació dels Metacrangonyctidae. Els principals objectius en aquest estudi han estat d'una banda (1) obtenir una filogènia fermament recolzada de la família en estudi, usant un mostreig geogràfic representatiu, i d'altra banda (2) estimar l'edat dels nodes principals de l'arbre filogenètic utilitzant tècniques de datació basades en el rellotge molecular i el

seu calibratge a partir de punts paleogeogràfics d'edat coneguda. L'objectiu últim ha estat provar la hipòtesi que suggereix que les espècies presents a costats oposats de l'Atlàntic tenen un origen vicariant, en contrast amb un origen a la regió Mediterrània i una subsegüent dispersió secundària al Carib. Els resultats exposats a continuació són consistents amb la hipòtesi de vicariança per tectònica de plaques, amb estimacions per a les diversificacions inicials datades entre 96 i 83 milions d'anys.

---

## Resultats

### *i. Filogènia preliminar amb els gens *cox1* i *rrnL*, i obtenció dels mitogenomes complets*

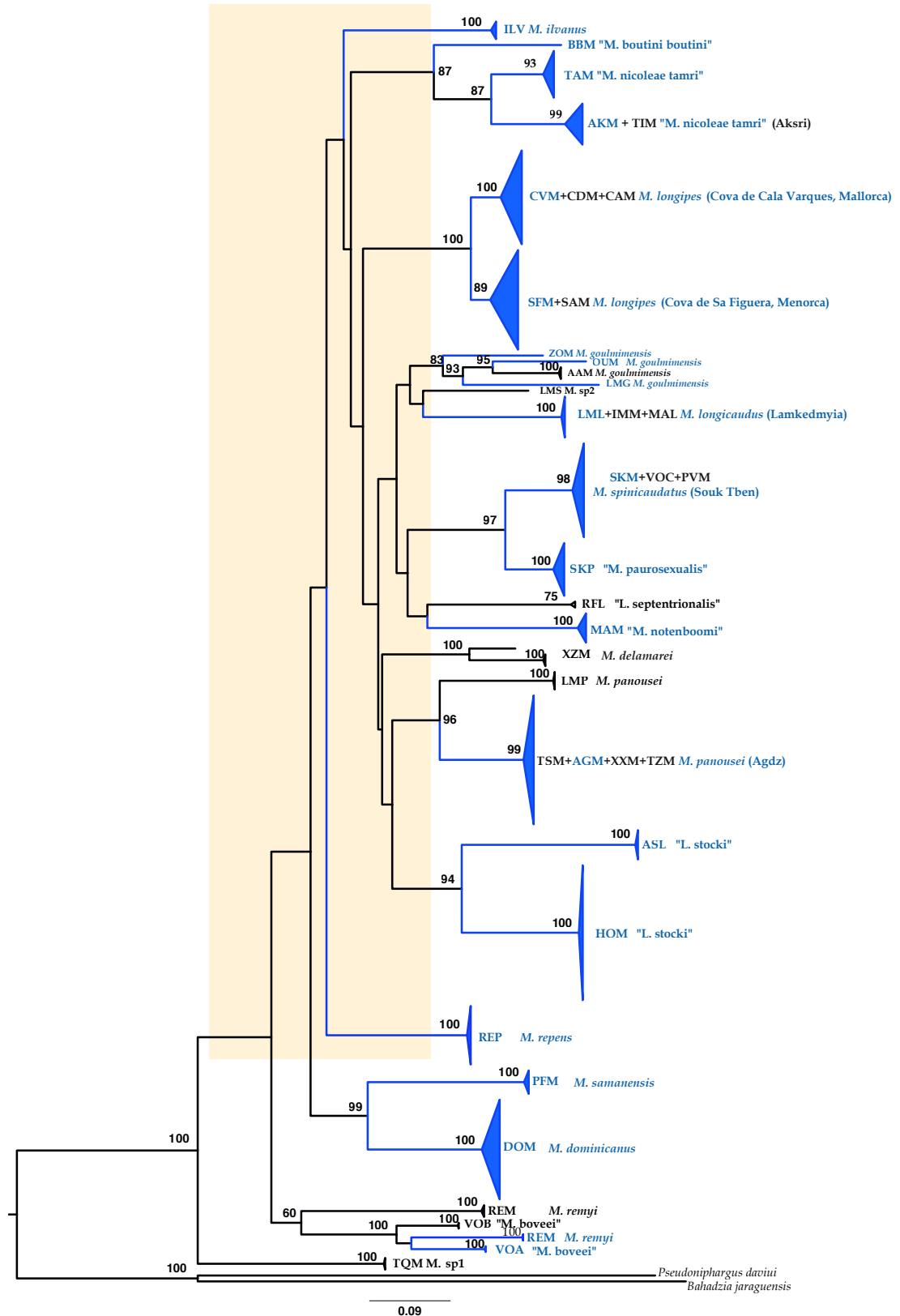
La filogènia obtinguda per màxima versemblança basada en les seqüències dels gens mitocondrials *cox1* (657 pb) i *rrnL* (441 pb) de 212 individus dels gèneres *Metacrangonyx* i *Longicacrangonyx*, les quals incloïen la majoria d'espècies i rangs geogràfics, va mostrar una bona resolució i suport pels nodes terminals i intermedis malgrat la resolució fou quasi nul·la per aquelles biparticions bassals (Fig. 3.20, Taula 3.7). Els nombres d'accés a les seqüències per a *cox1* són HE967026-HE967186 i per al marcador *rrnL* són HE967187-HE967277, i HE970657-HE970663. Les seqüències de *M. longipes* tenien els nombres d'accés anteriorment publicats a Bauzà-Ribot i col·laboradors (2011): FR846024-FR846060 per a *rrnL* i FR729731-FR729892 per a *cox1*. A partir dels resultats obtinguts amb aquesta anàlisi preliminar es destriaren els llinatges més divergents i representatius per elucidar les relacions filogenètiques i la biogeografia de la família estigobiont Metacrangonyctidae. No obstant els esforços, els mitogenomes d'alguns llinatges no es pogueren amplificar, com ara el de "*L. septentrionalis*" (RFL), *M. delamarei* (XZM), *M. panousei* (de Lamkedmya, LMP), *M. sp.1* (TQM) i *M. sp.2* (LMS). En total, s'aconseguien els mitogenomes complets de quinze dels vint llinatges nord-africans detectats i els cinc llinatges insulars (Fuerteventura, Menorca, Elba i dos a La Hispaniola) (Fig. 3.20), i la seqüència quasi completa del gen nuclear *SSU* exceptuant el cas de "*L. stocki*" (ASL) el qual manca d'aquest últim marcador. El set de dades es completà amb el mitogenoma prèviament obtingut de *Metacrangonyx longipes* (Cova de Cala Varques, Mallorca) (capítol II), així com el de dos grups externs: *Pseudoniphargus daviui* Jaume, 1991 (Pseudoniphargidae) (Illa de Cabrera) i *Bahadzia jaraquensis* Jaume i Wagner, 1998 (Hadziidae) (Illa de La Hispaniola). D'aquesta darrera espècie tampoc s'aconseguí amplificar el gen *SSU*. Els nombres d'accés dels 23 mitogenomes així com la del marcador nuclear *SSU* es poden trobar a la Taula 3.8.

**Taula 3.7.** Espècies de Metacrangonyx i Longipodacrangonyx incloses a l'estudi de la filogenètica preliminar. Les coordenades geogràfiques es presenten amb Datum WGS84(\*), i amb Datum ED50(#), les quals es registraren amb GPS. En aquelles localitats sense datum, els valors es van prendre arran de les imatges de Google Maps. El nom de les espècies que no estan en cursiva i entre cometes corresponen a les espècies que encara no s'han descrit formalment.

Espècie	Codi	Localitat	Coordenades	cox1	rrnL
"Longipodacrangonyx septentrionalis"	REFL	Sidi Abdellah, Taza (Marroc); pou	N32.55000° W7.83333°	2	0
"Longipodacrangonyx stocki"	ASL	Arbaa-Sahe, SW Tiznit (Marroc); pou	*N29°36.568' W9°55.187'	6	6
	HOM	Tafraut (Marroc); pou	*29R 502729 / 3288806	24	5
<i>M. spinicaudatus</i> Karaman & Pesce, 1980	SKM	Souk Tben, Haouz Plain, Marrakech (Marroc); pou	*N31°42.698' W8°04.257'	12	9
	PVM	Puit de la Porte Verte, Haouz Plain, Marrakech (Marroc); pou	*N31°39.594' W8°01.243'	3	3
"M. paurosexualis"	VOC	Vall de L'Ouirika, Alt Aïtes Occidental(Marroc); pou	*N31°22.551' W7°46.779'	1	1
	SKP	Souk Tben, Haouz Plain, Marrakech (Marroc); pou	*N31°42.698' W8°04.257'	9	3
"M. boveei"	VOA	Vall de L'Ouirika, Alt Aïtes Occidental(Marroc); pou	*N31°17.302' W7° 42.632'	2	2
	VOB	Vall de L'Ouirika, Alt Aïtes Occidental(Marroc); pou	*N31°18.560' W07°44.545'	2	1
<i>M. remyi</i> Ruffo, 1953	REM	Ijoukak, Alt Aïtes Occidental(Marroc); Font de la maison forestière (REM01-03)	*N30°59.895' W8°08.847'	3	3
		Riu de Agoundiss; a prop de Ijoukak (Marroc); font (REM04-05)	N/A	2	1
"M. nicoleae tamri"	TAM	Tamri, N Agadir (Marroc); pou	*N30°42.791' W9°46.761'	9	9
	AKM	Aksri, NW Agadir (Marroc); Font a prop de la cova de Talmat	N30°36.852' W9°28.129'	4	3
"M. boutini boutini"	TIM	Tizgni N'Chorfa, NW Agadir (Marroc); Font de Tifrit	N30°39.096' W9°21.637'	2	2
	BBM	Tizgni N'Chorfa, NW Agadir (Marroc); Cova de Win-Timdouine	N30°40'50" W9°20'42"	2	2
<i>M. dominicanus</i> Jaume & Christenson, 2001	DOM	Timzelite, Souss Massa NP, S Agadir (Marroc); pou	*29R 423825 / 3300524	1	1
		Juan Dolio (Dominican Rep.); pou (DOM01-06)	* N18°26.083' W69°25.860'	6	1
<i>M. samanensis</i> Jaume & Christenson, 2001	DOM	Juan Dolio (Dominican Rep.); pou (DOM07-DOM10)	*N18°26.076' W69°25.776'	4	3
		Juan Dolio (Dominican Rep.); pou (DOM11)	*N18°26.407' W69°25.551'	1	1
	Juan Dolio (Dominican Rep.); pou (DOM12)	*N18°26.434' W69°25.689'	1	0	
	Juan Dolio (Dominican Rep.); pou (DOM13-18)	N/A	N/A	6	0
	Platja del Frontón, Península de Samaná (Dominican Rep.); pou	*N19°17.799' W69°9.118'		5	1
<i>M. repens</i> Stock & Rondé-Broekhuizen, 1986	REP	Betancuria, Fuerteventura, Illes Canàries; pou a prop de l'Església (REP01)	*28R 592366 / 3144662	1	1
		Betancuria, Fuerteventura, Illes Canàries; pou (REP03-08; REP12-13)	*28R 591749 / 3142183	8	2
		Baranc del Cigarrón, Llanos de la Concepción, Fuerteventura, Illes Canàries (Espanya); pou (REP14-15)	*28R 590606 / 3151572	2	0

Espècie	Codi	Localitat	Coordenades	cox1	rrnL
<i>M. longipes</i> Chevreux, 1909	SFM	Sant Lluís, Menorca, Illes Balears; Cova de Ses Figueres	*31S 607475 / 4408145	13	4
	SAM	Ciutadella, Menorca, Illes Balears; Cova de S'Aigo	*31S 4424877 / 571259	5	3
	CVM	Porto Cristo, Mallorca, Illes Balears; Cova de Cala Varques C	# 31S 4372450 / 525550	8	3
	CDM	Ses Salines, Mallorca, Illes Balears; Cova des Dracs	*31S 4353625 / 512830	5	3
	CAM	Sant Llorenç, Mallorca, Illes Balears; Cova de S'Abisament	*N39°34.570'E03°22.233	4	3
	ILV	Illa d'Elba; Arxipèlag Tosca (Itàlia); pou	# 32T 610505 / 4737245	4	4
<i>M. itvanus</i> Stoch, 1997	XXM	Avant Agdz (Marroc); pou	N30.66950° W6.33145°	5	5
	AGM	Agdz (Marroc); pou	N30.691570° W6.447630°	3	3
	TZM	Tamzoulin Draa, SE d'Agdz (Marroc); pou	N30.508410° W6.102380°	1	1
	TSM	4 km abans de Tasla vinguent d'Agdz (Marroc); pou	*N30°34.613' W6°42.502'	15	4
	LMP	Lamkedmyia Meleh Jorf, NO d'Erfoud (Marroc); pou	N31.50063° W4.39498°	4	0
	AAM	Assif Anoulif, Oest Rich, NO d'Errachidia (Marroc); medi hiporreic	N32.26544° W4.809340°	3	3
<i>M. goulmimensis</i> Messouli, Boutin & Coineau, 1991	OUM	Ousroutou, E de Rich, N d'Errachidia (Marroc); pou	N32.258770° W4.533420°	1	1
	LMG	Lamkedmyia Meleh Jorf; NO d'Erfoud (Marroc); pou	N31.50063° W4.39498°	1	1
	ZOM	Zouala maisson, S d'Errachidia (Marroc); pou	N31.789380° W4.247950°	1	1
	LML	Lamkedmyia Meleh Jorf; NO d'Erfoud (Marroc); pou	N31.50063° W4.39498°	5	5
<i>M. longicaudus</i> Ruffo, 1954	IMM	Imiter, E d'Boulmane Daddès (Marroc); pou	N31.377701° W5.796114°	2	2
	MAL	Maadid, NE d'Erfoud (Marroc); pou	N31.48178° W4.21749°	1	1
<i>M. delamarei</i> Boutin & Coineau, 1991	XZM	Ouarzazat (Marroc); pou	N30.90687° W6.91112°	4	4
	MAM	Maadid, NE d'Erfoud (Marroc); pou	N31.48178° W4.21749°	6	4
"M. notenboomii" <i>Metacrangonyx</i> sp1 <i>Metacrangonyx</i> sp2	TQM	Tiqqi, NO d'Agadir (Marroc); Cova de Doussouille	N30°44.529' W9°19.803'	3	2
	LMS	Lamkedmyia Meleh Jorf; NO d'Erfoud (Marroc); pou	N31.50063° W4.39498°	1	1





**Figura 3.20.** Arbre de Màxima Versemblança dels 214 espècimens de metacranoníctids basat en els marcadors mitocondrials *cox1* i *rrrnL*. En blau, els llinatges que van ésser seleccionats per a la seqüenciació del mitogenoma complet. Només es mostren els valors de bootstrap superiors a 60. El color taronja assenyalava la manca de suport per a molts dels nodes basals.

Alguns dels tàxons nord-africans inclosos en aquest estudi no han estat encara descrits formalment. Tot i així, s'han citat prèviament amb binomis llatinitzats (Messouli 1988, 1994), pel que en aquesta tesi s'han conservat els mateixos noms per referir-se a aquests tàxons (tot i que es citen entre cometes, però no en cursiva, per poder destacar que no es troben formalment descrits).

Una petita part de la regió control no s'ha aconseguit seqüenciar en deu dels mitogenomes, mentre que a quatre d'ells ha mancat seqüenciar una part del fragment localitzat entre els gens *rrnL* i *rrnS*, molt probablement per mor del lliscament de la polimerasa a regions homopolimèriques, particularment aquelles formades per Adenines i Timines (Taula 3.8), molt freqüents en aquesta zona dels mitogenomes d'artròpodes. La grandària del mitogenoma ateny un rang de 14 i 15.5 kb, amb una mitjana de contingut en A+T als MPCGs del 72.73% -amb un rang d'entre 67.53 a 75.94%-. L'ordre gènic observat a totes les espècies de metacrangonictids és idèntic al trobat anteriorment a *M. longipes*, que al seu torn difereix del patró ancestral per a Pancrustacea (Boore et al., 1995; 1998; Bauzà-Ribot et al., 2009). Els dos grups externs emprats, el mitogenoma de *P. daviui* i el de *B. jaraquensis* han mostrat certes diferències quant a la reordenació gènica en comparació del mitogenoma de *Metacrangonyx* i del patró per a Pancrustacea (veure capítol anterior). Cal dir, emperò, que no s'ha aprofundit en les característiques genòmiques dels mitocondrials atès que no era un dels objectius d'aquesta tesi.

## ii. Anàlisis filogenètiques

Les anàlisis filogenètiques dels genomes mitocondrials s'han basat exclusivament en els 13 MPCGs, vist que en aquests gens és més senzill establir les identitats durant l'alineament per la presència de codons, la degeneració del codi genètic i l'afinitat d'aminoàcids amb característiques similars. Les seqüències nucleotídiques dels MPCGs individuals, un cop eliminats els codons d'aturada, es retroalinearen segons els patrons de traducció proteica (Abascal et al., 2010; Katoh & Toh, 2008) i foren posteriorment concatenats amb el programa *Phyutility* (Smith & Dunn, 2008). L'alineament obtingut es sotmeté al

Taula 3.8. Informació de les seqüències dels genomes mitocondrials per a les espècies incloses a les anàlisis.

Espècie	Localitat	Mida del mitogenoma (pb)	Mida del SSU (pb)	Mida del <i>read</i>	Cobertura	Núm. d'accés del mitogenoma	Núm. d'accés del SSU
<i>M. dominicanus</i> Jaume & Christenson, 2001	Juan Dolio, S Hispaniola (República Dominicana); pou	14,543 <sup>d</sup>	2511	625	4.3x <sup>a</sup>	HE860499	HE967299
<i>M. samauensis</i> Jaume & Christenson, 2001	Samaná, N Hispaniola (República Dominicana); pou	14,067 <sup>f</sup>	2413	551	5.7x <sup>a</sup>	HE860505	HE967297
<i>M. repeis</i> Stock & Rondé-Broekhuizen, 1986	Fuerteventura, Illes Canàries (Espanya); pou	14,355	2215	550	5.7x <sup>a</sup>	HE860495	HE967284
" <i>M. nicolae tamri</i> "	Aksri, NO d'Agadir (Marroc); deu a prop de la cova Talmat	13,517 <sup>e</sup>	1027+1268	411	145x <sup>3</sup>	HE860511	HE967292/3
" <i>M. nicolae tamri</i> "	Tamri, N d'Agadir (Marroc); pou	14,644	2415	595	59x <sup>b</sup>	HE860504	HE967294
" <i>M. boutini boutini</i> "	Timzelite, Souss Massa, S d'Agadir (Marroc); pou	13,301 <sup>e</sup>	2357	433	107x <sup>c</sup>	HE860497	HE967295
<i>M. panousei</i> Balazuc & Ruffo, 1953	Agdz (Marroc); pou	14,478 <sup>d</sup>	2295	453	80x <sup>c</sup>	HE860510	HE967289
<i>M. goulmimensis</i> Messouli, Boutin & Coineau, 1991	Lamkedmyia Meleh Jorf NO Erfoud (Marroc); pou	14,507 <sup>d</sup>	1173	413	178x <sup>c</sup>	HE860500	HE967279
<i>M. goulmimensis</i> Messouli, Boutin & Coineau, 1991	Oustroutou, E de Rich, N d'Errachidia (Marroc); pou	14,602	2179	454	100x <sup>c</sup>	HE860501	HE967278
<i>M. goulmimensis</i> Messouli, Boutin & Coineau, 1991	Zouala maison, S d'Errachidia (Marroc); pou	14,353	1922	423	165x <sup>c</sup>	HE860502	HE967280
<i>M. longicaudus</i> Ruffo, 1954	Lamkedmyia Meleh Jorf NO d'Erfoud (Marroc); pou	14,712 <sup>d</sup>	2272	403	135x <sup>c</sup>	HE860509	HE967281
" <i>M. notenboomii</i> "	Maadid, NE d'Erfoud (Marroc); pou	14,277 <sup>d</sup>	2275	420	168x <sup>c</sup>	HE860513	HE967298
" <i>M. paurosexualis</i> "	Souk Tben, Haouz Plain, Marràkeix (Marroc); pou	12,542 <sup>e</sup>	2370	424	94x <sup>c</sup>	HE860507	HE967291
<i>M. spinicaudatus</i> Karaman & Pesce, 1980	Souk Tben, Haouz Plain, Marràkeix (Marroc); pou	15,037	2338	534	124x <sup>b</sup>	HE860506	HE967290
<i>M. rennyi</i> Ruffo, 1953	Ijoukak, Alt Atlas Occidental (Marroc); deu a maison forestière	14,787 <sup>d</sup>	2246	454	281x <sup>b</sup>	HE860512	HE967287
" <i>M. boveei</i> "	Vall de L'Ouirika, Alt Atlas Occidental (Marroc); pou	15,012	2299	511	100x <sup>b</sup>	HE860498	HE967288
"Longipodacrangonyx stocki"	Taufraut (Marroc); pou	12,924 <sup>e</sup>	337+1210	395	160x <sup>c</sup>	HE860496	HE967282/3

Espècie	Localitat	Mida del mitogenoma (pb)	Mida del SSU (pb)	Mida del read	Cobertura	Núm. d'accés del mitogenoma	Núm. d'accés del SSU
"Longipodacrangonyx stockii"							
<i>M. longipes</i> Chevreux, 1909	Arbaa-Sahe, SO de Tiznit (Marroc); pou Mallorca, I. Balears; cova Cala Varques	13,006 <sup>f</sup> 14,113	N/A 2200	431 NA	103 <sup>xc</sup> NA	HE860508 AM944817	- HE967285
<i>M. longipes</i> Chevreux, 1909	Menorca, I. Balears; cova de Sa Figuera	14,117	2087	437	87 <sup>xc</sup>	HE861923	HE967286
<i>M. iloanus</i> Stoch, 1997	Illa d'Elba (Itàlia), pou	14,770 <sup>d</sup>	1173	540	78 <sup>xb</sup>	HE860503	HE967296
<i>Pseudoniphargus daviui</i> Jaume 1991	Illa de Cabrera, I. Balears; pou	15,155	1863	410	73 <sup>xb</sup>	FR872383	HE967300
<i>Bahadzia jaraguensis</i> Jaume & Wagner 1988	Oviedo; S d'Hispaniola (República Dominicana); cova	14,657	N/A	537	87 <sup>xb</sup>	FR872382	-

<sup>a</sup> Mètode de Sanger

<sup>b</sup> Roche FLX/454 amb marcatge

<sup>c</sup> Roche GS Junior sense marcatge

<sup>d</sup> Manca un bocí de la regió control

<sup>e</sup> Manca un bocí del fragment rrmL-rrmS

<sup>f</sup> Manca un bocí a la regió que inclou el rrmS i la regió control

test de Xia i Lemey (2009) per comprovar la presència de saturació a les diferents posicions nucleotídiques. Per a les primeres, segones i terceres posicions del triplets de MPCGs, el valors de  $I_{ss}$  (índex de saturació) han resultat ésser de 0.4644, 0.1609 i 0.6692, respectivament. Els valors de  $I_{ssc}$  (índex crític de substitució) pel nombre de taxa i la longitud de les seqüències de l'alineament analitzat ha estat de 0.8107-0.8197, la qual cosa indicaria que no hi ha saturació a les segones posicions i una saturació de moderada a baixa a les primeres i terceres posicions; això no obstant, els valors d' $I_{ss}$  obtinguts per a totes les posicions han estat significativament més baixos que els valors crítics d' $I_{ssc}$  (veure apartat 3.8 de Materials i Mètodes).

Les 21 seqüències nuclears ribosomals *SSU* s'alinearen considerant l'estructura secundària ribosomal (Kato & Toh, 2008) i s'analitzaren individualment i com una partició addicional per complementar el set de dades dels mitogenomes complets en algunes de les anàlisis. Les regions hipervariables i els blocs ambigüament alineats s'eliminaren amb el programa *Gblocks* v. 0.91b (Castresana, 2000) (veure Metodologia, apartat 3.3). Com a resultat s'obtingueren 49 blocs que contenen un màxim de 2159 caràcters (abans 2882) per seqüència.

**Taula 3.9.** Valors de BIC i BF pels diferents esquemes de particions emprats a les anàlisis.

Nombre particions	Esquema de particions	Ln*	K**	BIC***	LnP§	2LnBF§§
0		-149781,809	9	299647,484	-149977,39	-431,232
2	1+2 posicions vs. 3	-145801,879	16	291752,854	-145974,36	-844,244
3	1 vs. 2 vs. 3 posicions	<b>-144816,827</b>	26	289875,935	<b>-145002,387</b>	n.a.
13	per gen	-148918,591	124	298992,673	-149040,306	-3507,292
61	per codons§§§	-147746,836	67	296118,010	-145655,981	-567,704

En negreta s'ha marcat el valor més alt de versemblança.

\* Valor obtingut amb Garli.

\*\* K: número de paràmetres lliures

\*\*\* BIC =  $2 \text{Ln} + K \text{Ln} n$  ( $n$ = nombre de caràcters 11142 pb)

§ Valor de LnP (Model | Dades) obtingut amb MrBayes

§§ Un valor de  $2 \text{Ln BF} > 10$  es considera suficient per triar el model amb major nombre de paràmetres (Nylander et al., 2004; Kass & Raftery, 1995)

§§§ Model de Goldman & Yang (1994)

Els arbres bayesians obtinguts arran d'explorar els diferents sets de dades i els diferents possibles esquemes de particionament amb el millor model evolutiu implementat (segons valors de BIC obtinguts per MrAic) es mostren a la Figura 3.21. En el cas del set de dades per a tots els MPCGs el millor esquema de particionament segons el valor de BF (Factors de Bayes) i BIC (*Bayesian Information Criterion*) fou el de per posicions de codons -codons position- (veure A, tercer nombre, Fig. 3.21 i Taula 3.9). Tant per la resta d'esquemes de partició per al set de dades de MPCGs, com per al set

de dades per a aminoàcids –assumint els models mtArt i CAT- s’han obtingut arbres amb topologies similars, tot i que amb lleugeres discrepàncies pel que fa al monofiletisme del node B i C (Marroc Atlàntic), i amb alguns nodes amb un suport més baix (Fig. 3.21, A-C). La topologia basada en el set de dades de *SSU* (Fig. 3.21, D) ha resultat ésser concordant amb les obtingudes amb els MPCGs, tot i que amb valors de probabilitat posterior (*pp*) més baixos. L’addició del marcador nuclear *SSU* tractat com a quarta partició (Fig. 3.21, E) ha produït una topologia en gran manera concordant amb l’obtinguda amb el set de dades per a marcadors mitocondrials, afavorint emperò un parafiletisme del node que inclou els llinatges del Marroc Atlàntic. Els valors obtinguts als tests de S-H i al d’ELW indicarien que les topologies alternatives que s’han obtingut amb els diferents sets de dades (monofiletisme vs. parafiletisme del node B i C) no eren significativament diferents (Taula 3.10), pel que no es pot descartar cap de les dues topologies alternatives. Tot i així, atès que la topologia més freqüent obtinguda ha estat la que considera les espècies del Marroc Atlàntic com a parafilètiques, s’ha afavorit aquesta hipòtesi en aquest estudi.

**Taula 3.10.** Comparació de les diferents hipòtesis filogenètiques obtingudes per ML (monofiletisme vs. parafiletisme del node del Marroc Atlàntic).

Hipòtesis	Set de dades	lnL*	$\delta^{**}$	Test S-H***	ELW*
Monofiletisme	<i>SSU</i>	-8477,4867	2,3253	NS	0,8070
Parafiletisme		-8475,1616			
Monofiletisme	MPCGs	-125333,6521	-5,1383	NS	0,7518
Parafiletisme		-125338,7904			
Monofiletisme	MPCGs+ <i>SSU</i>	-133937,7229	0,6350	NS	0,5475
Parafiletisme		-133937,0880			

\* Valor del logaritme de Màxima Versemblança obtingut a RAxML.

\*\* Diferència entre el  $\ln L_{ML}$  del millor arbre obtingut per ML i el  $\ln L_i$  de la topologia alternativa a la qual es compara.

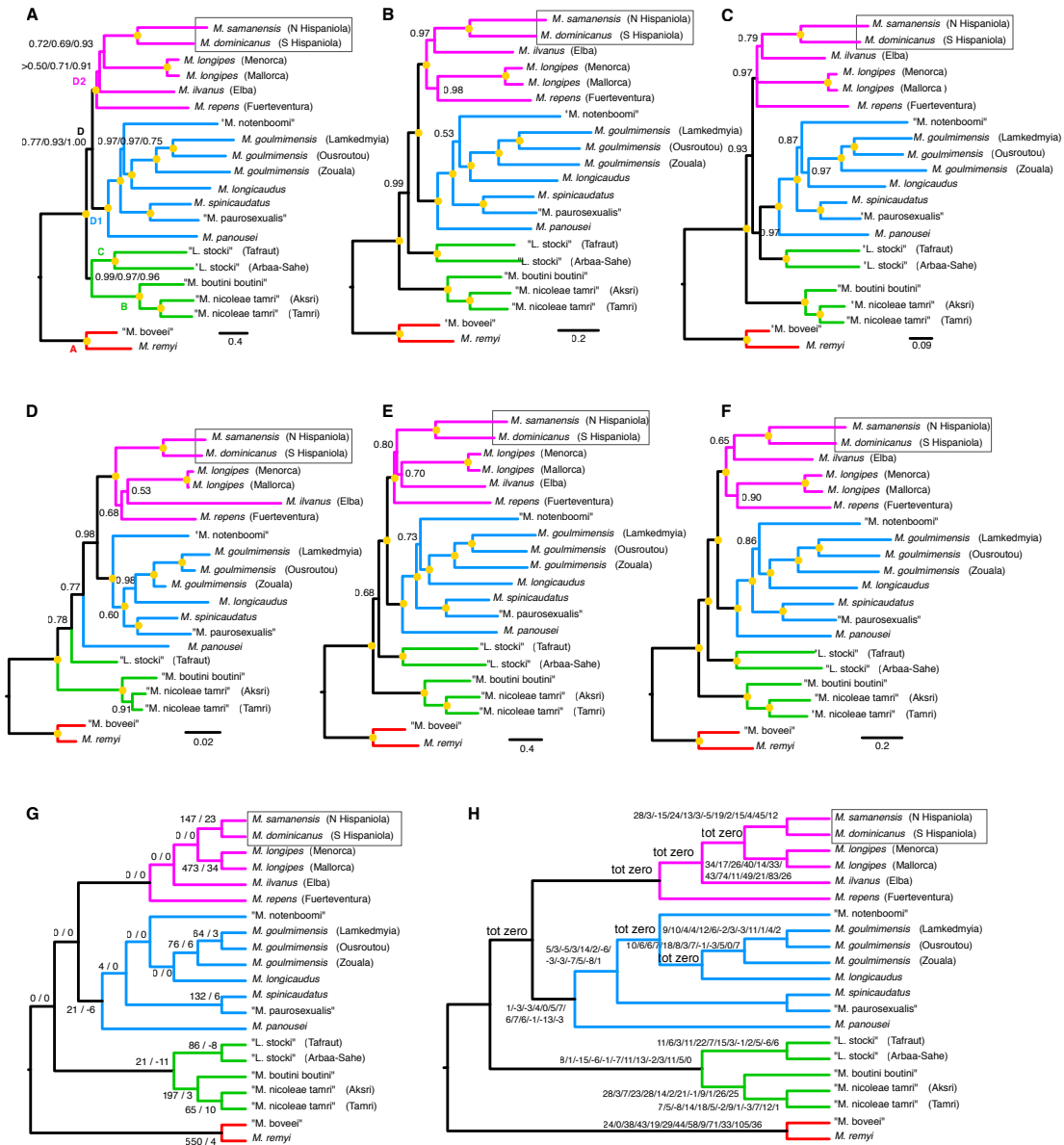
\*\*\* Resultat del test de S.H. NS significa que la diferència dels valors de lnL no és estadísticament significativa pel que no es permet descartar cap de les dues topologies testades (nivell de significació observat del 5%, 2% i 1%).

# Estadístic obtingut a partir del test de ELW que fa referència a la proporció de la versemblança de la topologia sobre la suma total de les versemblances de les dues topologies testades (és a dir, 1.0).

Els gràfics obtinguts arran de l’exploració dels arbres assolits per inferència bayesiana amb *AWTY* indiquen que les biparticions estan repartides aleatòriament al llarg de l’anàlisi, pel que les freqüències finals observades a cada bipartició no estan esbiaixades per mínims locals (Fig. 3.22).

A la Figura 3.23 es pot apreciar l’arbre bayesià obtingut a partir del set de dades combinat amb el millor model d’evolució segons BIC per a cadascuna de les tres particions (GTR+I+ $\Gamma$  per a les primeres i segones posicions, HKY+I+ $\Gamma$  per a les terceres

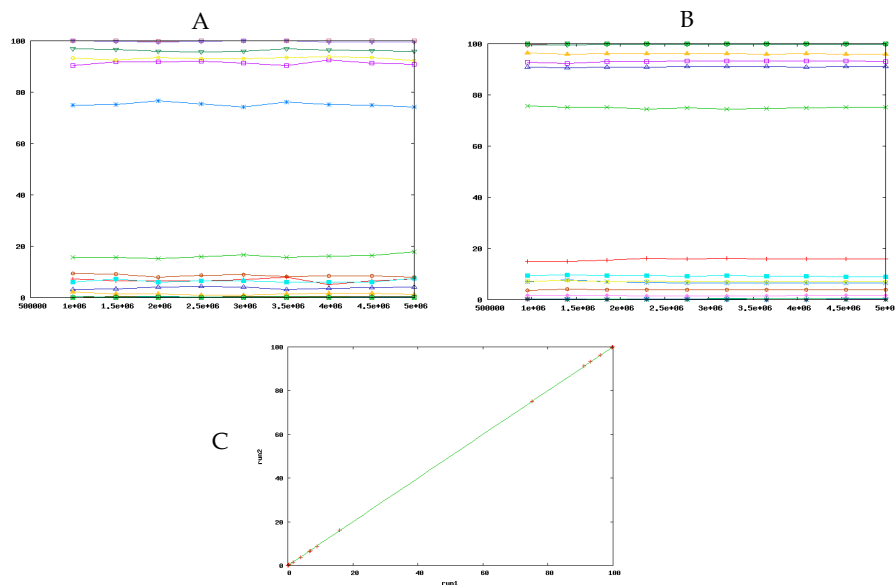
posicions, i GTR+I+ $\Gamma$  per a la partició de SSU). Aquest arbre es divideix en cinc clades principals altament suportats dins la família Metacrangonyctidae (A, B, C, D1 i D2), on cadascun d'ells mostra una clara delimitació geogràfica, tot i que B i C comparteixen la mateixa àrea al Marroc Atlàntic (Fig. 3.24).



**Figura 3.21.** Arbres filogenètics obtinguts amb els diferents esquemes de partició i diferents models evolutius. Els números en els nodes dels arbres A-F denoten els valors de probabilitat posterior inferiors a 1, mentre que els nodes amb un punt groc tenen un valor de probabilitat posterior d'1. S'ha destacat amb un quadre les espècies del Carib. **A)** Arbre obtingut a partir dels MPCGs tractats com a una sola partició (1r número), dues particions (1+2 posicions per codó vs. 3) (2n número), i tres particions (cada posició del codó es tractada com a una partició distinta) (3r número). **B)** Arbre obtingut arran de la implementació del model per codons de Goldman i Yang (1994). **C)** Arbre obtingut amb les seqüències aminoacídiques dels MPCGs. **D)** Arbre amb el set de dades per a SSU. **E)** Arbre obtingut a partir del set de dades combinat de MPCGs (particionat per posició de codons) \* SSU (com a 4a partició). **F)** Arbre amb els MPCGs amb el model per codons de Goldman i Yang (1994) + SSU. **G)** Valors obtinguts a l'anàlisi PBS amb el set de dades de tots els MPCGs vs. SSU. El primer número correspon al número de passos de Parsimònia pels MPCGs i el 2n número per a SSU. Els valors negatius indiquen un conflicte a la topologia representada per la partició de SSU en comparació de la topologia obtinguda a partir de tots els MPCGs. **H)** Valors de PBS per a cadascun dels MPCGs. Els valors són el número de passos de Parsimònia per a cada un dels MPCGs (ordre dels gens: atp6, atp8, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad4 i nad6).

Tant les seqüències mitocondrials com les del gen SSU dels individus pertanyents a la mateixa morfo-espècie han resultat ésser totes monofilètiques, malgrat que en certs

casos han estat altament divergents, com és el cas de *M. goulmimensis* i “*L. stocki*” (les distàncies genètiques no corregides per a *cox1* són: 0.164-0.176 i 0.159, respectivament). Totes les topologies assignen una divergència primerenca a “*M. boveei*” i *M. remyi* (clade A), que de fet es troben limitades a les valls del nord i deus de l’Alt Atlas occidental al Marroc (Figures 3.23 i 3.24). La resta de tàxons es distribueixen en tres clades principals, dos dels quals comprenen les espècies habitants de la costa atlàntica del Marroc (clade B i C). L’altre (clade D) es troba subdividit en dos sub-clades fortament suportats: el sub-clade D1, que es troba en ambdós costats de l’Alt Atlas al Marroc, i el sub-clade D2, que abasta les cinc espècies insulars de *Metacrangonyx*: *M. longipes* de Mallorca i Menorca, *M. ilvanus* d’Elba a la Mediterrània, *M. repens* de Fuerteventura a l’Atlàntic i *M. samanensis* i *M. dominicanus* de La Hispaniola al Carib, respectivament.

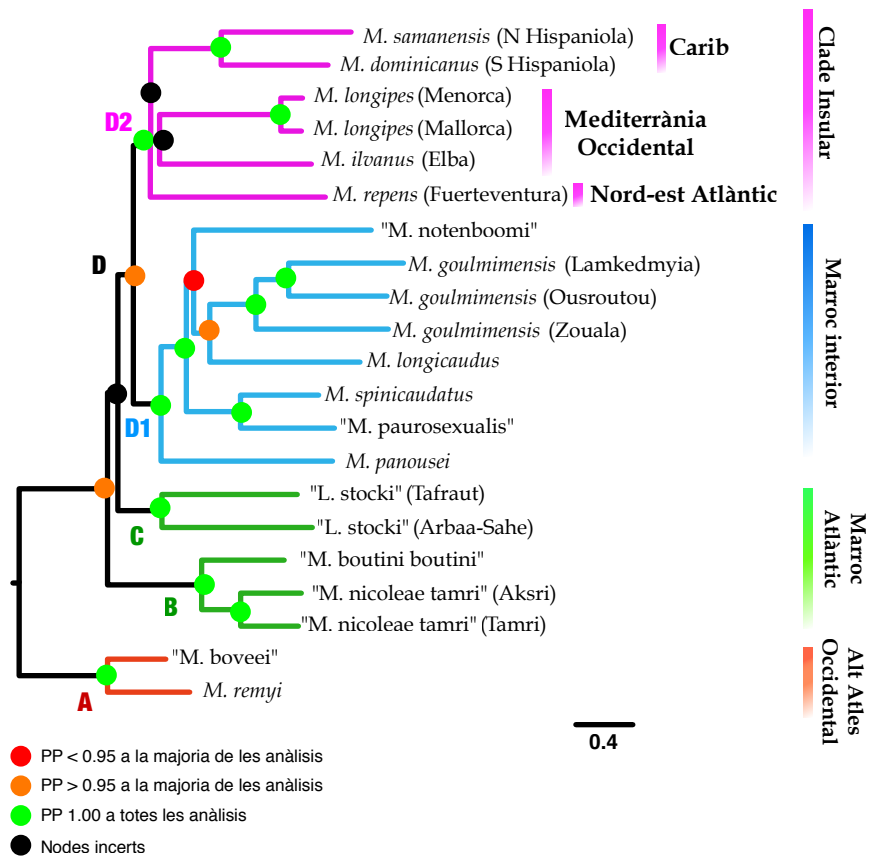


**Figura 3.22.** Estacionalitat (A i B) i convergència (C) de les topologies obtingudes a l’anàlisi bayesiana pel set de dades combinat (MPCGs + SSU) a *Metacrangonyctidae* calculades per AWTY (Nylander et al., 2008), amb els ordres slide (A), cumulative (B) i compare (C).

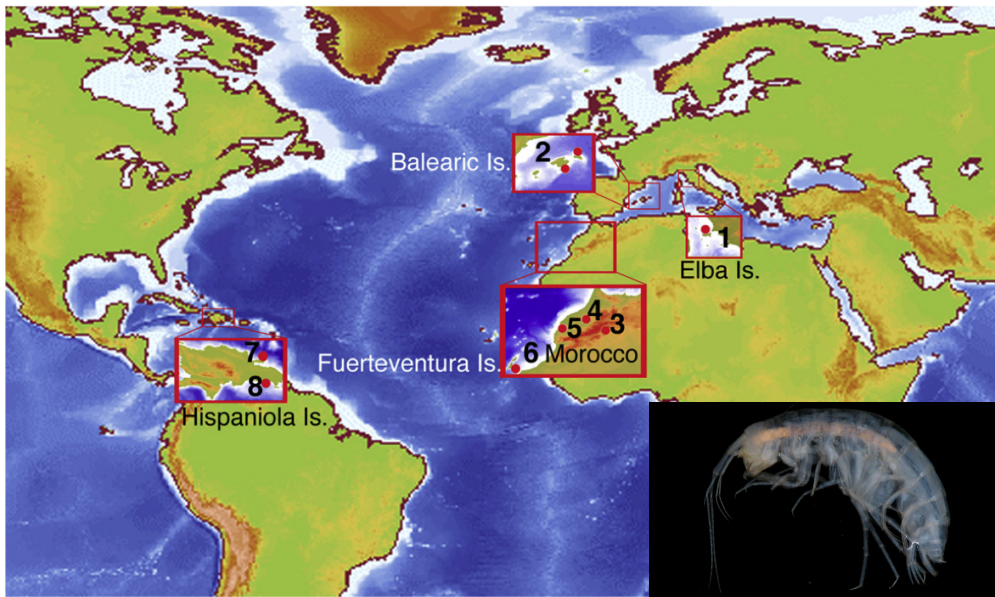
Les incerteses filogenètiques detectades s’han limitat a les posicions assignades a “*M. notenboomii*” de Maadid (sud-est de Marroc) del sub-clade D1, i la posició relativa de *M. ilvanus* de l’illa d’Elba (arxipèlag Toscana, Itàlia) amb els dos llinatges de *M. longipes* de les Balears dins els sub-clade D2. El test de PBS (*Partitioned Bremer Support*), el qual tracta a cadascun dels 13 MPCGs com una partició diferent, suggereix que la manca de resolució dins el sub-clade insular de D2 sorgeix de la falta de senyal filogenètic i no pas de la incongruència entre les diferents particions (Figura 3.21). Una ulterior anàlisi bayesiana realitzada a *Phycas* v.1.2.0. ha permès descartar la hipòtesi on



s'assumeix l'existència d'un arbre totalment resolt (totalment dicotòmic), afavorint la topologia que integra una "politomia vertadera" (*hard politomy vs. soft politomy*) (els valors de  $2xLnBF$  són de 879.54 comparant la topologia amb un prior de  $e^1$  vs. topologia completament dicotòmica i de 337.72 en el cas de comparar-ho amb un prior de  $e^2$ ). Un cop s'assumí un *prior* de politomia de  $e^1$  i  $e^2$ , s'observà una notable reducció dels valors de probabilitat posterior als nodes rellevants que inclouen les espècies insulars (D2) i la posició de l'espècie "M. notenboomi" dins el node D1, tot i que a la topologia més freqüent obtinguda a *Phycas* l'únic node que col·lapsa és l'insular D2 (Figura Addicional 2). Addicionalment, amb el programa DensiTree v.2.0.1. (Bouckaert, 2010) s'ha pogut visualitzar que la topologia dominant a la majoria dels arbres obtinguts per inferència bayesiana és la que assenyalava una politomia vertadera (Fig. 3.25) ensem amb la posició dubtosa de "M. notenboomi" dins el clade D1.

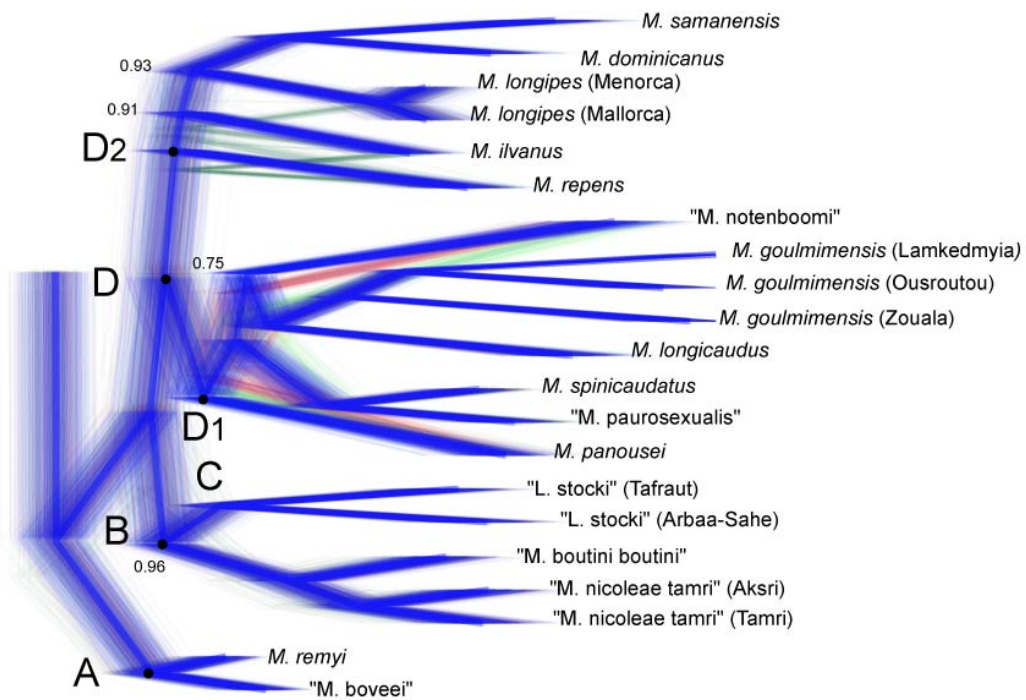


**Figura 3.23.** Arbre bayesià obtingut amb el set de dades de tots els MPCGs + SSU, amb el millor esquema de particionament i model evolutiu. El monofiletisme de *Metacrangonyctidae* està suportat amb un valor de probabilitat posterior d'1 a totes les anàlisis (no es mostren els grups externs). Els detalls de les diferents topologies obtingudes i els diferents valors de suport per a cadascuna de les anàlisis es mostren a la Figura 3.21).



**Figura 3.24.** Distribució de la família Metacrangonyctidae. El mapa mostra les localitzacions aproximades de les espècies les quals el seu mitogenoma ha sigut seqüenciat. 1) *Metacrangonyx ilvanus*; 2) *M. longipes*; 3) "*M. notenboomi*", *M. goulmimensis*, *M. longicaudus*, *M. spinicaudatus*, "*M. paurosexualis*", *M. panousei*; 4) *M. remyi*, "*M. boveei*"; 5) "*L. stocki*", "*M. boutini boutini*", "*M. nicoleae tamri*"; 6) *M. repens*; 7) *M. samanensis*; 8) *M. dominicanus*.

Foto: exemplar de *M. repens* per David Hernández Teixidor.

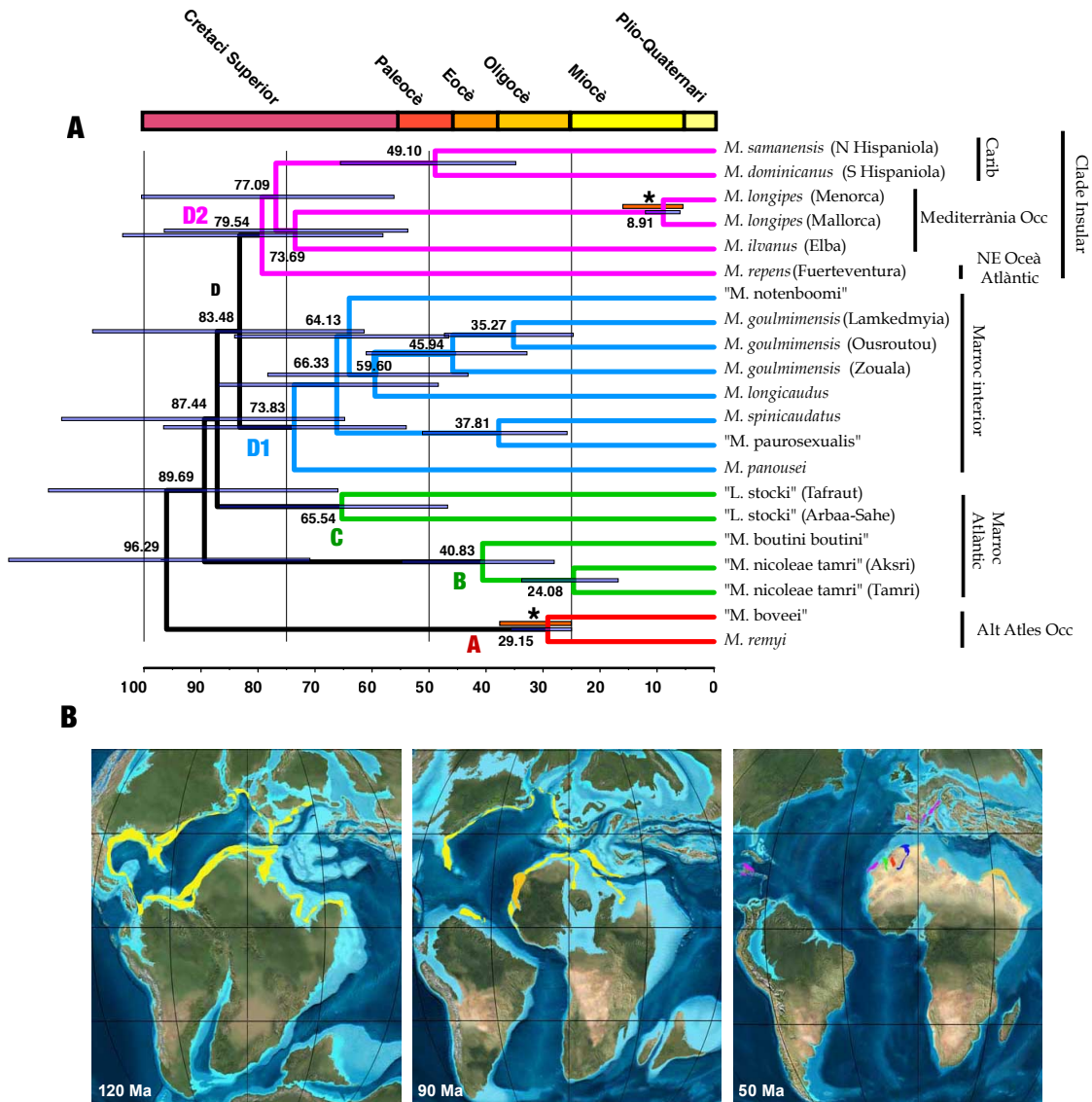


**Figura 3.25.** Representació de la densitat dels valors de probabilitat posterior dels arbres de Metacrangonyctidae obtinguda amb DensiTree. Els nombres indiquen els valors de probabilitats posteriors a l'arbre consens obtinguts a MrBayes amb el set de dades per a MPCGs. Aquelles àrees on la majoria dels arbres coincideixen quant a la topologia i/o longitud de branques presenten una coloració més densa (nodes ben suportats). Els diferents colors fan referència a les tres topologies alternatives obtingudes (el blau per la topologia més freqüent, el vermell per la següent topologia i el verd per la topologia més infreqüent).

### iii. *Estimacions del rellotge molecular*

Per a les estimacions es van excloure les seqüències nuclears de SSU perquè contenen *indels* llargs, els quals són indicadors de l'ocurrència d'un gran nombre d'esdeveniments altament estocàstics que resulten difícils de modelar en comparació amb les substitucions nucleotídiques. Un cop descartada la hipòtesi d'un rellotge molecular estricta, s'implementà un rellotge molecular relaxat (rellotge no correlacionat de taxes entre branques amb una distribució *log-normal*) (BF= 167.93,  $p < 0.001$ ; ucl.d.mean = 0.23, 0.28, 0.18 per a les primeres, segones i terceres posicions) a BEAST pel calibratge de la filogènia dels mitogenomes. Per aquesta comesa, es feren servir dos esdeveniments paleogeogràfics. D'una banda, s'assumí que la divergència dels dos llinatges de *M. longipes* a l'arxipèlag Balear estava relacionada amb el conjunt d'esdeveniments geològics que tingueren lloc a la Mediterrània occidental durant el Miocè Mitjà i Superior (des de 16 a 5.5 Ma). És a dir, l'origen del llinatge insular de *M. longipes* no podria ser anterior a la separació completa de les Illes Balears de la resta de microplaques continentals que formaven el cinturó peri-Hercinià d'Ibèria, la qual s'ha datat en 16 Ma (veure Apartat 7, Introducció), ni posterior al final de la Crisi de salinitat del Messinià fa 5.5 Ma, quan acabà la gran regressió que assecà total o parcialment la conca Mediterrània (Cuerda et al., 1969; Hsü et al., 1997; Pomar, 1979, 1991; Krijgsman et al., 1999; Duggen et al., 2003). D'altra banda, s'assumí que l'edat del node corresponent a les espècies germanes *M. remyi* i "*M. boveei*", les quals habiten a valls molt pròximes del vessant nord de l'Alt Atlas de Marràqueix al Marroc, no podia ser més jove que l'elevació d'aquesta serralada. D'acord amb Frizon et al. (2008), l'Atlas romangué submergit durant tot l'Eocè Mitjà (48.6-37.2 Ma); el primer aixecament data de l'Eocè Superior (37.2-33.9 Ma), mentre que el primer plegament important es produí durant l'Oligocè i s'estengué fins al començament del Miocè (25 Ma). Per aquest motiu, es proposà l'assignació d'un interval d'edat de 37.2 a 25 Ma per datar l'últim ancestre comú de les dues espècies germanes. El cronograma obtingut (Fig. 3.26) li assignà una edat de 95.8 Ma (amb un valor de HPD de 72.0-123.9 Ma) a la diversificació inicial que conduí als llinatges contemporanis de metacrangoníctids. La resta de les principals divisions a la filogènia del grup tingueren lloc en un marc de temps relativament estret, entre fa 88.0 (65.1-114.2) i 74.9 (55.2-97.6) Ma. El test de validació creuada va mostrar que les edats dels nodes calculats assumint per separat cadascun dels dos esdeveniments paleogeogràfics esmentats anteriorment, o bé implementat diferents models de diversificació i emprant les topologies alternatives -segons el monofiletisme-

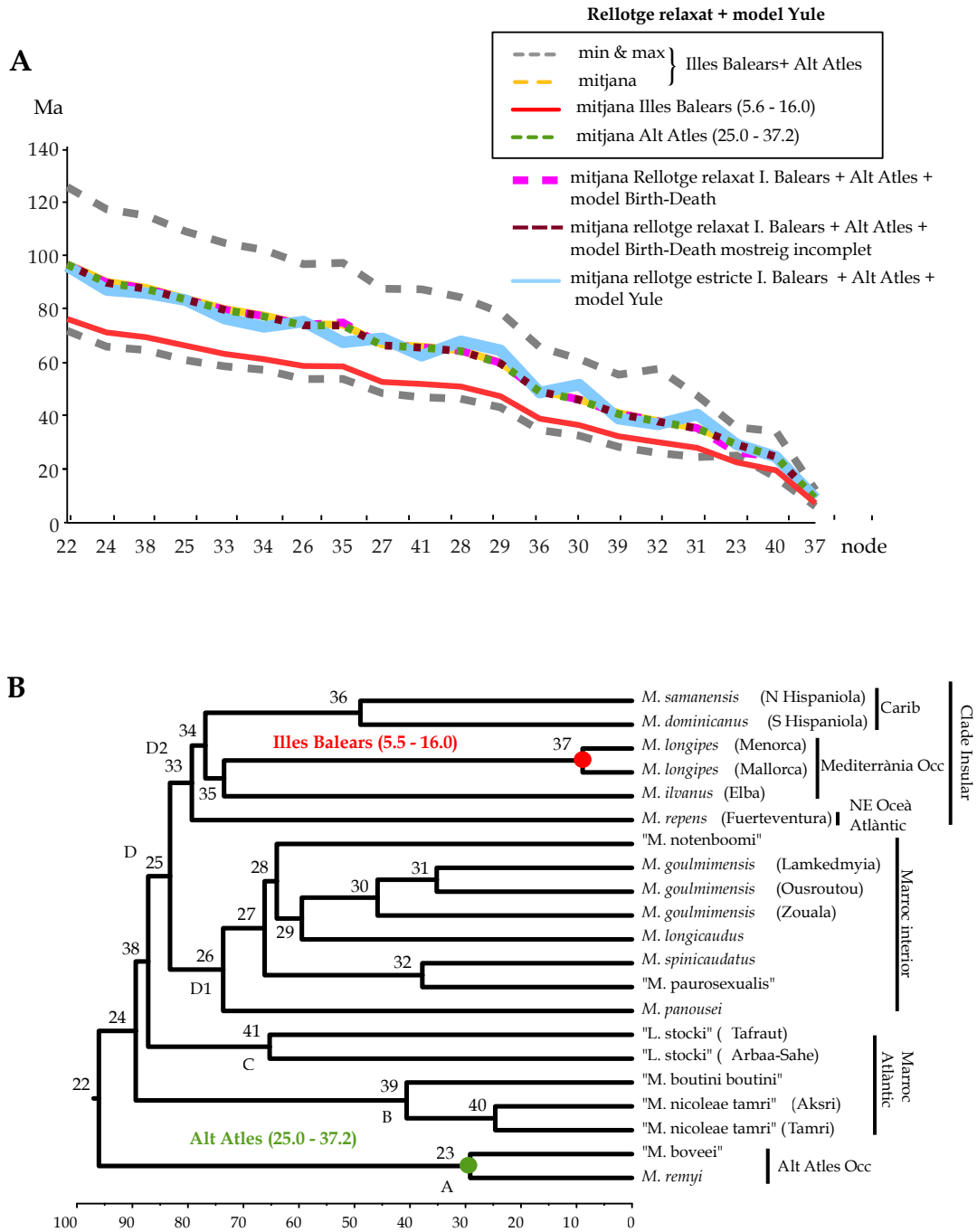
parafiletisme dels clades B i C-, resultaren ser similars i caigueren dins l'interval del 95% del valor de HPDs de les estimacions derivades després de considerar alhora els dos punts de calibratge (Fig. 3.27).



**Figura 3.26.** Marc temporal de la diversificació de la família *Metacrangonyctidae*. A) Estimacions de les edats de divergència a partir de l'anàlisi bayesiana de tots els MP CGs, amb dos punts de calibratge paleogeogràfic i un model de diversificació de Yule. Sobre els nodes s'indiquen els valors mitjans, mentre que les barres horitzontals representen els intervals de 95% de HPDs. Els asteriscs assenyalen els nodes fixats als punts de calibratge, amb llurs rangs d'edats en vermell.

B) Mapes de la paleogeografia global on s'indica la distribució putativa de *Metacrangonyctidae* en tres períodes diferents durant el Cretaci i Paleogen.





**Figura 3.27.** A) Comparació de les edats estimades pels nodes amb un o dos esdeveniments paleogeogràfics sota un rellotge estricte o relaxat i assumint dos models de diversificació diferents: el de Yule o el de Birth-Death. La separació dels dos llinatges balears de *M. longipes* es calibren amb un prior de 16 a 5.5 Ma, mentre que la divergència de *M. remyi* i "*M. boveei*" a l'Alt Atles es va fer amb un prior de 37.2-25 Ma. Tant les edats mínimes com les màximes es troben dins el 95% HPDs (gris), i els valors mitjans (groc), tenint en compte ambdós esdeveniments paleogeogràfics i un rellotge relaxat sota un model de Yule, es troben indicats amb línies discontinues. Els valors mitjans pels punts de calibratge per a les I. Balears i l'Alt Atles Occidental es mostren en vermell i en verd, respectivament. La mitjana de les edats obtingudes assumint un rellotge estricte i ambdós punts de calibratge baix un model de Yule s'indiquen en blau; en rosa per un model de Birth-Death i marró per un model Birth-Death de mostreig incomplet.

B) Arbre ultramètric que mostra la correspondència dels nodes amb els números assignats i els dos punts de calibratge emprats a les anàlisis.

## Discussió

Fins a la realització d'aquesta tesi solament s'havien fet servir les seqüències parcials del mitocondri i del genoma nuclear per explicar les distribucions disjunctes d'alguns crustacis i resoldre'n llur filogènia. Alguns exemples importants a esmentar són els estudis realitzats vers els decàpodes àtids (Page et al., 2008; Botello et al., 2012), els remipedis (Neiber, 2011) i els isòpodes cironàlids (Baratti et al., 2010). Encara que les filogènies basades en mitogenomes complets representin sols un locus únic i per tant, podrien no reconstruir necessàriament de forma correcte l'arbre de les espècies, compten amb un poder de resolució considerablement major que no pas les seqüències parcials de DNA -ja bé siguin mitocondrials o nuclears- a causa del gran nombre de posicions de nucleòtids considerades i l'alta taxa de mutació exhibida pels MPCGs (Duchene et al., 2011).

Les relacions filogenètiques inferides dins la família Metacrangonyctidae a partir dels 13 MPCGs i el marcador nuclear *SSU*, particionant les dades per posició dels codons, han resultat en un arbre gairebé resolt en la seva totalitat, car la politomia vertadera (*hard politomy*) que afecta al clade insular D2 (Fig. 3.23). Els resultats proven que la combinació de les filogènies de gens mitocondrials més la del gen nuclear *SSU* són compatibles amb altres hipòtesis d'arbres de gens basats només amb els MPCGs, encara que la combinació del set de dades ha aportat valors de suport estadístic menors a certs nodes (Fig. 3.21).

El nostre cronograma assigna una edat de 96 Ma a la diversificació inicial que sofrí la família Metacrangonyctidae i ubica al Cretaci Mitjà l'origen dels principals clades de la família (Fig. 3.26). Una antiga diversificació d'aquest llinatge ancestral tingué lloc a una escala geogràfica molt reduïda al Marroc, on concorren quatre (A, B, C, D1) dels cinc llinatges monofilètics reconeguts, un d'ells (D1) germà del cinquè clade insular (D2).

Les edats estimades al nostre estudi per la diferenciació dels cinc principals llinatges de Metacrangonyctidae es troben en concordança en termes generals amb Boutin (1994) (Figura 1.7, Introducció) en l'assignació d'una edat molt antiga per aquesta família, amb la diferenciació dels principals llinatges tenint lloc durant el Cretaci (entre 96-85 Ma). Aquest rang d'edat coincideix amb el gran cicle de

---

transgressió-regressió marina del Turonià-Cenomanià, al període del Cretaci Superior, el qual afectà gairebé tota la geografia del Marroc actual, i que s'ha suggerit anteriorment com a causa principal de la individualització en les aigües subterrànies continentals d'alguns llinatges marroquins (Boutin, 1994).

La nostra filogènia mostra que les espècies insulars de *Metacrangonyx* conformen un grup monofilètic fermament suportat (clade D2) àdhuc presentar una distribució extremadament disjunta. Sorprenentment, la espècie *M. repens* de Fuerteventura (Illes Canàries) pertany a aquest clade insular en lloc d'estar relacionada amb els tàxons marroquins, molt més propers geogràficament. S'han obtingut dos nodes no resolts en el clade insular: aquell que explicaria la relació entre els tàxons insulars mediterranis (l'illa d'Elba i els de les Illes Balears) amb les espècies de l'Atlàntic (Fuerteventura i La Hispaniola). Anteriorment, s'ha suggerit que la dificultat que rau en esbrinar les relacions filogenètiques d'amfípodes subterranis podria derivar de l'ocurrència d'episodis de diversificació molt sobtats (radiacions) associats a una colonització sincrònica de les aigües subterrànies per part de les diferents poblacions d'un mateix ancestre comú (Lefébure et al., 2007). La nostra troballa quant al clade insular -que es veu que ha divergit cap a la base com a una politomia forta o vertadera (*hard or true polytomy*)- estaria d'acord amb aquests postulats.

La hipòtesi vicariant es veu suportada en aquest estudi per l'edat calculada per la divergència entre el llinatge insular de La Hispaniola i el seu grup germà balear, la qual segons les nostres anàlisis rau en 75.3 Ma (55.5-98.1 Ma segons el 95% HPD) (Fig. 3.26). En aquella època, el Carib, l'Atlàntic oriental i la part de la mar de Tetis corresponent al que avui en dia seria la Mediterrània occidental, formà una via marítima contínua, on les seves costes i illes es situaven més a prop entre elles del que es troben entre si a l'actualitat (Smith et al., 1994). Cal esmentar que en aquella època l'illa La Hispaniola, Canàries i tota la conca de la Mediterrània occidental encara no s'havien format. De fet, la configuració actual del Carib no és més antiga que l'Eocè Superior (Iturralde-Vinent, 2006); el sorgiment de l'illa de Fuerteventura, la més antiga de l'arxipèlag canari, es remunta a 22 Ma enrere (Coello, 1992), mentre que tota la conca de la Mediterrània occidental es formà fa 20 Ma gràcies al desplaçament i rotació levogira del bloc Corso-Sard associat a l'obertura de la Conca Provençal (Alvarez, 1972). A més, hi ha evidències geològiques prou convincents que daten la presència d'arxipèlags i monts submarins al centre de l'Atlàntic oriental en almenys 60 Ma

---

(Geldmacher et al., 2001, 2005; Fernández-Palacios, 2011). Un altre factor a considerar és que aquestes illes paleo-Macaronèsiques es trobaven molt més a prop de la Mediterrània occidental que no pas a l'actualitat (Fernández-Palacios, 2011). D'altra banda, és molt probable l'existència d'illes efímeres que romangueren emergides uns pocs milions d'anys en el Proto-Carib (illes volcàniques, conques poc profundes) durant el Cretàcic Inferior (Iturralde-Vinent, 2006). Per tot això, és factible pensar que l'ancestre del llinatge insular de *Metacrangonyx* fos en realitat una espècie marina adaptada a aigües poc profundes que colonitzés les conques o zones de ribera d'aquests arxipèlags efímers i que no tindrien perquè coincidir exactament amb el lloc precís que ocupen les espècies actuals.

Per mitjà dels gens mitocondrials codificants per proteïnes hem calculat una taxa evolutiva a llarg termini de 10.9% per milió d'anys i per parell de seqüències. Aquest fet resulta sorprenent tenint en compte que aquesta taxa és gairebé cinc vegades més alta que la taxa estàndard de 2.3% trobada a mitogenomes d'artròpodes (Bower, 1994) i pels MPCGs a escarabats (Pons et al., 2010), o altres taxes estimades pel gen *cox1* a decàpodes marins (1.4%-2.6% per milió d'anys) les quals s'han fet servir sovint en estudis evolutius d'altres crustacis (Kornobis et al., 2010). En cas d'utilitzar la taxa estàndard de 2.3% a la família Metacrangonyctidae, les edats estimades que s'obtenen són encara més antigues (encara que poc creïbles), de manera que no es contradiu la conclusió principal que esdeveniments vicariants molt antics són els responsables del patró de distribució esbossat per aquesta família. L'única estimació similar de la qual es té coneixença és la taxa del 20% per milió d'anys a *cox1* calculada per a *Halocaridina rubra*, un decàpode habitant d'aigües subterrànies a l'arxipèlag de Hawaii, el qual sembla que presentaria una evolució accelerada a causa de la seva marcada estructura genètica i la freqüent aparició de colls d'ampolla poblacionals en aquesta espècie (Craft et al., 2008).



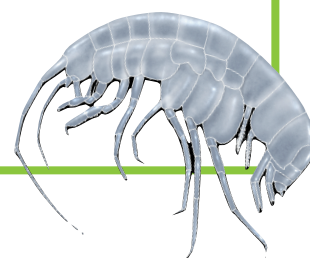
## Conclusions

La distribució amfi-atlànica de la família d'amfípodes estigobionts Metacrangonyctidae podria haver estat causada per vicariança d'un ancestre marí que es trobés àmpliament distribuït arreu del litoral de la Mar de Tetis o, alternativament, per una dispersió activa d'aquest ancestre marí a partir d'una àrea d'origen més localitzada (Stock, 1977; 1980; Holsinger, 1986; 1988; Rouch & Danielopol, 1987; Boxshall, 1989). En aquest darrer cas emperò, caldria esperar trobar a les aigües continentals llinatges amb diferents edats compartint el mateix territori, ja que haurien resultat de diferents episodis independents de colonització, independents entre si, i que la distribució general presentada per la família no es correspongués necessàriament amb l'antiga localització de la mar de Tetis. Per contra, a la nostra anàlisi molecular, ens trobem que el temps de divergència entre els diferents llinatges metacrangoníctids és extremadament antic i que les espècies insulars semblen tenir un origen sincrònic i molt antic.

Altres grups de crustacis estigobionts que han estat estudiats, com ara els llinatges de remipedis i alguns ostràcodes, copèpodes, termosbenacis i decàpodes també presenten presumptament aquesta distribució "tetiana" (Juan et al., 2010). La inferència de filogènies moleculars i la datació dels esdeveniments cladogenètics podrien sens dubte ajudar a clarificar l'origen dels seus patrons de distribució i llurs relacions amb la fragmentació de la Mar de Tetis. La gran quantitat de dades mitocondrials obtingudes a partir de la tecnologia de seqüenciació de DNA de nova generació resulta ésser una valuosa font d'informació genètica a l'hora d'elucidar el procés de diversificació i la seva datació en aquests grups de crustacis estigobionts.

---

**DISCUSSIÓ GENERAL**





Malgrat els avanços recents en el coneixement dels seus patrons i processos macro- i microevolutius, l'estudi de l'especiació i evolució de la fauna subterrània es troba encara en una fase inicial (Gibert & Deharveng, 2002; Juan et al., 2010). Alguns dels pocs estudis recents sobre l'origen i l'evolució de llinatges subterranis a artròpodes i que utilitzin dades moleculars són les aportacions de Ketmaier et al. (2003), Baratti et al. (2004), Arnedo et al. (2007), Ribera et al. (2010) i Juan i Emerson (2010). I no són poques les raons que expliquen aquesta mancança. D'una banda existeixen grans entrebancs a l'hora de localitzar i capturar espècimens, ja que ocupen ambients extremadament aïllats i de difícil accessibilitat (Culver et al. 2006). D'altra banda, qualsevol investigador que pretengui treballar amb aquesta fauna, probablement es topirà amb un gran nombre de llinatges críptics atès l'estasi morfològica sovint tan present, el que li causarà un bon garbuix (Proudlove & Wood, 2003; Lefébure et al., 2006). Menció apart mereix la gran dificultat per trobar marcadors moleculars prou fidedignes per aclarir la història evolutiva de llinatges que, en molts dels casos, divergiren (en termes evolutius) fa molt de temps. En aquest cas hom esperaria que una part considerable del senyal filogenètic hagués quedat esborrat per múltiples mutacions a les mateixes posicions nucleotídiques. Com poder respondre llavors la gran pregunta que ens ve immediatament al cap: quins esdeveniments històrics han modelat les distribucions actuals disjunctes de nombrosos tàxons subterranis relictos? Es deuen fonamentalment a processos de vicariança o bé a dispersió?

La present tesi doctoral ha pretès inferir la història de la colonització de les aigües subterrànies continentals per part dels membres de la família de crustacis amfípodes Metacrangonyctidae. S'ha estudiat la seva diversificació a partir de l'anàlisi d'una combinació de marcadors mitocondrials i nuclears, i de mitogenomes complets, i emprant diferents punts de calibratge basats en esdeveniments geològics pretèrits. L'ús d'aquestes eines moleculars i l'aplicació de rellotges moleculars han suposat un gran avanç en el coneixement de l'evolució d'aquest grup d'amfípodes subterranis, ja que fins ara es disposava tan sols de dades morfològiques per abordar aquests estudis.

Al capítol I es presenten el resultats obtinguts pel que fa a la història filogeogràfica de l'espècie endèmica de les Balears, *Metacrangonyx longipes*, la qual s'ha vist que es troba fortament estructurada geogràfica i genèticament, i on estimacions basades en el rellotge molecular assenyalen un origen del llinatge anterior a 6 Ma. Al

---

capítol II es descriu el primer genoma mitocondrial anotat per amfípodes fins aleshores, el de *M. longipes*, el qual ha resultat ésser el mitogenoma més petit dins d'aquest grup i el que té la regió control més curta dins crustacis, a més de presentar un ordre gènic diferent al descrit com el patró ancestral de Pancrustacea i no ésser compartit per cap altre crustaci. Al capítol III s'ha estudiat la biogeografia de la família Metacrangonyctidae, on gràcies a l'anàlisi de les seqüències nucleotídiques de mitogenomes i un marcador nuclear, s'ha pogut en gran part esbrinar la filogènia de la família i datar-ne els esdeveniments cladogenètics més importants. Al llarg d'aquesta tesi s'ha observat com els fenòmens d'aïllament geogràfic i esdeveniments vicariants - d'una banda "horizontals", com la formació de conques oceàniques i separació de plaques continentals, i d'altra banda "verticals", com processos orogènics i moviments eustàtics- han forjat la història evolutiva i els patrons de distribució actuals d'aquest grup d'amfípodes hipogeus.

### **1. Diversitat genètica a la fauna subterrània**

Arran de les primeres investigacions evolutives sobre organismes subterranis es proposaren dues hipòtesis alternatives quant a la diversitat genètica de la fauna hipogea (Kane, 1982). D'una banda, Barr (1968) suggerí que els animals hipogeus han experimentat nombrosos colls de botella poblacionals rere quedar aïllats dels seus ancestres epigeus, tot i que aquests episodis haurien estat de curta durada i haurien permès a les poblacions cavernícoles sobrevivents expandir-se cap a noves àrees subterrànies (veure Juan et al., 2010). Més endavant, Poulson i White (1969) proposaren una hipòtesi alternativa la qual suggeria que la fauna hipogea presentava una molt baixa variabilitat genètica atès haver sofert durant llargs períodes l'aïllament a ambients extrems (amb escassos recursos tròfics) i a l'adaptació a condicions ambientals molt estables. A més a més, segons aquests autors, la baixa variabilitat fenotípica observada en aquests organismes quant a la pèrdua de estructures visuals, de pigmentació i altres caràcters morfològics comuns, reflectien en realitat una diversitat genètica reduïda. Estudis recents que fan ús de la genètica molecular, han confirmat reiteradament que la fauna subterrània en realitat presenta una elevada diversitat genètica, tant intra- com interespecífica (Buhay & Crandall, 2005; Lejeune & Chevaldonné, 2006; Guzik et al., 2009). Els resultats presentats al capítol I d'aquesta memòria han corroborat també que les poblacions hipogees gaudeixen d'una gran variabilitat genètica, en mostrar que les diferents poblacions de *M. longipes* són en gran

---

mesura divergents i estan altament estructurades genèticament. Resultats similars s'han assolit al capítol III quant a l'estudi de la filogènia de la família Metacrangonyctidae, on es mostra l'elevat grau de divergència existent entre les espècies de la família i fins i tot a nivell intraespecífic, com és el cas entre diferents poblacions de les espècies *M. goulmimensis* o de "*L. stocki*" (distàncies no corregides igual o superiors a 0.16 per al gen *cox1* en ambdós casos). Seria bo esmentar el treball realitzat per Lefébure et al. (2006b), on es proposa un llindar molecular per a la delimitació d'espècies dins Crustacea de 0.16 substitucions/lloc a *cox1*, i que ha estat emprat ulteriorment per destriar entre possibles llinatges críptics a alguns tàxons de crustacis hipogeus (Trontelj et al., 2009; Juan et al., 2010; Camacho et al., 2011).

## 2. Antiguitat i "endemicitat" a la fauna estigobiont

Una característica comú de les faunes subterrànies és l'elevada proporció d'espècies endèmiques que hi trobem, sobretot en el cas d'hàbitats aquàtics subterranis. En aquests ecosistemes, caracteritzats per la seva estabilitat ambiental (Trontelj et al., 2009), la biodiversitat sembla estar fortament lligada a l'edat i als processos causants de l'aïllament/fragmentació de l'hàbitat (Gibert & Deharveng, 2002). Per aquest motiu, hom podria esperar trobar llinatges més joves a hàbitats recents com els oferts per illes volcàniques, i llinatges extremadament antics a zones tropicals, on podrien haver persistit amb nul·la o escassa influència dels eventuais canvis ecològics o climàtics ocorreguts a la superfície (Gibert & Deharveng, 2002). Tot i la gran homogeneïtat morfològica de les espècies hipogeus, estudis taxonòmics previs en aquesta investigació havien aconseguit identificar fins a 35 espècies diferents dins la família Metacrangonyctidae, de les quals 28 són exclusives del Marroc. Exemples notables de diversificació serien el parell d'espècies endèmiques *M. dominicanus* i *M. samanensis*, habitants de la mateixa illa -La Hispaniola-; o bé el cas de les espècies simpàtriques marroquines *M. paurosexualis* i *M. spinicaudatus* d'una banda, i *M. goulmimensis* i *M. longicaudus* de l'altra.

Trontelj i col·laboradors (2009) enumeren diferents causes que explicarien aquesta sorprenent riquesa d'endemismes, entre les quals s'hi troben la baixa capacitat de dispersió de la fauna estigobiont (Coineau & Boutin, 1992) i la fragmentació de l'hàbitat per mor de processos hidrològics i/o geològics que afavoririen l'aïllament i la posterior divergència a llarg termini, donant lloc a diferents espècies al·lopàtriques

---

(Barr & Holsinger, 1985; Kane et al., 1994; Gibert & Deharveng, 2002).

### 2.1. Absència d'episodis dispersius

Si es té en compte la naturalesa eurihalina de *M. longipes* (= capacitat per viure a aigües de diferent salinitat) i l'absència, com ja s'ha esmentat anteriorment, de diferenciació morfològica entre les poblacions de Mallorca i Menorca, el més plausible seria pensar que els diferents llinatges d'aquesta espècie balear han colonitzat les aigües subterrànies dolces de l'interior de forma recurrent des del medi perifèric costaner anquialí. No obstant això, els resultats obtinguts en aquesta tesi indiquen que *M. longipes* presenta una estructura poblacional extremadament fragmentada i molt antiga, descartant-se que esdeveniments dispersius hagin estat els principals responsables dels patrons de distribució actual d'aquest llinatge. Solament s'ha detectat flux gènic recent entre poblacions no allunyades més de 10 km, tot i que normalment aquest s'ha limitat a ocórrer entre poblacions no separades més de 2 km. Un cop fetes aquestes observacions a nivell filogeogràfic a *M. longipes*, quant l'alta divergència entre poblacions i la reduïda capacitat per dispersar-se a grans distàncies (Coineau & Boutin, 1992; Verovnik et al., 2003), hom podria esperar trobar resultats similars quant a l'alta divergència a nivell intra- i interespecífic a la resta d'espècies de la família Metarangonyctidae, i a la mancança d'una relació estricta entre distància genètica i geogràfica. I, en efecte, així ho han provat els resultats obtinguts als capítols I i III. L'evidència més notòria d'aquesta manca de dispersió la trobem en les relacions filogenètiques dels diferents llinatges insulars. Tant les espècies mediterrànies com l'espècie endèmica canària són més properes genèticament entre elles que no pas a les espècies geogràficament més pròximes del nord d'Àfrica i la costa atlàntica marroquina, respectivament. Al seu torn, s'agrupen millor amb espècies que habiten a milers de quilòmetres de distància, a l'altra costat de l'Atlàntic. A més a més, s'ha observat que totes les espècies de metacranoníctids presenten una elevada divergència tant inter- com intraespecífica, àdhuc tractar-se en alguns casos d'espècies morfològicament molt semblants com és el cas de *M. goulmimensis*, "*M. nicoleae tamri*" i "*L. stocki*" del Marroc, les quals presenten un alta proporció de polimorfismes com ja s'ha comentat a l'apartat anterior.

### 2.2. Fragmentació hidrogràfica i geològica

---

Els resultats tant pel que fa a la filogeografia de *M. longipes* com a la filogènia de la família Metacrangonyctidae, assenyalen que hi ha un alt grau de fragmentació entre els diferents llinatges. Això queda reflectit a la profunda estructuració genètica poblacional a *M. longipes* i l'elevada divergència intra i interespecífica dels metacrangoníctids del Marroc. En el cas de *M. longipes*, a més de l'elevada estructura poblacional, la fragmentació de l'hàbitat s'ha fet palesa en el vincle existent entre els diferents clades i determinades conques hidrogràfiques, a excepció d'algunes de les poblacions del centre de l'illa de Mallorca que ocupen la capçalera de tres conques hidrogràfiques diferents. No obstant això, aquestes tres conques molt probablement haurien format part del mateix sistema de paleodrenatge durant les darreres glaciacions, quan el nivell de la mar havia minvat considerablement. Tot plegat, suggereix que l'estructura poblacional a *M. longipes* s'ha vist profundament afectada per les fragmentacions-divisions sofertes per aquestes conques, que actuarien com a "unitats territorials independents" dificultant extremadament el flux gènic entre aquestes.

La profunda divergència observada entres els llinatges de Metacrangonyctidae del Marroc també suggereix l'existència enllà d'una elevada fragmentació d'hàbitat. És el cas de les diferents poblacions de *M. goulmimensis*, tan divergents entre si, i separades tan sols per 38 i 59 km, o el de les de "*M. nicoleae tamri*", les quals disten 35 km i presenten també una elevada divergència entre elles. O bé, el cas de les espècies germanes *M. remyi* i *M. boveei*, les quals es troben localitzades a valls veïnades de la vessant nord de l'Alt Atlas Occidental prop de Marràqueix, distants sols 52 km, i que molt probablement haurien començat el seu procés d'especiació arrel de l'aïllament sofert per l'orogènesi de l'Alt Atlas. Sens dubte emperò, el patró de distribució i fragmentació més sorprenent és l'observat en el cas de les cinc espècies endèmiques insulars que s'haurien originat a partir d'un únic llinatge ancestral àmpliament distribuït per les riberes de la mar de Tetis. Qui hagués esperat *a priori* que l'espècie endèmica canària *M. repens* fos més propera genèticament a espècies que es troben avui en dia a milers de quilometres a l'altre costat de l'Atlàntic, i no pas a espècies veïnes habitants de la costa marroquina? Si més no, costa *a priori* justificar el patró observat a les espècies endèmiques de la Mediterrània, les quals no formen un clade *per se* ben suportat i semblen estar igualment emparentades amb la de Fuerteventura i amb les caribenyes encara més allunyades. I per què no són aquestes espècies mediterrànies més properes genèticament a les habitants del Nord d'Àfrica, que no pas a les habitants

---



de l'altre costat de l'Atlàntic? I el que és més important si es cau: com explicar aquest resultat tenint en compte la mancança de potencial dispersiu a grans distàncies d'aquest tipus de fauna? Tant és així, que la hipòtesi més lògica en resposta de totes aquestes observacions seria la d'atribuir a la deriva continental el patró de distribució extremadament disjunt mostrat pels integrants d'aquest taxó subterrani.

### *3. Colonització del medi subterrani: un clar exemple d'especiació al·lopàtrica*

Una de les comeses fonamentals que ens férem en iniciar aquest estudi fou la d'esbrinar com s'havia produït l'ocupació del medi subterrani continental per part d'aquests tàxons talassoids (= d'origen marí). Més concretament, si aquesta ocupació era el resultat de (1) episodis de colonització activa del medi subterrani dulciaquícola per part d'organismes epigeus marins, seguits d'una adaptació posterior a la vida subterrània i dispersió pel medi hipogeu; o (2) d'esdeveniments de colonització passiva pel varament de poblacions marines ancestrals a aqüífers costaners, associats a regressions marines pretèrites. La primera fase d'ambdues hipòtesis faria referència a avantpassats marins que haurien envaït primer hàbitats marins de transició, com ara esquerdes submarines, surgències càrstiques submarines mixohalines, o l'ambient intersticial dels sediments d'arena i llim costaner, tots ells sotmesos a fortes variacions de salinitat (és a dir, a períodes d'exposició a aigües dessalades), i que d'alguna manera imiten les condicions ambientals trobades al medi subterrani d'aigua dolça (Notenboom, 1991). Pel que fa a la primera hipòtesi, aquests organismes marins ja pre-adaptats a les condicions hipogees, haurien ampliat el seu nínxol colonitzant activament els aqüífers d'aigua dolça, dispersant-se a través d'ells, i enllestit posteriorment l'adaptació al medi hipogeu; aquest procés hauria estat independent de l'esdeveniment de canvis ambientals, com ara glaciacions o regressions marines (Danielopol, 1980; Rouch & Danielopol, 1987). Un exemple d'aquest patró d'especiació parapàtrica el podrien representar certs ostracodes estigobionts dulciaquícoles europeus, molt propers evolutivament a tàxons eurihalins marins (Danielopol & Bonaduce, 1990). No obstant això, són nombrosos els estudis faunístics i biogeogràfics que afavoreixen la segona hipòtesi com a més probable per explicar l'origen de molts crustacis estigobionts (Page et al., 2008; Juan et al., 2010). Aquesta assumeix un escenari vicariant pel qual la colonització es produeix d'una forma passiva a través del

---

varament de poblacions ancestrals ocorregut durant episodis de regressió marina. Aquestes regressions marines podrien ésser degudes a canvis climàtics (glaciacions), o a aixecaments tectònics de la línia de costa en diferents períodes geològics, i serien responsables de l'aïllament gradual a aigües continentals subterrànies de poblacions de tàxons marins ancestrals, el que provocaria la seva ulterior diversificació per especiació al·lopàtrica (Notenboom, 1991; Botosaneanu & Holsinger, 1991; Holsinger, 1994).

S'ha vist que la filogeografia i la genètica de poblacions poden ser útils per destriar entre les dues anteriors hipòtesis esmentades (Juan et al., 2010). Com ja s'ha dit anteriorment, vam triar d'antuvi l'espècie endèmica eurihalina estigobiont *M. longipes* com a cas d'estudi perquè presentava un elevat nombre de poblacions distintes a quasi tot el territori insular de Mallorca i Menorca. En aquest cas, podríem parlar de dues fases diferents de colonització: la que es referiria al pas de l'ambient marí epigeu a l'hipogeu i el que suposaria l'arribada a aigües dolces des d'aigües anquialines. Si l'adaptació al medi subterrani es produí un sol cop, hom esperaria trobar que les diferents poblacions de *M. longipes* fossin monofilètiques, i no pas que cada una d'elles fos resultat d'un episodi independent de colonització des de la mar.

Els resultats obtinguts (capítol I) ens han portat a dos discerniments a recalcar. D'una banda, que totes les poblacions balears comparteixen un origen comú, el que és compatible amb la hipòtesi que un mateix llinatge d'origen marí, pre-adaptat a condicions hipogeues i àmpliament distribuït per la mar epicontinental que rodejava i/o cobria parcialment les illes Gimnèsies, donà lloc a les actuals poblacions de *M. longipes* de Mallorca i Menorca. D'altra banda, la filogeografia d'aquest llinatge suggereix que l'ocupació de les aigües continentals dolces de Mallorca, a partir d'ancestres ja adaptats a la vida subterrània (en aquest cas, des del medi anquialí), es realitzà mitjançant almenys tres episodis independents de colonització passiva. L'escassa connectivitat observada entre aquestes poblacions, i l'absència total d'haplotips compartits entre les diferents poblacions dulciaquícules de l'interior de l'illa i les poblacions anquialines (no hi ha flux gènic ni tan sols entre aquestes darreres), fa descartar gairebé completament la hipòtesi d'una colonització activa dels hàbitats subterrànies d'aigua dolça a partir de l'expansió via medi crevicular/intersticial d'ancestres marins litorals.

Esbrinar quina de les dues hipòtesis és la que explicaria millor l'ocupació del medi

---

subterrani continental per part dels Metacrangonyctidae, testada al capítol I, ha estat també d'ajuda per comprendre la filogènia intragenèrica obtinguda al capítol III, ja que ha permès explicar la sorprenent troballa que les espècies habitants d'illes (clade insular) estiguin més emparentades entre elles i no pas amb les més pròximes geogràficament. Podria dir-se que ens trobaríem davant d'un cas d'especiació peripàtrica, quan poblacions marginals d'una mateixa espècie vastament distribuïda (en el nostre cas, per les riberes de la primigènia mar de Tetis), van especiar originant descendents sense un patró filogeogràfic clar, és a dir, estant menys relacionats entre ells que no pas amb l'espècie ancestral comú.

Nogensmenys, per poder testar totes aquestes hipòtesis sobre l'origen i evolució dels diferents llinatges de metacranoníctids, és condició *sine qua non* l'ús d'eines que permetin datar els esdeveniments cladogenètics i comprovar si són compatibles amb les datacions realitzades en base a esdeveniments geològics rellevants. De fet, cada vegada són més els estudis que gràcies a l'aplicació de rellotges moleculars han pogut constatar la importància dels esdeveniments vicariants per explicar els patrons de distribució observats a certs grups de taxons, com ara els gecònids australians (Oliver et al., 2010), els rosegadors múrids troglobis de l'Àsia Oriental (Latinne et al., 2012) o els onicòfors de Nova Zelanda (Allwood et al., 2010). Sens dubte emperò, uns dels exemples més sorprenents és els descrit per Chakrabarty i col·laboradors (2012), els quals reporten un patró ancestral vicariant per a un taxó de teleostis hipogeus, arran de la deriva continental que afectà durant el Cretàcic mitjà al supercontinent Gondwana. Aquests autors suggereixen que el llinatge ancestral d'aquest taxó (conformat per dos gèneres, un present a les aigües subterrànies de Madagascar i l'altre al Nord-oest d'Àustràlia), adaptat a hàbitats càrstics subterranis del sud de Gondwana i amb grans limitacions pera dispersar-se llargs distàncies, hauria quedat subdividit arran de la ruptura oriental d'aquest supercontinent durant el Cretàcic mitjà. Un altre exemple a destacar és el del crustaci branquiòpode *Cyclestheria hislopi*. Schewentner i col·laboradors (2012) assenyalen com el que en un principi es pensava era una sola espècie circumtropical, vastament distribuïda arreu de Sud-Amèrica, Àsia i Austràlia, és en realitat un conglomerat d'espècies críptiques, que quedaren aïllades durant el Cretàcic, quan Gondwana es fragmentà aïllant el llinatge Sud-american de la resta. Moviments tectònics posteriors ocorreguts durant el Eocè-Miocè amb la col·lisió de la Índia i Àsia podrien haver estat els responsables dels patrons biogeogràfics dels altres llinatges habitants d'Àustràlia i Àsia.

---

#### 4. *Datació dels esdeveniments cladogenètics i vicariants a Metacrangonyctidae: utilitat dels mitogenomes complets*

Un cop descartada la dispersió com a causa de la distribució disjunta dels llinatges actuals, i observat l'elevat grau de divergència tant a nivell intra- com interespecífic, les següents qüestions que sorgiren foren: són realment els diferents llinatges de Mecrangonyctidae resultat d'episodis cladogenètics tan antics? I si així és el cas, com esperar topar amb marcadors moleculars amb un senyal filogenètic suficient i no saturat, per poder aplicar tècniques de rellotge molecular per datar aquests esdeveniments?

Estudis teòrics i empírics anteriors han suggerit que incrementant el nombre de gens o posicions nucleotídiques en un set de dades es millora la resolució filogenètica i el suport dels nodes inferits (Cummings et al., 1995; Rokas & Carrol, 2005; Vilstrup et al., 2011). Si s'inclou més d'un gen, es va reduint notablement la influència de qualsevol variació quant al senyal filogenètic contingut als diferents gens a causa de la selecció o a efectes estocàstics (Rohland et al., 2007). Addicionalment, són nombrosos els estudis que proven els guanys en l'ús de mitogenomes sencers com a marcadors moleculars per a estudis d'inferència filogenètica a diferents nivells taxonòmics (Ingman et al., 2000; Nardi et al., 2003; Carapelli et al., 2007; Meng et al., 2008; Miller et al., 2009; Singh et al., 2009; Wei et al., 2010; Duchene et al., 2011). Per aquests motius, i atès la gran divergència que esperàvem trobar dins Metacrangonyctidae (reflectida ja en l'anàlisi preliminar realitzat amb els marcadors *cox1* i *rrnL*), ens decantarem per l'ús de mitogenomes sencers, aportant d'aquesta manera el primer estudi filogenètic i biogeogràfic basat en mitogenomes emprés fins ara a Crustacea. Els resultats d'aquesta investigació han estat prou engrescadors ja que després d'una exploració exhaustiva de les dades obtingudes, s'ha pogut extreure prou senyal filogenètic dels MPGCs per elucidar la filogènia i datar els principals esdeveniments cladogenètics ocorreguts dins d'aquesta família.

Els resultats obtinguts en aquesta tesi apunten a que els principals esdeveniments cladogenètics dins del llinatge de *M. longipes* i de la família Metacrangonyctidae, amb molta probabilitat semblen respondre a la successió de moviments eustàtics en ambdós casos, i a processos associats a la deriva continental en el segon cas (tenint en compte

---

les limitacions i l'error associat a les estimacions del rellotge molecular i la incertesa de les datacions d'esdeveniments geològics). Esdeveniments vicariants haurien estat els principals responsables dels patrons i processos evolutius soferts pels metacranoníctids, i no pas els processos dispersius com s'ha observat a altres estudis de la fauna subterrània (veure Juan et al., 2010). Més concretament, els diferents cicles de transgressió-regressió marina que es succeïren entre 11.3 Ma (transgressió del Tortonà) i 5.7 Ma (Crisi de Salinitat del Messinià amb el posterior re-ompliment de la conca de la Mediterrània Occidental), haurien estat responsables dels principals esdeveniments cladogenètics a *M. longipes*. Les fragmentacions posteriors del llinatge, datades entre fa 0.5 i 2 Ma, les quals diferenciaren encara més les poblacions, podrien haver estat influïdes per diferents episodis d'eustatisme associats a cicles glaciar-interglaciar durant el Pliocè superior fins al Plistocè mitjà. Traslladant aquest patró evolutiu a un àmbit més global, hem observat que esdeveniments vicariants encara més dràstics i extrems, com és el cas de la deriva continental al Cretaci Mitjà o el major cicle de transgressió-regressió transcorregut durant el Cretaci Superior, haurien estat els principals responsables de la història biogeogràfica de la família Metacranonyctidae. L'edat estimada a la nostra investigació per a la divergència dels clades principals d'aquesta família és 96-80 Ma, i de 79-77 Ma pels llinatges insulars (Cretaci Mitjà-Superior). Al voltant d'aquesta època, els continents africà i sud-americà es trobaven molt més propers que no actualment i tot just començava a fragmentar-se per l'establiment d'aigües profundes entre mig la via marítima que havia mantingut connectades les aigües someres de la regió del Carib, l'Atlàntic Oriental i la part de la mar de Tetis localitzada a l'actual Mediterrània Occidental. Tant és així, que el llinatge primigeni de metacranoníctids, vastament distribuït arreu d'aquesta mar epicontinental s'anà fragmentant en distints llinatges, perdent el contacte per sempre més entre ells i donant lloc finalment a les espècies actuals per un procés d'especiació peripàtrica.

Arran de l'anàlisi dels MPCGs als metacranoníctids, s'ha estimat una taxa de substitució mitjana a llarg termini de 10.9% per milió d'anys i per parell de seqüències, mentre que per esdeveniments divergents més recents, com els reportats per la filogeografia de *M. longipes*, aquesta taxa seria més propera al 2.3%. No obstant això, a l'estudi anterior sobre les diferents poblacions de *M. longipes*, no s'han tingut en compte tots els MPCGs, així com sí s'ha fet per l'estudi global de la família Metacranonyctidae, alhora que els models de diversificació emprats en ambdós casos

---

han estat diferents (un model de coalescència per *M. longipes*, i un model d'especiació de Yule per a la família Metacrangonyctidae).

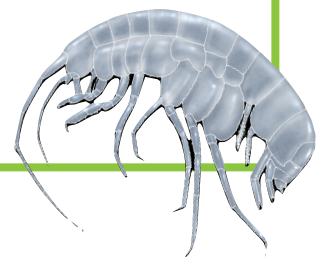
### 5. Radiacions sincròniques

Un cop analitzades les relacions filogenètiques a la família Metacrangonyctidae i la filogeografia de *M. longipes*, s'ha observat que certs nodes dels arbres presenten uns valors de suport estadístic inferiors als que es consideren òptims per tenir un node satisfactòriament resolt. En aquest estudi emperò, hem demostrat que no tots els casos de politomies són resultat d'inferències incongruents o d'artefactes metodològics. Concretament, en el cas del clade insular que agrupava tots els llinatges insulars de *Metacrangonyx*, el patró de diversificació correspon a una politomia veritable; i el mateix cal dir dels principals nodes de *M. longipes*. Aquesta dificultat per resoldre les relacions filogenètiques a amfípodes subterranis ja s'havia trobat anteriorment en altres estudis, i s'ha vinculat al desencadenament de radiacions (diversificacions) sobtades associades amb episodis de colonització sincrònica de les aigües subterrànies per part de diferents poblacions d'un mateix ancestre (Notenboom, 1991; Lefébure et al., 2007).

Així doncs, seria bo tenir en compte aquestes observacions en estudis futurs sobre tàxons que hagin pogut originar-se per especiació peripàtrica, com en el cas dels membres de la família Metacrangonyctidae. Aquestes politomies explicarien els patrons de diversificació comuns a diversos llinatges de crustacis subterranis talassoids (gambes, termosbenacis, copèpodes misofrioids, etc.), que podrien haver estat afectats pels mateixos esdeveniment vicariants, o bé ser producte d'episodis d'aïllament i especiació tan propers en el temps que no permetrien la fixació de substitucions nucleotídiques informatives suficients, el que es tradueix en topologies amb branques internes extremadament curtes que dificulten molt les estimacions dels temps de coalescència. És el cas d'alguns llinatges de lepidòpters (Kodandaramaiah et al., 2010; Rota, 2011) i de rèptils (Stanley et al., 2011; Townsend et al., 2011), als quals s'ha comprovat que el que eren en un principi relacions filogenètiques no resoltes eren en realitat clades politòmics veritables producte de radiació explosiva.



**CONCLUSIONS GENERALS**







1. Les poblacions de l'espècie endèmica balear *M. longipes* mostren una marcada estructuració geogràfica, limitades per conques hidrogràfiques i una elevada divergència genètica, malgrat trobar-se geogràficament molt properes. Aquest patró implica una mancança de flux genètic entre les poblacions i les distintes conques durant un prolongat lapse de temps.
  2. La reconstrucció d'hàbitats ancestrals indica que la colonització del medi dulciaquícola subterrani ocorregué en tres episodis independents a diferents parts de Mallorca a partir de poblacions habitants del medi subterrani anquialí, i no viceversa.
  3. Els esdeveniments cladogenètics més rellevants a la història evolutiva del llinatge coincideixen amb canvis eustàtics que haurien induït a l'expansió, contracció i finalment a la fragmentació de *M. longipes*. Concretament, la separació del llinatge ancestral de *M. longipes* s'estima que va ocórrer al voltant de fa 5.4-6.2 Ma, tot just coincidint amb la dràstica regressió marina de la Crisi de Salinitat Messiniana de la Mediterrània. El altres dos processos cladogenètics principals coincideixen amb el cicle de transgressions-regressions marines del Pliocè Mitjà-Superior, datats entre 4.2 i 2.7 Ma, i els cicles de glaciació i desglaciació del Plistocè que tingueren lloc fa uns 2.0 i 0.5 Ma.
  4. En aquesta investigació s'ha presentat la primera seqüència completa d'un genoma mitocondrial d'un amfípode, el de *M. longipes*, i el segon per als Peracarida, superordre de Crustacea Malacostraca, aquest últim escassament representat a la base de dades de genomes mitocondrials. L'obtenció d'aquest primer mitogenoma va ser completada amb 20 mitogenomes més de la família Metacrangonyctidae i dos més de famílies properes.
  5. L'ordre gènic es troba conservat dins la família Metacrangonyctidae tot i que difereix de la resta de peracàrides i del patró ancestral proposat per a Pancrustacea. La presència d'una transposició comú al gen de *trnG* als metacrangoníctids i a la resta d'amfípodes anotats posteriorment, assenyalaria aquest canvi com a una apomorfia de l'ordre Amphipoda. Altres característiques d'aquests mitogenomes són el seu alt contingut en AT, la seva
-

extremada compactació, i que algunes espècies tenen la regió control més curta de la que es tingui referència dins els Crustacea, amb una longitud de només 38-40 pb.

6. La filogènia dels metacrangoníctids obtinguda agrupa les espècies analitzades en cinc llinatges evolutius independents: el clade basal de l'Alt Atlas Occidental, els dos clades parafilètics del Marroc Atlàntic, i els dos subclades que inclouen per una banda el llinatge de l'interior del Marroc, i per l'altra el llinatge insular, aquest darrer de distribució amfi-atlànica extremadament disjunta.
  7. En aquesta investigació s'han testat dues hipòtesis alternatives (vicariança vs. dispersió) per explicar la distribució global observada dels diferents llinatges metacrangoníctids. Tant la baixa correlació entre distància geogràfica i genètica com així les edats obtingudes arran d'assumir un rellotge molecular relaxat, assenyalen la vicariança com a principal factor modulador dels patrons de distribució exhibits pels membres aquesta família. Aquestes anàlisis apunten que l'origen dels principals clades metacrangoníctids tingué lloc al Cretaci Mitjà, fa uns 96-83 Ma, quan la deriva continental i els cicles de transgressions-regressions marines datats en aquella època, portaren a la diversificació inicial d'aquesta família.
  8. Les politomies obtingudes per diferents nodes i amb diferents metodologies tant a la filogènia del llinatge de *M. longipes* com a la de la família Metacrangonyctidae, sobretot al clade insular, corroborarien que la colonització del medi insular ha succeït de forma simultània o en un espai de temps relativament curt. Aquest patró filogenètic és característic dels processos d'especiació al·lopàtrica, i més concretament, d'especiació peripàtrica.
  9. La família Metacrangonyctidae ha proporcionat un marc idoni per a la recerca del procés d'especiació al medi hipogeu atès les limitades oportunitats de dispersió dels seus membres entre els diferents sistemes subterranis, l'estabilitat ambiental, l'alt grau d'aïllament característic d'aquests ambients i la seva distribució extremadament disjunta.
-

## General Conclusions

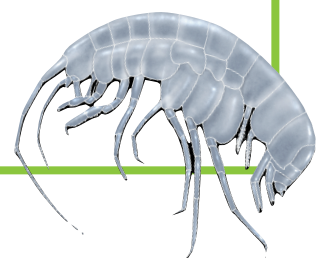
1. The Balearic Islands *Metacrangonyx longipes* populations display an ancient fragmented structure. They show, even at geographically close localities, high genetic divergences and a structuring congruent with hydrographic catchments. This pattern is compatible with a limited gene flow between populations/catchments and with a long-term populations isolation.
  2. The reconstruction of habitat character state in the *M. longipes* phylogeny points out that colonization of fresh inland groundwaters has occurred by at least three independent episodes from primary anchialine, brackish water ancestors at different Mallorcan localities.
  3. *M. longipes* major cladogenetic events deduced are roughly coincident with a succession of eustatic changes that probably acted as vicariant events. These changes may have induced recurrent range expansions, contractions and the fragmentation of *M. longipes* ancestral populations. The estimated separation of the two major lineages at 5.4-6.2 million years ago (Mya) may be related to the "Messinian Salinity Crisis", a mega-marine regression that almost dried up the Mediterranean basin. The other two main cladogenetic events in the obtained phylogeny might be linked to the upper-middle Pliocene transgression-regression cycle dated back to 4.2-2.7 Mya and to the subsequent glacial-interglacial cycles during the Pleistocene (2.0-0.5 Mya).
  4. We have obtained and analysed in the present study the first complete nucleotide mitochondrial sequence of an amphipod species (*M. longipes*). It represents the second mitochondrial genome reported for Peracarida, a superorder of the crustacean order Malacostraca, which is under-represented in genome datasets. In addition, 20 extra mitogenomes of the family Metacrangonyctidae and two from species of other related amphipod families, have been obtained within the frame of the present PhD thesis.
  5. We have found a conserved mitochondrial gene order within Metacrangonyctidae but differing from the one of the other peracarids analysed so far and from the putative pancrustacean ancestral pattern. A transposition of
-

the *trnG* gene, present in all the Metacrangonyctidae, is shared with the amphipod mitogenomes that have been annotated so far. This indicates that this change can be an apomorphy of Amphipoda. All the mitogenomes analysed in the present study have a high AT content and are very compact. We have identified the shortest control region known in Crustacea, with a length of 38-40pb.

6. The Metacrangonyctidae phylogeny based on mitochondrial protein coding genes and nuclear RNA shows that species are grouped into five independent evolutionary lineages: a basal clade with species from the Western High Atlas in Morocco, two paraphyletic clades that are distributed over Atlantic Morocco, and two additional subclades, one present at both sides of the High Atlas in inland Morocco and the second made of several insular species having an extremely disjunct amphi-atlantic distribution.
  7. We have examined alternative hypotheses of vicariance and dispersal as explanations of the global disjunct metacrangonyctid distribution. We favour vicariance by plate tectonics based on a) the low correlation between geographic and genetic distances obtained in the dataset, and b) molecular dating of the Metacrangonyctidae phylogeny using relaxed molecular clock methods. This analysis assigns a mid-Cretaceous origin to the main clades of the family at 83-96 Mya, an estimate compatible with plate tectonics leading to the initial diversification of the family.
  8. Evidence of true politomies found both within the *M. longipes* and the Metacrangonyctidae phylogenies, in particular for the insular clade, point out to that colonization of island environments took place almost synchronously or over a short period of time. This phylogenetic pattern is compatible with an allopatric or, in some cases, peripatric speciation process.
  9. The extremely disjunct geographic distribution of metacrangonyctids, their limited dispersal capabilities, in conjunction with the isolation and environmental stability of the habitat they occupy make the family a valuable and excellent study-group to understand historic evolutionary processes in subterranean fauna.
-

---

## **BIBLIOGRAFIA**





- Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104-2105.
- Abascal F, Zardoya R, Telford M J (2010) TranslatorX: multiple alignments of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**: W7-W13.
- Abdo Z, Minin V N, Joyce P, Sullivan J (2005) Accounting for uncertainty in the tree topology has little effect on the decision theoretic approach to model selection in phylogeny estimation. *Molecular Biology and Evolution* **22**: 691-703.
- Acosta J, Canals M, López-Martínez J, Muñoz A, Herranz P, Urgeles R, Palomo C, Casamor J L (2002) The Balearic Promontory geomorphology (Western Mediterranean): morphostructure and active processes. *Geomorphology* **49**: 177-204.
- Acosta J, Canals M, Carbó A, Muñoz A, Urgeles R, Muñoz-Martín A, Uchupi E (2004) Sea floor morphology and Plio-Quaternary sedimentary cover of the Mallorca Channel, Balearic Islands, western Mediterranean. *Marine Geology* **206**: 165-179.
- Aitkin M (1991) Posterior Bayes factors. *Journal of the Royal Statistical Society B* **53**: 111-142.
- Akaike H (1973) Information theory as an extension of the maximum likelihood principle. Dins Petrov B N, Csaki F (eds.), *Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.
- Albrecht-Buehler G (2006) Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17828-17833.
- Aldous D J (2001) Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statistical Science* **16**: 23-34.
- Altschul S F, Madden T L, Schaffer A A, Zhang J H, Zhang Z, Miller W and Lipman, D J (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.
- Alvarez W (1972) Rotation of the Corsica-Sardinia microplate. *Nature* **235**: 103-105.
- Allwood J, Gleeson D, Mayer G, Daniels S, Beggs J R, Buckley T R (2010) Support for vicariant origins of the New Zealand Onychophora. *Journal of Biogeography* **37**: 669-681.
- Andersen N M, Cheng L, Damgaard J, Sperling F A H (2000) Mitochondrial DNA sequence variation and phylogeography of oceanic insects (Hemiptera: Gerridae: *Halobates* spp.). *Marine Biology* **136**: 421-430.
- Andrew V, Brower Z (2006) The how and why of branch support and partitioned branch support, with a new index to assess partition incongruence. *Cladistics* **22**: 378-386.
- Antonelli A, Sanmartín I (2011) Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches *Systematic Biology* **60**: 596-615.
- Arbogast B S, Edwards S V, Wakeley J, Beerli P, Slowinski J B (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology & Systematics* **33**: 707-740.
- Aris-Brosou S, Yang Z (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Systematic Biology* **51**: 703-714.
- Arnedo M, Oromí P, Murriá C, Macías-Hernández N, Ribera C (2007) The dark side of an island radiation: systematics and evolution of troglobitic spiders of the genus *Dysdera* Latreille (Araneae:
-



- Dysderidae) in the Canary Islands. *Invertebrate Systematics* **21**: 623–660.
- Avise J C, Arnold J, Ball R M, Bermingham E, Lamb T, Neigel J E, Reeb C A, Saunders N C (1987) Intra-specific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology & Systematics* **18**: 489–522.
- Avise J C (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**: 371–379.
- Avise J C, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London, Series B* **265**: 457–463.
- Avise J C (2000) *Phylogeography: The History and Formation of Species*. 447 pp. Boston: Harvard University Press.
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard M A, Alekseyenko der V (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology & Evolution* **29**: 2157–2167.
- Baker R H, Yu X, DeSalle R (1998) Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Molecular Phylogenetics & Evolution* **9**: 427–436.
- Balakrishnan R (2005) Species concepts, species boundaries and species identification: A view from the tropics. *Systematic Biology* **54**: 689–693.
- Balazuc J, Ruffo S (1953) Due nuove species del genere *Metacrangonyx* Chevreux delle acque interne del Nord Africa Francese. *Museo Civico di Storia Naturale di Verona* **4**: 25–33.
- Banda E, Santanach P (1992) The Valencia trough (western Mediterranean): an overview. *Tectonophysics* **208**: 183–202.
- Baratti M, Filippelli M, Nardi F, Messana G (2010) Molecular phylogenetic relationships among some stygobitic cirolanid species (Crustacea, Isopoda). *Contributions to Zoology* **79**: 57–67.
- Barr T C Jr (1967) Ecological studies in the Mammoth Cave system of Kentucky. I The biota. *International Journal of Speleology* **3**: 47–204.
- Barr T C Jr (1968) Cave ecology and evolution of troglobites. *Evolutionary Biology* **2**: 35–102.
- Barr T C Jr, Holsinger J R (1985) Speciation in cave faunas. *Annual Review of Ecology & Systematics* **16**: 313–337.
- Barracough T G, Hogan J E, Vogler A P (1999) Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae) *Proceeding of the Royal Society of London, Series B* **266**: 1061–1067.
- Bayes T (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**: 370–418.
- Bell C D, Soltis D E, Soltis P S (2010) The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**: 1296–1303.
- Benton M J, Ayala F J (2003) Dating the tree of life. *Science* **300**: 1698–1700.
- Bérard S, Bergeron A, Chauve C, Paul C (2007) Perfect sorting by reversals is not always difficult. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **4**: 4–16.
- Bermingham E, Rohwer S, Freeman S, Wood C (1992) Vicariance biogeography in the Pleistocene and speciation in North American wood warblers, a test of Mengel's model. *Proceedings of the National Academy of Sciences (USA)* **89**: 6624–6628.
- Bernt M, Merkle D, Ramsch K, Fritsch G, Perseke M, Bernhard D, Schlegel M, Stadler P F, Middendorf M (2007) CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics* **23**: 2957–

- 2958.
- Blake W J, KÄern M, Cantor C R, Collins J J (2003) Noise in eukaryotic gene expression. *Nature* **422**: 633–637.
- Bonnet E, Van de Peer Y (2002) ZT: a software tool for simple and partial Mantel tests. *Journal of Statistical Software* **7**: 1-12.
- Boore J L, Collins T M, Stanton D, Daehler L L, Brown W M (1995) Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**: 163–165.
- Boore J L, Lavrov D, Brown W M (1998) Gene translocation links insects and crustaceans. *Nature* **392**: 667–668.
- Boore J L (1999) Animal mitochondrial genomes. *Nucleic Acids Research* **27**: 1767-1780.
- Boore J L, Macey J R, Medina M (2005) Sequencing and comparing whole mitochondrial genomes of animals. *Methods in Enzymology* **395**: 311-348.
- Börner G V, Yokobori S, Mörl M, Dörner M, Pääbo S (1997) RNA editing in metazoan mitochondria: staying fit without sex. *FEBS Letters* **409**: 320–324.
- Bortolotti V, Principi G (2005) Tethyan ophiolites and Pangea break-up. *The Island Arc* **14**: 442-470.
- Botello A, Iliffe T M, Alvarez F, Juan C, Pons J, Jaume D (2012) Historical biogeography and phylogeny of *Typhlatya* cave shrimps (Decapoda: Atyidae) based on mitochondrial and nuclear data. *Journal of Biogeography* doi:10.1111/jbi.12020.
- Botosaneanu L (1986) A faunistic, distributional and ecological synthesis of the world fauna inhabiting subterranean waters (including the marine interstitial). Dins *Stygofauna Mundi* (pp. 47-56). Leiden: EJ Brill/Dr W Backhuys.
- Botosaneanu L, Holsinger J R (1991) Some aspects concerning colonization of the subterranean realm, especially of subterranean waters: a response to Rouch & Danielopol, 1987. *Stylogologia* **6**: 11-39.
- Bouckaert R R (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**: 1372-1373.
- Boutin C, Messouli M (1988a) *Longipodacrangonyx maroccanus* n. gen. n. sp., nouveau représentant du groupe *Metacrangonyx* dans les eaux souterraines du Maroc. *Crustaceana* **13**: 256-271.
- Boutin C, Messouli M (1988b) *Metacrangonyx gineti* n. sp. d'une source du Haut- Atlas marocain et la famille des *Metacrangonyctidae* n. fam. (Crustacés Amphipodes stygobies). *Vie Milieu* **38**: 67-84.
- Boutin C, Coineau N (1990) Regression Model, Modél biphasé d'évolution et origine des microorganismes stygobies interstitiels continentaux. *Revue de Micropaléontologie* **33**: 303-322.
- Boxshall G A (1989) Colonization of inland marine caves by misophrioid copepods. *Journal of Zoology* **219**: 521-526.
- Boutin C (1994) Phylogeny and biogeography of metacrangonyctid amphipods in North Africa. *Hydrobiologia* **287**: 49-64.
- Braby M F, Trueman J W H, Eastwood R (2005) When and where did troidine butterflies (Lepidoptera: Papilionidae) evolve? Phylogenetic and biogeographic evidence suggests an origin in remnant Gondwana in the Late Cretaceous. *Invertebrate Systematics* **19**: 113–143.
- Bremer K (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**: 795-803.
- Britten R J (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**: 1393-1398.
- Bromham L, Penny D (2003) The modern molecular clock. *Nature Reviews Genetics* **4**: 216-224.
- Brower A V Z (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National*
-

- Academy of Sciences (USA)* **91**: 6491-6495.
- Brown J W, Rest J S, Garcia-Moreno J, Sorenson M D, Mindell D P (2008) Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biology* **6**: 6.
- Brown T A, Cecconi C, Tkachuk A, Bustamante C, Clayton D A (2005) Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. *Genes & Development* **19**: 2466-2476.
- Brown W M, George M Jr, Wilson A C (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences (USA)* **76**: 1967-1971.
- Brown W M, Prager E M, Wang A, Wilson A C (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *Journal of Molecular Evolution* **18**: 225-239.
- Brown W M (1983) Evolution of animal mitochondrial DNA. Dins Nei M, Koehn R K (eds.), *Evolution of genes and proteins* (pp. 62-88) Sunderland, Mass: Sinauer.
- Buckley T R, Cunningham C W (2002) The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. *Molecular Biology and Evolution* **19**: 394-405.
- Buhay J E, Crandall K A (2005) Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Molecular Ecology* **14**: 4259-4273.
- Burnham K P, Anderson D R (2001) Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research* **28**: 111-119.
- Burnham K P, Anderson D A (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (pp. 488) 2a edició. New York: Springer-Verlag.
- Burnham K P, Anderson D R (2004) Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research* **33**: 261-304.
- Caccone A, Sbordoni V (2001) Molecular biogeography of cave life: a study using mitochondrial DNA from bathysciine beetles. *Evolution* **55**: 122-130.
- Camacho A I, Dorda B A, Rey I (2011) Identifying cryptic speciation across groundwater populations: first COI sequences of Bathynellidae (Crustacea, Syncarida). *Graellsia* **67**: 7-12.
- Cameron S L, Miller K B, D'Haese C A, Whiting M F, Barker S C (2004) Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). *Cladistics* **20**: 534-557.
- Cameron S L, Lambkin C L, Barker S C, Whiting M F (2007) A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology* **32**: 40-59.
- Cameron S L, Whiting M F (2007) Mitochondrial genomic comparisons of the subterranean termites from the genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). *Genome* **50**: 188 - 202.
- Carapelli A, Liò P, Nardi F, van der Wath E, Frati F (2007) Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal parafly of Hexapoda and Crustacea. *BMC Evolutionary Biology* **7**: S8.
- Carapelli A, Comandi S, Convey P, Nardi F, Frati F (2008) The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda, Collembola). *BMC Genomics* **9**: 315.
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology & Evolution* **17**: 540-552.
- Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: A thriving
-

- tower of Babel. *Annual Review of Entomology* **45**: 1–54.
- Cavalli-Sforza L L, Edwards A W F (1964) Phylogenetic analysis: models and estimation procedures. *Evolution* **21**: 550-570.
- Clayton D A (1982) Replication of animal mitochondrial DNA. *Cell* **28**: 693-705.
- Clayton D A (1996) Mitochondrial DNA replication. Dins DePamphilis M L (ed.), *DNA Replication in Eukaryotic Cells* (pp. 1015–1027). Nova York: Cold Spring Harbor Laboratory Press.
- Coello J, Cantagrel J M, Hernan F, Fuster J M, Ibarrola E, Ancochea E, Casquet C, Jamond C, Diaz De Teran J R, Cendrero A (1992) Evolution of the eastern volcanic ridge of the Canary Islands based on new K-Ar data. *Journal of Volcanology and Geothermal Research* **53**: 251-274.
- Coineau N (1984) Quelques aspects de l'adaptation des Crustacés Isopodes, Amphipodes et Syncarides à la vie dans le milieu interstitiel. *Mémoires de Biospéologie* **11**: 7-15.
- Coineau N, Boutin C (1992) Biological processes in space and time: colonization, evolution and speciation in interstitial stygobionts. Dins Camacho A I (ed.), *The Natural History of Biospeleology Museo Nacional de Ciencias Naturales & CSIC, Monographias 7*, Madrid 1992:423-452.
- Coineau N (1994) Evolutionary biogeography of the microparasellid isopod *Microcharon* (Crustacea) in the Mediterranean Basin. *Hydrobiologia* **287**: 77–93.
- Colgan D J, McLauchlan A, Wislon G DF, Livingston S, Macaranas J, Edgecombe G D, Cassis G, Gray M R (1998) Molecular phylogenetics of the Arthropoda: relationships based on histone H3 and U2 snRNA DNA sequences. *Australian Journal of Zoology* **46**: 419–437.
- Cook C E, Yue Q, Akam M (2005) Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proceedings of the Royal Society of London, Series B* **272**: 1295–1304.
- Cooper S J B, Bradbury J H, Saint K M, Leys R, Austin A D *et al.* (2007) Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia. *Molecular Ecology* **16**: 1533-1544.
- Cox C B, Moore P D (2005) *Biogeography: An Ecological and Evolutionary Approach* (7a edició). Blackwell.
- Craft J D, Russ A D, Yamamoto M N, Iwai T Y, Jr Hau S, Kahiapo J, Chong C T, Ziegler-Chong S, Muir C, Fujita Y, Polhemus D A, Kinzie III R A, Santos S R (2008) Islands under islands: The phylogeography and evolution of *Halocaridina rubra* Holthuis, 1963 (Crustacean: Decapoda: Atyidae) in the Hawaiian archipelago. *Limnology & Oceanography* **53**: 675–689.
- Craw R C, Grehan J R, Heads M J (1999) *Panbiogeography: tracking the history of life*. Editorial Oxford University Press.
- Crease T (1999) The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* **233**: 89-99.
- Crisci J V, Katinas L, Posadas P (2000) *Introducción a la Teoría y Práctica de la Biogeografía Histórica*. Bons Aires: Sociedad Argentina de Botánica.
- Crisci J V (2001) The voice of historical biogeography. *Journal of Biogeography* **28**: 157-168. Crisci J V, Katinas L, Posadas P (2003) *Historical Biogeography: An Introduction*. Harvard University Press.
- Crisci J V, Sala O E, Katinas L, Posadas P (2006) Bridging historical and ecological approaches in biogeography. *Australian Systematic Botany* **19**: 1–10.
- Crisp M D, Trewick S A, Cook L G (2011) Hypothesis testing in biogeography. *Trends in Ecology and Evolution* **26**: 66-72.
- Croizat L (1952) *Manual of Phytogeography*. Junk, The Hague.
- Croizat L (1958) Panbiogeography: or, an introductory synthesis of zoogeography, phytogeography, and
-

- geology: with notes on evolution, systematics, ecology, anthropology, etc. Léon Croizat, Caracas, Venezuela.
- Croizat L (1964) *Space, Time, Form: The Biological Synthesis*. Léon Croizat, Caracas, Venezuela.
- Croizat L, Nelson G, Rosen D E (1974) Centres of origin and related concepts. *Systematic Zoology* **23**: 265–287.
- Cruickshank R (2002) Molecular markers for the phylogenetics of mites and ticks. *Systematic & Applied Acarology* **7**: 3-14
- Cuerda J, Sacares J, Colom G (1969) Hallazgos de terrazas pliocénicas, marinas en la región de Lluçmajor. *Acta Geológica Hispánica* **4**: 35-37.
- Culver D C (1970) Analysis of simple cave communities: niche separation and species parking. *Ecology* **51**: 949-958.
- Culver, D (1982) *Cave Life. Evolution and Ecology*. Harvard University Press, Cambridge, MA.
- Culver D C, Deharveng L, Bedos A, Lewis J J, Madden M, Reddell J R, Sket B, Trontelj P, White D (2006) The mid-latitude biodiversity ridge in terrestrial cave fauna. *Ecography* **29**: 120–128.
- Culver D C, Pipan T, Schneider K (2007) Vicariance, dispersal and scale in the aquatic subterranean fauna of karst regions. *Freshwater Biology* **54**: 918–929.
- Culver D C, Pipan T (2009) *Biology of Caves and Other Subterranean Habitats*. Regne Unit, Oxford: Oxford University Press.
- Cummings M P, Otto S P, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology & Evolution* **12**: 814-822.
- Cummings M P, Handley S A, Myers D S, Reed D L, Rokas A, Winka K (2003) Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology* **52**: 477–487.
- Cvetkov L (1968) Un filet phréatobiologique. *Bulletin de l'Institut de Zoologie et Musée*, **22**: 215-219.
- Chakrabarty P, Davis M P, Sparks J S (2012) The first record of a trans-oceanic sister-group relationship between obligate vertebrate troglobites. *PLoS One* **7**: e44083.
- Chaudhuri K, Chen K, Mihaescu, R, Rao S (2006) On the tandem duplication-random loss model of genome rearrangement. *SODA* 564-570.
- Chevreaux E (1909) Amphipodes (première série). *Archives de Zoologie Experimentale et Generale* **5**: 27-42.
- Choe C P, Hancock J M, Hwang U W, Kim W (1999) Analysis of the primary sequence and secondary structure of the unusually long SSU rRNA of the soil bug, *Armadillidium vulgare*. *Journal of Molecular Evolution* **49**: 798-805.
- Choubert G, Faure-Muret A (1962) Evolution du domaine atlasique marocain depuis les temps paléozoïques. Dans *Livre a la Mémoire du Professeur Paul Fallot* (pp. 447 - 527). Paris: Bulletin de la Société Géologique de France.
- Christiansen K, Culver D C (1987) Biogeography and the distribution of cave Collembola. *Journal of Biogeography* **14**: 459-477.
- Dalens H, Deharveng L, Bedos A, Souqual M C (2002) Les isopodes terrestres endémiques de l'Europe de l'Ouest. *Endemica Ecocart CD-ROM*. Forthcoming.
- Danielopol D L (1980) An essay to assess the age of the freshwater interstitial ostracods of Europe. *Bijdragen tot de Dierkunde* **50**: 243-291.
- Danielopol D L, Bonaduce G (1990) The origin and distribution of the interstitial Ostracoda of the species group *Xestoleberis arcturi* Triebel (Crustacea). *Courier Forschungsinstitut Senckenberg* **123**: 69-86.
- Danielopol D L, Marmonier P, Boulton A J, Bonaduce G (1994) World subterranean ostracod
-

- biogeography: dispersal or vicariance. *Hydrobiologia* **287**: 119–129.
- Danielopol D L (2000) Biodiversity in groundwater: a large-scale view. *TREE* **15**: 223–224.
- Dañoibeitia J J, Arguedas M, Gallart J, Banda E, Makris J (1992) Deep crustal configuration of the Valencia trough and its Iberian and Balearic borders from extensive refraction and wide-angle reflection seismic profiling. *Tectonophysics* **302**: 37–55.
- Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle of Life*. John Murray, London.
- de Bruyn M, Nugroho E, Hossain M M, Wilson J C, Mather P B (2005) Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seaway. *Heredity* **94**: 370–378.
- De Candolle A P (1820) Géographie Botanique. Dins Cuvier F (ed.), *Dictionnaire des Sciences Naturelles, tom XVIII* (pp. 359–436). Paris: Levrault .
- Desalle R, Egan M G, Siddall M (2005) The unholy trinity: taxonomy, species delimitation, and DNA barcoding. *Philosophical Transactions of the Royal Society of London, Series B* **360**: 1905–1916.
- Dickerson R E (1971) The structure of cytochrome c and the rates of molecular evolution. *Journal of Molecular Evolution* **1**: 26–45.
- Dillon W P, Odale R N (1968) Late Quaternary sea-level curve: reinterpretation based on glaciotectionic influence. *Geology* **6**: 56–60.
- Dirheimer G, Keith G, Dumas P, Westhof E (1995) Primary, secondary, and tertiary structures of tRNAs. Dins Soll D, Rajbhandary U (eds.) *tRNA: structure, biosynthesis, and function* (pp. 93–126). Washington DC: American Society for Microbiology Press.
- Do C B, Woods D A, Batzoglou S (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **22**: e90–e98.
- Doolittle R F, Feng D-F, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. *Nature* **271**: 470–477.
- Drummond A J, Ho S Y W, Phillips M J, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: 699–710.
- Drummond A J, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.
- Duchêne S, Archer F I, Vilstrup J, Caballero S, Morin P A (2011) Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One* **6**: e27138.
- Duggen S, Hoernle K, van den Bogaard P, Rupke L, Morgan J P (2003) Deep roots of the Messinian Salinity Crisis. *Nature* **422**: 602–606.
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**: 2571–2581.
- Eddy S R, Durbin R (1994) RNA Sequence Analysis Using Covariance Models. *Nucleic Acids Research* **22**: 2079–2088.
- Edgar R C (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Edgar R C, Batzoglou S (2006) Multiple sequence alignment. *Current Opinion in Structural Biology* **16**: 368–373.
- Emerson B C (2007) Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Systematic Biology* **56**: 337–345.
-



- Excoffier K, Smouse P E, Quattro J M (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Journal of Evolutionary Bioinformatics Online* **1**: 47-50.
- Fairbridge R W (1961) Eustatic changes in sea level. *Physics & Chemistry of the Earth* **4**: 99-185.
- Fairbridge R W (1962) World sea-level and climatic changes. *Quaternaria* **6**: 111-134.
- Fan Y, Wu R, Chen M-H, Kuo L, Lewis P O (2011) Choosing among partition models in Bayesian phylogenetics. *Molecular Biology & Evolution* **28**: 523-532.
- Felsenstein J (1973) Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* **22**: 240-249.
- Felsenstein J (1978) The number of evolutionary trees. *Systematic Zoology* **27**: 27-33.
- Felsenstein J (1981) Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution* **17**: 368-376.
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.
- Felsenstein J (2004) *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.
- Feng D-F, Doolittle F (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* **25**: 351-360.
- Fernández-Palacios J M, de Nascimento L, Otto R, Delgado J D, García-del-Rey E, Arevalo J R, Whittaker R J (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography* **38**: 226-246.
- Fichant G A, Burks C (1991) Identifying potential tRNA genes in genomic DNA sequences. *Journal of Molecular Biology* **220**: 659-671.
- Field K G, Olsen G J, Lane D J, Giovannoni S J, Ghiselin M T, Raff E C, Pace N R, Raff R A (1988) Molecular phylogeny of the animal kingdom. *Science* **239**: 748-753.
- Finston T L, Johnson M S, Humphreys W F, Eberhard S M, Halse S A (2007) Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. *Molecular Ecology* **16**: 355-365.
- Finston T, Francis C, Johnson M (2009) Biogeography of the stygobitic isopod *Pygolabis* (Malacostraca: Tainisopidae) in the Pilbara, Western Australia: evidence for multiple colonisations of the groundwater. *Molecular Phylogenetics and Evolution* **52**: 448-460.
- Fitch W M (1970) Distinguishing homologous from analogous proteins. *Systematic Zoology* **19**: 99-113.
- Fleischer R C, McIntosh C E, Tarr C L (1998) Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Molecular Ecology* **7**: 533-545.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R C (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology & Biotechnology* **3**: 294-299.
- Fontboté J M, Guimerà J, Roca E, Sàbat F, Santanach P, Fernández-Ortigosa F (1990) The Cenozoic geodynamic evolution of the Valencia trough (western Mediterranean). *Revista de la Sociedad Geológica de España* **3**: 249-259.
- Fornós J J, Gelabert B, Ginés A, Ginés J, Tuccimei P, Vesica P L (2002) Phreatic overgrowths on
-

- speleothems: a useful tool in structural geology in littoral karstic landscapes. The example of Eastern Mallorca (Balearic Islands). *Geodinamica Acta* **15**:113-125.
- Fornós J J, Gelabert B (2011) Condicionants litològics i estructurals del carst a les Illes Balears. *Endins* **17**: 37-52.
- Francino M P, Ochman H (1997) Strand asymmetries in DNA evolution. *Trends in Genetics* **13**: 240-245.
- Frati F, Simon C, Sullivan J, Swofford D L (1997) Gene evolution and phylogeny of the mitochondrial cytochrome oxidase gene in collembola. *Journal of Molecular Evolution* **44**: 145-158.
- Frizon de Lamotte D, Zizi M, Missenard Y, Hafid M, El Azzouzi M (2008) in *Continental Evolution: The Geology of Morocco*, eds Michard A, Chalouan A, Saddiqi O, Frizon de Lamotte D (Springer, Heidelberg), pp 133-202.
- Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. **147**: 915-925.
- Fuglsang A (2008) Impact of bias discrepancy and amino acid usage on estimates of the effective number of codons used in a gene, and a test for selection on codon usage. *Gene* **410**: 82-88.
- Gatesy J, DeSalle R, Wahlberg N (2007) How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Systematic Biology* **56**: 355-363.
- Gaut B, Lewis P (1995) Success of maximum likelihood in the four-taxon case. *Molecular Biology & Evolution* **12**: 152-162.
- Gelabert B, Sàbat F, Rodríguez-Perea A (1992) A structural outline of the Serra de Tramuntana of Mallorca (Balearic Islands). *Tectonophysics* **203**: 167-183.
- Geldmacher J, Hoernle K, van den Bogaard P, Zankl G, Garbe-Schönberg D (2001) Earlier history of the > 70-Ma-old Canary hotspot based on temporal and geochemical evolution of the Selvagens Archipelago and neighbouring seamounts in the eastern North Atlantic. *Journal of Volcanology and Geothermal Research* **111**: 55-87.
- Geldmacher J, Hoernle K, van den Bogaard P, Duggen S, Werner R (2005) New 40K/39Ar age and geochemical data from seamounts in the Canary and Madeira volcanic provinces: Support for the mantle plume hypothesis. *Earth & Planetary Science Letters* **237**: 85-101.
- Gibert J, Danielopol D L, Stanford J A eds (1994) *Groundwater Ecology*. San Diego (CA): Academic Press.
- Gibert J, Malard F, Turquin M J, Laurent R (2000) Karst ecosystems in the Rhône river basin. Dins Wilkens H, Culver F, Humphreys W F (eds.) *Subterranean Ecosystems – Ecosystems of the World* (pp. 533- 558). Amsterdam: Elsevier.
- Gibert J, Deharveng L (2002) Subterranean Ecosystems: A Truncated Functional Biodiversity. *BioScience* **52**: 473-481.
- Gillespie J H, Langley C H (1979) Are evolutionary rates really variable? *Journal of Molecular Evolution* **13**: 27-34.
- Gillespie J H (1984) The molecular clock may be an episodic clock. *Proceedings of National Academy of Sciences (USA)* **81**: 8009-8013.
- Giribet G, Wheeler W C (1999) On gaps. *Molecular Phylogenetics & Evolution* **13**: 132-143.
- Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **101**: 301-320.
- Goetze E (2003) Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proceedings of the Royal Society of London, Series B* **270**: 2321-2331.
-



- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology & Evolution* **11**: 725-736.
- Goldman N, Anderson J P, Rodrigo A G (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**: 652-670.
- Golubchik T, Wise M J, Easteal S, Jermin L S (2007) Mind the gaps: Evidence of bias in estimates of multiple sequence alignments. *Molecular Biology & Evolution* **24**: 2433-2442.
- Gotoh O (1996) Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. *Journal of Molecular Biology* **264**: 823-838.
- Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research* **10**: 7055-7064.
- Gràcia F, Clamor B, Landreth R, Vicens D, Watkinson P (2001) Evidències geomorfològiques dels canvis del nivell marí. *Monografies de la Societat d'Història Natural de les Balears* **9**: 91-119.
- Grantham R C, Gautier C, Gouy M, Mercier R, Pavé A (1980) Codon catalog usage and the genome hypothesis. *Nucleic Acids Research* **8**: 49-79.
- Grasso C, Lee C (2004) Combining partial order alignment and progressive multiple sequence alignment increases alignment speed and scalability to very large alignment problems. *Bioinformatics* **20**: 1546-1556.
- Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics* **20**: 80-86.
- Green P J (1995) Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711-732.
- Guzik M T, Abrams K M, Cooper S J B, Humphreys W F, Cho J-L, Austin A (2008) Phylogeography of the ancient Parabathynellidae (Crustacea, Bathynellacea) from the Yilgarn region of Western Australia. *Invertebrate Systematics* **22**: 205-216.
- Hall T A (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Research* **41**: 95-98.
- Harzhauser M, Kroh A, Mandic O, Piller W E, Göhlich U, Reuter M, Berning B (2007) Biogeographic responses to geodynamics: A key study all around the Oligo-Miocene Tethyan Seaway. *Zoologischer Anzeiger* **246**: 214-256.
- Hasegawa M, Kishino H (1989) Confidence limits on the maximum likelihood estimate of the hominoid tree from mitochondrial DNA sequences. *Evolution* **43**: 672-677.
- Hassanin A, Leger N, Deutsch J (2005) Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Systematic Biology* **54**: 277-298.
- Hassanin A (2006) Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. *Molecular Phylogenetics & Evolution* **38**: 100-116.
- Hays J D, Imbrie J, Shackleton N J (1976) Variations in the Earth's orbit: Pacemaker of the ice ages. *Science* **194**: 1121-1131.
- Heads M (2005) Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* **21**: 62-78.
- Hedges S B, Chen H, Kumar S, Wang D Y, Thompson A S, Watanabe H (2001) A genomic timescale for the
-

- origin of eukaryotes. *BMC Evolutionary Biology* **1**: 4.
- Hedges S B, Parker P H, Sibley C G, Kumar S (1996) Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**:226-229.
- Hennig W (1966) *Phylogenetic systematics*. University of Illinois Press, Urbana, Chicago, London.
- Hewitt G M (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of The Royal Society of London, Series B* **359**: 183-195.
- Hillaire-Marcel C, Garipey C, Ghaleb B, Goy J L, Zazo C, Cuerda, J (1996) U-series measurements in tyrrhenian deposits from Mallorca. Further evidence for two last interglacial high sea levels in the Balearic Islands. *Quaternary Science Reviews* **15**: 53-62.
- Hillis D M, Dixon M T (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology* **66**: 411-453.
- Hillis D M, Mable B K, Moritz C (1996) Applications of molecular systematics: the state of the field and a look to the future. Dins Hillis D, Moritz C, Mable B K (eds.) *Molecular Systematics* (pp. 515-543, 655). Sunderland, MA: Sinauer Associates.
- Hipp A L, Hall J C, Sysma K J (2004) Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. *Systematic Biology* **53**: 81-89.
- Ho Y W S, Phillips M J, Drummond A J, Cooper A (2005) Accuracy of rate estimation using relaxed-clock models with a critical focus on the early Metazoan radiation. *Molecular Biology & Evolution* **22**: 1355-1363.
- Ho Y W S, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS ONE* **3**: e1615.
- Ho S Y W (2009) An examination of phylogenetic models of substitution rate variation among lineages. *Biology letters* **5**: 421-424.
- Hoffman R J, Boore J L and Brown W M (1992) A novel mitochondrial genome organization for the Blue Mussel, *Mytilus edulis*. *Genetics* **131**: 397-412.
- Holder M T, Lewis P O (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics* **4**: 275-284.
- Holsinger J R (1986) Zoogeographic patterns of North American subterranean amphipod crustaceans. Dins Gore R H, Heck K L (eds.), *Crustacean Biogeography*. Rotterdam: Balkema.
- Holsinger J R (1988) Trogllobites: the evolution of cave-dwelling organisms. *American Scientist* **76**: 146-153.
- Holsinger J R (1991) What can vicariance biogeographic models tell us about the distributional history of subterranean amphipods? *Hydrobiologia* **223**: 43-45.
- Holsinger J R (1993) Biodiversity of subterranean amphipod crustaceans: global patterns and zoogeographic implications. *Journal of Natural History* **27**: 821-835.
- Holsinger J R (1994) Pattern and process in the biogeography of subterranean amphipod crustaceans. *Hydrobiologia* **287**: 131-145. In D. C. Culver and J. R. Holsinger (eds.), *Biogeography of Subterranean Crustaceans: the effects of different scales*.
- Holsinger J R (2005) *Vicariance and dispersalist biogeography*, in Culver D C, and White W B, eds., *Encyclopedia of Caves*, Elsevier Academic Press, p. 591-599.
- Holthuis L B (1973) Caridean shrimps found in landlocked saltwater pools at four Indo-West Pacific localities (Sinai Peninsula, Funafuti Atoll, Maui, and Hawaiian Islands), with the description of one new genus and four new species. *Zoologische Verhandelingen* **128**: 1-48.
- Hsü K J, Bernoulli D (1978) Genesis of the Tethys and the Mediterranean. In: Hsü J, Montadert L, et al. *Init*
-

- Rep D S D P, Washington, USA 42: 943-949.
- Hsü K J, Montadert L, Bernoulli D, Cita M B, Erickson A, Garrison R E, Kidd R B, Melieres F, Müller C, Wright R (1977) History of the mediterranean salinity crisis. *Nature* **267**: 399-403.
- Huelsenbeck J P, Larget B, Swofford D (2000) A compound Poisson process for relaxing the molecular clock. *Genetics* **154**: 1879-1892.
- Huelsenbeck J P, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Huelsenbeck J P, Larget B, Miller R E, Ronquist F (2002) Potential applications and pitfalls of bayesian inference of phylogeny. *Systematics Biology* **35**: 673-688.
- Humphreys W F (2000) *Relict faunas and their derivation*. In H Wilkens, D C Culver and W F Humphreys (Eds). *Ecosystems of the World 30 Subterranean Ecosystems*, Amsterdam (Elsevier): 399-415.
- Hwang U W, Kim W (1999) General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *The Korean Journal of Parasitology* **37**: 215-28.
- Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology* **151**: 389-409.
- Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* **2**: 13-34.
- Iliffe T M, Hart C W and Manning R B (1983) Biogeography and the caves of Bermuda. *Nature* **302**: 141-142.
- Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708-713.
- Ito A, Aoki M N, Yokobori S-I, Wada H (2010) The complete mitochondrial genome of *Caprella scaura* (Crustacea, Amphipoda, Caprellidea), with emphasis on the unique gene order pattern and duplicated control region. *Mitochondrial DNA* **21**: 183-190.
- Iturralde-Vinent M A, MacPhee R D E (1999) Paleogeography of the Caribbean region: implications for cenozoic biogeography. *Bulletin of the American Museum of Natural History* **238**: 1-95.
- Iturralde-Vinent M A (2006) Meso-Cenozoic Caribbean paleogeography: Implications for the historical biogeography of the region. *International Geology Review* **48**: 791-827.
- Jaume D, Christenson K (2001) Amphi-Atlantic distribution of the subterranean amphipod family
- Jaume D, Vonk R (2012) Discovery of *Metacrangonyx* in inland groundwaters of Oman (Amphipoda: Gammaridea: Metacrangonyctidae). *Zootaxa* **54**: 68.
- Jeffreys H (1935) Some tests of significance, treated by the Theory of Probability. *Proceedings of the Cambridge Philosophy Society* **31**: 203-222.
- Jia W, Higgs P G (2008) Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Molecular Biology & Evolution* **25**: 339-351.
- Jobb G, von Haeseler A, Strimmer K (2004) Treefinder: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology* **4**: 18.
- Jolivet, L Augier R, Robin C, Suc J P, Rouchy J M (2006) Lithospheric-scale geodynamic context of the Messinian salinity crisis. *Sedimentary Geology* **188-189**: 9-33.
- Jones E J W, Cande S C, Spathopoulos F (1995) Evolution of a major oceanographic pathway: the equatorial Atlantic. *Journal of the Geological Society of London* **90**: 199-213.
- Juan C, Emerson B C (2010) Evolution underground: shedding light on the diversification of subterranean
-

- insects. *Journal of Biology* **9**: 17.
- Juan C, Guzik MT, Jaume D, Cooper SJB (2010) Evolution in caves: Darwin's 'wrecks of ancient life' in the molecular era. *Molecular Ecology* **19** (18): 3865-3880.
- Juberthie C, Decu V eds (1994) *Encyclopaedia Biospeologica*. Soc Biospéologie, Tome 1. 2. Société de Biospéologie, Moulis and Bucarest(France).
- Jukes T H, Cantor C R (1969) Evolution of protein molecules. Dins Munro H N (ed.), *Mammalian Protein Metabolism* (pp. 21-132). Nova York: Academic Press.
- Kadereit J, Griebeler E M, Comes H P (2004) Quaternary diversification in European alpine plants: pattern and process. *Philosophical Transactions of the Royal Society B* **359**: 265-274.
- Kane T C (1982) Genetic patterns and population structure in cave animals. Dins Mossakowski D i Roth G (eds.). *Environmental Adaptation and Evolution* (pp. 131- 149). Nova York: Gustav Fisher & Stuttgart.
- Kane T C, Culver D C, Mathieu J (1994) Biotic fluxes and gene flow. Dins Gibert J, Danielopol D, Stanford J (eds.), *Groundwater Ecology* (pp. 245-270). San Diego, CA: Academic Press.
- Karaman G S (1989) *Metacrangonyx ortali* n. sp., a new subterranean member of the family Crangonyctidae, from the Dead Sea region. *Studia Marina* **20**: 33-49
- Karaman G S, Pesce G L (1979) Researches in Africa by the Zoological Institute of l'Aquila, Italy. V. On three subterranean amphipods from North Africa (Amphipoda: Gammaridea). *Bulletin Zoölogisch Museum, Universiteit van Amsterdam* **7**: 197- 207.
- Kass R E, Raftery A E (1995) Bayes factors. *Journal of American Statistical Association* **90**: 773-795.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059-3066.
- Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511-518. Katoh K, Asimenos G, Toh H (2009) Multiple Alignment of DNA Sequences with MAFFT. In *Bioinformatics for DNA Sequence Analysis*, Ed. David Posada. **537**: 39-64.
- Katoh K, Toh H (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* **9**: 212.
- Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**: 1899-1900.
- Keddie E M, Higazi T, Unnasch T R (1998) The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Molecular Biochemical Parasitology* **95**: 111-127.
- Ketmaier V, Argano R, Caccone A (2003) Phylogeography and molecular rates of subterranean aquatic Stenassellid Isopods with a peri-Tyrrhenian distribution. *Molecular Ecology* **12**: 547-555.
- Ki J-S, Hop H, Kim S-J, Kim I-C, Park H G, Lee J-S (2010) Complete mitochondrial genome sequence of the Arctic gammarid, *Onisimus nanseni* (Crustacea; Amphipoda): Novel gene structures and unusual control region features. *Comparative Biochemistry and Physiology* **5**: 105-115.
- Kilpert F, Podsiadlowski L (2006) The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics* **7**:241.
- Kilpert F, Podsiadlowski L (2010) The mitochondrial genome of the Japanese skeleton shrimp *Caprella mutica* (Amphipoda: Caprellidea) reveals a unique gene order and shared apomorphic translocations with Gammaridea. *Mitochondrial DNA* **21**: 77-86.
- Kilpert F, Held C, Podsiadlowski L (2012) Multiple rearrangements in mitochondrial genomes of Isopoda
-

- and phylogenetic implications. *Molecular Phylogenetics and Evolution* **64**: 106-117.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**: 624-626.
- Kingsford M J, Choat J H (1985) The fauna associated with drift algae captured with a plankton-mesh purse seine net. *Limnology & Oceanography* **30**: 618-630.
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* **29**: 170-179.
- Kishino H, Hasegawa M (1990) Converting distance to time: Application to human evolution. In *Methods in Enzymology* **183** (Doolittle R. F. ed.): 550-570.
- Kishino H, Thorne J L, Bruno W (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology & Evolution* **18**: 352-361.
- Kjer K M (1995) Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Molecular Phylogenetics and Evolution* **4**: 314-330.
- Kjer K M, Honeycutt R L (2007) Site specific rates of mitochondrial genome and phylogeny of eutheria. *BMC Evolutionary Biology* **7** (8).
- Kluge A G (1997) Testability and the refutation and corroboration of cladistic hypotheses. *Cladistics* **13**: 81-96.
- Knowlton N, Weigt L A (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society B* **265**: 2257-2263.
- Kodandaramaiah U, Peña C, Braby M F, Grund R, Müller C J, Nylin S, Wahlberb N (2010) Phylogenetics of Coenonymphina (Nymphalidae: Satyrinae) and the problem of rooting rapid radiations. *Molecular Phylogenetics & Evolution* **54**: 363-394.
- Koenemann S, Holsinger J R (1999) Phylogenetic analysis of the amphipod family Bogidiellidae s. lat., and revision of taxa above the species level. *Crustaceana* **72**: 781-816.
- Kohne D E (1970) Evolution of higher organism DNA. *Quarterly Reviews of Biophysics* **33**: 327-375.
- Kornobis E, Pálsson S, Kristjánsson B K, Svavarsson J (2010) Molecular evidence of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland. *Molecular Ecology* **19**: 2516-2530.
- Kornobis E, Pálsson S, Sidorov D A, Holsinger J R, Kristjánsson B K (2011) Molecular taxonomy and phylogenetic affinities of two groundwater amphipods, *Crangonyx islandicus* and *Crymostygius thingvallensis*, endemic to Iceland. *Molecular Phylogenetics and Evolution* **58**: 527-539.
- Krebes L, Bastrop R (2012) The mitogenome of *Gammarus duebeni* (Crustacea Amphipoda): A new gene order and non-neutral sequence evolution of tandem repeats in the control region. *Genomics & Proteomics* **7**: 201-211.
- Krijgsman W, Hilgen F J, Raffi I, Sierro F J, Wilson D S (1999) Chronology, causes and progression of the Messinian salinity crisis. *Nature* **400**: 652-655.
- Kullback S, Leibler R A (1951) On information and sufficiency. *Annals of Mathematical Statistics* **22**: 79-86.
- Kumar S (2005) Molecular clocks: Four decades of evolution. *Nature Reviews Genetics* **6**: 654-662.
- Kumar S, Blaxter M L (2010) Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11**: 571.
- Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *Journal of Molecular Evolution* **37**: 380-398.
-

- Kvist L, Martens J, Nazarenko A, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution* **20**: 243–247.
- Lanteri A A, Confalonieri V A (2003) Filogeografía: objetivos, métodos y ejemplos. Dins Llorente J, Morrone J (eds.), (pp:185-193). *Una perspectiva latinoamericana de la biogeografía*. Mèxic: Facultad de Ciencias, UNAM.
- Lartillot N, Philippe H (2004) A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**: 1095-1109.
- Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology* **55**: 195–207.
- Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3, a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286-2288.
- Latinne A, Waengsothorn S, Rojanadilok P, Eiamampai K, Sribuarod K, Michaux J R (2012) Combined mitochondrial and nuclear markers revealed a deep vicariant history for *Leopoldamys neilli*, a cave-dwelling rodent in Thailand. *Plos One* **7**: e47670.
- Laurenti J L (1768) Specimen medicum, exhibens synopsis reptilium emendatum cum experimentis circa venena et antidota Reptilium Austriacorum. Joan. Tomae, Vienna.
- Lavrov D, Lang F (2005) Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. *Systematic Biology* **54**: 651-659.
- Lee M S Y (2001) Unalignable sequences and molecular evolution. *Trends in Ecology and Evolution* **16**: 681–685.
- Lefébure T, Douady C J, Gouy M, Trontelj P, Briolay J, Gibert J (2006a) Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular Ecology* **15**: 1797-1806.
- Lefébure T, Douady C J, Gouy M, Gibert J (2006b) Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution* **40**: 435-4467.
- Lefébure T, Douady C J, Malard F, Gibert J (2007) Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Molecular Phylogenetics and Evolution* **42**: 676-686.
- Lejeusne C, Chevaldonné P (2006) Brooding crustaceans in a highly fragmented habitat: the genetic structure of Mediterranean marine cave-dwelling mysid populations. *Molecular Ecology* **15**: 4123–4140.
- Lemmon A R, Moriarty E C (2004) The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology* **53**: 265–277.
- Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Molecular Biology & Evolution* **24**: 2669-2680.
- Letsch H O, Küch P, Stocsits R R, Misof B (2010) The impact of rRNA secondary structure consideration in alignment and tree reconstruction: simulated data and a case study on the phylogeny of Hexapods. *Molecular Biology and Evolution* **27**: 2507-2521.
- Lewis P O, Holder M T, Holsinger K E (2005) Polytomies and Bayesian phylogenetic inference. *Systematic Biology* **54**: 241–253.
- Li W-H (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.
-



*Bioinformatics* **25**: 1451-1452.

Lin Q, Cui P, Ding F, Hu S, Yu J (2012) Replication-associated Mutational Pressure (RMP) governs strand biased compositional asymmetry (SCA) and gene organization in animal mitochondrial genomes. *Current Genomics* **13**: 28-36.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**: 709-715.

Lomolino M V, Riddle B R, Brown J H (2006) *Biogeography*. Sinauer, Sunderland, MA.

Lowe T M, Eddy S R (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**: 955- 964.

Luo A, Qiao H, Zhang Y, Shi W, Ho S YW, Xu W, Zhang A, Zhu C (2010) Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evolutionary Biology* **10**: 242.

Lynch M (1996) Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Molecular Biology and Evolution* **13**: 209-220.

Machida R J, Miya M U, Nishida M, Nishida S (2002) Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). *Marine Biotechnology* **4**: 406-417.

Maddison W P (1989) Reconstructing character evolution on polytomous cladograms. *Cladistics* **5**: 365-377.

Maddison D R, Maddison W P (2005) *MacClade*. Sinauer Associates, Sunderland, Massachusetts.

Maddison W P, Maddison D R (2011) Mesquite: a modular system for evolutionary analysis. V. 2. 75 <http://mesquiteproject.org>.

Mantel N A (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209-220.

Mantel N A, Valand R S (1970) A technique of nonparametric multivariate analysis. *Biometrics* **26** : 547-558.

Marcadé I, Cordaux R, Doublet V, Debenest C, Bouchon D, Raimond R (2007) Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). *Journal of Molecular Evolution* **65**: 651-659.

Margalef R (1952) Materiales para la hidrobiología de la isla de Menorca. *Publicacions de l'Institut de Biologia Aplicada* **11**: 5-112.

Margulies M, Egholm M, Altman W E, Attiya S, Bader J S, Bemben L A, Berka J, Braverman M S, Chen Y J, Chen Z, et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.

Marmonier P, Vervier P, Gibert J, Dole-Olivier M J (1993) Biodiversity in ground waters. *Trends in Ecology and Evolution* **8**: 392-395.

Martin A P, Palumbi S R (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences (USA)* **90**: 4087-4091.

Martin A P (1999) Substitution rates of organelle and nuclear genes in sharks: Implicating metabolic rate (again). *Molecular Biology & Evolution* **16**: 996-1002. Martin J W, Davis G E (2001) *An Updated Classification of the Recent Crustacea*. Natural History Museum of Los Angeles County, 132 pp.

Martin, A. P. and T. M. Burg (2002). Perils of Paralogy: Using HSP70 Genes for Inferring Organismal Phylogenies. *Syst. Biol.* **51**, 570-587.

Marzo M, Pomar L, Ramos E, Rodríguez-Perea A (1983) El Paleogeno del SW de la Sierra Norte de Mallorca. Dins Pomar L, Obrador A, Fornós J J, Rodríguez-Perea A (eds.), *El Terciario de las Baleares (Mallorca Menorca)*. *Guía de las excursiones del X Congreso Nacional de Sedimentología* (pp. 75-90). Illes

---

- Balears, Menorca: Inst. Est. Baleàrics, Universitat de Palma de Mallorca.
- Masta S E (2000) Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TYC arm. *Molecular Biology and Evolution* **17**: 1091-1100.
- Mayr E (1982) Processes of Speciation in Animals. Dins Barigozi C (ed.), *Mechanisms of speciation* (pp. 1-19). Nova York: Alan R. Riss.
- McCaskill J S (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**: 1105-1119.
- McGrath D, Myers A A (1989) The drift amphipod *Hyale grimaldii* in Irish and British waters. *Journal of the Marine Biological Association of the United Kingdom* **69**: 913-918.
- Meng X F, Shi M I N, Chen X (2008) Population genetic structure of *Chilo suppressalis* (Walker) (Lepidoptera: Crambidae): strong subdivision in China inferred from microsatellite markers and mtDNA gene sequences. *Molecular Ecology* **17**: 2880-2897.
- Messouli M (1988) Les Crustacés Amphipodes Souterrains du Groupe *Metacrangonyx*. Répartition, Systématique et Phylogénie. Ms Thesis. Université de Marrakech, Morocco, 234 pp.
- Messouli M (1994) Evolution, Phylogénie et Biogéographie Historique des Metacrangonyctidae, Crustacés Amphipodes Stygobies du Nord de l'Afrique et des Régions Voisines. Ph.D Thesis. Université de Marrakech, Morocco, 312 pp.
- Messouli M, Boutin C, Coineau N (1991) Phylogénie et biogéographie évolutive d'un groupe de Metacrangonyctidae, Crustacés Amphipodes stygobies du Maroc. Les espèces du groupe *panousei*. *Mémoires de Biospéologie* **25**: 247-262.
- Meyer, A., and R. Zardoya. 2003. Recent advances in the (molecular) phylogeny of vertebrates. *Annual Review of Ecology, Evolution & Systematics* **34**: 311-318.
- Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M (2007) Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research* **35**: e97.
- Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform. *Nature Protocols* **3**: 267-278.
- Miller K G, Kominz M A, Browning J V, Wright J D, Mountain G S, Katz M E, Sugarman P J, Cramer B S, ChristieBlick N, Pekar S F (2005) The Phanerozoic record of global sea-level change. *Science* **310**: 1293-1298.
- Miller K G, Mountain G S, Wright J D, Browning J V (2011) A 180-million-year record of sea level and ice volume variations from continental margin and deep-sea isotopic records. *Oceanography* **24**: 40-53.
- Miller W, Drautz D I, Janecka J E, Lesk A M, Ratan A, Tomsho L P, Packard M, Zhang Y, McClellan L R, Qi J, et al. (2009) The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research* **19**: 213-220.
- Moreau C S, Bell C D (2011) Fossil Cross-Validation of the dated Ant phylogeny (Hymenoptera: Formicidae). *Entomologica Americana* **117**: 127-133.
- Morrone J J and Crisci J V (1995) Historical Biogeography: Introduction to Methods. *Annual Review of Ecology, Evolution and Systematics* **26**: 373-401.
- Mullis K, Faloona F (1987) Specific synthesis of DNA in vitro via a polymeerage-catalyzed chain reaction. *Methods of Enzymology* **155**: 807-814.
- Nardi F, Spinsanti G, Boore J L, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: monophyletic or paraphyletic?. *Science* **299**: 1887-1889.
- Near T J, Sanderson M J (2004) Assessing the quality of molecular divergence time estimates by fossil
-



- calibrations and fossil based model selection. *Philosophical Transactions of the Royal Society of London B* **359**: 1477-1483.
- Near T J, Meylan P A, Shaffer H B (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *American Naturalist* **165**: 137-146.
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York.
- Neiber M T, Hartke T R, Stemme T, Bergmann A, Rust J, Iliffe T M, Koenemann S (2011) Global biodiversity and phylogenetic evaluation of Remipedia (Crustacea). *PLoS One* **6**: e19627.
- Nelson G (1969) The problem of historical biogeography. *Systematic Zoology* **18**: 243-246.
- Nelson G, Platnick N (1981) *Systematics and biogeography: cladistics and vicariance*. Columbia University Press, New York.
- Newton M A, Raftery A E (1994) Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical methodology* **56**: 3-48.
- Nikolaou C, Almirantis Y (2006) Deviations from Chargaff's second parity rule in organellar DNA - insights into the evolution of organellar genomes. *Gene* **381**: 34-41.
- Notenboom J (1988a) Phylogenetic relationships and biogeography of the groundwater-dwelling amphipod genus *Pseudoniphargus*, with emphasis to the Iberian species. *Bijdragen tot de Dierkunde* **58**: 159-204.
- Notenboom J (1991) Marine regressions and the evolution of groundwater dwelling amphipods (Crustacea). *Journal of Biogeography* **18**: 437-454.
- Nylander J A A (2004) MrAIC.pl. Program distributed by the author. *Evolutionary Biology Centre*, Uppsala University.
- Nylander J A A, Wilgenbusch J C, Warren D L, Swofford D L (2008) AWTY (Are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24**: 581-883.
- Nylander J A, Ronquist F, Huelsenbeck J P, Nieves-Aldrey J L (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology* **53**: 47-67.
- Ochman H, Wilson A C (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution* **26**: 74-86.
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**: 470-474.
- Ojala D, Merkel C, Gelfand R, Attardi G (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* **22**: 393-403.
- Oliver P M, Adams M, Doughty P (2010) Molecular evidence for ten species and Oligo-Miocene vicariance within a nominal Australian gecko species (*Crenadactylus ocellatus*, Dyplogadactylidae). *BMC Evolutionary Biology* **10**: 386.
- Olivet J L, Bonnin J, Beuzart P, Auzende J M (1984) Cinématique de l'Atlantique nord et central. Dins *Rapports scientifiques*. CNEX,O. 108 pp. Paris.
- Oromí P, Martín J L, Medina A L, Izquierdo I (1991) The evolution of the hypogean fauna in the Canary Islands. *The Unity of Evolutionary Biology Congress* (Maryland, EEUU) **2**: 380-395.
- Oulbazz Z, Coineau N and Boutin C (1998) *Metacrangonyx knidirii* n. sp. et *Metacrangonyx remyi* Balazuc et Ruffo, 1953 (Crustaces Amphipodes Metacrangonyctidae) des eaux souterraines du Maroc. *Mémoires de Biospéologie* **25**: 35-43.
- Oxford, UK.

- Page, R. D. M. (1994). Maps between trees and cladistic analysis of historical associations among genes, organisms and areas. *Syst. Biol.* **43**, 58-77.
- Page R D M, Holmes E C (2000) *Molecular Evolution. A Phylogenetic Approach*. Blackwell Science Ltd.,
- Page R, Holmes E (2000) *Molecular Evolution, A Phylogenetic Approach*. Blackwell Science.
- Page T J, Humphreys W F, Hughes J M (2008) Shrimps down under: Evolutionary relationships of subterranean crustaceans from Western Australia (Decapoda: Atyidae: *Stygiocaris*). *PLoS ONE* **3**: e1618.
- Palomino D, Tomás Vázquez J, Díaz del Río V, Fernández-Salas L M (2009) Estudio de los procesos sedimentarios recientes de la Bahía de Palma a partir del análisis de la morfología y la respuesta acústica (Islas Baleares, Mediterráneo Occidental). *Revista de la Sociedad Geológica de España* **22** : 79-93.
- Papadopoulou A, Anastasiou I, Vogler A P (2010) Revisiting the insect mitochondrial molecular clock: The mid-Aegean trench calibration. *Molecular Biology & Evolution* **27**: 1659-1672.
- Payton C E (1977) *Seismic Stratigraphy: Applications to Hydrocarbon Exploration*. American Association of Petroleum Geologists Memoir **26**.
- Peltier W R, Fairbanks R G (2006) Global glacial ice volume and Last Glacial Maximum duration from an extended Barbados sea level record. *Quaternary Science Reviews* **25**: 322-3,337.
- Perna N T, Kocher T D (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* **41**: 353-358.
- Perseke M, Fritsch G, Ramsch K, Bernt M, Merkle D, Middendorf M, Bernhard D, Stadler PF, Schlegel M (2008) Evolution of mitochondrial gene orders in echinoderms. *Molecular Phylogenetics and Evolution* **47**: 855-864.
- Pierre C, Caruso A, Blanc-Valleron M M, Rouchy J M, Orszag-Sperber F (2006) Reconstruction of the paleo environmental changes around the Miocene-Pliocene boundary along a West-East transect across the Mediterranean. *Sedimentary Geology* **188-189**: 319-340.
- Podsiadlowski L, Bartolomaeus T (2006) Molecular Major rearrangements characterize the mitochondrial genome of the isopod *Idotea baltica* (Crustacea: Peracarida). *Phylogenetics and Evolution* **40**: 893-899.
- Pol D (2004) Empirical problems of hierarchical Likelihood Ratio Test for model selection. *Systematic Biology* **53**: 949-962.
- Pomar L (1979) La evolución tectonosedimentaria de las Baleares: Análisis crítico. *Acta Geològica Hispànica* **14**: 293-310.
- Pomar L, Rodríguez-Perea A (1983) El Neógeno inferior de Mallorca: Randa. En: L. Pomar, A. Obrador, J. J. Fornos y A. Rodríguez-Perea (Eds.), *El Terciario de las Baleares (Mallorca Menorca)*. Guía de las excursiones del X Congreso Nacional de Sedimentología, Menorca. Institut d'Estudis Balearics, Universitat de Palma de Mallorca, 115-137.
- Pomar L (1991) Reef geometries, erosion surfaces and high-frequency sea-level changes, upper miocene Reef Complex, Mallorca, Spain. *Sedimentology* **38**: 243-269.
- Pons J, Vogler A P (2006) Size, frequency, and phylogenetic signal of multiple-residue indels in sequence alignment of introns. *Cladistics* **22**: 144-156.
- Pons J, Ribera I, Bertranpetit J, Balke M (2010) Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Molecular Phylogenetics & Evolution* **56**: 796-807.
- Porter M (2007) Subterranean biogeography: what have we learned from molecular techniques? *Journal of Cave and Karst Studies* **69**: 179-186.
-

- Posada D, Crandall K A (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.
- Posada D, Crandall K A (2001) Selecting the best-fit model of nucleotide substitution. *Systematic Biology* **50**: 580-601.
- Posada D, Buckley T R (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793-808.
- Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**: 1253-1256.
- Posadas P, Crisci J V, Katinas L (2006) Historical biogeography: A review of its basic concepts and critical issues. *Journal of Arid Environments* **66**: 389-403.
- Poulson T L, White W B (1969) The cave environment. *Science* **165**: 971-981.
- Proudlove G, Wood P J (2003) The blind leading the blind: cryptic subterranean species and DNA taxonomy. *Trends in Ecology & Evolution* **18**:272-273.
- Racovitz E G (1907) Essai sur les problemas biospeologiques. *Archives de Zoologie Experimentale et Generale (Biospeologica I)* **4**: 371-488.
- Rambaut A, Drummond A J (2007). Tracer v. 1.4. <http://beastbioedacuk/Tracer/>.
- Ramos-Onsins S E, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**: 2092-2100.
- Rand K N (1996) Crystal violet can be used to visualize DNA bands during gel electrophoresis and to improve cloning efficiency. *Elsevier Trends Journals Technical Tips Online*, T40022.
- Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution* **43**: 304-311.
- Renner S S (2005) Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science* **10**: 550-558.
- Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Molecular Biology & Evolution* **15**: 957-966.
- Riba O (1981) Aspectes de la geologia marina de la conca Mediterrània Balear durant el Neògen. Dins: *Memòries de la Reial Acadèmia de Ciències i Arts de Barcelona* (pp.85). Barcelona.
- Riba O (1983) Las Islas Baleares en el marco geológico de la Cuenca Mediterránea occidental durante el terciario. *X Congreso Nacional de Sedimentología, Menorca*: 5-20.
- Riddle B R, Dawson M N, Hadly E A, Hafner D J, Hickerson M J, Mantooth S J, Yoder A D (2008) The role of molecular genetics in sculpting the future of integrative biogeography *Prog in Physical Geography* **32**: 173-202.
- Ridley M (1996) *Evolution* 2nd ed. Cambridge, MA: Blackwell Science, 508-533.
- Ripplinger J, Sullivan J (2008) Does choice in model selection affect Maximum Likelihood analysis? *Systematic Biology* **57**: 76-85.
- Roca E (1994) La evolución geodinámica de la Cuenca Catalano-Balear y áreas adyacentes desde el Mesozoico hasta la actualidad. *Acta Geologica Hispanica* **29**: 3-25.
- Rocha E P C (2004) The replication-related organization of bacterial genomes. *Microbiology* **150**: 1609-1627.
- Rodrigo A G, Kelly-Borges M, Bergquist P R, Bergquist P L (1993) A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *New Zealand Journal of Botany* **31**: 257-268.
- Rodríguez-Perea A, Ramos-Guerrero E, Pomar L, Paniello X, Obrador A, Martí J (1987) El Triásico de las

- Baleares. *Cuadernos de Geología Ibérica* 1987, **11**: 295-321.
- Rohland N, Malaspinas A S, Pollack J L, Slatkin M, Matheus P, Hofreiter M (2007) Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biology* **5**: e207.
- Rokas A, Holland P W (2000) Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution* **15**: 454-459.
- Rokas A, Carroll S B (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology & Evolution* **22**: 1337-1344.
- Ronquist F, Huelsenbeck J P (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.
- Ronquist F, Huelsenbeck J P, van der Mark P (2005) MrBayes3 manual.
- Rosen D E (1978) Vicariant patterns and historical explanation in biogeography. *Systematic Zoology* **27**: 159-188.
- Rosenfeld J A, Payne A, DeSalle R (2012) Random roots and lineage sorting. *Molecular Phylogenetics & Evolution* **64**: 12-20.
- Rota J (2011) Data partitioning in Bayesian analysis: molecular phylogenetics of metalmark moths (Lepidoptera: Choreutidae). *Systematic Entomology* **36**: 317-329.
- Rouch R, Danielopol D L (1987) L'origine de la faune aquatique souterraine, entre le paradigme du refuge et le modèle de la colonisation active. *Stygologia* **3**: 345-372.
- Ruffo S (1954) *Metacrangonyx longicaudus* n. sp. (Amphipoda Gammaridae) delle acque sotterranee del Sahara Marocchino. *Museo Civico di Storia Naturale di Verona* **4**: 127-130.
- Ruffo S (1982) Una nouvelle espèce de *Metacrangonyx* Chevreux (Amphipoda: Gammaridae) du désert du Sinaï. *Israel Journal of Zoology* **31**: 151-156.
- Saiki R K, Gelfand D H, Stoffel S, Scharf S J, Higuchi R, Horn G T, Mullis K B, Erlich H A (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.
- Saitou N & Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.
- Salvato P, Simonato M, Battisti A, Negrisolò E (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* **9**: 331.
- Sanderson M J (1995) Objections to bootstrapping phylogenies: a critique. *Systematic Biology* **44**: 299-320.
- Sanderson M J (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology & Evolution* **14**: 1218-1231.
- Sanderson M J, Kim J (2000) Parametric phylogenetics? *Systematic Biology* **49**: 817-829.
- Sanderson M J (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology & Evolution* **19**: 101-109.
- Sanger F, Nicklen S, Coulson A R (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences (USA)* **74**: 5463-5467.
- Sboner A, Habegger L, Pflueger D, Terry S, Chen D Z, Rozowsky J S, Tewari A K, Kitabayashi N, Moss B J, Chee M S, Demichelis F, Rubin M A, Gerstein M B (2010) FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biology* **11**: R104.
- Slater J G, Hellinger S and Tapscott C (1979) The paleobathymetry of the Atlantic Ocean from the Jurassic to the Present. *Journal of Geology* **85**: 509-552.
-

- Schattner P, Brooks A N, Lowe T M (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33**: W686-W689.
- Schettino A, Turco E (2009) Breakup of Pangea and plate kinematics of the central Atlantic and Atlas regions. *Geophysical Journal International* **178**: 1078-1097.
- Schettino A, Turco E (2011) Tectonic history of the Western Tethys since the Late Triassic. *Geological Society of America* **123**: 89-105.
- Schneider A, Ebert D (2004) Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. *Journal of Molecular Evolution* **59**: 90-96.
- Schneider A, Cannarozzi G M (2009) Support patterns from different outgroups provide a strong phylogenetic signal. *Molecular Biology & Evolution* **26**: 1259-1272.
- Schneider S, Kueffer J M, Roessli D, Excoffier L (1997) ARLEQUIN 1.1: A software for population genetic analysis. Geneva: University of Geneva, Genetics and Biometry Laboratory.
- Schwarz G (1978) Estimating the dimension of a model. *Annals of Statistics* **6**: 461-464.
- Schwentner M, Clavier S, Fritsch M, Olesen J, Padhye S, Timms B V, Richter S (2013) *Cyclestheria hislopi* (Crustacea: Branchiopoda): A group of morphologically cryptic species with origins in the Cretaceous. *Molecular Phylogenetics & Evolution* **66**: 800-810.
- Sengör, A M C (1998) Die Tethys vor hundert Jahren und heute. *Mitteilungen der Österreichischen Geologischen Gesellschaft* **89**: 5-177.
- Seung Chul Shin, Jin Cho, Jong Kyu Lee, Do Hwan Ahn, Hyoungseok Lee, and Hyun Park. (2012) Complete mitochondrial genome of the Antarctic amphipod *Gondogeneia antarctica* (Crustacea, amphipod) *Mitochondrial DNA* **23**: 25-27.
- Shackleton N J, Opdyke N D (1973) Oxygen isotope and palaeomagnetic stratigraphy of Equatorial Pacific core V28-238: Oxygen isotope temperatures and ice volumes on a 105 year and 106 year scale. *Quaternary Research* **3**: 39-55.
- Shackleton N J, Backman H, Zimmerman H, Kent D V, Hall M A, Roberts D G, Schnitker D, Baldauf J G, Desprairies A, Homrighausen R, et al. (1984) Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* **307**: 620-623.
- Shapiro L H, Strazanac J S, Roderick G K (2006) Molecular phylogeny of Banza (Orthoptera: Tettigoniidae), the endemic katydids of the Hawaiian Archipelago. *Molecular Phylogenetics & Evolution* **41**: 53-63.
- Sharp P M, Li W H (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. *Nucleic Acids Research* **14**: 7737-7749.
- Shavit L, Penny D, Hendy M D, Holland B R (2007) The problem of rooting rapid radiations. *Molecular Biology and Evolution* **24**: 2400-2411.
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114-1116.
- Simmons M P, Ochoterena H (2000) Gaps as Characters in Sequence-Based Phylogenetic Analyses. *Systematics Biology* **49**: 369-381.
- Simmons M P, Ochoterena H, Carr T G, (2001) Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Systematics Biology* **50**: 454-462.
- Simon C, Frati F, Beckenback A T, Crespi B, Liu H and Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Annals of the Entomological Society of America* **87**: 651-701.
-

- Singh T R, Tsagkogeorga G, Delsuc F, Blanquart S, Shenkar N, Loya Y, Douzery E J P, Huchon D (2009) Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics* **10**: 534.
- Sket, B (1996) Biotic diversity of hypogean habitats in Slovenia and its cultural importance. *Proceedings of the International Biodiversity Seminar*; **30**. Eslovenia.
- Sket B (1997) Distribution of *Proteus* (Amphibia: Urodela: Proteidae) and its possible explanation. *Journal of Biogeography* **24**: 263–280.
- Sket B (1999) The nature of biodiversity in hypogean waters and how it is endangered. *Biodiversity and Conservation* **8**: 1319-1338.
- Smith A B, Peterson K J (2002) Dating the time of origin of major clades. *Annual Review of Earth and Planetary Sciences* **30**: 65–88.
- Smith A E, Marker K A (1968) N-Formylmethionyl transfer RNA in mitochondria from yeast and rat liver. *Journal of Molecular Biology* **38**: 241-243.
- Smith A G, Smith D G, Funnell B M (1994) *Atlas of Mesozoic and Cenozoic Coastlines*, Cambridge: Cambridge University Press.
- Smith S A, Dunn C W (2008) Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**: 715-716.
- Sneath P H A and Sokal R R (1973) *Numerical Taxonomy*, Freeman W H (ed.) (pp 230-234). San Francisco: San Francisco Company.
- Sorenson M D, Franzosa E A (2007) TreeRot, version 3 Boston University, Boston, MA.
- Sosdian S, Yair Rosenthal (2009) Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* **325**: 306-310.
- Sperling F A H, Spence J R, Andersen N M (1997) Mitochondrial DNA, allozymes, morphology, and hybrid compatibility in *Limnoporus* water striders (Heteroptera: Gerridae): do they all track species phylogenies?. *Annals of Entomological Society of America* **90**: 401–415.
- Sprinzi M, Hartmann T, Weber J, Blank J, Zeidler R (1989) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research* **17**: 1–172.
- Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**: 456–463.
- Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics* **22**: 2688–2690.
- Stampfli G M, Borel G D (2002) A plate tectonic model for the Paleozoic and Mesozoic constrained by dynamic plate boundaries and restored synthetic oceanic isochrons. *Earth and Planetary Science Letters* **196**: 17-33.
- Stanley E L, Bauer A M, Jackman T R, Branch W R, Mouton P Le Fras N (2011) Between a rock a hard politomy: Rapid radiation in the rupicolous girdled lizards (Squamata: Cordylidae). *Molecular Phylogenetics & Evolution* **58**: 53-70.
- Sterrerr W (1973) Plate tectonics as a mechanism for the dispersal and speciation in interstitial sand fauna. *Netherlands Journal of Sea Research* **7**: 200-222.
- Stock J H (1977) The taxonomy and zoogeography of the hadziid Amphipoda with emphasis on the West Indian taxa. *Studies on the Fauna of Curaçao and other Caribbean Islands* **55**: 1-130.
- Stock J H (1980) Regression model evolution as exemplified by the genus *Pseudoniphargus* (Amphipoda). *Bijdragen tot de Dierkunde* **50**: 105-144.
-



- Stock J H, Rondé-Broekhuizen B (1986) A new species of *Pygocrangonyx*, an amphipod genus with African affinities, from Fuerteventura. *Bijdragen Tot de Dierkunde* **56**: 247-266.
- Stock J H (1993) Some remarkable distribution patterns in stygobiont Amphipoda. *Journal of Natural History* **27**: 807-819.
- Stoch F (1995) The ecological and historical determinants of crustacean diversity in groundwaters, or: why are there so many species?. *Memoires de Biospeologie* **22**: 139-160.
- Stoch F (1997) *Metacrangonyx ilvanus* n. sp., the first Italian representative of the family Metacrangonyctidae (Crustacea: Amphipoda). *Annals of Limnology* **33**: 255-262.
- Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. *Proceeding of the Royal Society London B* **269**: 137-142.
- Suchard M A, Weiss R E, Sinsheimer J S (2001) Bayesian selection of continuous-time markov chain evolutionary models. *Molecular Biology and Evolution* **18**: 1001-1013.
- Suchard M A, Rambaut A (2009) Many-core algorithms for statistical phylogenetics. *Bioinformatics* **25**: 1370-1376.
- Sueoka N, Kawanishi Y (2000) DNA G+C content of the third codon position and codon usage biases of human genes. *Gene* **261**: 53-62.
- Suess E (1893) Are great ocean depths permanent? *Natural Science* **2**: 180-187.
- Suess E (1906) *The face of the Earth (Der Antlitz der Erde)*. Claredon Press, Oxford, 2 vols.
- Sullivan J, Joyce P (2005) Model selection in phylogenetics. *Annual Review of Ecology, Evolution and Systematics* **36**: 445-466.
- Sullivan J, Swofford D L (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *Journal of Mammalian Evolution* **4**: 77-86.
- Sullivan J, Swofford D L (2001) Should we use model-based methods for phylogenetic inference when we know assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Systematic Biology* **50**: 723-729.
- Supek F, Vlahoviček (2004) INCA: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics* **20**: 2329-2330.
- Sutovsky P, Moreno R, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999) Ubiquitin tag for sperm mitochondria. *Nature* **402**, 371-372.
- Suzuki Y, Glazko G V, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of National Academy of Sciences (USA)* **99**: 16138-16143.
- Swofford D L, Olsen G K, Waddell P J, Hillis D M (1996) *Phylogeny reconstruction*. Pp: 407-514 In *Molecular systematics*, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.
- Swofford D L (2002) PAUP 4.0 b10, Phylogenetic analysis using parsimony, (Sunderland: Sinauer Associates).
- Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics* **105**: 437-460.
- Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematics Biology* **56**: 564-577.
- Talavera G, Vila Roger (2011) What is the phylogenetic signal limit form mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evolutionary Biology* **11**: 315.
- Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-

- joining method. *Proceedings of the National Academy of Sciences (USA)* **101**:11030-11035.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596-1599.
- Thiede J (1979) Paleogeography and Paleobathymetry of the Mesozoic and Cenozoic North Atlantic Ocean. *GeoJournal* **3**: 263-272.
- Thorne J L, Goldman N, Jones D T (1996) Combining protein evolution and secondary structure. *Molecular Biology and Evolution* **13**: 666-673.
- Thorne J L, Kishino H, Painter I S (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology & Evolution* **15**: 1647-1657.
- Thorne J L, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51**: 689-702.
- Timmermans M J T N, Dodsworth S, Culverwell C L, Bocak L, Ahrens D, Littlewood D T J, Pons J, Vogler A P (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Research* **38**: e197.
- Townsend T M, Mulcahy D G, Noonan B P, Sites Jr J W, Kuczynski C A, Wiens J J, Reeder T W (2011) Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. *Molecular Phylogenetics & Evolution* **61**: 363-380.
- Trontelj P, Douady C J, Fiser C, Gibert J, Spela G, Lefébure T, Sket B, Zaksek V (2009) A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts?. *Freshwater Biology* **54**:727-744.
- Tuccimei P, Fornós J J, Ginés A, Ginés J, Gràcia F, Mucedda M (2003) Sea level change at Capo Caccia (NW Sardinia) and Mallorca (Balearic Islands) during oxygen isotope substage 5e, based on Th/U datings of phreatic overgrowths on speleothems. *Societat d'Història Natural de les Balears*, 121-135.
- Tuccimei P, Fornós J J, Ginés A, Ginés J, Gràcia F, Mucedda M (2006) Sea level change at Capo Caccia (NW Sardinia) and Mallorca (Balearic Islands) during oxygen isotope substage 5e, based on Th/U datings of phreatic overgrowths on speleothems. In: Pons G X & Vicens D (eds) Geomorfologia litoral; Quaternari. Homenatge a Joan Cuerda Barceló. Monografies de la Societat d'Història Natural de les Balears, Palma, 14: 121-135.
- Vandel A (1964) *Biospéologie. La biologie des animaux cavernicoles*. Paris, Gauthier-Villars Éditeur.
- Verovnik R, Sket B, Trontelj P (2004) Phylogeography of subterranean and surface populations of water lice *Asellus aquaticus* (Crustacea: Isopoda). *Molecular Ecology* **13**: 1519-1532.
- Vesica P L, Tuccimei P, Turi B, Fornós J J, Ginés A, Ginés J (2000) Late Pleistocene paleoclimates and sea-level change in the Mediterranean as inferred from stable isotope and U-series studies of overgrowths on speleothems, Mallorca, Spain. *Quaternary Science Reviews* **19**: 865-879.
- Vilstrup J T, Ho S Y W, Foote A D, Morin P A, Kreb D, Krützen M, Parra G J, Robertson K M, de Stephanis R, Verborgh P, Willerslev E, Orlando L, Gilbert M T (2011) Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evolutionary Biology* **11**: 65.
- Vogler AP, Welsh A, Hancock JM (1997) Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). *Mol Biol Evol* **14**:6-19
- Vonk R, Schram F R (2003) *Ingolfiellidea* (Crustacea, Malacostraca, Amphipoda): a phylogenetic and biogeographic analysis. *Contributions to Zoology* **72**: 39-72.
- Wägele J W (1985) On the Tethyan origin of the stygobiont Anthuridea *Curassanthura* and *Cyathura*
-

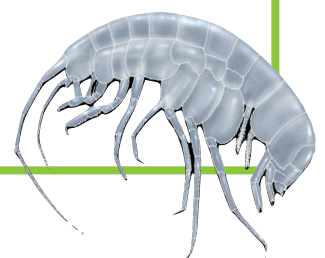


- (Stygocyathura), with description of *Curassanthura canariensis* n. sp. from Lanzarote (Crustacea, Isopoda). *Stygologia* **1**: 258-269.
- Walsh H E, Kidd M G, Moum T, Friesen V L (1999) Polytomies and the power of phylogenetic inference. *Evolution* **53**: 932-937.
- Wallace A R (1855) On the law which has regulated the introduction of new species. *Annual Magazine of Natural History Series 2* **16**: 184-196.
- Wallace A R (1876) *The Geographical Distribution of Animals*. Macmillan, London.
- Wan X F, Xu D, Kleinhofs A, Zhou J (2004) Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evolutionary Biology* **4**: 19.
- Wegener A (1929) *Die Entstehung der Kontinente und Ozeane*, Sammlung Vieweg, Brunswick.
- Wei S-J, Shi M, Chen X-X, Sharkey M J, van Achterberg C, Ye G-Y, He J-H (2010) New views of strand asymmetry in insect mitochondrial genomes. *PLoS One* **5**: e12708.
- Wei S-J, Shi M, Sharkey M J, van Achterberg C, Chen X-X (2010) Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. *BMC Genomics* **11**: 371.
- Wheeler W C (1990) Nucleic acid sequence phylogeny and random outgroups. *Cladistics* **6**: 363-367.
- White D J, Wolff J N, Pierson M, Gemmel N J (2008) Revealing the hidden complexities of the mtDNA inheritance. *Molecular Ecology* **17**: 4925-4942.
- Wiens J J, Kuczynski C A, Smith S A, Mulcahy D, Sites Jr J W, Townsend T M, Reeder T W (2008) Branch length, support, and congruence. Testing the phylogenomic approach with 20 nuclear loci in snakes. *Systematic Biology* **57**: 420-431.
- Willis K J, Niklas K J (2004) The role of Quaternary environmental change in plant macroevolution: the exception or the rule? *Philosophical Transactions of the Royal Society B* **359**: 159-172.
- Woese C R, Fox G E (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Journal of Molecular Evolution* **9**: 369-371.
- Wolstenholme D R (1992) Animal mitochondrial DNA: structure and evolution. *International Review of Cytology* **141**: 173-216.
- Wolstenholme D R, Macfarlane J L, Okimoto R, Clary D O, Wahleithner J A (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proceedings of the National Academy of Sciences (USA)* **84**: 1324-1328.
- Wright F (1990) The 'effective number of codons' used in a gene. *Gene* **87**: 23-29.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics* **1**: 323-334.
- Wyman S K, Jansen R K, Boore J L (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics Applications Note* **20**: 3252-3255.
- Xia X (2001) *Data analysis in molecular biology and evolution*. Boston: Kluwer Academic Publishers.
- Xia X, Xie Z (2001) DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity* **92**: 371-373.
- Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE. In *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*, 2nd edition, P. Lemey, M. Salemi and A.M. Vandamme, eds. (Cambridge: University Press), pp. 615-630.
- Xie W, Lewis P O, Fan Y, Kuo L, Chen M-H (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60**: 150-160.
- Xiong B, Kocher T D (1991) Comparison of mitochondrial DNA sequences of seven morphospecies of

- black flies (Diptera: Simuliidae). *Genome* **34**: 306-311.
- Yang Z H, Goldman N, Friday A (1995) Maximum likelihood trees from DNA sequences: A peculiar statistical estimation problem. *Systematic Biology* **44**: 384-399.
- Yang Z H (2004) A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zoologica Sinica* **50**: 645-656.
- Yang Z H (2007) Paml 4: phylogenetic analysis by maximum likelihood. *Molecular Biology & Evolution* **24**: 1586-1591.
- Yang Z H, Rannala B (2005) Branch-length prior influences Bayesian posterior probability of phylogeny. *Systematic Biology* **54**: 455-470.
- Yang J S, Yang W J (2008) The complete mitochondrial genome sequence of the hydrothermal vent galatheid crab *Shinkaia crosnieri* (Crustacea: Decapoda: Anomura): A novel arrangement and incomplete tRNA suite. *BMC Genomics* **9**: 257.
- Yoder A D, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology & Evolution* **17**: 1081-1090.
- Yokobori S, Pääbo S (1995) tRNA editing in metazoans. *Nature* **377**: 490.
- Yokoyama Y, Esat T M (2011) Global climate and sea level: Enduring variability and rapid fluctuations over the past 150,000 years. *Oceanography* **24**: 54-69.
- Zakšek V, Sket B, Gottstein S, Franjević D, Trontelj P (2009) The limits of cryptic diversity in groundwater: phylogeography of the cave shrimp *Troglocaris anophthalmus* (Crustacea: Decapoda: Atyidae). *Molecular Ecology* **18**: 931-946.
- Zhang D X, Hewitt G M (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* **25**: 99-120.
- Zhang D X, Szymura J M, Hewitt G M (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution* **40**: 382-391.
- Zuekerkandl E, Pauling L (1962) *Molecular disease, evolution, and genie heterogeneity*. Dins Kasha M, Pullman B (eds.) *Horizons in biochemistry* (pp. 189). Nova York: Academic Press.
- Zuekerkandl E and Pauling L (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**: 357-366.
- Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis, University of Texas at Austin, TX.
-



**MATERIAL ADICIONAL**





Taula Addicional 1. Oligonucleòtids emprats per completar el mitogenoma sencer de *M. longipes*.

Nom encebador	Sentit	Seqüència	Posició	Regió
Cob-long-R1	R	CCCTTATGTCACCTCAATTTCATATTCAACCTG	355	<i>cob</i>
Met-fill-10304	R	TAAACCCCTTATGTCACCTCC	359	<i>cob</i>
Prolong-cob-F2	R	AGCACTAGGAGATGACGAG	391	<i>cob</i>
Prolong-cob-F1	R	GAAACAGGCTCTAAAAACCCATTG	531	<i>cob</i>
Met-fill-rev-10014	F	AAACGCATAATAGTGGGAGTG	602	<i>cob</i>
Prolong-cob-R1	F	TTAGGTTAGGACCGATATAAAG	675	<i>cob</i>
Met-prolong-nd2f-R1	F	TACTACTTGCCCTTCTACACC	989	<i>cob</i>
Met-nd2f-R1	R	CCTTTTATTATTTCTGTAGTTAT	1614	<i>Nad2</i>
Met-nd2f-R2	R	CTCAAAGTCTGGATAAAAAAATAC	1755	<i>Nad2</i>
Met-OS-cox-p1	R	GATAGTATATTTACCCCTAACCC	2277	<i>Nad2</i>
Met-OS-cox-p2	R	GGAACAAATCCACTCAAAGG	2313	<i>Nad2</i>
Met-cox1-R1	R	CCCTTATAGCAGTTCCTACTATCCTAGC	2776	<i>Cox1</i>
Met-cox1-F	F	TAGTAGGAACGTCTATAAAGG	2756	<i>Cox1</i>
Met-cox1-R2	R	GAAGAAGTCAGTTACCGAACCCCTCC	2923	<i>Cox1</i>
Met-cox1-F3	F	TAAGAAGAATAGTAGAAAGAGGGGTTGGC	3014	<i>Cox1</i>
Met-cox1-F2	F	TATACGAGTTGGGATAATAGGAATAGACC	3211	<i>Cox1</i>
Met-cox1-F4	F	TAACGACCCGAAACCTAAATACCTC	3302	<i>Cox1</i>
Met-cox1-R333	R	CCTAACCCACTAAAAATCTTAATCCAGTAGG	3647	<i>Cox1</i>
Met-cx1-r2	R	CTGTGTTTGGTATTTTGTCTGGATTTAGACATTGG	3862	<i>Cox1</i>
Met-cx2-R1	R	GCACAGCTCAAGAATGAATCACATCCCTAGCCC	4769	<i>Cox2</i>
Met-fill-2713	F	TTATCATAGTAGAAATACTTCCC	4991	<i>tRNA-Lys</i>
Met-fill-R4160	R	CAAGTAACCTAAGTCTGAAACC	6334	<i>Cox3</i>
Pro-prolong-f2	F	CAGGAGTTAGAGTTACTTGAGC	6316	<i>Cox3</i>
Pro-prolong-f1	F	GAGCACATAACTCTATACTTCAAGGAC	6334	<i>Cox3</i>
Met-fill-4495	F	CATTTTGTAGATGTTGTTTGG	6624	<i>Cox3</i>
Prolong-Meta-H5	F	GTTTTTATTACTGATGAGGC	6657	<i>Cox3</i>
Prolong-newMetfill-4494	R	AAAAGGATCAAACCCACATTC	6822	<i>Nad3</i>
Pro-prolong6-R	R	ACACTATAAAAAACAACAGGATTACTC	7218	<i>tRNA-Asn/Glu</i>
Met-prolong6-R	R	TTTATGAGAGTTAGAAGGGG	7810	<i>Nad5</i>
Met-fill-rev5681	R	TTAGAAGAATAATTGGGGTAGG	7955	<i>Nad5</i>
Met-fill-6094	F	TCCTGATAGCAGTATAGTTAAAG	8315	<i>Nad5</i>
Met-fill-rev-8225	R	CTACATATTAGAGAGAACTACATC	8577	<i>Nad5</i>
Met-fill-8430	F	TAAGTCCTAACCCATCCCATCC	8713	<i>Nad5</i>
Pro-prolong-f1	F	AAAATTATAGATCCTCTAACTGG	9820	<i>Nad4</i>
Meta-polyT-nad4-long	F	TTAAATTTAACAAAGAAAAATATTTTTTTTTTT	10193	<i>Nad4</i>
Met-RO-F73	F	TTCATAGGGGTCATAATTGC	10818	<i>Nad6</i>
Meta-prolongA-R1	R	GAAATAATACTCTGAATATGGC	11031	<i>Nad6</i>
Meta-prolongA-R2	R	GAATAGACACTACCAGAGAATAC	11134	<i>Nad6</i>
Met-RO-F594	F	ATCATAACGAAAACCGAGGTA	11358	<i>Nad1</i>
Met-RO-F743	F	TAAAGTAAATCCTCCCCCTC	11520	<i>Nad1</i>
Met-nd1-F776	F	GTATTAATCCTGACACAAGCTCTGAACC	11529	<i>Nad1</i>
Met-RO-R851	R	GCTGAAACTAATCGAACACC	11597	<i>Nad1</i>
Prolong-Meta-A3-R	R	CATTAAGAATAATTGGTTAGCCTC	11629	<i>Nad1</i>
Prolong-Meta-A3-F	F	GTCCGCAAAAGGTTGAAGTATG	12014	<i>Nad1</i>
Met-RO-R1306	R	AAACAGGATTTTGAGGCATA	12052	<i>Nad1</i>
Met-RO-R1370	R	TTCTGTTCCTGTTGCTTTT	12130	<i>Nad1</i>
Met-16S-F2	F	CTATCCAAAATTATTACGCTGTTATCCC	12433	<i>rrnL</i>
Met-16S-F1	F	TTCATAAATAGTTCACCTTCATTCCAG	12691	<i>rrnL</i>
Met-16S-R2	R	TGAAAAATAGAAAAGTATAGCCTGCCC	12819	<i>rrnL</i>
Met-OS-16S-p2	F	GCCCAACTAGACTAATCTTCTACAA	13052	<i>rrnL</i>
Met-OS-16S-p1	F	TAACCCGATACAAAAGGTAATATAC	13232	<i>rrnL</i>
Met-12S-F1	F	CTAATCCTAGTCCATTTACTTATTTATC	13740	<i>rrnS</i>
Met-12S-pT	F	AAAAAAAAAAAAAAAAWAAAAAWTTYCMCC	13778	<i>rrnS</i>
Met-polyT-R1	R	GAGGGTTTTTCTATAAATGAT	13918	<i>rrnS</i>
Met-polyT-R2	R	ATCTTTTTTTTATTAGAAGTCC	13969	<i>rrnS</i>
Met-polyT2nd-F	F	CAACAATTAAGTTTTTTTTTTTTTTTWA	14027	<i>rrnS-Regió control</i>

**Taula Addicional 2.** Oligonucleòtids emprats per completar el mitogenoma sencer de *M. repens*, *M. samanensis* i *M. dominicanus*.

Nom encebador	Sentit	Seqüència	Posició	Regió
<i>M. samanensis</i>				
Sam-gap2-F1	F	AGAGATGATTGGTCTTGAGC	1517	<i>cob</i>
Sam-gap2-R1	R	TGTGAGCAACAGAAAGAGAAAG	2503	<i>Nad2</i>
Sam-gap-R2	R	CAATCATCTCTGTGAGCAAC	2514	<i>Nad2</i>
Sam-short1882-R	R	CCAGGCACCTAGCACAAAATACAAAG	3157	<i>Cox1</i>
Sam-cox1-F1	F	CTGGGCTAGGATAGTGGGTACTGC	3154	<i>Cox1</i>
Sam-cox1-R1	R	TAGCAGTACCCACTATCCTAGCCCAG	3179	<i>Cox1</i>
Sam-cox1-F4	F	ACGAGTACTGGGATAAAAATGGACC	3595	<i>Cox1</i>
Sam-cox1-R4	R	GGTCCAATTTTATCCCACTAAGCTCG	3620	<i>Cox1</i>
Sam-short-3457-R	R	CGAAGAGAGGGTAAAAGCGATTAT	4927	<i>Cox2</i>
Sam-long-2665-F	F	AAAGGTTTCGTTTATTGGAGACG	5074	<i>Cox2</i>
Sam-gap34-R	R	ATAAGCATAAGAAAATTGAAAACCTCAG	5715	<i>Atp6</i>
Sam-short-383-F	F	ATAGAAGATAGAAACCGACCTGGCTTG	12678	<i>rrnL</i>
Sam-long-3275-R	R	AGCCAGGTCGGTTTCTATCTTC	12702	<i>rrnL</i>
Sam-long-489-F	F	TTCAACATCGAGGTCGTAATCAG	12789	<i>rrnL</i>
Sam-16s-F7	F	TATCCAGAACTATTACGCTGTTATCCCTAGAG	12828	<i>rrnL</i>
Sam-16s-R7	R	CTCTAGGGATAACAGCGTAATAGTTCGGATA	12859	<i>rrnL</i>
Sam-16s-F6	F	GCTGTATCCCTAGAGTAACTTAATCCTTTCC	12844	<i>rrnL</i>
Sam-16s-R6	R	AAATAAAAAGGAAAGGATTAAGTACTCTAGGG	12844	<i>rrnL</i>
Sam-long-3603-R2	R	AGGGACGATAAGACCCTAGAAGC	13030	<i>rrnL</i>
Sam-short-879-F	F	AAATCTCACTGGGCAGGCTTAC	13179	<i>rrnL-rrnS</i>
Sam-short-176-R	R	GAAAGTAAGACCTGCCAGTGAG	13205	<i>rrnL</i>
<i>M. dominicanus</i>				
Dom-short-1520-R	R	AGCATTAGAGGAAGAAGCCAATTTC	3375	<i>Cox1</i>
Dom-short-cox1-R4	R	CCTACCCCTCTCAACTATCTCC	3483	<i>Cox1</i>
MetDom-cox1-F4	F	GAGAATAGTTGAGAGGGGGGTAGGG	3460	<i>Cox1</i>
Dom-short-2046-R	R	AGCCTCTGTCTAACCACATGAGAG	3901	<i>Cox1</i>
MetDom-cox2-F1	F	CCAAGACAAAACCATTGAGTCTATCTG	4884	<i>Cox2</i>
MetDom-cox2-R1	R	GTTCAGATAGACTCAATGGTTTGCTCTTGG	4913	<i>Cox2</i>
Dom-gap3-R	R	GGTTATGAGTCCTTATCCTTATGTG	12345	<i>Nad1</i>
Dom-short-573-F	F	TTGGTAGAAGATAGAAACCGACCTG	12713	<i>rrnL</i>
MetDom-16s-F2	F	TTATCCCTAGAGTAACTTTTTCTCCCTT	12886	<i>rrnL</i>
MetDom-16s-R2	R	AAGGGAGGAAAAAGTACTCTAGGGATAA	12914	<i>rrnL</i>
MetDom-16s-F3	F	TTGTTAGACCTTCATTCCAGTTTCA	13134	<i>rrnL</i>
MetDom-16s-R3	R	TGAAAACCTGGAATGAAGGGTCTAACAA	13160	<i>rrnL</i>
MetDom-1086-F	F	CGGGCAGGCTTACTTCTATTATC	13225	<i>rrnL</i>
Dom-AT1-F	F	AATTAGCTGACCAACTAGACTTATC	13475	<i>rrnL</i>
<i>M. repens</i>				
Rep-AT2-R	R	CAAGGCAAGACTTCCAATACAC	211	<i>cob</i>
Repens-cox1-R9	R	CAACCAGTTCCTACTCCACTCTCCAC	3056	<i>Cox1</i>
Repens-cox1-F8	F	GGTCATAGTGGTIGTTCTGTAGATTGG	3088	<i>Cox1</i>
Repens-cox1-F11	F	TTTTGTCTTACCTGTATTGGCTGG	3266	<i>Cox1</i>
Rep-Gap1-F	F	TTGTCTATGGGCTGTATTG	3820	<i>Cox1</i>
Rep-Gap2-F	F	CGCTATCTGTACGTTTATGTGC	5687	<i>Atp6</i>
Rep-Gap2-R	R	CTGCTGTGTATGAAGACCTTG	6379	<i>Cox3</i>
Rep-Gap3-F	F	TAATCCAAGACCATCTCAACC	8720	<i>Nad5</i>
Rep-Gap3-R	R	ATTGCCTATCTTCAGTGTGC	9659	<i>Nad4</i>
Repens-16s-R12	R	AATGGGATGTAAGTACTCTAGGGATA	12483	<i>rrnL</i>
Rep-AT1-F	F	GCTGGACCAAGTTATTATTTC	13883	<i>rrnS</i>



**Taula Addicional 3.** Llistat dels taxons i nombres d'accés dels mitogenomes de Crustacea i Hexapoda emprats per a l'anotació gènica i les anàlisis de divergència gènica.

Espècies	Nombre d'accés	Taxonomia
<b>CRUSTACEA</b>		
<i>Argulus americanus</i>	NC_005935	Maxillopoda Branchiura
<i>Armadillidium vulgare</i>	EF643519	Malacostraca Peracarida Isopoda
<i>Armillifer armillatus</i>	NC_005934	Pentastomida
<i>Artemia franciscana</i>	NC_001620	Branchiopoda Anostraca
<i>Callinectes sapidus</i>	NC_006281	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Cherax destructor</i>	NC_011243	Malacostraca Eucarida Decapoda Pleocyemata
<i>Daphnia pulex</i>	NC_000844	Branchiopoda Anomopoda
<i>Eriocheir sinensis</i>	NC_006992	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Euphausia superba</i>	AB084378	Malacostraca Eucarida Euphausiacea
<i>Fenneropenaeus chinensis</i>	NC_009679	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Gothelphusa dehaani</i>	NC_007379	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Gonodactylus chiragra</i>	NC_007442	Malacostraca Hoplocarida Stomatopoda
<i>Halocaridina rubra</i>	NC_008413	Malacostraca Eucarida Decapoda Pleocyemata Caridea
<i>Harpisquilla harpax</i>	NC_006916	Malacostraca Hoplocarida Stomatopoda
<i>Hutchinsoniella macracantha</i>	NC_005937	Cephalocarida
<i>Idotea baltica</i>	DQ442915	Malacostraca Peracarida Isopoda
<i>Lepeophtheirus salmonis</i>	NC_007215	Maxillopoda Copepoda
<i>Ligia oceanica</i>	NC_008412	Malacostraca Peracarida Isopoda Oniscidea
<i>Litopenaeus vannamei</i>	NC_009626	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Lysiosquillina maculata</i>	NC_007443	Malacostraca Hoplocarida Stomatopoda
<i>Macrobrachium rosenbergii</i>	NC_006880	Malacostraca Eucarida Decapoda Pleocyemata Caridea
<i>Marsupenaeus japonicus</i>	NC_007010	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Megabalanus volcano</i>	NC_006293	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Pagurus longicarpus</i>	NC_003058	Malacostraca Eucarida Decapoda Anomura
<i>Panulirus japonicus</i>	NC_004251	Malacostraca Eucarida Decapoda Pleocyemata
<i>Parhyale hawaiiensis</i>	AY639937	Malacostraca Peracarida Amphipoda
<i>Pennaeus monodon</i>	NC_002184	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Pollicipes mitella</i>	NC_008742	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Pollicipes polymerus</i>	NC_005936	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Portunus trituberculatus</i>	NC_005037	Malacostraca Eucarida Decapoda Brachyura
<i>Pseudocarcinus gigas</i>	NC_006891	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Pseudosquilla ciliata</i>	AY947836	Malacostraca Hoplocarida Stomatopoda
<i>Speleonectes tulumensis</i>	NC_005938	Remipedia
<i>Squilla empusa</i>	NC_007444	Malacostraca Hoplocarida Stomatopoda
<i>Squilla mantis</i>	NC_006081	Malacostraca Hoplocarida Stomatopoda
<i>Tetraclita japonica</i>	NC_008974	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Tigriopus californicus</i>	NC_008831	Maxillopoda Copepoda
<i>Tigriopus japonicus</i>	NC_003979	Maxillopoda Copepoda Harpacticoida
<i>Triops cancriformis</i>	NC_004465	Branchiopoda Notostraca
<i>Triops longicaudatus</i>	NC_006079	Branchiopoda Notostraca
<i>Vargula hilgendorffii</i>	NC_005306	Ostracoda Myodocopa
<i>Metacrangonyx longipes</i>	AM944817	Malacostraca Peracarida Amphipoda
<b>HEXAPODA</b>		
<i>Aleurodicus dugesii</i>	NC_005939	Insecta Hemiptera Hemimetabola
<i>Anopheles gambiae</i>	NC_002084	Insecta Diptera Holometabola
<i>Antheraea pernyi</i>	NC_004622	Insecta Lepidoptera Holometabola
<i>Apis mellifera ligustica</i>	NC_001566	Insecta Hymenoptera Holometabola
<i>Bombyx mori</i>	NC_002355	Insecta Lepidoptera Holometabola
<i>Ceratitidis capitata</i>	NC_000857	Insecta Diptera Holometabola
<i>Chrysomya putoria</i>	NC_002697	Insecta Diptera Holometabola
<i>Crioceris duodecimpunctata</i>	NC_003372	Insecta Coleoptera Holometabola
<i>Drosophila melanogaster</i>	NC_001709	Insecta Diptera Holometabola
<i>Gomphiocephalus hodgsoni</i>	NC_005438	Collembola
<i>Gryllotalpa orientalis</i>	NC_006678	Insecta Orthoptera Hemimetabola
<i>Haematobia irritans irritans</i>	NC_007102	Insecta Diptera Holometabola
<i>Heterodoxus macropus</i>	NC_002651	Insecta Phthiraptera Hemimetabola

Espècies	Nombre d'accés	Taxonomia
<b>HEXAPODA</b>		
<i>Homalodisca coagulata</i>	NC_006899	Insecta Hemiptera Hemimetabola
<i>Japyx solifugus</i>	NC_007214	Diplura
<i>lepidopsocid RS-2001</i>	NC_004816	Insecta Psocoptera Hemimetabola
<i>Locusta migratoria</i>	NC_001712	Insecta Orthoptera Hemimetabola
<i>Melipona bicolor</i>	NC_004529	Insecta Hymenoptera Holometabola
<i>Nesomachilis australica</i>	NC_006895	Insecta Archaeognatha Ametabola
<i>Onychiurus orientalis</i>	NC_006074	Collembola
<i>Orthetrum triangulare melania</i>	AB126005	Insecta Odonata Hemimetabola
<i>Ostrinia nubilalis</i>	NC_003367	Insecta Lepidoptera Holometabola
<i>Pachyptylla venusta</i>	NC_006157	Insecta Hemiptera Hemimetabola
<i>Periplaneta fuliginosa</i>	NC_006076	Insecta Dictyoptera Hemimetabola
<i>Philaenus spumarius</i>	NC_005944	Insecta Hemiptera Hemimetabola
<i>Podura aquatica</i>	NC_006075	Collembola
<i>Pteronarcys princeps</i>	NC_006133	Insecta Plecoptera Hemimetabola
<i>Pyrocoelia rufa</i>	NC_003970	Insecta Coleoptera Holometabola
<i>Schizaphis graminum</i>	NC_006158	Insecta Hemiptera Hemimetabola
<i>Thermobia domestica</i>	NC_006080	Insecta Thysanura Ametabola
<i>Thrips imaginis</i>	NC_004371	Insecta Thysanoptera Hemimetabola
<i>Triatoma dimidiata</i>	NC_002609	Insecta Hemiptera Hemimetabola
<i>Tribolium castaneum</i>	NC_003081	Insecta Coleoptera Holometabola
<i>Tricholepidion gertschi</i>	NC_005437	Insecta Thysanura Ametabola
<i>Xenos vesparum</i>	DQ364229	Strepsiptera Holometabola

Taula Addicional 4. Mitogenomes de peracàrids emprats per a l'anotació dels MPCGs de *Metacrangonyctidae*.

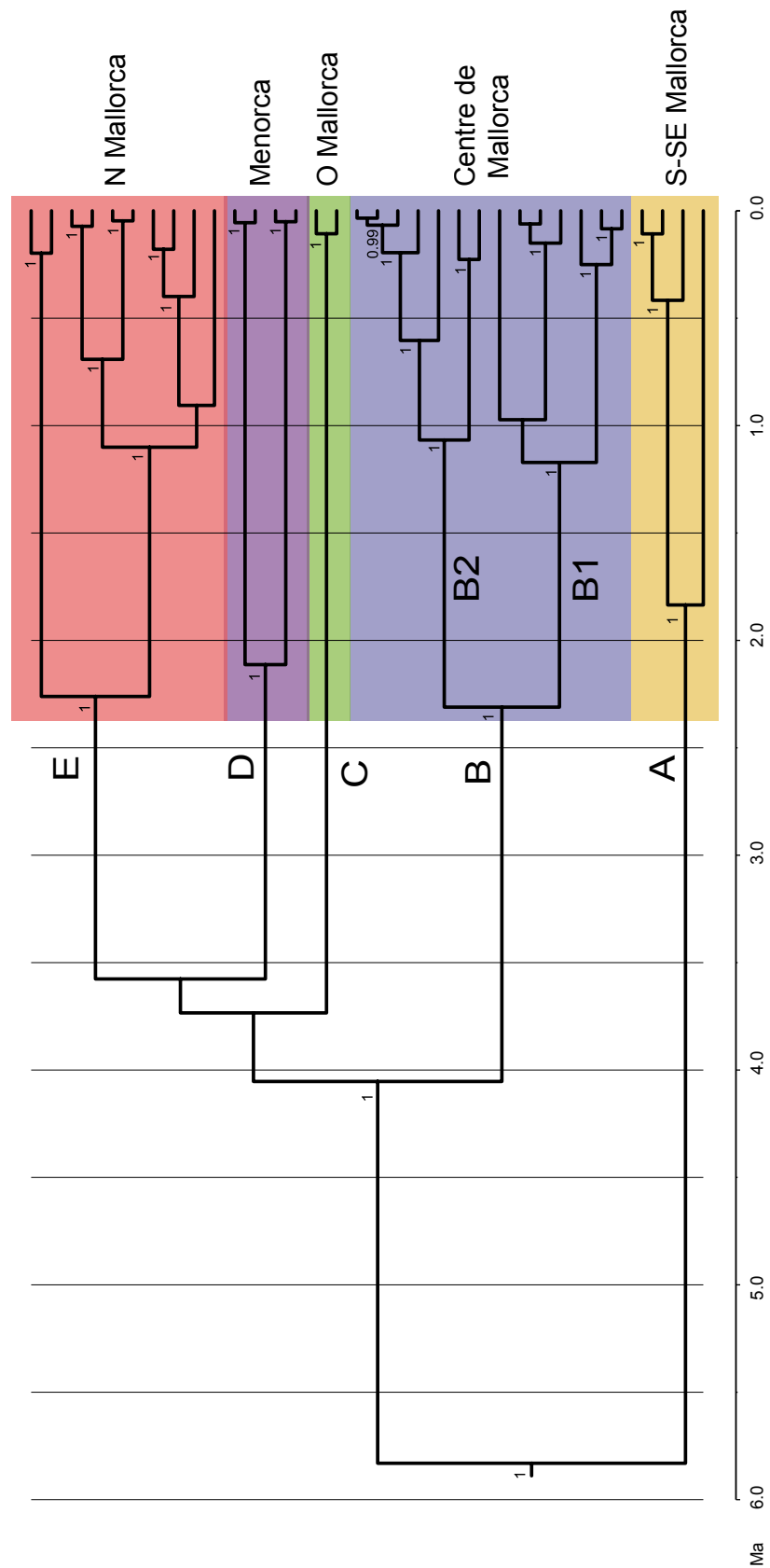
Espècie	Nombre d'accés	Taxonomia	Seqüenciació mitogenoma	AT biaix	GC biaix
<i>Caprella scaura</i>	AB539699	Amphipoda; Caprelliidae; Caprellida; Caprellidae; Caprella	complet	-0.02	-0.13
<i>Caprella mutica</i>	GU130250	Amphipoda; Caprelliidae; Caprellida; Caprellidae; Caprella	complet	-0.02	-0.17
<i>Parhyale hawaiiensis</i>	AY639937/FM1957525/ FM1957526	Amphipoda; Gammaridae; Talitroidea; Hyalidae; Parhyale	parcial	-0.02	-0.12
<i>Gondogeneia anatarctica</i>	JN827386	Amphipoda; Gammaridae; Eusiroidea; Pontogeneiidae; Gondogeneia	complet	-0.01	-0.29
<i>Onisimus nanseni</i>	FJ555185	Amphipoda; Gammaridae; Lysianassoidea; Lysianassidae; Onisimus	complet	0.00	-0.20
<i>Eophtreotocus sp.</i>	FJ790313	Isopoda; Phreatoicidae; Phreatoicidae; Eophtreotocus.	complet	-0.10	0.25
<i>Asellus aquaticus</i>	GU130252	Isopoda; Asellota; Aselloidea; Asellidae; Asellus	parcial	0.00	-0.12
<i>Eurydice pulchra</i>	GU130253	Isopoda; Cirolanidae; Eurydice	parcial	-0.05	0.20
<i>Glyptonotus cf. antarcticus</i>	GU130254	Isopoda; Valvifera; Chaetiliidae; Glyptonotus	parcial	-0.03	0.04
<i>Janira maculosa</i>	GU130255	Isopoda; Asellota; Janiroidea; Janiridae; Janira	parcial	-0.04	-0.03
<i>Sphaeroma serratum</i>	GU130256	Isopoda; Sphaeromatidae; Sphaeroma	complet	-0.07	0.22
<i>Ligia oceanica</i>	NC_008412	Isopoda; Oniscidea; Diplocheira; Ligidae; Ligia	complet	-0.04	0.13
<i>Armadillidium vulgare</i>	EF643519	Isopoda; Oniscidea; Crinocheira; Armadillidiidae;	parcial	0.13	-0.22
<i>Idotea baltica</i>	DQ442915	Armadillidium Isopoda; Valvifera; Idoteidae; Idotea	parcial	-0.08	0.16

Taula Addicional 5. Freqüències nucleotídiques, valors de biaix per a AT i GC per posició del codó del mitogenoma de *M. longipes*.

Gen	%A1	%T1	%C1	%G1	A1T1 biaix	G1C1 biaix	%A2	%T2	%C2	%G2	A2T2 biaix	G2C2 biaix	%A3	%T3	%C3	%G3	A3T3 biaix	G3C3 biaix
<i>cob</i>	32,98	35,09	14,25	17,68	-0,03	0,11	24,01	43,80	21,90	10,29	-0,29	-0,36	33,51	50,92	11,35	-0,19	-0,21	-0,46
<i>nad2</i>	35,95	40,19	7,55	16,31	-0,06	0,37	17,58	56,36	12,12	13,94	-0,52	0,07	41,21	45,46	5,45	0,21	-0,05	0,18
<i>cox1</i>	31,12	30,16	13,62	25,10	0,02	0,30	17,51	44,16	21,01	17,32	-0,43	-0,10	36,38	50,78	6,81	0,08	-0,17	-0,06
<i>cox2</i>	31,00	34,06	14,85	20,09	-0,05	0,15	26,32	42,53	17,11	14,04	-0,24	-0,10	38,16	49,12	5,70	0,04	-0,13	0,10
<i>atp8</i>	39,62	39,62	11,33	9,43	0,00	-0,09	33,96	47,17	13,21	5,66	-0,16	-0,40	54,71	41,51	1,89	-0,22	0,14	0,00
<i>atp6</i>	35,43	38,11	11,21	15,25	-0,04	0,15	16,59	50,67	17,94	14,80	-0,51	-0,10	39,91	47,08	6,28	0,02	-0,08	0,03
<i>cox3</i>	28,52	32,70	15,59	23,19	-0,07	0,20	21,29	44,11	17,49	17,11	-0,35	-0,01	30,80	52,08	8,75	0,08	-0,26	-0,02
<i>nad3</i>	33,90	33,90	9,32	22,88	0,00	0,42	22,03	53,39	13,56	11,02	-0,42	-0,10	33,05	51,70	5,08	0,22	-0,22	0,33
<i>nad5</i>	39,27	35,95	8,90	15,88	0,04	0,28	21,64	47,47	15,53	15,36	-0,37	-0,01	41,19	44,67	6,46	0,11	-0,04	0,09
<i>nad4</i>	34,62	38,27	8,66	18,45	-0,05	0,36	20,55	49,77	13,01	16,67	-0,42	0,12	41,55	47,04	5,25	0,21	-0,06	0,08
<i>nad4L</i>	42,11	37,89	7,37	12,63	0,05	0,26	23,16	53,69	5,26	17,89	-0,40	0,55	45,26	44,21	4,21	0,37	0,01	0,20
<i>nad6</i>	36,09	42,61	9,47	11,83	-0,08	0,11	21,89	54,45	15,38	8,28	-0,43	-0,30	30,18	55,62	10,65	-0,20	-0,30	-0,50
<i>nad1</i>	34,20	36,16	10,10	19,54	-0,03	0,32	19,54	48,54	14,33	17,59	-0,43	0,10	42,67	46,58	3,91	0,22	-0,04	0,27
Tots els PCGs	34,47	35,85	11,08	18,60	-0,02	0,25	20,84	48,10	16,26	14,80	-0,40	-0,05	38,43	48,29	6,69	0,08	-0,11	-0,01

**Taula 3.6.** Nombre d'aminoàcids a cadascun dels MPCGs, freqüències nucleotídiques, contingut en A+T i valors de biaix per a AT i GC als gens del mitogenoma de *M.longipes*

Gen	Cadena	Mida (aa)	%A	%C	%G	%C	% A+T	AT biaix	GC biaix
<i>cob</i>	-	378	30.17	15.83	10.73	43.27	73.44	-0.18	-0.19
<i>nad2</i>	+	330	31.58	8.38	12.71	47.33	78.91	-0.20	0.21
<i>cox1</i>	+	513	28.34	13.81	16.15	41.7	70.04	-0.19	0.08
<i>cox2</i>	+	228	31.83	12.55	13.72	41.9	73.73	-0.14	0.04
<i>atp8</i>	+	52	42.77	8.8	5.66	42.77	85.54	0.00	-0.22
<i>atp6</i>	+	222	30.64	11.81	12.26	45.29	75.93	-0.19	0.02
<i>cox3</i>	+	262	26.87	13.94	16.22	42.97	69.84	-0.23	0.08
<i>nad3</i>	+	117	29.66	9.32	14.69	46.33	75.99	-0.22	0.22
<i>nad5</i>	-	572	34.03	10.3	12.97	42.7	76.73	-0.11	0.11
<i>nad4</i>	-	438	32.24	8.97	13.76	45.03	77.27	-0.17	0.21
<i>nad4L</i>	-	94	36.84	5.61	12.28	45.27	82.11	-0.10	0.37
<i>nad6</i>	+	168	29.39	11.83	7.89	50.89	80.28	-0.27	-0.20
<i>nad1</i>	-	306	32.14	9.45	14.66	43.76	75.9	-0.15	0.22
<i>rrnL</i>	-	n/a	39.93	7.92	13.81	38.34	78.27	0.02	0.27
<i>rrnS</i>	-	n/a	35.83	7.63	12.66	43.88	79.71	-0.10	0.25
Regió control	+	n/a	40.79	7.89	7.89	43.43	84.22	-0.03	0.00
Tots els PCGs	n/a	n/a	31.25	11.34	13.33	44.08	75.33	-0.17	0.08
Tots els tRNAs	n/a	n/a	40.95	9.01	12.98	37.06	78.01	0.05	0.18
Genoma complet	+	n/a	37.36	12.41	11.56	38.67	76.03	-0.02	-0.04



**Figura Addicional 1.** Cronograma basat en l'arbre ultramètric bayesià amb el set de dades mitocondrials combinat (cox1 i cob) de *Metacrangonyx longipes* obtingut a partir d'un rellotge relaxat no correlacionat amb distribució log-normal i baix un model d'especiació de Yule. La datació dels clades principals es realitzà baix l'assumpció d'una taxa fixa de substitució del 2.3% per parell de bases per milió d'anys.

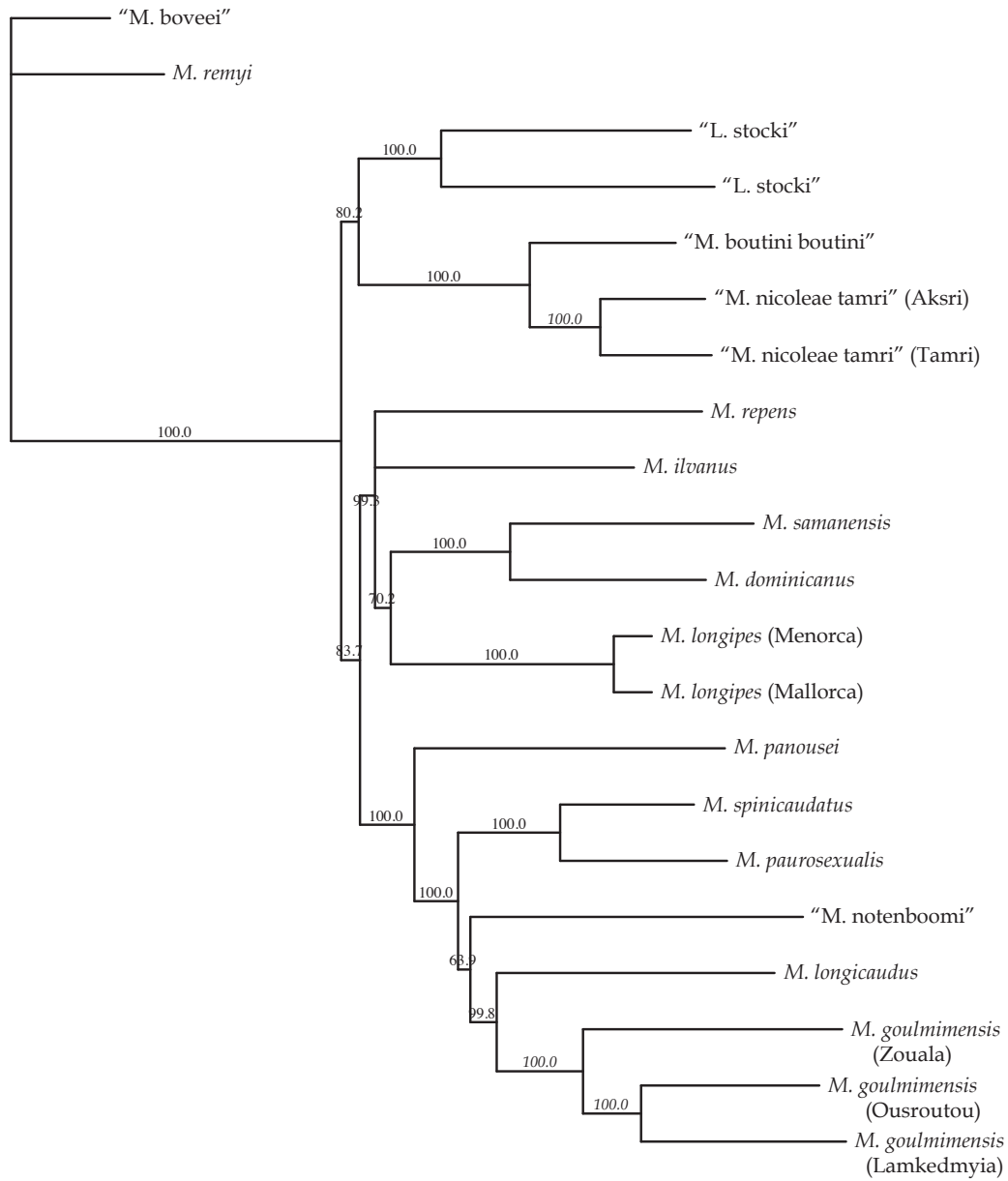
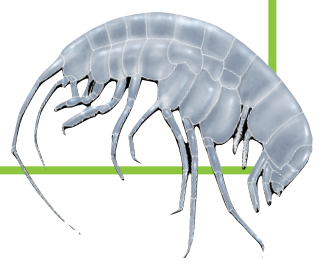


Figura Addicional 2. Arbre consens obtingut per la regla de la majoria del 50% amb els valors de probabilitat posterior obtinguts amb Phycas.





**PUBLICACIONES**





# The complete mitochondrial genome of the subterranean crustacean *Metacrangonyx longipes* (Amphipoda): A unique gene order and extremely short control region

MARIA M. BAUZÀ-RIBOT<sup>1</sup>, DAMIÀ JAUME<sup>2</sup>, CARLOS JUAN<sup>1</sup>, & JOAN PONS<sup>2</sup>

<sup>1</sup>Departament de Biologia, Universitat de les Illes Balears, Palma de Mallorca, Balearic Islands, Spain, and

<sup>2</sup>IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, Esporles, Balearic Islands, Spain

(Received 12 November 2008; revised 3 March 2009; accepted 13 April 2009)

## Abstract

Metazoan mitochondrial genomes usually consist of the same gene set, but some taxonomic groups show a considerable variety in gene order and nucleotide composition. The mitochondrial genomes of 37 crustaceans are currently known. Within the malacostracan superorder Peracarida, only three partial mitogenome sequences and the complete sequence of *Ligia oceanica* (Isopoda) are available. Frequent translocation events have changed the mitochondrial gene order in crustaceans, providing an opportunity to study the patterns and mechanisms of mitogenome rearrangement and to determine their impact on phylogenetic reconstructions. Here we report the first complete nucleotide sequence of an amphipod species, *Metacrangonyx longipes*, belonging to a phylogenetically enigmatic family occurring in continental subterranean waters. The genome has 14,113 bp and contains the usual 13 protein coding genes and two rRNA subunits, but only 21 out of the typical 22 tRNA genes of Metazoa. This is the shortest mitogenome described thus far for a crustacean and also one of the richest in AT (76.03%). The genome compactness results from a very small control region of 76 bp, the occurrence of frequent gene overlap, and the absence of large non-coding fragments. Six of the protein-coding genes have unusual start codons. Comparison of individual protein coding genes with the sequences known for other crustaceans suggests that *nad2*, *nad6*, *nad4L* and *atp8* show the highest divergence rates. *M. longipes* shows a unique crustacean mitogenome gene order, differing even from the condition found in *Parhyale hawaiiensis* (Amphipoda), whose coding sequence has also been completed in the present study.

**Keywords:** *Metacrangonyx longipes*, *Amphipoda*, *mitochondrial genome*, *control region*, *gene order*

## Introduction

The mitochondrial genome (mitogenome) of metazoans generally comprises a circular double-stranded DNA molecule of 12–20 kb with a highly conserved gene content. It includes 13 protein-coding, two ribosomal and up to 22 transfer RNA genes (Wolstenholme 1992). The Crustacea have more than 52,000 described species, with a range in body plan not matched in any other group of metazoans (Martin and Davis 2001). They include the six recognized classes: Branchiopoda, Cephalocarida, Malacostraca, Maxillopoda, Ostracoda and Remipedia (Martin and Davis 2001). The mitogenome sequences of 37 species

of Crustacea have been completed thus far (<http://www.ncbi.nlm.nih.gov/genomes>), of which 15 correspond to malacostracan decapods (Carapelli et al. 2007; Yang and Yang 2008). Within the malacostracan peracarid order Amphipoda, only a partial mitogenome sequence is currently available in sequence databases: that of *Parhyale hawaiiensis*, although it lacks of about 3 kb including the *rrnS* gene and parts of *rrnL* and *nad2*, and also the control region (Cook et al. 2005). In addition, in the peracarid order Isopoda, one entire (*Ligia oceanica*) and two partial mitogenomes (*Armadillidium vulgare* and *Idotea baltica*) are known (Kilpert and Podsiadlowski 2006;

Correspondence: J. Pons, IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, Miquel Marqués 21, Esporles, 07190 Balearic Islands, Spain. Tel: 34 971 611332. Fax: 34 971 611761. E-mail: joan.ponspons@uib.es

Podsiadlowski and Bartolomaeus 2006; Marcadé et al. 2007). The taxon sampling set for crustacean mitogenomes is quite poor because only 30 out of about 800 known crustacean families are represented (Martin and Davis 2001). Despite this, two important insights into pancrustacean phylogenetics are based on mitogenome data. First, phylogenetic analyses of protein-coding genes (PCGs) including the more intensively sampled mitochondrial genomes of Hexapoda suggest a mutual paraphyly of Crustacea and Hexapoda (Cook et al. 2005; Carapelli et al. 2007). Second, frequent translocation events have apparently changed the mitochondrial gene order in crustaceans compared with the putative ancestral pattern (Kilpert and Podsiadlowski 2006; Yang and Yang 2008). This gene order results from a common inversion of a *trnL2* gene present in a large number of crustaceans and insects, that translocated from a location inferred to be the primitive state as it is found in chelicerates, myriapods, onychophorans, tardigrades, as well as in Pogonophora, Annelida, Echiura, and Mollusca (Boore et al. 1995, 1998). Gene order is not conserved within the superorder Peracarida (for which only information on Isopoda and Amphipoda is currently available), nor is it even conserved within the Isopoda. Despite those differences, the mitogenome of the isopods *L. oceanica*, *I. baltica* and *A. vulgare* share some gene rearrangements (i.e. putative isopod synapomorphies), compared with the arthropod pattern and that of the amphipod *P. hawaiiensis* (Kilpert and Podsiadlowski 2006).

The Metacrangonyctidae (Boutin and Messouli 1988) represent a small family of amphipod crustaceans with two genera: *Metacrangonyx* (Chevreux 1909) (17 species) and *Longipodacrangonyx* (Boutin and Messouli 1988) (monotypic). All members of the family occur only in continental subterranean waters and represent a phylogenetically enigmatic lineage of marine origin showing an extremely disjunct geographic distribution. Two species are found in the Dominican Republic (Hispaniola) (Jaume and Christenson 2001), one is known from Fuerteventura in the Canary Islands (Stock and Rondé-Broekhuizen 1986), 11 from Morocco (Balazuc and Ruffo 1953; Ruffo 1954; Karaman and Pesce 1979; Boutin and Messouli 1988a,b; Messouli et al. 1991; Oulbaz et al. 1998), one from Elba Island, Italy (Stoch 1997), one from the Balearic Islands (Chevreux 1909; Margalef 1952) and two from the Middle East (Ruffo 1982; Karaman 1989). Whereas most species live in interstitial freshwater associated with springs, wells or alluvial sediments, some taxa occur in brackish or athalassohaline waters. Only *Metacrangonyx longipes* (Chevreux 1909) from the Balearics and the two Hispaniolan species are ordinary cave dwellers, living in fresh to marine subterranean waters (Jaume and Christenson 2001).

It has been proposed that the Metacrangonyctidae derive from marine littoral ancestors that colonized

the continental ground waters during episodes of marine regression (Boutin and Coineau 1990). Although first supposed to be no older than the opening of the northern Atlantic ocean (Boutin 1994), the discovery of *Metacrangonyx* in the Greater Antilles (Jaume and Christenson 2001) suggests a much older origin for the genus: at least before the opening of the northern Atlantic (110 million years before present). Thus its current distribution would be the result of vicariance by plate tectonics and of peripatric speciation associated with episodes of regression in the paleocoastline of Tethys.

We present here the first complete sequence of a mitochondrial genome of an amphipod. We have used the mitogenome of *M. longipes* to compare its gene order with those of other crustaceans, as well as its nucleotide composition and tRNA structure. We especially focus in comparisons on other peracarids such as the amphipod *P. hawaiiensis*, for which we have almost completed the whole mitogenome (except approximately 500 bp of the control region that has not been sequenced because of technical problems), and the isopods *L. oceanica*, *I. baltica* and *A. vulgare*.

## Materials and methods

### Sampling and DNA extraction

A 3 mm long specimen of *M. longipes* preserved in absolute ethanol was used for DNA extraction by means of the DNeasy Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol for total genomic DNA purification. The specimen was collected in Cala Varques cave (Mallorca Island, Spain) during fall 2007.

### PCR primers and conditions

Gene fragments at opposing ends of the mitochondrial genome were amplified using standard protocols outlined elsewhere (Balke et al. 2005) and universal primers (Table I). Based on the sequence obtained, we designed species-specific long primers (Table I) of about 25–29 bp targeted at the *cox1/rrnL* genes to amplify two long fragments of about 4.5 and 10 kb covering the whole circular mitochondrial genome. Long-range PCR amplifications were performed using TaKaRa LA *Taq* polymerase (Takara Bio, Inc., Tokyo, Japan) according to the manufacturer's specifications. The general reaction mixture for each 50 µl was: 5 µl of 10 × LA PCR buffer, 5 µl of 25 mM MgCl<sub>2</sub>, 8 µl of dNTP mixture (2.5 mM each), 2.5 µl forward primer (10 µM), 2.5 µl reverse primer (10 µM), 0.5 µl Takara LA *Taq* (5 U/µl), 24.5 µl distilled H<sub>2</sub>O and 2 µl genomic DNA. PCR cycles were as follows: after an initial denaturation step of 94°C for 90 s, 14 cycles were performed at 94°C for 30 s, 57–62°C (depending on primers) for 30 s and 68°C for 5–15 min depending

Table I. Universal and long PCR primers designed to amplify the mitochondrial fragments of *M. longipes*.

Primer name	Forward/reverse	Gene	5' → 3' sequence	Reference
16Sa (aka M14)	Forward	<i>rrnL</i>	CGCCTGTTTAAACAAAAACAT	Xiong and Kocher (1991)
16Sb (aka M74)	Reverse	<i>rrnL</i>	CTCCGGTTTGAACCTCAGATCA	Xiong and Kocher (1991)
CB3	Forward	<i>cob</i>	GAGGAGCAACTGTAATTACTAA	Barraclough et al. (1999)
CB4	Reverse	<i>cob</i>	AAAAGAAARTATCATTGAGGTTGAAT	Barraclough et al. (1999)
HCO2198	Reverse	<i>cox1</i>	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
LCO1490	Forward	<i>cox1</i>	GGTCAACAAATCATAAAGATATTGG	Folmer et al. (1994)
Metcox1_F2	Forward	<i>cox1</i>	TATACGAGTTGGGATAATAGGAATAGACC	Present study
Metcox1_F4	Forward	<i>cox1</i>	TAACTGACCGAAACCTAAATACCTCTT	Present study
Met16s_R2	Reverse	<i>rrnL</i>	TGAAAAATAGAAAGTATAGCCTGCC	Present study
Met16S_F1	Forward	<i>rrnL</i>	TTCATAAATAGTTCTACCTTTCATTCCAG	Present study
Metcox1_R1	Reverse	<i>cox1</i>	CCCTTATAGCAGTTCCTACTATCCTAGC	Present study
Met16S_F2	Forward	<i>rrnL</i>	CTATCCAAAATATTACGCTGTATCCC	Present study
Metcox1_R2	Reverse	<i>cox1</i>	GAAGAAGTCAGTTACCGAACCTCC	Present study

on the expected fragment size. This was followed by 16 cycles at 94°C for 30 s, 57–62°C for 30 s and 68°C for 5–15 min (increasing by 15 s each cycle) and a final extension for 10 min at 72°C.

#### Cloning and sequencing

Long mitochondrial fragments were digested independently with *Dra*I, *Rsa*I and *Taq*I restriction enzymes according to the manufacturer's specifications. DNA digestions showed fragments ranging from 150 pb to 1.5 kb when checked on 2% agarose gels. DNA fragments from the three digestions were pooled and purified using the MinElute PCR Purification Kit (Qiagen) and were then cloned into a pJET blunt cloning vector (Fermentas, Glen Burnie, MD, USA) according to the specifications of the manufacturer. One-shot competent *Escherichia coli* cells from Invitrogen (Madison, WI, USA) were used for transformation. Ninety-six recombinant colonies were screened by PCR amplification for inserts of a minimum of 300 bp, and 63 were sequenced in both directions using the pJET vector sequencing primers. Sequences obtained from clones were then used to design specific primers to sequence the long PCR fragments directly by primer walking (list of primers available upon request) to obtain a full contig of the mitogenome. Additional primers were designed to close particular gaps in the sequence. The forward and reverse strands of small PCR amplicons or the long PCR fragments were cycle-sequenced using the ABI BigDye Terminator Cycle Sequencing Kit on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

#### Gene annotation and sequence analysis

Analyses of the quality of chromatograms and contig construction to obtain the whole mitochondrial sequence were performed with the software CodonCode Aligner v2.0 (CodonCode Corp., Denham, MA, USA). Ambiguous nucleotide positions

were validated by direct inspection of the chromatograms. Preliminary gene identification was determined by BLAST searching on GenBank databases (<http://www.ncbi.nlm.nih.gov>) and making multiple alignments to other crustacean nucleotide and amino acid sequences (see Additional File 1 for a list of species and accession numbers). Definitive annotations were performed using the DOGMA webserver (Dual Organellar GenoMe Annotator; <http://www.bugmaster.jgi-psf.org/dogma>). The 5' and 3' ends of protein and ribosomal genes were refined manually by comparison with the complete genes of other crustaceans. Transfer RNA genes were determined with tRNAscan-SE Search Server v1.21 (<http://www.lowelab.ucsc.edu/tRNAscan-SE>) using a tRNA covariance model (Lowe and Eddy 1997) and by inspection of anti-codon sequences and the predicted secondary structures. Nucleotide frequencies of protein coding and RNA genes were calculated with the DAMBE software package (Xia and Xie 2001), while the effective number of codons was determined according to INCA v1.20 (Supek and Vlahovicek 2004).

#### Divergence in protein coding genes

Mean nucleotide divergences of individual PCGs were estimated from pairwise comparisons among the complete mitogenomes of crustaceans and were subsequently compared with the values obtained for 35 species representing all major Hexapoda orders for which there are data available. MEGA v4.0.2 (Tamura et al. 2007) was used to calculate corrected distances using the Maximum Composite Likelihood model (Tamura et al. 2004) and among-sites rate variation following a gamma distribution with a shape parameter of 0.4 as estimated in RAxML v7.0.4 (Stamatakis 2006). Gapped positions were not considered in the analysis of each pairwise comparison. Mean divergence values were normalized by dividing the value obtained for each gene by the value of the gene with the highest rate.

### Gene rearrangement analyses

We used the program CREx (Bernt et al. 2007) to deduce gene rearrangement scenarios in crustacean mitogenomes based on the detection of strong interval trees on the CREx webserver (<http://www.pacosy.informatik.uni-leipzig.de/crex>). The strong interval trees reflect genes that appear consecutively in several of the input gene orders; that is, given two gene orders, a set of genes is a common interval if the genes in that set appear consecutively in both gene orders. A certain subset of all common intervals, the “strong common intervals”, can be represented as the nodes of a special type of tree. The descendants of a node (strong common interval) are simply the strong common intervals that it includes entirely. If the descendants of a node appear in the same order in both input gene orders, the node is called “linear increasing” (+); if the children of a node appear in exactly the opposite order, it is “linear decreasing” (−); otherwise, the node is called prime (Bernt et al. 2007).

## Results and discussion

### Genome organization

The mitochondrial sequence of *M. longipes* has an overall length of 14,113 bp (EMBL accession number: AM944817) and shows the usual circular organization found in most metazoans (Figure 1). To our

knowledge, this is the smallest mitogenome described so far for a crustacean: close to that of *Tigriopus californicus* (Copepoda, Harpacticoida, Maxillopoda; 14,546 bp) (Machida et al. 2002). Gene annotation reveals the presence of the typical 13 PCGs and the 2 rRNA subunits of metazoan mitochondrial genomes (Table II), but only 21 tRNA genes instead of the typical 22; this is similar to the condition found in the isopod *L. oceanica* (Kilpert and Podsiadlowski 2006). The compactness of the genome is due to the occurrence of frequent gene overlap, since more than 20 genes share borders. These overlapping regions range in size from just 1 bp (several cases) to a maximum of 63 bp (in the gene coding for tRNA-Val, which overlaps with 44 bp of the 5' end of *rrnL* and with 19 bp of the 3' end of *rrnS*). Small non-coding sequences or intergenic spacers (range 1–17 bp; see Table II) are also evident in the mitochondrial DNA (mtDNA). A further region of non-coding DNA comprising 76 bp, placed between the *rrnS* and *cob* genes and with an AT content of 84.22% presumably corresponds to the control region and contains the origin of mtDNA replication. The region has a putative secondary structure folding into a hairpin, with a stem of 15 paired nucleotides plus a short loop of four nucleotides (Figure 2). This is similar to other stem-loop structures known to occur in insect mitochondrial control regions (Zhang et al. 1995) and that are presumed to be the origin of replication of mtDNA. The 3'-flanking sequence around the stem region shows

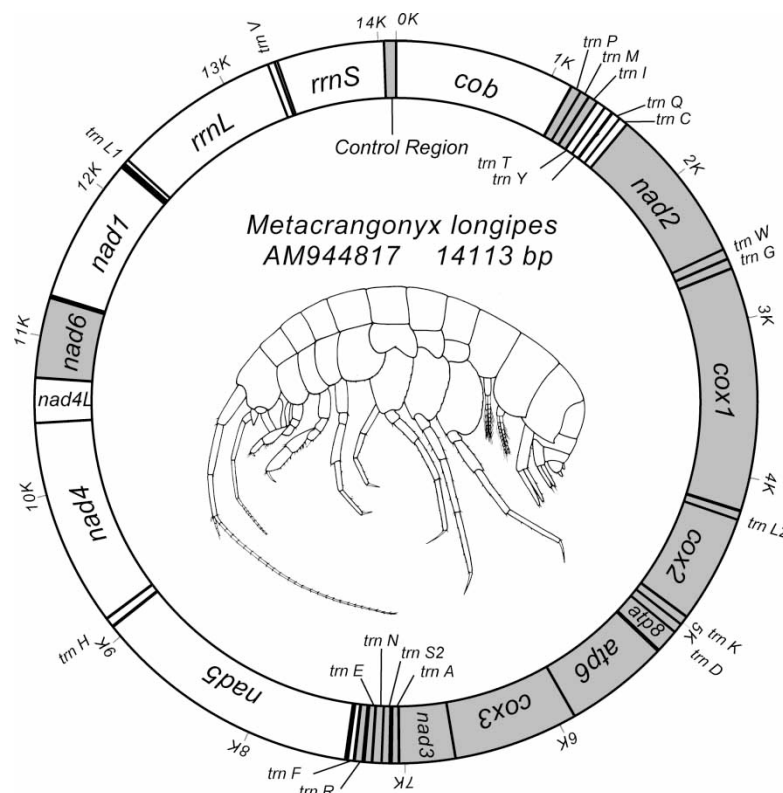


Figure 1. Map of the mitochondrial genome of *M. longipes*. Note: Grey and white segments indicate genes coded on the + strand and − strand, respectively.



Table II. Gene content of the mitochondrial genome of *M. longipes*.

Gene	Position	Strand	Size (bp)	Start codon	Stop codon
<i>cob</i>	1–1137	–	1137	ATG	TAA
Intergenic spacer 1	1138–1140	+	3	n.a.	n.a.
<i>trnP</i>	1141–1204	+	64	n.a.	n.a.
<i>trnM</i>	1205–1265	+	61	n.a.	n.a.
<i>trnI</i>	1264–1326	+	63	n.a.	n.a.
<i>trnT</i>	1325–1383	–	59	n.a.	n.a.
Intergenic spacer 2	1384–1384	+	1	n.a.	n.a.
<i>trnY</i>	1385–1443	–	59	n.a.	n.a.
<i>trnQ</i>	1439–1498	–	60	n.a.	n.a.
<i>trnC</i>	1499–1556	–	58	n.a.	n.a.
<i>nad2</i>	1556–2546	+	991	ATT	T
<i>trnW</i>	2547–2610	+	64	n.a.	n.a.
Intergenic spacer 3	2611–2612	+	2	n.a.	n.a.
<i>trnG</i>	2613–2674	+	62	n.a.	n.a.
<i>cox1</i>	2677–4218	+	1542	ATT	TAA
<i>trnL1</i>	4214–4273	+	60	n.a.	n.a.
<i>cox2</i>	4274–4958	+	685	ATT	T
<i>trnK</i>	4957–5016	+	60	n.a.	n.a.
<i>trnD</i>	5017–5077	+	61	n.a.	n.a.
<i>atp8</i>	5078–5236	+	159	ATC	TAA
<i>atp6</i>	5227–5895	+	669	ATG	TAA
<i>cox3</i>	5895–6683	+	789	ATG	TAA
Intergenic spacer 4	6684–6684	+	1	n.a.	n.a.
<i>nad3</i>	6685–7038	+	354	ATT	TAG
<i>trnA</i>	7037–7097	+	61	n.a.	n.a.
<i>trnS2</i>	7084–7146	+	63	n.a.	n.a.
<i>trnN</i>	7147–7207	+	61	n.a.	n.a.
<i>trnE</i>	7205–7265	+	61	n.a.	n.a.
<i>trnR</i>	7254–7320	+	67	n.a.	n.a.
<i>trnF</i>	7315–7374	–	60	n.a.	n.a.
<i>nad5</i>	7358–9076	–	1719	ATT	TAG
Intergenic spacer 5	9077–9082	+	6	n.a.	n.a.
<i>trnH</i>	9083–9142	–	60	n.a.	n.a.
<i>nad4</i>	9140–10,454	–	1315	ATA	T
<i>nad4L</i>	10,454–10,738	–	285	ATA	TAA
<i>nad6</i>	10,737–11,243	+	507	ATA	TAA
Intergenic spacer 6	11,244–11,260	+	17	n.a.	n.a.
<i>nad1</i>	11,261–12,181	–	921	ATA	TAA
<i>trnL2</i>	12,176–12,235	–	60	n.a.	n.a.
<i>rrnL</i>	12,206–13,342	–	1137	n.a.	n.a.
<i>trnV</i>	13,299–13,361	–	63	n.a.	n.a.
<i>rrnS</i>	13,343–14,037	–	695	n.a.	n.a.
Control region	14,038–14,113	+	76	n.a.	n.a.
All protein coding genes	n.a.	n.a.	11,073	n.a.	n.a.
All tRNAs	n.a.	n.a.	1287	n.a.	n.a.
Complete genome	1–14,113	+	14,113	n.a.	n.a.

the conserved motif GACT present also in the isopod *L. oceanica* and the hoplocarid malacostracan *Squilla mantis* (Kilpert and Podsiadlowski 2006), but the TATA element found in many hexapods at the 5'-flanking region is here replaced by an AATT motif. The low level of non-coding sequences found in the mitogenome of *M. longipes* (< 1%) and the occurrence of frequent gene overlap are indicative of an extremely compact mitogenome.

*Protein coding genes: Nucleotide composition and codon usage*

The AT content of the protein genes of *M. longipes* is 75.33% (A = 31.25%, C = 11.34%, G = 13.33%

and T = 44.08%), while that of the complete mitogenome (+ strand) is 76.03%; this is one of the highest percentages reported in crustaceans and similar to those frequently found in Hexapoda mitochondrial genomes. *Argulus americanus* (Branchiura, Maxillopoda) has the highest AT content found so far in any crustacean at 77.80% (Machida et al. 2002), while the nearly complete amphipod mitochondrial sequence of *P. hawaiiensis* reaches 73.66% (Cook et al. 2005) (authors' own data).

Six of the 13 PCGs of *M. longipes* display unusual start codons for an arthropod mtDNA. The codon ATT is present in five genes (Table II), including *cox1*,

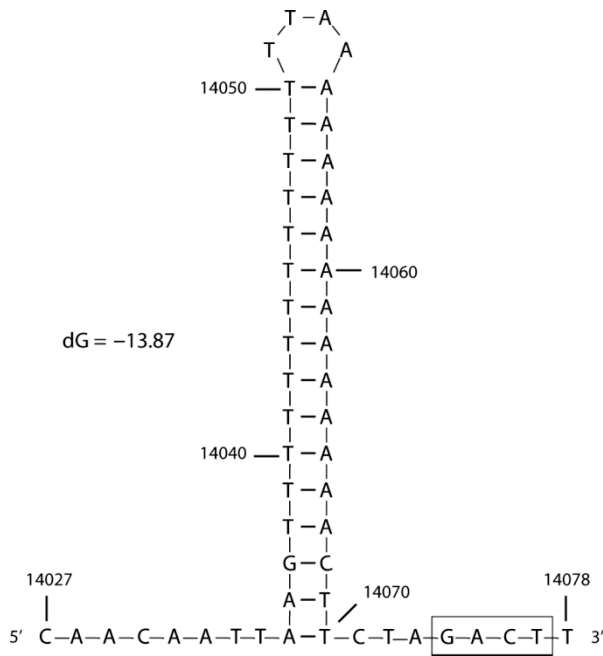


Figure 2. Putative secondary structure of the mitochondrial control region of *M. longipes*. Note: The box indicates the conserved GACT motif present also in the isopod *L. oceanica* and the hoplocarid malacostracan *S. mantis*.

which starts with ACG in other malacostracans (Kilpert and Podsiadlowski 2006 and references therein). In addition, the gene *atp8* starts with the non-canonical codon ATC (Table II). In turn, three of the PCGs show truncated stop codons (Table II). The genes for *nad2*, *nad4* and *cox2* end in a single T. As shown elsewhere, these truncated stop codons are likely to be completed by post-transcriptional polyadenylation, with final transcripts having functional UAA terminal codons (Ojala et al. 1981).

The *M. longipes* mitogenome shows a clear bias in nucleotide frequencies, with similar values in both strands (Table III). Strand bias reflected by GC skew (Perna and Kocher 1995) is slightly negative but close to zero in the genes encoded by the + strand, in contrast to the peracarid isopods studied so far, which show positive values (Kilpert and Podsiadlowski 2006). This has been attributed to the occurrence of inversions of the control region containing the replication origin in *L. oceanica* and *I. baltica*, since most crustaceans show moderate to high negative GC skews in the + strand (Hassanin et al. 2005; Hassanin 2006; Kilpert and Podsiadlowski 2006).

The effective number of codons was calculated for the PCGs. This is a simple measure of codon usage, ranging from 20 when only one codon is used for each amino acid, to 61 (or even 62 in the invertebrate mitochondrial genetic codes since UGA codes there for tryptophan) when all synonymous codons are equally in use (Wright 1990). In the *M. longipes* mitogenome the PCGs show low effective number of codons values ( $35.38 \pm 2.84$ ), so they use about one-half of the possible codons. There is a positive correlation between effective number of codons values and GC content in third codon positions ( $r^2 = 0.464$ ;  $P < 0.01$ ), as described elsewhere (Kilpert and Podsiadlowski 2006). However, the genes *nad2*, *nad3* and *nad6* use a lower number of effective codons than could be expected from their relatively high GC content at third codon positions (18% for *nad2*, and 23% for both *nad3* and *nad6*). Compared with isopods, *M. longipes* displays a considerably lower mean number of effective codons (and hence lower GC content at synonymous sites), and shows values similar to those found in the amphipod *P. hawaiiensis*

Table III. Number of amino acids in the protein coding genes, nucleotide frequencies, AT content, and AT and AG skew for mitochondrial genes of *M. longipes*.

Gene	Strand	Size (amino acids)	%A	%C	%G	%T	%A + T	AT skew	GC skew
<i>cob</i>	-	378	30.17	15.83	10.73	43.27	73.44	-0.18	-0.19
<i>nad2</i>	+	330	31.58	8.38	12.71	47.33	78.91	-0.20	0.21
<i>cox1</i>	+	513	28.34	13.81	16.15	41.7	70.04	-0.19	0.08
<i>cox2</i>	+	228	31.83	12.55	13.72	41.9	73.73	-0.14	0.04
<i>atp8</i>	+	52	42.77	8.8	5.66	42.77	85.54	0.00	-0.22
<i>atp6</i>	+	222	30.64	11.81	12.26	45.29	75.93	-0.19	0.02
<i>cox3</i>	+	262	26.87	13.94	16.22	42.97	69.84	-0.23	0.08
<i>nad3</i>	+	117	29.66	9.32	14.69	46.33	75.99	-0.22	0.22
<i>nad5</i>	-	572	34.03	10.3	12.97	42.7	76.73	-0.11	0.11
<i>nad4</i>	-	438	32.24	8.97	13.76	45.03	77.27	-0.17	0.21
<i>nad4L</i>	-	94	36.84	5.61	12.28	45.27	82.11	-0.10	0.37
<i>nad6</i>	+	168	29.39	11.83	7.89	50.89	80.28	-0.27	-0.20
<i>nad1</i>	-	306	32.14	9.45	14.66	43.76	75.9	-0.15	0.22
<i>rrnL</i>	-	n.a.	39.93	7.92	13.81	38.34	78.27	0.02	0.27
<i>rrnS</i>	-	n.a.	35.83	7.63	12.66	43.88	79.71	-0.10	0.25
Control region	+	n.a.	40.79	7.89	7.89	43.43	84.22	-0.03	0.00
All protein coding genes	n.a.	n.a.	31.25	11.34	13.33	44.08	75.33	-0.17	0.08
All tRNAs	n.a.	n.a.	40.95	9.01	12.98	37.06	78.01	0.05	0.18
Complete genome	+	n.a.	37.36	12.41	11.56	38.67	76.03	-0.02	-0.04



and the more AT-rich mitogenome of *A. americanus* (Branchiura, Maxillopoda) (Machida et al. 2002).

#### Divergence in protein coding genes

We used the complete dataset of mtDNA sequences of crustacean taxa plus a representation of all major Hexapodan orders for which data are available (35 taxa; see Additional File 1) to assess the relative divergence of individual PCGs. The genes showing lower corrected divergences across Crustacea and Hexapoda were *cox1*, *cox2*, *cox3* and *cob*, while *nad2*, *nad6*, *nad4L* and *atp8* displayed about twice the mean divergence values (Figure 3). There seems to be an association between gene variation and length and, perhaps, strand location, because shorter genes, often present on the – strand (such as *atp8* and *nad4L*), are the most divergent. Nevertheless, *nad2* is placed on the + strand in Hexapoda, and in most crustaceans also shows a high substitution rate. As noted elsewhere (Cameron and Whiting 2007; Salvato et al. 2008), both variability and codon usage analyses of individual PCGs of Isoptera and Lepidoptera reveal that some of the genes most used in molecular systematics, such as *cox1* and *cox2*, have the lowest variability, while the neglected genes *nad2*, *nad3*, *nad4* and *nad6* may prove to be very useful for systematics given their variability and informative nature. Our results show that this could be extended to crustaceans, which show an underlying substitution pattern similar to hexapods at protein coding genes.

#### Transfer RNA genes

We identified 17 tRNA genes in a general search on the *M. longipes* mitogenome using tRNAscan-SE, and other four (*trnS1*, *trnN*, *trnF* and *trnV*) were inferred from less stringent specific searches in non-coding regions (COVE cut-off score of –20). Despite this,

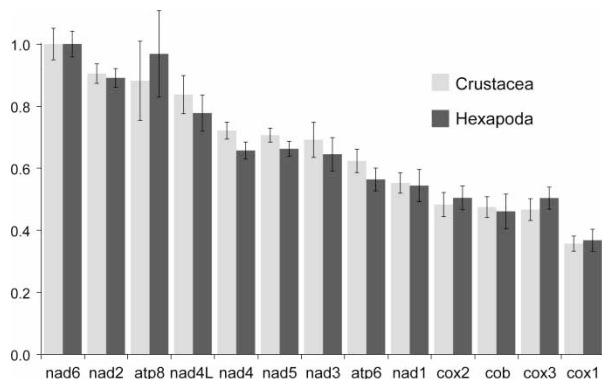


Figure 3. Mean relative corrected divergences of protein coding genes of Crustacea and Hexapoda. Note: DNA divergences of individual genes were estimated from pairwise comparisons among the complete mitogenomes of crustaceans and 35 species representing all major Hexapoda orders.

the *trnS2* gene (tRNA-Ser<sub>AGN</sub>) was not found, although it could almost completely overlap with either the *trnG* or *trnW* genes (COVE scores of +0.30 and –3.54, respectively). The *trnS2* gene shows unusual characteristics in many arthropods, such as the lack of the DHU arm (Kilpert and Podsiadlowski 2006 and references therein). In addition, in *M. longipes* the tRNA-Thr shows an unusual secondary structure, lacking completely the TΨC arm, whereas the tRNA-Gln lacks the loop normally present at this arm (Figure 4). Nucleotide mismatches were evident in the acceptor stem for tRNA-Gln, tRNA-Arg and tRNA-Ile, and in the anticodon stem for tRNA-Lys (Figure 4). Many cases of mismatches in stems have been described in mitochondrial tRNAs, and are supposed to be modified by RNA editing (Ojala et al. 1981; Xiong and Kocher 1991; Yokobori and Paabo 1995; Kilpert and Podsiadlowski 2006). The tRNA genes are present in both strands although most of them (13 genes) are located in the + strand (Table II and Figure 1).

#### Ribosomal RNA genes

The *rrnS* and *rrnL* genes are approximately 695 and 1137 bp in length, respectively (Table II), and around 78% AT-rich, thus being considerably shorter than in other crustaceans. This further explains the extreme compactness of the *M. longipes* mitogenome. The *rrnL* gene of *M. longipes* is closest to those of the amphipods *P. hawaiiensis* and *Niphargus rhenorhodanensis* (accession number: EF028415) (75% sequence identity), while the *rrnS* gene does not show any significant similarity to the sequences of other crustaceans. Not enough information on crustacean 12S and 16S rRNAs secondary structure is available to attempt reconstructing their structure based on comparative analyses.

#### Gene order

*M. longipes* shows a mitochondrial gene order not found in any other crustacean so far analysed (Figure 5). Although the pancrustacean position of *trnL2* between *cox1* and *cox2* is conserved (Boore et al. 1995), many rearrangements in the *Metacrangonyx* genome compared with the ancestral pattern can be deduced (Boore et al. 1995, 1998). At least three transpositions involving genes *trnR*, *trnG* and *trnC* separately, two shifts of strand (reversals) – one involving the gene *cob* and another the segment including *trnP* and *trnT* – and three complex tandem duplications with subsequent random losses are needed to account for the pattern observed in *M. longipes* compared with the pancrustacean ancestral pattern using heuristic analyses of strong common intervals with CREx (Additional File 2). Alternatively,

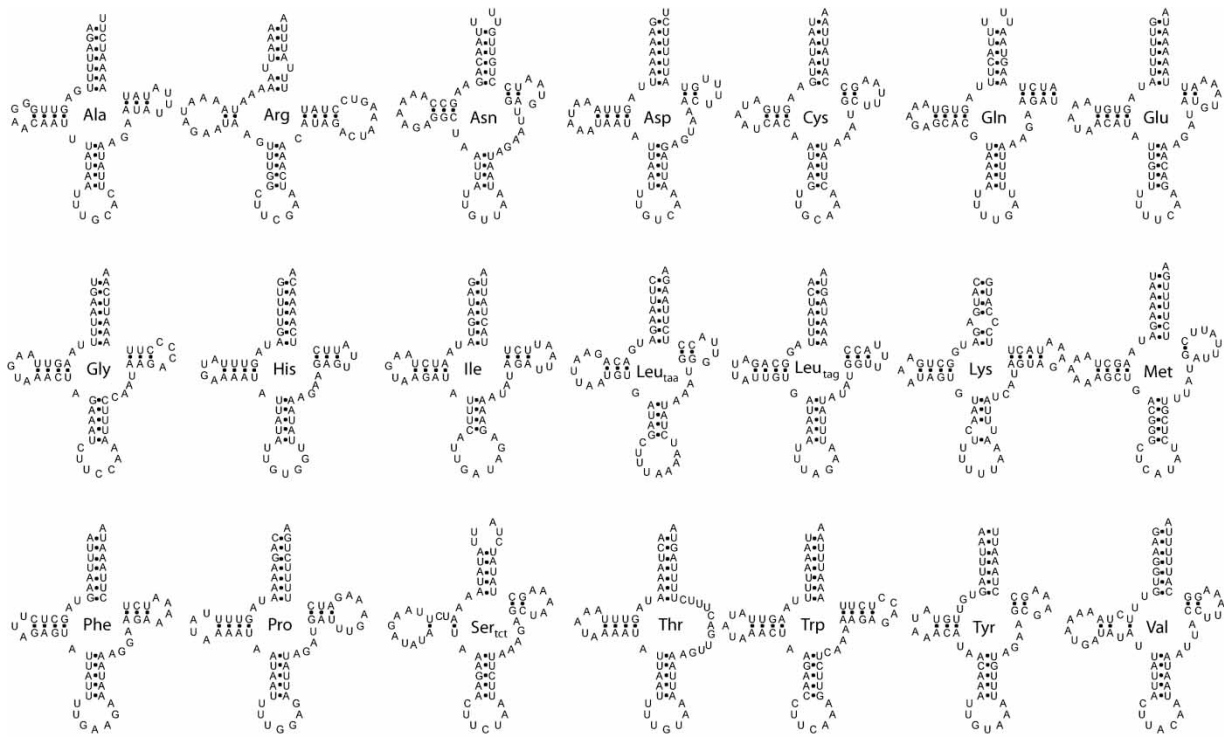


Figure 4. Putative secondary structures of mitochondrial tRNAs in *M. longipes*.

one single reversal of the ancestral pancrustacean segment including *cob nad6 trnP trnT*, followed by a new reversal of the gene *nad6* from the – to the + strand, plus three tandem duplications with subsequent random losses could have produced the *M. longipes* mitogenome gene order. The gene order also differs from the pattern found in the only other amphipod analysed thus far, *P. hawaiiensis* (Cook et al. 2005) (authors’ our own data). We have almost completed the sequence for the mitogenome of this species except for a short part of the gene *rrnS* and the control region (EMBL accession numbers: FM957525 and FM957526), annotating the genes for *trnV*, partial *rrnS*, *trnM*, *trnY*, *trnC*, and locating the gene *trnH* between *nad5* and *nad4*, which was absent in the previous annotation (Cook et al. 2005). In addition, based on tRNAscan results, we

reannotated the tRNA genes *trnW* and *trnG* previously annotated as *trnC* and *trnW*, respectively (Cook et al. 2005) (accession number: AY639937). In *P. hawaiiensis*, at least 10 of the tRNA genes show positional changes with respect to the pancrustacean pattern. The occurrence of identical transpositions of *trnR* and *trnG* in both *P. hawaiiensis* and *M. longipes* mitogenomes with respect to the ancestral arrangement suggests they could have arisen in the common ancestor of amphipods. The other peracarid mitogenomes known, those of the isopods *L. oceanica* and the incomplete ones of *I. baltica* and *A. vulgare*, show quite different translocations from the assumed ancestral pancrustacean gene order (Figure 5), with apparently no common shifts derived from the peracarid ancestor being able to explain the observed patterns (Kilpert and Podsiadlowski 2006).

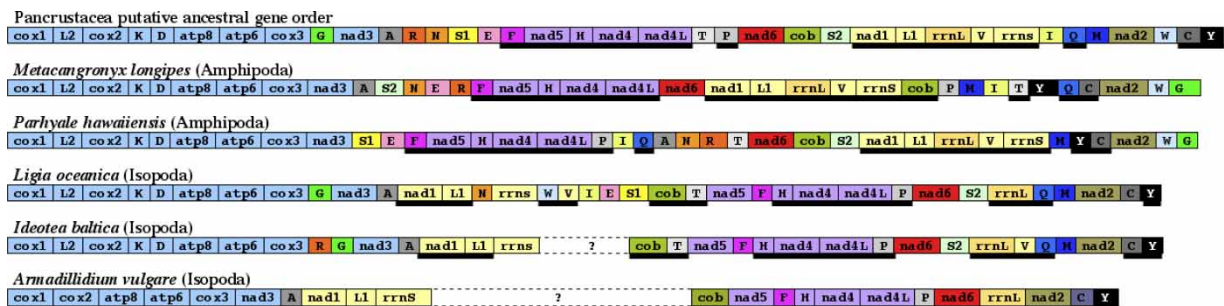


Figure 5. Mitochondrial gene order in Peracarida (Isopoda + Amphipoda) mitogenomes compared with the pancrustacean ancestral pattern. Note: Different colours are used to identify particular conserved and rearranged segments or genes. Genes underlined are present at the – strand.

## Conclusions

The sequence of *M. longipes* introduced herein is the first complete mitogenome of a crustacean amphipod and the second for a peracarid thus far obtained, a superorder that is under-represented in the crustacean mitochondrial genome datasets currently available. The mitogenome is very compact, with a short control region, and it appears to be the shortest mitogenome described for a crustacean. Its AT content is high (76.03%), and gene order is not conserved compared with the other four peracarids whose complete or nearly complete mitogenomes are known: the isopods *L. oceanica*, *I. baltica* and *A. vulgare* and the amphipod *P. hawaiiensis*. Common transpositions of *trnR* and *trnG* in both *P. hawaiiensis* and *M. longipes* mitogenomes with respect to the ancestral pancrustacean arrangement suggest that they were present in the common ancestor of these two amphipods. Many differences in gene order are remarkable compared with the condition displayed in isopods. Thus, no inverted strand bias of nucleotide frequencies is found in *M. longipes*, contrary to what is reported for the mitogenomes of *L. oceanica* and *I. baltica* (Kilpert and Podsiadlowski 2006). The data presented herein not only expand the sampling within the crustacean mitochondrial genomes but also will help, when congeneric species from different geographic areas are sequenced, to solve the phylogenetic position and historical biogeography of this enigmatic family found exclusively in subterranean waters.

## Acknowledgements

The present work has been financed by the research project CGL2006-01365 of the Spanish Ministerio de Ciencia y Innovación and European Union FEDER funds. MBR and JP are supported by FPI and “Ramon y Cajal” grants from the Ministerio de Ciencia e Innovación of Spain, respectively.

**Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## References

Balazuc J, Ruffo S. 1953. Due nuove species del genere *Metacrangonyx* Chevreux delle acque interne del Nord Africa Francese. Mem Mus Civ Stor Nat Verona 4:25–33.

Balke M, Ribera I, Beutel RG. 2005. The systematic position of Aspidytidae, the diversification of Dytiscoidea (Coleoptera, Adephaga) and the phylogenetic signal of third codon positions. J Zool Syst Evol Res 43:223–242.

Barraclough TG, Hogan JE, Vogler AP. 1999. Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). Proc R Soc Lond B Biol Sci 266:1061–1067.

Bernt M, Merkle D, Ramsch K, Fritsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M. 2007. CREx:

Inferring genomic rearrangements based on common intervals. Bioinformatics 23:2957–2958.

Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376:163–165.

Boore JL, Lavrov D, Brown WM. 1998. Gene translocation links insects and crustaceans. Nature 392:667–668.

Boutin C. 1994. Phylogeny and biogeography of metacrangonyctid amphipods in North Africa. Hydrobiologia 287:49–64.

Boutin C, Coineau N. 1990. Regression model. Modèle biphasé d'évolution et origine des micro-organismes stygobies interstitiels continentaux. Rev Micropaléontol 33:303–322.

Boutin C, Messouli M. 1988a. *Metacrangonyx gineti* n. sp. d'une source du Haut-Atlas marocain et la famille des Metacrangonyctidae n. fam. (Crustacés Amphipodes stygobies). Vie Milieu 38:67–84.

Boutin C, Messouli M. 1988b. *Longipodacrangonyx maroccanus* n. gen. n. sp., nouveau représentant du groupe *Metacrangonyx* dans les eaux souterraines du Maroc. Crustaceana 13:256–271.

Cameron SL, Whiting MF. 2007. Mitochondrial genomic comparisons of the subterranean termites from the genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). Genome 50:188–202.

Carapelli A, Liò P, Nardi F, van der Wath E, Frati F. 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol Biol 7:S8.

Chevreaux E. 1909. Amphipodes (1ère série). Arch Zool Exp Gén 5:27–42.

Cook CE, Yue Q, Akam M. 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proc R Soc Lond B Biol Sci 272:1295–1304.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek RC. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299.

Hassanin A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Mol Phylogenet Evol 38:100–116.

Hassanin A, Leger N, Deutsch J. 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst Biol 54:277–298.

Jaume D, Christenson K. 2001. Amphi-Atlantic distribution of the subterranean amphipod family Metacrangonyctidae (Gammaridea). Contrib Zool 70:99–125.

Karaman GS. 1989. *Metacrangonyx ortalii* n sp, a new subterranean member of the family Crangonyctidae, from the Dead Sea region. Stud Mar 20:33–49.

Karaman GS, Pesce GL. 1979. Researches in Africa by the Zoological Institute of l'Aquila, Italy V: On three subterranean amphipods from North Africa (Amphipoda: Gammaridea). Bull Zool Mus Univ Amsterdam 7:197–207.

Kilpert F, Podsiadlowski L. 2006. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. BMC Genom 7:241.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 25:955–964.

Machida RJ, Miya MU, Nishida M, Nishida S. 2002. Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). Mar Biotechnol 4:406–417.

Marcadé I, Cordaux R, Doublet V, Debenest C, Bouchon D, Raimond R. 2007. Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). J Mol Evol 65:651–659.



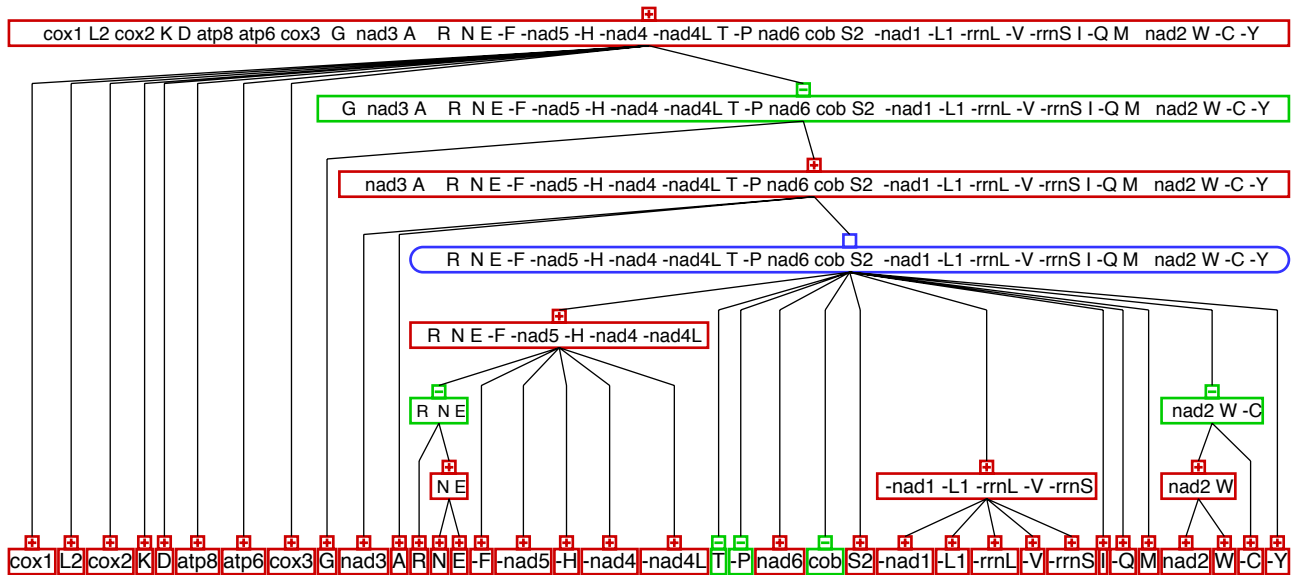
- Margalef R. 1952. Materiales para la hidrobiología de la isla de Menorca. *Publ Inst Biol Apl* 11:5–112.
- Martin JW, Davis GE. 2001. An updated classification of the recent Crustacea. Los Angeles, CA: Natural History Museum of Los Angeles County. p 123.
- Messouli M, Boutin C, Coineau N. 1991. Phylogénie et biogéographie évolutive d'un groupe de Metacrangonyctidae, Crustacés Amphipodes stygobies du Maroc. I. Les espèces du groupe *panousei*. *Mém Biospéol* 25:247–262.
- Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470–474.
- Oulbaz Z, Coineau N, Boutin C. 1998. *Metacrangonyx knidirii* n sp et *Metacrangonyx remyi* Balazuc et Ruffo, 1953 (Crustacés Amphipodes Metacrangonyctidae) des eaux souterraines du Maroc. *Mém Biospéol* 25:35–43.
- Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol* 41:353–358.
- Podsiadlowski L, Bartolomaeus T. 2006. Major rearrangements characterize the mitochondrial genome of the isopod *Idotea baltica* (Crustacea: Peracarida). *Mol Phylogenet Evol* 40: 893–899.
- Ruffo S. 1954. *Metacrangonyx longicaudus* n sp (Amphipoda Gammaridae) delle acque sotterranee del Sahara Marocchino. *Mem Mus Civ Stor Nat Verona* 4:127–130.
- Ruffo S. 1982. Una nouvelle espèce de *Metacrangonyx* Chevreux (Amphipoda: Gammaridae) du désert du Sinaï. *Isr J Zool* 31: 151–156.
- Salvato P, Simonato M, Battisti A, Negrisolò E. 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genom* 9:331.
- Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stoch F. 1997. *Metacrangonyx ilvanus* n sp, the first Italian representative of the family Metacrangonyctidae (Crustacea: Amphipoda). *Ann Limnol* 33:255–262.
- Stock JH, Rondé-Broekhuizen B. 1986. A new species of *Pygocrangonyx*, an amphipod genus with African affinities, from Fuerteventura. *Bijdr Dierk* 56:247–266.
- Supek F, Vlahovick K. 2004. INCA: Synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics* 14:2329–2330.
- Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 101:11030–11035.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
- Wolstenholme DR. 1992. Animal mitochondrial DNA: Structure and evolution. *Int Rev Cytol* 141:173–216.
- Wright F. 1990. The “effective number of codons” used in a gene. *Gene* 87:23–29.
- Xia X, Xie Z. 2001. DAMBE: Software package for data analysis in molecular biology and evolution. *J Hered* 92:371–373.
- Xiong B, Kocher TD. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera: Simuliidae). *Genome* 34:306–311.
- Yang JS, Yang WJ. 2008. The complete mitochondrial genome sequence of the hydrothermal vent galatheid crab *Shinkaita crosnieri* (Crustacea: Decapoda: Anomura): A novel arrangement and incomplete tRNA suite. *BMC Genom* 9:257.
- Yokobori S, Paabo S. 1995. tRNA editing in metazoans. *Nature* 377:490.
- Zhang DX, Szymura JM, Hewitt GM. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. *J Mol Evol* 40:382–391.

## Additional Files

Additional File 1. Taxon names and EMBL accession numbers of the crustacean and hexapod mitogenomes used for gene annotation and gene divergence analyses.

Species	Accession no.	Taxonomy
<b>CRUSTACEA</b>		
<i>Argulus americanus</i>	NC_005935	Maxillopoda Branchiura
<i>Armadillidium vulgare</i>	EF643519	Malacostraca Peracarida Isopoda
<i>Armillifer armillatus</i>	NC_005934	Pentastomida
<i>Artemia franciscana</i>	NC_001620	Branchiopoda Anostraca
<i>Callinectes sapidus</i>	NC_006281	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Cherax destructor</i>	NC_011243	Malacostraca Eucarida Decapoda Pleocyemata
<i>Daphnia pulex</i>	NC_000844	Branchiopoda Anomopoda
<i>Eriocheir sinensis</i>	NC_006992	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Euphausia superba</i>	AB084378	Malacostraca Eucarida Euphausiacea
<i>Fenneropenaeus chinensis</i>	NC_009679	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Geothelphusa dehaani</i>	NC_007379	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Gonodactylus chiragra</i>	NC_007442	Malacostraca Hoplocarida Stomatopoda
<i>Halocaridina rubra</i>	NC_008413	Malacostraca Eucarida Decapoda Pleocyemata Caridea
<i>Harpisquilla harpax</i>	NC_006916	Malacostraca Hoplocarida Stomatopoda
<i>Hutchinsoniella macracantha</i>	NC_005937	Cephalocarida
<i>Idotea baltica</i>	DQ442915	Malacostraca Peracarida Isopoda
<i>Lepeophtheirus salmonis</i>	NC_007215	Maxillopoda Copepoda
<i>Ligia oceanica</i>	NC_008412	Malacostraca Peracarida Isopoda Oniscidea
<i>Litopenaeus vannamei</i>	NC_009626	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Lysiosquillina maculata</i>	NC_007443	Malacostraca Hoplocarida Stomatopoda
<i>Macrobrachium rosenbergii</i>	NC_006880	Malacostraca Eucarida Decapoda Pleocyemata Caridea
<i>Marsupenaeus japonicus</i>	NC_007010	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Megabalanus volcano</i>	NC_006293	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Pagurus longicarpus</i>	NC_003058	Malacostraca Eucarida Decapoda Anomura
<i>Panulirus japonicus</i>	NC_004251	Malacostraca Eucarida Decapoda Pleocyemata
<i>Parhyale hawaiiensis</i>	AY639937	Malacostraca Peracarida Amphipoda
<i>Penaeus monodon</i>	NC_002184	Malacostraca Eucarida Decapoda Dendrobranchiata
<i>Pollicipes mitella</i>	NC_008742	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Pollicipes polymerus</i>	NC_005936	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Portunus trituberculatus</i>	NC_005037	Malacostraca Eucarida Decapoda Brachyura
<i>Pseudocarcinus gigas</i>	NC_006891	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
<i>Pseudosquilla ciliata</i>	AY947836	Malacostraca Hoplocarida Stomatopoda
<i>Speleonectes tulumensis</i>	NC_005938	Remipedia
<i>Squilla empusa</i>	NC_007444	Malacostraca Hoplocarida Stomatopoda
<i>Squilla mantis</i>	NC_006081	Malacostraca Hoplocarida Stomatopoda
<i>Tetraclita japonica</i>	NC_008974	Maxillopoda Thecostraca Cirripedia Thoracica
<i>Tigriopus californicus</i>	NC_008831	Maxillopoda Copepoda
<i>Tigriopus japonicus</i>	NC_003979	Maxillopoda Copepoda Harpacticoida
<i>Triops cancriformis</i>	NC_004465	Branchiopoda Notostraca
<i>Triops longicaudatus</i>	NC_006079	Branchiopoda Notostraca
<i>Vargula hilgendorffii</i>	NC_005306	Ostracoda Myodocopa
<i>Metacrangonyx longipes</i>	AM944817	Malacostraca Peracarida Amphipoda
<b>HEXAPODA</b>		
<i>Aleurodicus dugesii</i>	NC_005939	Insecta Hemiptera Hemimetabola
<i>Anopheles gambiae</i>	NC_002084	Insecta Diptera Holometabola
<i>Antheraea pernyi</i>	NC_004622	Insecta Lepidoptera Holometabola
<i>Apis mellifera ligustica</i>	NC_001566	Insecta Hymenoptera Holometabola
<i>Bombyx mori</i>	NC_002355	Insecta Lepidoptera Holometabola
<i>Ceratitis capitata</i>	NC_000857	Insecta Diptera Holometabola
<i>Chrysomya putoria</i>	NC_002697	Insecta Diptera Holometabola
<i>Crioceris duodecimpunctata</i>	NC_003372	Insecta Coleoptera Holometabola
<i>Drosophila melanogaster</i>	NC_001709	Insecta Diptera Holometabola
<i>Gomphiocephalus hodgsoni</i>	NC_005438	Collembola
<i>Grylotalpa orientalis</i>	NC_006678	Insecta Orthoptera Hemimetabola
<i>Haematobia irritans irritans</i>	NC_007102	Insecta Diptera Holometabola
<i>Heterodoxus macropus</i>	NC_002651	Insecta Phthiraptera Hemimetabola
<i>Homalodisca coagulata</i>	NC_006899	Insecta Hemiptera Hemimetabola
<i>Japyx solifugus</i>	NC_007214	Diplura
<i>lepidopsocid RS-2001</i>	NC_004816	Insecta Psocoptera Hemimetabola
<i>Locusta migratoria</i>	NC_001712	Insecta Orthoptera Hemimetabola
<i>Melipona bicolor</i>	NC_004529	Insecta Hymenoptera Holometabola
<i>Nesomachilis australica</i>	NC_006895	Insecta Archaeognatha Ametabola

Species	Accession no.	Taxonomy
<i>Onychiurus orientalis</i>	NC_006074	Collembola
<i>Orthetrum triangulare melania</i>	AB126005	Insecta Odonata Hemimetabola
<i>Ostrinia nubilalis</i>	NC_003367	Insecta Lepidoptera Holometabola
<i>Pachypsylla venusta</i>	NC_006157	Insecta Hemiptera Hemimetabola
<i>Periplaneta fuliginosa</i>	NC_006076	Insecta Dictyoptera Hemimetabola
<i>Philaenus spumarius</i>	NC_005944	Insecta Hemiptera Hemimetabola
<i>Podura aquatica</i>	NC_006075	Collembola
<i>Pteronarcys princeps</i>	NC_006133	Insecta Plecoptera Hemimetabola
<i>Pyrocoelia rufa</i>	NC_003970	Insecta Coleoptera Holometabola
<i>Schizaphis graminum</i>	NC_006158	Insecta Hemiptera Hemimetabola
<i>Thermobia domestica</i>	NC_006080	Insecta Thysanura Ametabola
<i>Thrips imaginis</i>	NC_004371	Insecta Thysanoptera Hemimetabola
<i>Triatoma dimidiata</i>	NC_002609	Insecta Hemiptera Hemimetabola
<i>Tribolium castaneum</i>	NC_003081	Insecta Coleoptera Holometabola
<i>Tricholepidion gertschi</i>	NC_005437	Insecta Thysanura Ametabola
<i>Xenos vesparum</i>	DQ364229	Strepsiptera Holometabola



Additional File 2. Rearrangement steps deduced using detection of strong interval trees to account for the gene order of *M. longipes* mitogenome compared with the ancestral pancrustacean order.

Downloaded By: [Pons, Joan] At: 09:30 10 June 2009

RESEARCH ARTICLE

Open Access

# Islands beneath islands: phylogeography of a groundwater amphipod crustacean in the Balearic archipelago

Maria M Bauzà-Ribot<sup>1</sup>, Damià Jaume<sup>2</sup>, Joan J Fornós<sup>3</sup>, Carlos Juan<sup>1</sup> and Joan Pons<sup>2\*</sup>

## Abstract

**Background:** Metacrangonyctidae (Amphipoda, Crustacea) is an enigmatic continental subterranean water family of marine origin (thalassoid). One of the species in the genus, *Metacrangonyx longipes*, is endemic to the Balearic islands of Mallorca and Menorca (W Mediterranean). It has been suggested that the origin and distribution of thalassoid crustaceans could be explained by one of two alternative hypotheses: (1) active colonization of inland freshwater aquifers by a marine ancestor, followed by an adaptative shift; or (2) passive colonization by stranding of ancestral marine populations in coastal aquifers during marine regressions. A comparison of phylogenies, phylogeographic patterns and age estimations of clades should discriminate in favour of one of these two proposals.

**Results:** Phylogenetic relationships within *M. longipes* based on three mitochondrial DNA (mtDNA) and one nuclear marker revealed five genetically divergent and geographically structured clades. Analyses of cytochrome oxidase subunit 1 (*cox1*) mtDNA data showed the occurrence of a high geographic population subdivision in both islands, with current gene flow occurring exclusively between sites located in close proximity. Molecular-clock estimations dated the origin of *M. longipes* previous to about 6 Ma, whereas major cladogenetic events within the species took place between 4.2 and 2.0 Ma.

**Conclusions:** *M. longipes* displayed a surprisingly old and highly fragmented population structure, with major episodes of cladogenesis within the species roughly correlating with some of the major marine transgression-regression episodes that affected the region during the last 6 Ma. Eustatic changes (vicariant events) -not active range expansion of marine littoral ancestors colonizing desalinated habitats-explain the phylogeographic pattern observed in *M. longipes*.

## Background

Subterranean fauna provides unique opportunities for the study of evolutionary mechanisms and speciation processes [1]. In recent years, phylogeographic analyses have revealed unprecedented cases of cryptic speciation, restricted distribution and presumed sympatric speciation among different cave-dwelling animal groups [2]. Nevertheless, the occurrence of extensive morphological conservatism in subterranean fauna frequently hampers the establishment of phylogenetic inferences based solely on morphological features. In this context, homoplasy

arises from common exposure to the particular selective pressures inherent to cave life (i.e., darkness and oligotrophy) or from the lack of directional selection [3,4]. Conversely, isolation in caves can lead these morphologically undifferentiated subterranean organisms to display high levels of genetic divergence [4-6].

Geological and hydrological processes, in particular shifts in water tables, can lead to the isolation or connection of aquifers, with consequent effects on gene flow between populations of subterranean aquatic organisms [6]. In the same way, marine regressions are suggested to have played a major role in the isolation of many marine relicts in continental groundwaters [7-10]. Recent molecular phylogenetic and phylogeographic studies on subterranean amphipods emphasize the role

\* Correspondence: jpons@imedea.uib-csic.es

<sup>2</sup>IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, c/Miquel Marqués, 21, 07190-Esporles, Balearic Islands, Spain

Full list of author information is available at the end of the article

played by historical factors (i.e., glacial or drought episodes) in the pattern of genetic diversification and distribution displayed by these animals [5,6,11]. Likewise, [12] considered the influence of contingency, i.e., whether the colonization event involved a single localized surface ancestor or multiple, geographically separated ancestors, on the shaping of these patterns. In addition, larval life history traits, such as feeding mode (planktotrophic vs. lecithotrophic) can play a determinant role in crustacean distribution, as they control the duration of the dispersive phase [13,14]. However, stygobiont amphipods have a comparatively reduced dispersal potential (as do all peracarid crustaceans), as the females carry offspring in a marsupium and these are brooded and not released into the water column until metamorphosed into diminutive non-natatory adults [15].

Among the obligate dwellers of subterranean waters (stygobionts), a high number belong to so-called thalassoid lineages, organisms that are derived directly from marine ancestors [7]. Thalassoid forms are known to occur among a vast array of faunistic groups, especially the Crustacea [16,8]. The ancestors of thalassoid animals presumably inhabited marine transitional habitats, such as submarine fissures, mixohaline submarine karstic springs or the interstitial medium developed in sandy and gravelly coastal sediments, where sharp variations in salinity (i.e., periodical exposure to desalinated waters) and other environmental conditions mimic, in some way, those found in fresh groundwaters [7]. Colonization of inland freshwater aquifers by this preadapted marine fauna might have proceeded as a natural extension of their primary niche, followed by an adaptive shift; this process would be independent of the occurrence of environmental constraints, such as episodes of glaciation, drought or marine regression [17,18]. This hypothesis provides a plausible explanation for the origin of some freshwater stygobiont ostracods closely related to marine euryhaline taxa [19]. However, most faunistic and biogeographic evidence favours an alternative vicariant scenario by which colonization occurs passively via stranding of ancestral populations during episodes of marine regression [7-10]. Accordingly, sea withdrawal or tectonic uplift at different geological periods could have led to the gradual isolation of populations of ancestral marine taxa in inland groundwaters, triggering their ulterior diversification and speciation. This hypothesis explains satisfactorily the distribution of many stygobiont crustaceans and is testable by collating a phylogenetic framework and molecular-clock-age estimates of relevant clades, with their respective geographic distributions [2,20].

Here, we studied the phylogeography of *Metacrangonyx longipes* Chevreux, 1909, a euryhaline stygobiont amphipod crustacean that is endemic to Mallorca and

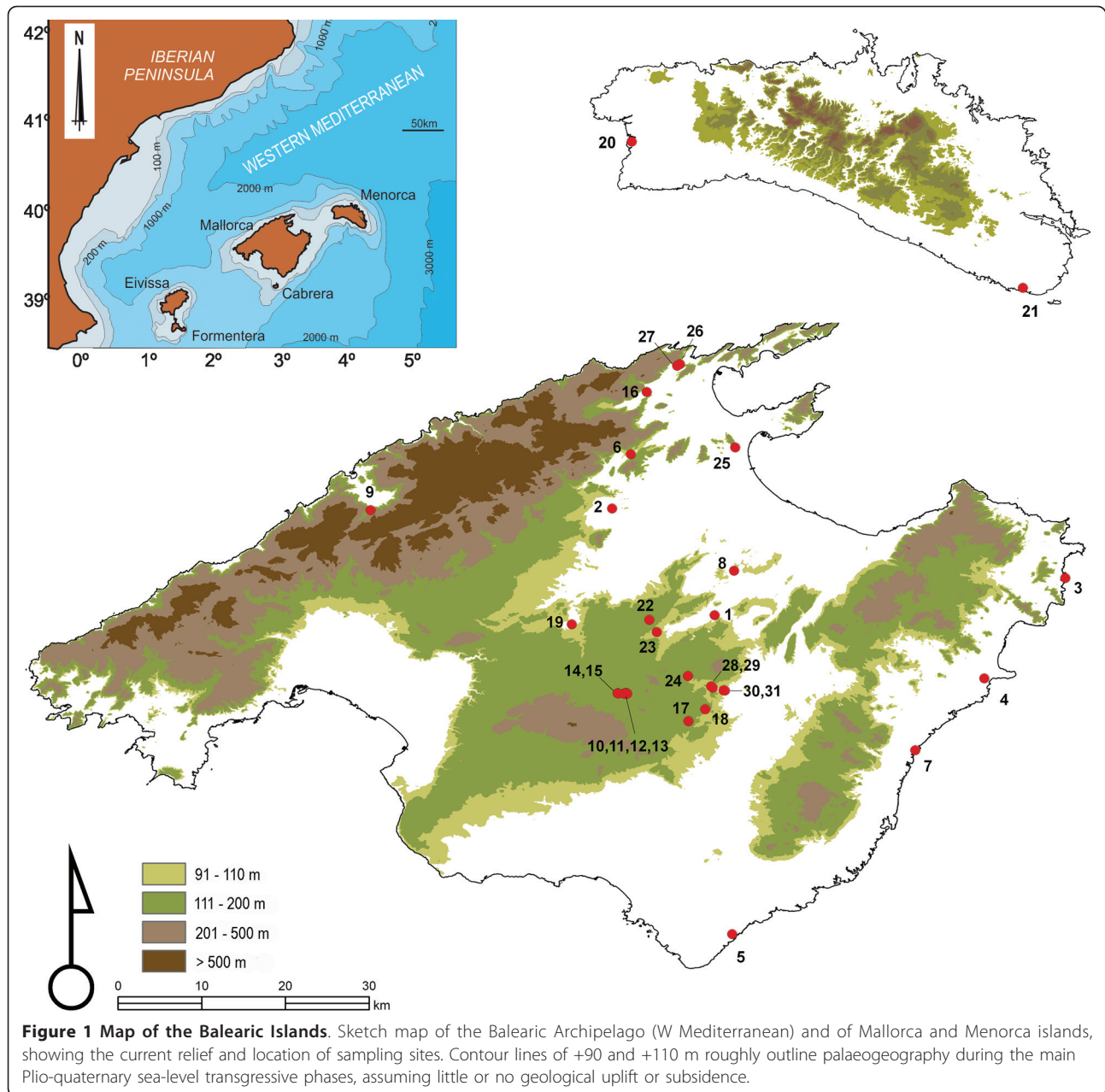
Menorca (Balearic Islands; W Mediterranean). On Mallorca, it occurs in various types of groundwater habitats, from coastal anchialine caves (*sensu* [21]) of raised salinity to freshwater inland wells, caves and springs. On Menorca, the species is restricted to coastal anchialine caves and wells and is absent from fresh inland groundwaters. On both islands, the species is limited to lowlands and is absent in apparently suitable habitats located at elevations higher than 125 m above sea level. The Metacrangonyctidae is a strictly inland water subterranean family with no close relatives; however, several lines of evidence strongly suggest its marine origin: (1) its members are known only from continental regions that were covered by ancient epicontinental seas [22,23]; and (2) several species still maintain ties with the marine environment (i.e., they live in anchialine wells and caves in coastal areas; [23]).

In this study, we used the sequences of three mitochondrial and one nuclear gene of *M. longipes* and of several congeneric species to perform a phylogenetic analysis of the species and infer population divergence times. Moreover, we use sequences of the cytochrome oxidase subunit 1 gene from a more comprehensive data set to perform a phylogeographic analysis and to examine the population structure of this taxon. Given the manifested euryhalinity of *M. longipes* and the absence of any appreciable morphological differentiation between its populations on the two islands, our initial prediction was that the species could have dispersed across the groundwater environment of the islands using the virtually continuous peripheral coastal anchialine pathway, from which it could have colonized inland freshwater habitats recurrently. If this was the case, we could expect a pattern of considerable gene flow and shallow genetic divergences within each island, with genetic signatures of inland populations deriving from coastal ones. However, our study revealed that this amphipod displays a remarkably ancient and highly fragmented population structure, with episodes of cladogenesis that could be related to major sea-level changes that affected the islands during the last 6 Ma.

## Results

Four gene fragments—three mitochondrial (cytochrome oxidase subunit 1 (*cox1*), cytochrome b (*cob*) and 16S rRNA (*rrnL*)) and one nuclear (Histone H3A)—with a total sequence length of about 1.7 Kb were sequenced from 34 *Metacrangonyx longipes* specimens and the outgroups *Metacrangonyx ilvanus*, *M. remyi* and *M. sp* (details on sampling localities appear in Additional file 1, Figure 1 and in the Methods section). These mitochondrial sequences were assumed not to correspond to nuclear pseudogenes, as the mtDNA protein-coding genes considered did not include stop codons or



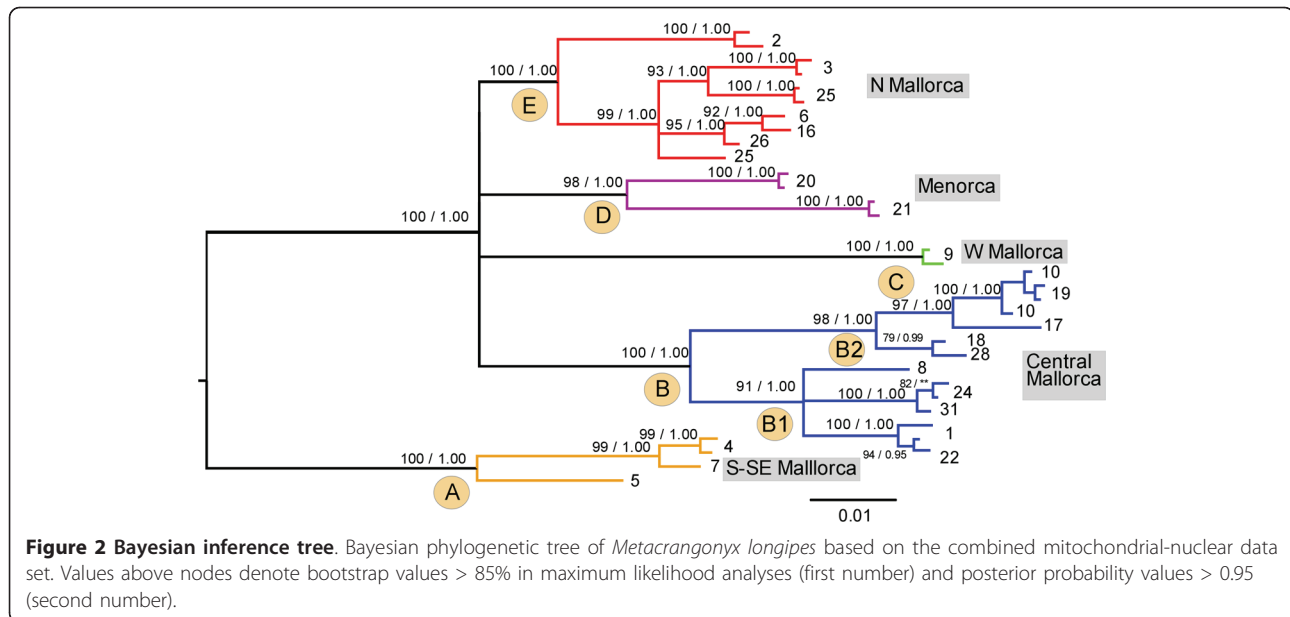


frameshift mutations and no double peaks appeared in the corresponding chromatograms. Moreover, the separate analyses of each marker gave essentially congruent tree topologies, with Partition Bremer Support (PBS) positive values for most of the tree nodes (not shown). Few nodes with low support showed PBS values close to zero, suggesting that their low phylogenetic signal is not due to incongruence among markers. Most of the variation is contained in the mitochondrial genes: *cox1*, *cob* and *rrnL* had 120, 78 and 43 parsimony informative positions, respectively. Histone H3A sequences rendered five haplotypes only, with six parsimony informative

sites and two fixed substitutions in *M. longipes* with respect to the outgroup species.

#### Phylogenetic analyses and genetic distances

Bayesian and maximum likelihood (ML) analyses of the combined mitochondrial and nuclear data set yielded a similar topology, in which five divergent monophyletic lineages of *M. longipes* not showing geographical overlap were clearly recognized (see Figure 1 for a map of Mallorca and Menorca and the corresponding sampling sites, and Figure 2 for the Bayesian tree). A clade comprising three anchialine caves from the S and SE of



Mallorca (clade A; localities 4, 5 and 7) was highly supported and recovered as sister to the remaining clades after rooting the tree with congeneric species. The remaining populations formed four highly supported clades, although the evolutionary relationships among them remain unresolved: clade D (Menorcan, corresponding to anchialine caves 20 and 21); two divergent Mallorcan clades, one located on the west side (clade C, corresponding to freshwater cave 9) and the other on the north side of the island (clade E; localities 2, 3, 6, 16, 25 and 26, corresponding to both anchialine caves and freshwater wells); and clade B, comprising wells located far inland in the Mallorcan central area. The latter cluster was, in turn, subdivided into two genetic groups (clade B1: localities 1, 8, 22, 24, 31; and clade B2: localities 10, 17, 18, 19 and 28) showing an approximate NE-SW geographical segregation. A more comprehensive data set comprising the *cox1* gene fragment from 162 specimens was used in population analyses (Table 1). Bayesian and ML analyses performed on the *cox1* data set resulted in phylogenetic trees that were compatible with those derived from the combined analyses mentioned above; however, ambiguous or non-supported relationships persisted among clades, such as the position of the Menorcan populations with respect to their Mallorcan counterparts. In addition, the relationship among Mallorcan clades E and C was only weakly supported (Additional file 2). Parsimonious reconstructions of habitat type based on the *cox1* or the total evidence mtDNA phylogenetic analysis showed at least three transitions to fresh inland groundwaters from anchialine brackish habitats (see Additional file 3).

*Cox1* uncorrected distances between collection sites ranged from a minimum of 0.5% between those located close to each other (viz., 18 and 28, only 3.5 km apart) to a maximum of 8.9% (corrected to 9.8% using a GTR model) between populations from the two islands (viz., localities 4 and 20, separated by 68 km) or between some Mallorcan populations. As deduced from the phylogenetic analyses, clade A was the most divergent (7.8-8% mean uncorrected genetic distance with respect to the remaining clades), whereas the distance between the other clades fell between 6.3-7.5%. The distance between subclades B1 and B2 averaged 5.5%.

#### Population genetic structure and genetic diversity

Fifty different *cox1* haplotypes were identified in the sampled specimens from the 31 populations analysed (EMBL accession numbers FR729731-FR729892) (Table 1). Four haplotypes were shared between neighbouring populations: haplotypes H27 and H29 in several central Mallorcan localities (Montuiri/Ruberts; stations 10, 14 and 19), whereas haplotypes H19 and H22 were present in two Sineu wells (stations 22 and 23). Table 1 summarizes the standard intra-population diversity estimated for *cox1*. Six populations included only one haplotype ( $h = 0$ ), although in three of them only two specimens (the only ones collected) were analysed. In sharp contrast, three populations showed maximum diversity indices, as every individual bore a different haplotype ( $h = 1$ ). The rest of populations attained low-to-moderate  $h$  values, in the range of 0.25-0.86. Diversity was much lower in the two Menorcan populations (three haplotypes per 28 individuals) compared with the

**Table 1 Summary of *cox1* MtDNA population statistics**

Sampling locality	N° indiv.	N° Haplotypes	Hap. Diversity $\pm$ SD	Nuc. Diversity ( $\pi$ ) $\times 10^2 \pm$ SD
1	6	H17 (3), H18 (2), H20 (1)	0.733 $\pm$ 0.155	0.480 $\pm$ 0.334
2	4	H11 (1), H12 (2), H13 (1)	0.833 $\pm$ 0.222	0.235 $\pm$ 0.210
3	19	H2 (15), H3 (4)	0.351 $\pm$ 0.111	0.055 $\pm$ 0.063
4	4	H43 (1), H44 (1), H45 (2)	0.833 $\pm$ 0.222	0.157 $\pm$ 0.155
5	6	H40 (6)	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
6	2	H6 (1), H9 (1)	1.000 $\pm$ 0.500	0.313 $\pm$ 0.383
7	8	H41 (7), H42 (1)	0.250 $\pm$ 0.180	0.039 $\pm$ 0.056
8	12	H25 (10), H26 (2)	0.303 $\pm$ 0.148	0.047 $\pm$ 0.059
9	14	H37 (8), H38 (4), H39 (2)	0.615 $\pm$ 0.102	0.110 $\pm$ 0.099
10	3	H27 (1), H29 (1), H33 (1)	1.000 $\pm$ 0.272	0.417 $\pm$ 0.375
11	3	H27 (2), H32 (1)	0.667 $\pm$ 0.314	0.313 $\pm$ 0.295
12	1	H30 (1)	NC	NC
13	1	H49 (1)	NC	NC
14	3	H27 (1), H29 (1), H50 (1)	1.000 $\pm$ 0.272	0.313 $\pm$ 0.295
15	1	H31 (1)	NC	NC
16	1	H1 (1)	NC	NC
17	1	H34 (1)	NC	NC
18	1	H35 (1)	NC	NC
19	7	H27 (2), H28 (1), H29 (2), H48 (2)	0.857 $\pm$ 0.102	0.283 $\pm$ 0.212
20	5	H14 (4), H15 (1)	0.400 $\pm$ 0.237	0.063 $\pm$ 0.080
21	13	H16 (13)	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
22	2	H19 (2)	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
23	2	H19 (2)	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
24	5	H21 (5)	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
25	23	H4 (1), H5 (20), H8 (1), H46 (1)	0.249 $\pm$ 0.117	0.245 $\pm$ 0.169
26	4	H7 (3), H10 (1)	0.500 $\pm$ 0.265	0.391 $\pm$ 0.315
27	1	H47 (1)	NC	NC
28	6	H22 (4), H23 (1), H36 (1)	0.600 $\pm$ 0.215	1.701 $\pm$ 1.043
29	1	H22 (1)	NC	NC
30	1	H24 (1)	NC	NC
31	2	H22 (2)	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000

Number of specimens of *Metacrangonyx longipes* per collection site, *cox1* mtDNA haplotypes, haplotype and nucleotide diversities with corresponding standard deviations.

Mallorcan populations (47 haplotypes per 144 individuals). Nucleotide diversity (mean number of pair-wise differences  $\pi$ ) were low at most locations ( $\pi < 0.5\%$ ); population 28 alone exhibited a  $\pi$  value  $> 1\%$  because of the presence of an individual bearing a divergent haplotype. Neutrality tests were non-significant in all cases, with the exception of populations 7 (Fu's  $F_s = -0.182$ ,  $P < 0.05$ ) and 8 (Ramos-Onsins and Rozas  $R_2 = 0.51$ ,  $P < 0.01$ ).

The Mantel test revealed a moderately significant correlation ( $r = 0.467$ ,  $P < 0.001$ ) between genetic and geographic distances. In addition, *cox1* pair-wise population  $F_{ST}$  values were generally high ( $F_{ST} = 0.5-0.99$ ) and highly significant, with the exception of between neighbouring populations (viz., 19 vs. 10-15, 24 vs. 28). Accordingly, SAMOVA showed that the optimum number of population groups necessary to maximize  $F_{CT}$

values was  $K = 19-20$  ( $F_{CT} = 0.961-0.963$ ), whereas the corresponding  $F_{SC}$  values were the lowest in the  $K$  series tested, as expected [24]. This structure grouped wells 10-15 and 19 into a single population, whereas wells at Campanet (6) and Pollença (16) on one side, and wells at Vilafranca (28-31) and Cala Sant Vicenç (26, 27) on the other, formed three different populations. Most sites harbouring the same population were less than 2 km apart, with the exception of sites 10 and 19, which were separated by about 10 km. The remaining sites each represented a single, distinct population group, suggesting the occurrence of a high population subdivision with gene flow occurring exclusively among wells located less than 10 km apart. The occurrence of an alternative geographical structuring that yielded higher  $K$  values was explored using AMOVA, as this method allows recognition of population groups with any number of  $K$ . This

analysis yielded a similar geographical setting, but identified wells 16, 19 and 26 as different populations; AMOVA showed that  $F_{CT}$  values reached a plateau at 0.96 at  $K = 19-27$ , with the highest value attained at  $K = 23$ .

#### Estimation of coalescence time

The coalescence of the mitochondrial sequences of *M. longipes* was estimated via Bayesian analyses using the *cox1* population data set and implementing a relaxed molecular clock with a substitution rate fixed at 0.0115 per year per lineage [25], or the range 0.007-0.013 estimated elsewhere for crustaceans [26]. Tree root ages fell between 5.4 and 6.2 Ma depending on the assumed rate, while other node ages were remarkably similar in both instances, although the crustacean mitochondrial rate range rendered slightly older estimates and broader confidence intervals (Table 2 and Figure 3). Estimations using the Yule model on the combined mitochondrial data set and a standard 2.3% rate fell also in the same range (Table 2). Based on the coalescent model, divergence of the Mallorcan clade B -comprising localities from the central area of the island- can be traced back at 2.3-2.7 Ma, whereas that of clade E - occupying the N and NE of the island- seems to have occurred at 2.0-2.4 Ma. In both cases, 95% highest posterior densities (HPDs) fell within the range 3.7-1.2 Ma (Figure 3 and Table 2). Seemingly, the node corresponding to clade D (Menorca) was dated at 2.1-2.3 Ma, whereas that of clade A (comprising S and SE Mallorcan sites) was dated at 1.4-1.6 Ma (95% HPD, 3.5-0.6 Ma in both cases). Nodes corresponding to the two Mallorcan sister subclades B1 and B2 were dated at 1.1-1.2 and 1.0-1.3 Ma, respectively (95% HPD, 1.8-0.6 Ma). In contrast, the

coalescence of monophyletic sequences from particular Mallorcan caves or wells was much more recent, with estimates falling within 0.1-0.2 Ma.

#### Discussion

The thalassoid condition of *Metacrangonyx longipes* is supported in our study as we can deduce at least three independent episodes of colonization of fresh inland groundwaters from primary anchialine, brackish water ancestors. *M. longipes* populations appeared split into five deep genetic lineages devoid of any relevant morphological differentiation. Gene flow between populations did not exceed 10 km and was frequently limited to occur in a radius of less than 2 km. Therefore, our results do not support an active colonization of fresh inland groundwater habitats by expansive crevicular/interstitial marine littoral ancestors (although past episodes of dispersal during favourable conditions can not be ruled out completely) [17-19]. If that was the case, we should have found evidence of substantial connectivity between the populations of *M. longipes* established far inland in completely fresh waters and those of the coastal anchialine medium, and among anchialine population themselves.

Some of the *M. longipes* clades were linked to particular or neighbouring hydrographic catchments and were found nowhere else (Figure 4). Thus, clade C was found exclusively at the Torrent de Sóller catchment, whereas clade B2 was restricted to the head-waters of Torrent de Muro (localities 10-15 and 19) and to some vicine stations at the Torrent de Na Borges catchment (localities 17-18 and 28). Likewise, clade B1 (localities 1, 8, 22-24 and 28-31) was found only at the head-waters of three different catchments: Torrent de Na Borges, Son Bauló and Son Real (Figure 4); nevertheless, these three torrents became recurrently confluent and formed a single palaeodrainage system in past glacial periods with lower sea-level, when the shallow shelf between Mallorca and Menorca was completely exposed sub-aerially (see below). These results suggest that quartering within and displacement along the hyporheic medium associated with these water-courses played a role in structuring the populations of the species. Even limited dispersal across the watershed of adjacent catchments seems possible, as shown above: the plains where the watershed between Torrent de Muro and Torrent de Na Borges is located harbours small, shallow perched aquifers that probably form a continuum in winter, when the area is soaked and attracts important numbers of waders and other waterbirds (D. J., personal observation).

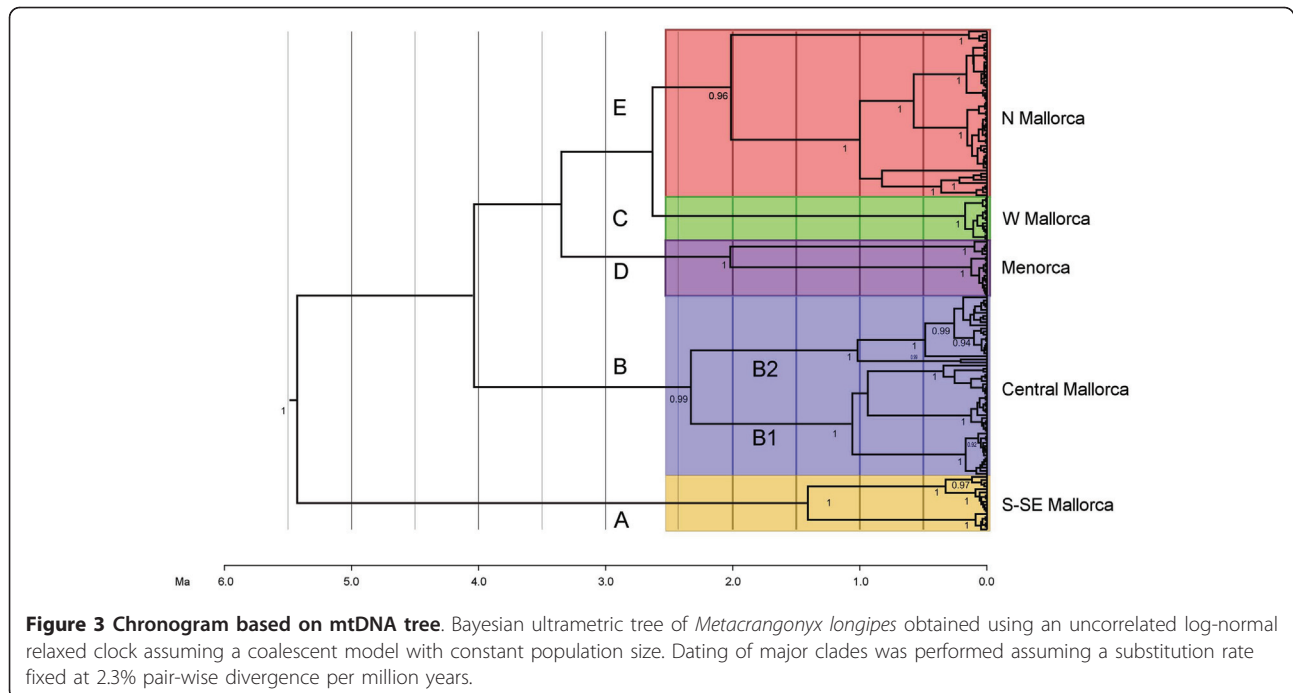
In a study on hyalid and crangonyctoid stygobiont amphipods from W Australian calcrete aquifers, Cooper et al. [5] showed that the major mitochondrial *cox1* lineages were restricted to a single isolated calcrete,

**Table 2 Estimation of coalescence times**

Clade	Coalescence Model	Coalescence Model	Yule Model
	Arthropod fixed 2.3%	Crustacean 1.4-2.6%	Mit. Combined Arthropod fixed 2.3%
Node A	1.37 (0.64-2.18)	1.57 (0.76-2.48)	1.83 (1.24-2.48)
Node B	2.33 (1.44-3.19)	2.66 (1.68-3.70)	2.31 (1.78-2.83)
Node C	0.17 (0.05-0.32)	0.19 (0.05-0.36)	0.11 (0.02-0.22)
Node D	2.07 (1.04-3.08)	2.35 (1.14-3.49)	2.11 (1.47-2.77)
Node E	2.04 (1.24-2.83)	2.36 (1.46-3.27)	2.26 (1.68-2.88)
Node B1	1.14 (0.68-1.62)	1.30 (0.76-1.84)	1.17 (0.83-1.54)
Node B2	1.05 (0.59-1.56)	1.21 (0.68-1.22)	1.07 (0.74-1.41)
Tree root	5.38 (3.45-7.46)	6.22 (4.10-8.70)	5.83 (4.46-7.09)

Age of the major clades shown in Figure 3, estimated using a Bayesian non-correlated relaxed molecular clock assuming a Yule tree prior, or a coalescent constant population size model and alternative calibration rates. Mean and 95% HPD values are given in million years.



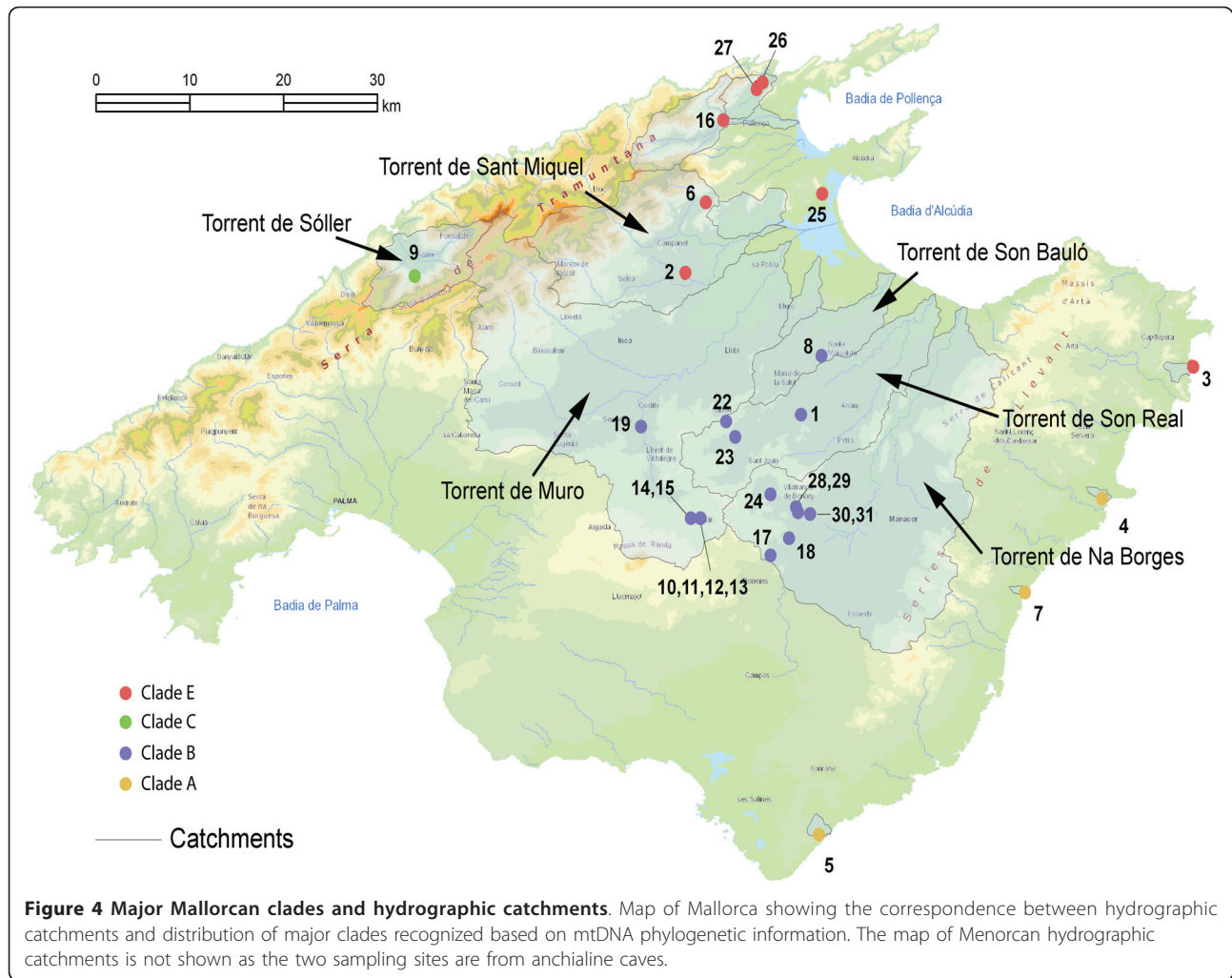


whereas most of the genetic variation occurred between calcretes. Although some populations from neighbouring calcretes placed in the same palaeodrainage channel are genetically similar (suggesting the occurrence of gene flow in the past), populations do not appear necessarily clustered according to palaeodrainage channel. This pattern could result from the occurrence of gene flow or range expansion between palaeodrainages in the past, before populations became isolated in particular calcretes. The ulterior isolation of populations could be associated with a major period of aridification that affected the region between 10 and 4 Ma [5].

Major cladogenetic events in *M. longipes* can be related to the succession of past sea-level changes in the Mediterranean (with the caveat of the limitations and errors associated with molecular-clock estimations). During the Tortonian (11.3 Ma), Mallorca and Menorca were invaded by an epicontinental sea that reduced the former to a cluster of small islands roughly corresponding to its current uplands, whereas the southern half of Menorca was probably completely submerged (see Figure 1) [27]. We assume that a single *M. longipes* population was then distributed along the entire continental shelf of the archipelago. This ancestral population overcame the phase of deposition of evaporites of the so-called “Messinian Salinity Crisis”, which was dated precisely at 5.96-5.59 Ma [28] and was coeval with a generalized marine regression episode that could have dried up the Mediterranean completely at that epoch. This mega-regression was probably the ultimate cause of the

split of the species into two major lineages: the former clade A, corresponding to the population that remained associated and followed the receded sea coastline towards the SE; and the remaining clades, which were presumably derived from the portion of the population that followed the receded coastline towards the N (see Figure 1). The age of the most recent common ancestor of clade A and its sister group (the remaining populations) has been estimated in our analyses at ca. 5.4-6.2 Ma using a relaxed molecular clock based on *cox1* sequences and a coalescent model.

Sometime between 4.2 and 2.7 Ma (upper-middle Pliocene; Figure 3), the populations from Menorca (node D), the Mallorcan central zone (node B), N Mallorca (node E) and W Mallorca (node C) became separated. The corresponding cladogenetic events might be linked to a single major marine transgression-regression cycle, such as that triggered by the upper Pliocene refilling of the depressed basins in the W Mediterranean area. The upper Pliocene transgression, which took place immediately after the Salinity Crisis, probably reached ca. +100 m above the current sea level in the Balearic area [29]. This might have enabled the species to reach the current central zone of Mallorca (Figure 1). More recently, our phylogeny shows that at the beginning of the late Pliocene, clades B, D and E experienced further splits that were followed by a differentiation of populations, with major secondary bifurcations occurring between 2.0 and 0.5 Ma. The uncertainties and large stochastic errors associated with the molecular



clock estimations preclude the correlation of the tree node ages with the datings of particular geological and climate transitional episodes. However, it is remarkable that the obtained tree topology is in agreement with the documented chronological succession of changes in the late Pliocene to mid-Pleistocene sea-level record in the North Hemisphere [30-32]. We suggest that recent cladogenetic events in *M. longipes* can be linked to two major cooling events roughly dated back at 2.5 to 3 and 1.2 to 0.85 Ma, respectively.

### Conclusions

Our data suggest that marine transgression-regression cycles (eustatic changes) may have induced the repeated range expansion, contraction and fragmentation of populations of *M. longipes*, which appears currently split into several isolated and genetically divergent lineages adapted to a broad spectrum of salinity conditions. This scenario could explain the difficulty in resolving the phylogenetic relationships among different lineages of

this amphipod, regardless of the method or sequence data set used: the rapid isolation and almost synchronous diversification of peripheral populations of the same ancestor in inland aquifers may have led to this situation. This hypothesis has been proposed to account for the distribution of particular anchialine and fresh groundwater taxa at various taxonomic levels and at larger geographical scales [33]. Our study stressed the importance of changes in sea level as a cause of deep intra-specific genetic divergence in thalassoid subterranean amphipods, a pattern that was apparently not accompanied by remarkable morphological differentiation [32].

### Methods

#### Sampling

One hundred and sixty-two specimens of *M. longipes* were collected from seven anchialine and one freshwater cave, and from 23 freshwater wells spanning the entire geographic range of the species (Figure 2), using a modified Cvetkov net [34] and hand-held plankton nets.

Individuals were preserved in 95% ethanol in the field and conserved at -20°C for subsequent molecular analyses. The sampling locations (with their geographical coordinates) and the number of individuals analysed for three mitochondrial and one nuclear marker are reported in Additional file 1. Three congeneric species were used as out-groups: the Moroccan *Metacrangonyx* sp. and *M. remyi* Balazuc & Ruffo, 1953 were collected in a well at Tamri (the coast of Agadir) and at the type locality located 1280 m above sea-level in the High Atlas, respectively. *M. ilvanus* Stoch, 1997 was collected in a well at Elba Island (Italy).

### Sequencing

Genomic DNA was isolated from whole specimens using the DNeasy Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendations. PCR was used to amplify a fragment of ~650 bp of the mitochondrial *cox1* gene using the primers described in [35] or, in some cases, using the specific primers metacoxF2 (5'-GAACTTAGATACCCWGGTAATTTGATYGG-3') and metacoxR2 (5'-TCAGTTAATAAYATAGTAATAGCYCC-3'). Fragments of three other genes were also amplified in a subset of 34 individuals: 400 bp of the 16S rRNA (*rrnL*) gene were amplified using the specific primers 16SmetaF (5'-RGTATTTTGACCGTGCTAAGG-3') and 16SmetaR (5'-TGTA AAAAATTAARGTTGAA-CAAAC-3'), 360 bp of the cytochrome b (*cob*) gene were amplified using the primers described in [36], and 325 bp of the nuclear gene Histone H3A were amplified using the primers from [37]. EMBL accession numbers for the *M. longipes* individuals and outgroup species for *rrnL*, *cob* and Histone H3A are FR846024-FR846060, FR846061-FR846096 and FR846097-FR846133, respectively.

PCR was performed on a PTC-100 thermocycler (MJ Research) using a reaction volume of 25 µl and amplification conditions consisted of one cycle at 94°C for 2 min and 40 cycles of 94°C for 30 s, 47-55°C for 30 s and 72°C for 1 min, followed by a final incubation step at 72°C for 10 min. Amplified products were purified with Invitex columns (Invitex GMBH, Berlin, Germany), according to the manufacturer's instructions. The fragments were sequenced in both directions using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit v. 2.0 and electrophoresed and detected on an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA, USA). Alignments were performed using MAFFT <http://www.ebi.ac.uk/Tools/mafft/index.html>, with default parameters.

### Phylogenetic analyses

Partition Bremer Support values were estimated using TreeRot v. 3 [38] and PAUP 4.0b10 [39]. Phylogenetic Bayesian analyses were conducted using MrBayes v.

3.1.2 [40]. We selected the model that fit the data best for each partition in the jModelTest [41] using the Bayesian information criterion. Models were tested for each of the three codon positions. The HKY+I model was selected for the first and second positions, and the GTR+G model for the third position in the case of the mitochondrial-protein-coding genes, whereas the HKY+I and F81+I models were used for *rrnL* and Histone H3A, respectively. Competing partition strategies were compared using Bayesian Information Criterion [42]. In the combined mitochondrial and nuclear data set, four partitions were favoured (first + second codon positions of *cox1* and *cob*, third codon positions of *cox1* and *cob*, *rrnL* and Histone H3A as separate partitions), whereas two partitions were selected in the case of *cox1*-only data sets (first + second vs third codon positions). Two independent runs were performed for each Bayesian search with default prior values, random trees and three heated and one cold Markov chains running for five million generations and sampled at intervals of 1000 generations. All parameters were unlinked and rates were allowed to vary freely over partitions. The burn-in and convergence of runs were assessed by examining the plot of generations against likelihood scores using the *sump* command in MrBayes. The convergence of all parameters in the two independent runs was also assessed using the Tracer program, v. 1.4 [43]. Trees resulting from the two independent runs (once burn-in samples were discarded) were combined in a single majority consensus topology using the *sumt* command in MrBayes, and the frequencies of the nodes in a majority rule tree were taken as *a posteriori* probabilities [40]. Maximum likelihood analyses using the above-mentioned partition schemes were performed using RAxML v. 7.0.4 implementing a fast bootstrapping algorithm [44]. Finally, we used Mesquite v. 2.74 [45] to reconstruct the *M. longipes* habitat character state at ancestral nodes (inland fresh vs. brackish groundwaters) using parsimony. In this analysis, we used the *cox1* phylogenetic tree (as it represents a full population sampling) and the observed habitat distribution among populations to minimize the number of steps of habitat change.

### Population analyses

A Mantel test was performed on genetic (*cox1*) and geographic distances of populations using the ZT program [46], to check for the occurrence of isolation by distance. Population diversity indices for the *cox1* data set, such as number of haplotypes, haplotype and nucleotide diversity, and pair-wise  $F_{ST}$  distances and their significance based on 10,000 permutations were obtained using ARLEQUIN v. 3.01 [47]. Populations represented by only one sequenced individual were excluded from the analyses. SAMOVA v. 1.0 [24] was used to identify geographical groupings that maximized genetic variance



between groups of populations ( $F_{SC}$ ). The method calculates F statistics ( $F_{SC}$ ,  $F_{ST}$  and  $F_{CT}$ ) using AMOVA [48] and identifies the optimum number of population groups for a set of sampled populations given a geographic distribution. We used 100 simulated annealing processes for each value of  $K$  from  $K = 2$  to  $K = 20$ . Neutrality tests were performed for individual populations calculating Fu's  $F_S$  [49] and the parameter  $R_2$  [50] using ARLEQUIN v. 3.01 and DnaSP v. 5.10.1 [51], respectively, with the latter assuming no recombination and 10,000 replicates. Simulations have shown that  $R_2$  and  $F_S$  are better at detecting population growth compared with other tests, the former being superior for small sample sizes [50].

### Estimation of divergence time

Two different strategies were explored to estimate population divergence times. First we enforced the standard mitochondrial arthropod rate fixed at 2.3% pair-wise divergence per million years (0.0115 substitutions per year and lineage [25], and secondly we implemented a mitochondrial rate range of 1.4 to 2.6% substitutions per million years, that was previously estimated for marine decapods and has been frequently applied to other crustaceans [6,26,52]. In both approaches, the *cox1* data set of the 162 sampled individuals was used applying an uncorrelated log-normal clock, assuming a coalescent model with constant population size as the best model fitting the data. BEAST [53] analyses were run starting from a random tree and using the models and partitions described for the MrBayes analyses. The remaining parameters (nucleotide frequencies and substitution model across partitions) and the rate-heterogeneity models were unlinked and estimated from the data. The search was set to 20 million generations, sampling every 1000 generations. The prior for the crustacean mitochondrial range in clock rate was implemented as a normal distribution with a mean of 0.01 substitutions per year per lineage, with maximum and minimum values of 0.013 and 0.007, respectively. The outputs of two independent runs were analysed using Tracer v. 1.4 after discarding the first 2 million generations. In another analysis, a reduced data set comprising the three combined mitochondrial genes from 34 individuals representing the major lineages was used and applied the fixed standard arthropod mitochondrial clock mentioned above but assuming a Yule model.

### Additional material

**Additional file 1: List of sampling sites.** Population labels, sampling sites, island, geographical position and number of specimens analysed for three mtDNA and one nuclear marker of *Metacrangonyx longipes*.

**Additional file 2: Bayesian *cox1* mtDNA tree.** Bayesian phylogenetic tree of *Metacrangonyx longipes* based on the *cox1* mitochondrial data set.

Values above nodes correspond to bootstrap values > 85% in maximum likelihood analyses (first number) and to posterior probability values > 0.95 (second number).

**Additional file 3: Ancestral habitat tracing on the Bayesian *cox1* mtDNA tree.** Parsimonious reconstruction of *M. longipes* habitat at ancestral nodes. Inland fresh groundwater and brackish groundwater populations are indicated in blue and yellow, respectively.

### Acknowledgements

We greatly appreciate support provided by Joan R. Bosch, Rafel Mas and Antoni Martínez Taberner to locate suitable wells in the Pollença, Búger and Ruberts areas, respectively, and by Lluç García, Alejandro Botello, Fernando Cánovas and Bartomeu Cañellas during fieldwork. Marta Fuster prepared the maps. The constructive criticism and suggestions made by Jean-François Flot and two anonymous reviewers considerably improved the final version of the manuscript. Research has been supported by Spanish grants CGL2006-01365, CGL2009-08256 and CGL2010-18616 of the Spanish Ministry of Science and Innovation and European Union FEDER funds. MMRB benefited from a FPI fellowship from the Spanish Ministry of Science and Innovation.

### Author details

<sup>1</sup>Departament de Biologia, Universitat de les Illes Balears, Edifici Guillem Colom, Campus Universitari, ctra. Valldemossa, km 7.5, 07122-Palma de Mallorca, Balearic Islands, Spain. <sup>2</sup>IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, c/Miquel Marquès, 21, 07190-Esporles, Balearic Islands, Spain. <sup>3</sup>Karst and Littoral Geomorphology Research Group, Universitat de les Illes Balears, Edifici Guillem Colom, Campus Universitari, ctra. Valldemossa, km 7.5, 07122-Palma de Mallorca, Balearic Islands, Spain.

### Authors' contributions

MMBR performed the laboratory work. MMRB, CJ and JP carried out the molecular genetic analyses and participated in sampling. CJ drafted the manuscript. JJF participated in geological analyses. DJ participated in sampling and produced the last version of the manuscript with CJ. DJ, CJ and JP conceived the study. All authors read and approved the final manuscript.

Received: 22 March 2011 Accepted: 26 July 2011

Published: 26 July 2011

### References

1. Poulson TL, White WB: **The cave environment.** *Science* 1969, **165**:971-981.
2. Juan C, Guzik MT, Jaume D, Cooper SJB: **Evolution in caves: Darwin's 'wrecks of ancient life' in the molecular era.** *Mol Ecol* 2010, **19**:3865-3880.
3. Proudlove G, Wood PJ: **The blind leading the blind: cryptic subterranean species and DNA taxonomy.** *Trends Ecol Evol* 2003, **18**:272-273.
4. Lefébure T, Douady CJ, Gouy M, Trontelj P, Briolay J, Gibert J: **Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments.** *Mol Ecol* 2006, **15**:1797-1806.
5. Cooper SJB, Bradbury JH, Saint KM, Leys R, Austin AD, Humphreys WF: **Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia.** *Mol Ecol* 2007, **16**:1533-1544.
6. Finston TL, Johnson MS, Humphreys WF, Eberhard SM, Halse SA: **Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape.** *Mol Ecol* 2007, **16**:355-365.
7. Notenboom J: **Marine regressions and the evolution of groundwater dwelling amphipods (Crustacea).** *J Biogeogr* 1991, **18**:437-454.
8. Botosaneanu L, Holsinger JR: **Some aspects concerning colonization of the subterranean realm - especially of subterranean waters: a response to Rouch & Danielopol, 1987.** *Stygologia* 1991, **6**:11-39.
9. Boutin C, Coineau N: **"Regression model", "Modèle biphasé" d'évolution et origine des microorganismes stygobies interstitiels continentaux.** *Rev Micropaléont* 1990, **33**:303-322.
10. Holsinger JR: **Pattern and process in the biogeography of subterranean amphipods.** *Hydrobiologia* 1994, **287**:131-145.

11. Murphy NP, Adams M, Austin AD: Independent colonization and extensive cryptic speciation of freshwater amphipods in the isolated groundwater springs of Australia's Great Artesian Basin. *Mol Ecol* 2009, **18**:109-122.
12. Trontelj P, Douady CJ, Fiser C, Gibert J, Spela G, Lefébure T, Sket B, Zaksek V: A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? *Freshwater Biol* 2009, **54**:727-744.
13. Kano Y, Kase T: Genetic exchange between anchialine cave populations by means of larval dispersal: the case of a new gastropod species *Neritilla cavernicola*. *Zool Scr* 2004, **33**(5):423-437.
14. Russ AD, Santos SR, Muir C: Genetic population structure of an anchialine shrimp, *Metabetaeus lohena* (Crustacea: Alpheidae), in the Hawaiian Islands. *Rev Biol Trop* 2010, **58**(1):159-170.
15. Calman WT: Crustacea- In: Lankester ER ed., A treatise on Zoology. *Adam & Charles Black, London* 1909, **7**(3):1-346.
16. Botosaneanu L: Stygofauna Mundi. A faunistic, distributional and ecological synthesis of the world fauna inhabiting subterranean waters (including the marine interstitial). *E J Brill, Leiden* 1986.
17. Danielopol DL: An essay to assess the age of the freshwater interstitial ostracods of Europe. *Bijdr Dierkd* 1980, **50**:243-291.
18. Rouch R, Danielopol DL: L'origine de la faune aquatique souterraine, entre le paradigme du refuge et le modèle de la colonisation active. *Stylogologia* 1987, **2**:345-372.
19. Danielopol DL, Bonaduce G: The origin and distribution of the interstitial Ostracoda of the species group *Xestoleberis arcturi* Triebel (Crustacea). *Cour Forsch Senck* 1990, **123**:69-86.
20. Page TJ, Humphreys WF, Hughes JM: Shrimps down under: Evolutionary relationships of subterranean crustaceans from Western Australia (Decapoda: Atyidae: *Stygocarids*). *PLoS ONE* 2008, **3**:e1618.
21. Stock JH, Illiffe TM, Williams D: The concept "anchialine" reconsidered. *Stylogologia* 1986, **2**:90-92.
22. Boutin C: Phylogeny and biogeography of metacrangonyctid amphipods in North Africa. *Hydrobiologia* 1994, **287**:49-64.
23. Jaume D, Christenson K: Amphibio-Atlantic distribution of the subterranean amphipod family Metacrangonyctidae (Crustacea, Gammaridea). *Contrib Zool* 2001, **70**:99-125.
24. Dupanloup I, Schneider S, Excoffier L: A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 2002, **11**:2571-2581.
25. Brower AVZ: Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc Natl Acad Sci USA* 1994, **91**:6491-6495.
26. Knowlton N, Weigt LA: New dates and new rates for divergence across the Isthmus of Panama. *Proc R Soc Lond B* 1998, **265**:2257-2263.
27. Pomar L: Reef geometries, erosion surfaces and high-frequency sea-level changes, upper miocene Reef Complex, Mallorca, Spain. *Sedimentology* 1991, **38**:243-269.
28. Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS: Chronology, causes and progression of the Messinian salinity crisis. *Nature* 1999, **400**:652-655.
29. Cuerda J, Sacarés J, Colom G: Hallazgo de terrazas marinas en la región de Lluchmayor (Mallorca). *Acta Geol Hisp* 1969, **4**:25-37.
30. Shackleton NJ, Opydyke ND: Oxygen isotope and palaeomagnetic stratigraphy of Equatorial Pacific core V28-238: Oxygen isotope temperatures and ice volumes on a 10<sup>5</sup> year and 10<sup>6</sup> year scale. *Quat Res* 1973, **3**:39-55.
31. Shackleton NJ, Backman H, Zimmerman H, Kent DV, Hall MA, Roberts DG, Schnitker D, Baldauf JG, Desprairies A, Homrighausen R, et al: Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* 1984, **307**:620-623.
32. Sosdian S, Yair Rosenthal: Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* 2009, **325**:306-310.
33. Lefébure T, Douady CJ, Mallard F, Gibert J: Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Mol Phylogenets Evol* 2007, **42**:676-686.
34. Cvetkov L: Un filet phréatobiologique. *Bull Inst Zool Mus Acad Bulgare Sci Sofia* 1968, **27**:215-218.
35. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek RC: DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 1994, **3**:294-299.
36. Barraclough TG, Hogan JE, Vogler AP: Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). *Proc R Soc Lond B* 1999, **266**:1061-1067.
37. Colgan DJ, Mclachlan A, Wilson GDF, Livingston S, Edgecombe GD, Macaranas J, Cassis G, Gray MR: Molecular phylogenetics of the Arthropoda: relationships based on histone H and U2 snRNA DNA sequences. *Aust J Zool* 1998, **46**:419-437.
38. Sorenson MD, Franzosa EA: *TreeRot*, version 3 Boston University, Boston, MA; 2007.
39. Swofford D: *PAUP\*: Phylogenetic analysis using Parsimony\* (and other methods)*, 4.0b10 Sunderland, MA: Sinauer Associates; 2002.
40. Huelsenbeck JP, Ronquist F: MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001, **17**:754-755.
41. Posada D: jModelTest: Phylogenetic Model Averaging. *Mol Biol Evol* 2008, **25**:1253-1256.
42. Schwarz G: Estimating the dimension of a model. *Ann Stat* 1978, **6**: 461-464.
43. Rambaut A, Drummond AJ: *Tracer v. 1.4*. 2007 [http://tree.bio.ed.ac.uk/software/Trace].
44. Stamatakis A, Ludwin T, Meier H: RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 2005, **21**:456-463.
45. Maddison WP, Maddison DR: Mesquite: a modular system for evolutionary analysis. Version 2.7.1. 2009 [http://mesquiteproject.org].
46. Bonnet E, Van de Peer Y: zt: A software tool for simple and partial Mantel tests. *J Stat Softw* 2002, **7**:1-12.
47. Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005, **1**:47-50.
48. Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992, **131**:479-491.
49. Fu Y-X: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997, **147**:915-925.
50. Ramos-Onsins SE, Rozas J: Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 2002, **19**:2092-2100.
51. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, **25**:1451-1452.
52. Kornobis E, Pálsson S, Kristjánsson BK, Svavarsson J: Molecular evidence of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland. *Mol Ecol* 2010, **19**:2516-2530.
53. Drummond AJ, Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007, **7**:214.

doi:10.1186/1471-2148-11-221

Cite this article as: Bauzà-Ribot et al.: Islands beneath islands: phylogeography of a groundwater amphipod crustacean in the Balearic archipelago. *BMC Evolutionary Biology* 2011 **11**:221.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit



# Mitogenomic Phylogenetic Analysis Supports Continental-Scale Vicariance in Subterranean Thalassoid Crustaceans

Maria M. Bauzá-Ribot,<sup>1</sup> Carlos Juan,<sup>1</sup> Francesco Nardi,<sup>2</sup> Pedro Oromí,<sup>3</sup> Joan Pons,<sup>4,\*</sup> and Damià Jaume<sup>4</sup>

<sup>1</sup>Departament de Biologia, Universitat de les Illes Balears, 07122 Palma, Spain

<sup>2</sup>Department of Evolutionary Biology, University of Siena, via Aldo Moro 2, 53100 Siena, Italy

<sup>3</sup>Departamento de Biología Animal, Universidad La Laguna, 38205 La Laguna, Tenerife, Canary Islands, Spain

<sup>4</sup>IMEDEA (CSIC-UIB), Mediterranean Institute for Advanced Studies, c/ Miquel Marqués 21, 07190 Esporles, Spain

## Summary

Many continental subterranean water crustaceans (“stygo-bionts”) display extreme disjunct distributions, where different species in the same genus are isolated on continents or islands separated by broad oceanic expanses [1]. Despite their freshwater habitat, most of these taxa appear to be most closely related to typical marine groups (“thalassoid” origin) [2]. Among the hadzioids—thalassoid amphipods including the stygobiont families Hadziidae, Pseudoniphargidae, and Metacrangonyctidae—several genera are restricted to inland groundwaters ranging from the Caribbean region to the Mediterranean and Middle East, including interspersed oceanic islands [3]. This distribution might have arisen from Tethyan vicariance [4–7] triggered by the sequential occlusion of the former Tethys Sea, a vast circumtropical ocean existing from the Middle Jurassic up to 20 million years ago (mya). Previous studies have been based on morphological analyses or limited DNA sequence data, making it difficult to test this hypothesis [8–10]. We used complete mitochondrial protein-coding gene sequences, mainly obtained by next-generation sequencing methods and a nuclear ribosomal gene to resolve the phylogeny and to establish a time frame for diversification of the family Metacrangonyctidae (Amphipoda). The results were consistent with the plate tectonics vicariance hypothesis, with major diversifications occurring between 96 and 83 mya.

## Results and Discussion

For many years, a key question in zoogeography has been the origin of the extremely disjunct distribution patterns of stygobiont crustaceans after the discovery of Caribbean lineages related to Mediterranean taxa [1]. This distribution pattern exhibited by many genera of thalassoid stygobiont crustaceans is currently explained as a vicariant process whereby plate tectonics caused the fragmentation of a marine ancestor’s range, once continuously distributed along the shores of ancient seas [4–6]. This might have been followed by secondary isolation and subsequent speciation in brackish or limnic groundwaters, a process triggered by episodes of sea-level oscillation or of tectonic uplift at coastal areas

[5, 7]. Other alternative hypotheses, such as broad-range, open-water marine dispersal are also possible for crustaceans with free-swimming larval stages [1, 11]. In addition, deep-sea dispersal along the crevicular medium associated with the circumglobal system of spreading zones has also been suggested to explain the presence of some of these taxa in geologically young oceanic islands [8, 12].

Testing among these alternative hypotheses requires robust phylogenies and accurate calibration points to derive a reliable estimation of divergence times [8, 9]. Transoceanic dispersal cannot be discarded a priori in stygobiont groups exhibiting a presumed Tethyan distribution but with shallow genetic divergences. Thus, to lend credence to the ancient vicariant origin hypothesis, divergence times between phylogenetic sister lineages placed at opposite sides of the Atlantic should be older than the establishment of deep-water conditions between Iberia and North America at about 95–110 million years ago (mya) [13, 14].

Here we studied the phylogeny of the Metacrangonyctidae, a family of stygobiont amphipod crustaceans with representatives in Hispaniola (Antilles), the Canary Islands, and around the Mediterranean region and the Middle East [15] (Figure 1). This monophyletic taxon [16] comprises two genera, *Metacrangonyx* Chevreux, 1909 and *Longipodacrangonyx* Boutin and Messouli, 1988, including a total of 18 species formally described—11 of which are limited to Morocco—plus at least 18 additional Moroccan taxa still awaiting formal description [16, 17]. The presence of two species of *Metacrangonyx* in the Caribbean region [18] offers an unmatched opportunity to test through phylogenetic analyses the role of dispersal and vicariance in the establishment of disjunct transoceanic distribution patterns in stygobiont crustaceans. We have sequenced the complete mitochondrial genome (~16 kb) (mitogenome) and the nuclear *SSU* ribosomal gene of 21 divergent lineages within 16 metacrangonyctid species and two outgroup taxa (Table 1; see Supplemental Experimental Procedures available online). The data set includes the recently characterized mitogenome of *Metacrangonyx longipes* Chevreux, 1909, from Mallorca (Balearic Islands) [19]. Species were chosen to cover the major genetic and geographic lineages revealed in a preliminary mitochondrial phylogenetic analysis based on cytochrome oxidase subunit 1 (COI) and 16S RNA (*rml*) gene fragments (Figure S1; Table S1). Most of the mitogenomes considered herein were obtained with 454 sequencing technology, using multiplexing by parallel-tagged libraries or by pooling untagged amplicons [20, 21]. Our main aims were (1) to derive a strongly supported phylogeny of the Metacrangonyctidae, sampling their full geographic distribution and (2) to estimate tree node ages using molecular dating techniques and paleogeographic calibration points. We aimed to test the hypothesis that species at opposite sides of the Atlantic had a vicariant origin, not a Mediterranean source followed by a secondary dispersal to the Caribbean.

## Mitogenomic Phylogenies Solve Evolutionary Relationships within the Metacrangonyctidae

A Bayesian tree (Figure S2) was built using the full set of mitochondrial protein-coding genes (MPCGs) by implementing the

\*Correspondence: [jpons@imedea.uib-csic.es](mailto:jpons@imedea.uib-csic.es)



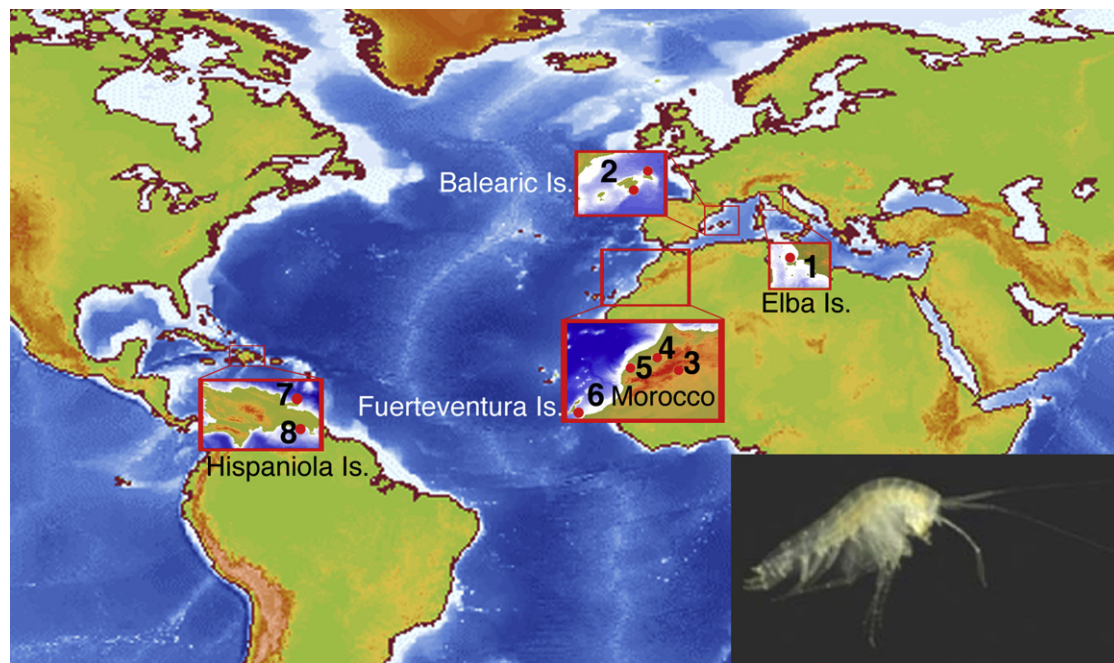


Figure 1. Metacrangonyctid Species Distribution

Map showing approximate geographic sampling locations of metacrangonyctid species for which whole mitochondrial genomes have been sequenced. See Table 1 and Table S1 for precise distribution of each taxon. 1, *Metacrangonyx ilvanus*; 2, *M. longipes*; 3, “*M. notenboomi*,” *M. goulmimensis*, *M. longicaudus*, *M. spinicaudatus*, “*M. paurosexualis*,” *M. panousei*; 4, *M. remyi*, “*M. boveei*”; 5, “*L. stocki*,” “*M. boutini boutini*,” “*M. nicoleae tamri*”; 6, *M. repens*; 7, *M. samanensis*; and 8, *M. dominicanus*. Inset shows *M. dominicanus* (photo by T.M. Iliffe). See also Figure S1.

best partition scheme and evolutionary models selected using the Bayesian Information Criterion (BIC; see Supplemental Experimental Procedures). Similar tree topologies were obtained using other partition schemes or the protein data set, but with discrepancies in the relationship among the taxa from Atlantic Morocco (they appeared as monophyletic in some analyses but paraphyletic in others) and in nodes relating the five insular species (Figure S2). The tree topology based on SSU sequences showed congruence with that obtained with MPCGs, but posterior probability values were low, particularly for basal nodes (Figure S2). The combined mitochondrial + SSU analyses supported paraphyly of the Atlantic Morocco lineages (Figure 2; Figure S2), but a Shimodaira-Hasegawa test showed that alternative topologies were not significantly different.

The tree recovered in the combined analysis (Figure 2) that we regard as our best phylogenetic hypothesis shows five strongly supported major clades within the family Metacrangonyctidae (here named A, B, C, D1, and D2). Each of these has a clearly delimited geographic projection, although B and C share the same overall area in Morocco. All tree topologies agreed on the assignment of an early divergence to “*M. boveei*” (tentative binomen) and *M. remyi* (clade A), which occur in the northern valleys and springs of the Western High Atlas in Morocco (Figures 1 and 2). Clade D appeared further subdivided into two strongly supported subclades: D1, present at both sides of the High Atlas in Morocco; and D2, embracing the five insular *Metacrangonyx* species from Mallorca-Menorca and Elba in the Mediterranean, Fuerteventura in the Canary Islands, and Hispaniola in the Caribbean. A Partition Bremer Support test taking each of the 13 MPCGs as a different partition suggested that the lack of resolution within the island subclade D2 arose from an absence of phylogenetic

signal and not from incongruence among gene partitions (Figure S2). A further Bayesian analysis implementing a polytomy prior in Phycas (see Supplemental Experimental Procedures) also led to the rejection of a fully resolved tree in favor of a hard polytomy.

Several recent attempts to resolve the phylogeny and to explain the disjunct distributions of some atyid shrimp [9], remipedes [8], and cirrolanid isopods [10] have been based only on partial mitochondrial and nuclear DNA sequences. Although mitogenome-based phylogenies represent a single locus and do not necessarily reflect the correct species tree, they have a considerably higher resolution power than partial (mitochondrial or nuclear) DNA sequences because of the large number of nucleotide positions considered and the high mutation rate exhibited by MPCGs. Our analysis of the phylogenetic relationships within the Metacrangonyctidae based on the 13 MPCGs with a partitioning by codon position plus the SSU nuclear marker produced an almost fully resolved tree except for the hard polytomy affecting insular clade D2.

#### Time Frame for Metacrangonyctid Diversification

Following the rejection of a strict molecular clock, we estimated node ages enforcing a relaxed molecular clock on the combined analysis topology assuming three independent substitution rates for each mitochondrial codon position. Two paleogeographic events were used to calibrate the tree. The divergence of the two lineages of *M. longipes* present on the Balearic archipelago was assumed to be associated with the complex geologic events that occurred in the Western Mediterranean from the Middle to the Late Miocene. Namely, the age for the complete separation of the Balearic Islands from other continental microplates detached from the Iberian

**Long-Term Vicariance in Cave Amphipods**

3

Table 1. Collection Sites and Mitochondrial Genome Sequence Information for Species Included in Analyses

Species	Locality	Mitogenome Length (bp)	SSU Length (bp)	Average Read Length (bp)	Coverage	Acc. N. Mitogenome	Acc. N. SSU
<i>M. dominicanus</i> Jaume and Christenson, 2001	Juan Dolio, S. Hispaniola (Dominican Rep.); well	14,543 <sup>d</sup>	2,511	625	4.3× <sup>a</sup>	HE860499	HE967299
<i>M. samanensis</i> Jaume and Christenson, 2001	Samaná, N. Hispaniola (Dominican Rep.); well	14,067 <sup>f</sup>	2,413	551	5.7× <sup>a</sup>	HE860505	HE967297
<i>M. repens</i> (Stock and Rondé-Broekhuizen, 1986)	Fuerteventura, Canary Is. (Spain); well	14,355	2,215	550	5.7× <sup>a</sup>	HE860495	HE967284
" <i>M. nicolae tamri</i> "	Aksri, NW Agadir (Morocco); spring near Talmat cave	13,517 <sup>e</sup>	1,027+1268	411	145× <sup>c</sup>	HE860511	HE967292-3
" <i>M. nicolae tamri</i> "	Tamri, N. Agadir (Morocco); well	14,644	2,415	595	59× <sup>b</sup>	HE860504	HE967294
" <i>M. boutini boutini</i> "	Timzelite, Souss Massa NP, S. Agadir (Morocco); well	13,301 <sup>e</sup>	2,357	433	107× <sup>c</sup>	HE860497	HE967295
<i>M. panousei</i> Balazuc and Ruffo, 1953	Agdz (Morocco); well	14,478 <sup>d</sup>	2,295	453	80× <sup>c</sup>	HE860510	HE967289
<i>M. goulmimensis</i> Messouli, Boutin, and Coineau, 1991	Lamkedmyia Meleh Jorf, NW Erfoud (Morocco); well	14,507 <sup>d</sup>	1,173	413	178× <sup>c</sup>	HE860500	HE967279
<i>M. goulmimensis</i> Messouli, Boutin, and Coineau, 1991	Ousroutou, E Rich, N. Errachidia (Morocco); well	14,602	2,179	454	100× <sup>c</sup>	HE860501	HE967278
<i>M. goulmimensis</i> Messouli, Boutin, and Coineau, 1991	Zouala maïsson, S. Errachidia (Morocco); well	14,353	1,922	423	165× <sup>c</sup>	HE860502	HE967280
<i>M. longicaudus</i> Ruffo, 1954	Lamkedmyia Meleh Jorf, NW Erfoud (Morocco); well	14,712 <sup>d</sup>	2,272	403	135× <sup>c</sup>	HE860509	HE967281
" <i>M. notenboomii</i> "	Maadid, NE Erfoud (Morocco); well	14,277 <sup>d</sup>	2,275	420	168× <sup>c</sup>	HE860513	HE967298
" <i>M. paurosexualis</i> "	Souk Tben, Haouz Plain, Marrakech (Morocco); well	12,542 <sup>e</sup>	2,370	424	94× <sup>c</sup>	HE860507	HE967291
<i>M. spinicaudatus</i> Karaman and Pesce, 1980	Souk Tben, Haouz Plain, Marrakech (Morocco); well	15,037	2,338	534	124× <sup>b</sup>	HE860506	HE967290
<i>M. remyi</i> Ruffo, 1953	Ijoukak, Western High Atlas (Morocco); spring at maison forestière	14,787 <sup>d</sup>	2,246	454	281× <sup>b</sup>	HE860512	HE967287
" <i>M. boveei</i> "	L'Ourika valley, Western High Atlas (Morocco); well	15,012	2,299	511	100× <sup>b</sup>	HE860498	HE967288
" <i>Longipodacrangonyx stocki</i> "	Tafraut (Morocco); well	12,924 <sup>e</sup>	337+1210	395	160× <sup>c</sup>	HE860496	HE967282-3
" <i>Longipodacrangonyx stocki</i> "	Arbaa-Sahe, SW Tiznit (Morocco); well	13,006 <sup>f</sup>	N/A	431	103× <sup>c</sup>	HE860508	N/A
<i>M. longipes</i> Chevreux, 1909	Mallorca, Balearic Is. (Spain); Cala Varques Cave	14,113	2,200	NA	NA	AM944817	HE967285
<i>M. longipes</i> Chevreux, 1909	Menorca, Balearic Is. (Spain); Cala Figuera cave	14,117	2,087	437	87× <sup>c</sup>	HE861923	HE967286
<i>M. ilvanus</i> Stoch, 1997	Elba Island (Italy); well	14,770 <sup>d</sup>	1,173	540	78× <sup>b</sup>	HE860503	HE967296
<i>Pseudoniphargus daviui</i> Jaume 1991	Cabrera, Balearic Is. (Spain); well	15,155	1,863	410	73× <sup>b</sup>	FR872383	HE967300
<i>Bahadzia jaraguensis</i> Jaume and Wagner 1988	Oviedo; S. Hispaniola (Dominican Rep.); cave	14,657	N/A	537	87× <sup>b</sup>	FR872382	N/A

See Table S1 for precise locations. Species names in quotes and not in italics denote taxa not formally described yet [17].

<sup>a</sup>Sanger method.

<sup>b</sup>Roche FLX/454 with tagging.

<sup>c</sup>Roche GS Junior with no tagging.

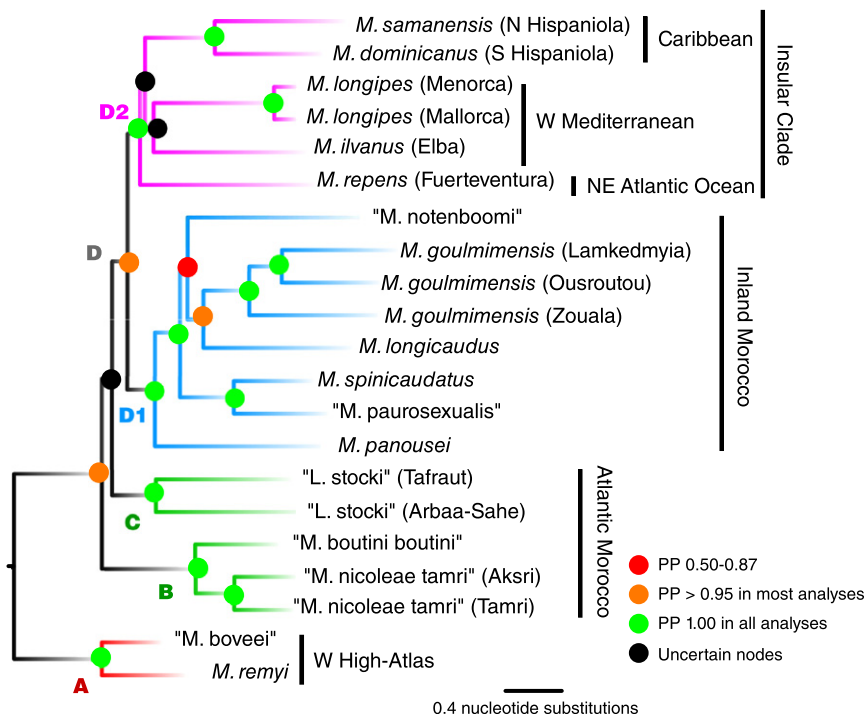
<sup>d</sup>Partial mitogenome due to A-T rich region was not completely sequenced.

<sup>e</sup>Partial mitogenome due to fragment comprising *rnlL-trnS* region was not completely sequenced.

<sup>f</sup>Partial mitogenome due to fragment including *rnlS* gene and A-T rich region was not completely sequenced.

rim (16 mya) and the Messinian Salinity Crisis (5.5 mya) were taken as the upper and lower bounds, respectively ([22, 23] and references therein). Furthermore, we assumed that the age of the node corresponding to the sister species *M. remyi* and "*M. boveei*," which dwell in neighboring valleys of the northern slopes of the Marrakech High Atlas in Morocco, cannot be younger than the uplift of this mountain range. According to [24], the Atlas domain remained submerged until the Middle Eocene (48.6–37.2 mya), the earliest uplift being dated as Late Eocene (37.2–33.9 mya), whereas the first significant folding extended through the entire Oligocene until the Early Miocene (25 mya). Thus, we propose to assign an age interval of 37.2–25.0 million years (my) to the last common ancestor of the two species.

Our chronogram (Figure 3) assigned an age of 96 my (with 95% higher posterior densities [HPDs] of 71–125 my) to the initial diversification that led to the contemporary metacrangonyctid lineages and suggests a mid-Cretaceous origin for the main clades of the family. Cross-validation showed that node ages estimated assuming the two foregoing paleogeographic events separately or by implementing different diversification models fell within the 95% HPDs of the estimations derived after considering the two calibration points altogether (Figure S3). An old lineage diversification at a remarkably small geographic scale took place in Morocco, where four of the five recognized monophyletic lineages concur, one of them resulting to be sister to the insular clade. Our estimated ages agree in general terms with those of Boutin [25] in the



**Figure 2. Mitogenomic + SSU Phylogenetic Tree of the Family Metacrangonyctidae**

The topology was obtained from Bayesian analysis of combined mitochondrial protein-coding genes and 18S rRNA (SSU). The best partitioning scheme was selected based on Bayesian Information Criterion (GTR+I +  $\Gamma$  in first/second positions as different partitions, HKY+I +  $\Gamma$  for third mitochondrial codon positions and GTR+I +  $\Gamma$  for SSU) (see Table S2). Monophyly of the Metacrangonyctidae was supported with maximum posterior probability in all analyses (outgroups not shown). Dots of different colors at nodes summarize posterior probability support values obtained using different methods (details on topologies and support values for each analysis are shown in Figure S2).

the position of the Western Mediterranean formed a continuous seaway, and their shores and islands were placed much closer to each other than at present [27]. Note that Hispaniola, the Canary Islands and the entire Western Mediterranean basin had not yet formed at that time. The current Caribbean configuration is no older than the Late Eocene in age [28];

assignment of a remarkably old age to the Metacrangonyctidae, with the differentiation of the major lineages in our phylogenetic reconstruction corresponding to the Cretaceous, c. 96–83 mya. This age range is coincident with the major Late Cretaceous Cenomano–Turonian transgression–regression cycle that affected most of the current Moroccan geography, and it was suggested to be the cause of individualization in continental groundwaters of some Moroccan lineages [25]. Our phylogeny shows that the insular species of *Metacrangonyx* (including the Caribbean and Mediterranean taxa) formed a strongly supported monophyletic group in all analyses (subclade D2), despite their extremely disjunct distribution. Remarkably, *M. repens* from Fuerteventura (Canary Islands) belongs to this insular clade and not to any of the geographically closer Moroccan clades. Two unresolved nodes account for the relationship between the Mediterranean insular species (Elba and Balearic Islands) and those of the Atlantic (Fuerteventura and Hispaniola). It has been suggested elsewhere that the difficulty in resolving phylogenetic relationships in subterranean amphipods might arise from the occurrence of sudden radiations associated with the synchronous colonization of groundwater by different populations of the same ancestor [2,26]. Our finding that the insular clade of *Metacrangonyx* diversified basally as a true polytomy is in agreement with the hypothesis of a simultaneous divergence ultimately leading to speciation in the isolated populations. The initial diversification of the insular clade is estimated to have occurred at 79 mya (95% HPDs 60–108 mya), a time frame compatible with the plate tectonics vicariance hypothesis if we consider the uncertainties associated with both the tectonic reconstruction of Tethys history and molecular clock estimations. Our estimated age for the divergence between the metacrangonyctids of Hispaniola and their Balearic sister group (77 mya; 95% HPDs 57–101 mya) lends additional support to this hypothesis. During that epoch, the Caribbean, the East Atlantic, and the portion of the Tethys Sea placed at

Fuerteventura, the oldest island of the Canary archipelago, dates back to 22 mya [29], whereas the entire Western Mediterranean basin formed at c. 20 mya [30]. However, there is compelling geologic evidence for the presence of drowned archipelagos and seamounts in the central East Atlantic since at least 60 mya [31–33]. These Paleo-Macaronesian islands were located much closer to the Western Mediterranean than today [33]. Ephemeral islands likely lasting a few million years each have been present in the Proto-Caribbean (volcanic islands, shallow banks and ridges) since the early Cretaceous [28]. Thus, the existence of these vanished archipelagos makes it likely that the ancestor of the insular lineage of *Metacrangonyx* was a shallow-water marine species that populated islands, shallow banks or strips of coast placed in this overall area, but not in the precise locations occupied by modern species.

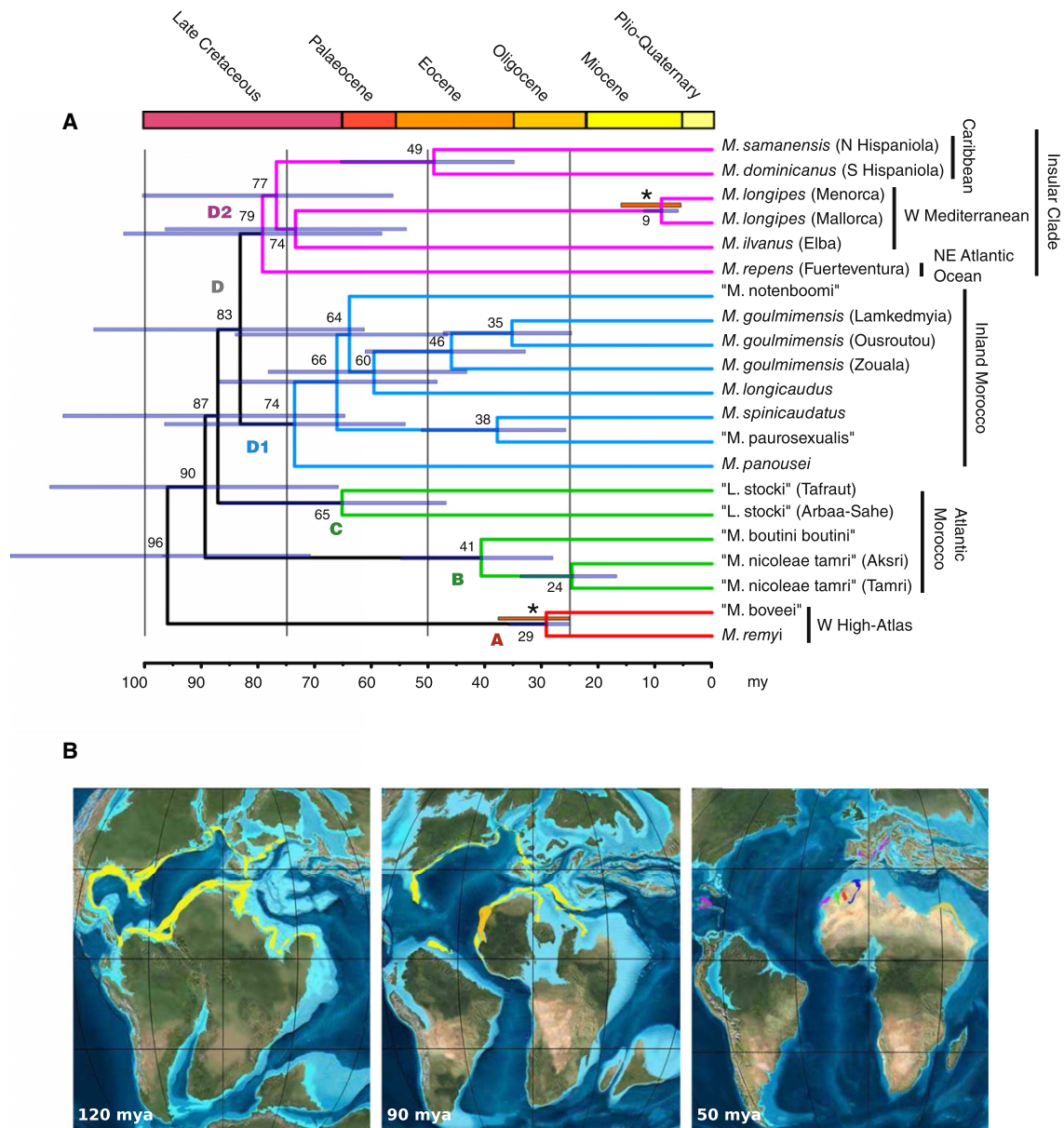
Across the Metacrangonyctid mitochondrial protein-coding genes, we found an average long-term evolutionary rate of pairwise sequence divergence of 10.9% per million years. This is almost five times higher than the “standard” 2.3% of arthropod mitogenomes [34] and beetle MPCGs [35], or other rates estimated for the COI gene in marine decapods (1.4%–2.6% per million years) that have been applied frequently to other crustaceans [36]. If the “standard” 2.3% rate is used the estimated ages would be even older, thus not contradicting the main conclusion of an ancient vicariance. The only similar estimate to our knowledge is the 20% COI rate per million years obtained in the Hawaiian stygobiont decapod *Halocaridina rubra*, an accelerated evolution that has been related to the strong genetic structure and to the frequent occurrence of population bottlenecks in this species [37].

Other stygobiont crustacean groups such as remipedes and some ostracod, copepod, thermosbaenacean, and decapod lineages also display a presumed Tethyan distribution [38]. A clarification of their molecular phylogenies and timing of cladogenesis could shed light on the origin of their distribution



**Long-Term Vicariance in Cave Amphipods**

5



**Figure 3. Time Frame for Diversification of the Metacrangonyctidae**

(A) Shown are the divergence times for the Metacrangonyctidae estimated from Bayesian analysis of all mitochondrial protein-coding gene sequences based on two paleogeographic calibration points and a Yule diversification model (see main text and [Supplemental Experimental Procedures](#)). Mean values are indicated on nodes, whereas horizontal bars across nodes represent the 95% highest probability density intervals. Asterisks identify node constraints, with their respective age ranges in red, implemented as flat priors in the analysis. See also [Figure S3](#) for ages estimated using single calibration points, strict clock, and Birth-Death models.

(B) Maps show global paleogeography at three different periods with corresponding putative metacrangonyctid distributions. Maps modified from Ron Blakey, NAU Geology (<http://jan.ucc.nau.edu/~rcb7/globaltext2.html>).

patterns in the context of the fragmentation of the Tethys Sea. Extensive mitochondrial data sets in combination with multiple nuclear loci obtained by next-generation DNA sequencing are a promising source of genetic information to unravel the process and timing of diversification of these stygobiont crustaceans.

**Accession Numbers**

Sequences obtained for this paper were deposited under EMBL accession numbers HE967026–HE967186 for COI, HE967187–HE967277 and HE970657–HE970663 for *rnrL*, FR872382–FR872383, HE860495–HE860513, and HE861923 for mitogenomes, and HE967278–HE967300 for *SSU*.

**Supplemental Information**

Supplemental Information includes three figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2012.09.012>.

**Acknowledgments**

We are greatly indebted to C. Boutin, N. Coineau, M. Messouli, M. Yacoubi-Khebiza, and M. Boulanour for support during fieldwork in Morocco, aid in species identification, and fruitful discussions. J.A. Ottenwalder and J.A. Alcover shared fieldwork in the Dominican Republic. J.M. Bichain and



A. Faille loaned Moroccan specimens collected in the frame of expedition Win-Timdouine 2008. We also thank F. Frati for hosting M.M.B.-R. at Department of Evolutionary Biology of the University of Siena. B. Emerson kindly advised and made comments on an earlier version of the manuscript. This work was supported by Spanish MCINN grant CGL2009-08256 and CSIC Intramural grant 2009301141, partially financed with EU FEDER funds. M.M.B.-R. benefited of a PhD Spanish FPI fellowship during the completion of this study.

Received: June 15, 2012

Revised: August 16, 2012

Accepted: September 5, 2012

Published online: October 11, 2012

## References

1. Stock, J.H. (1993). Some remarkable distribution patterns in stygobiont Amphipoda. *J. Nat. Hist.* 27, 807–819.
2. Notenboom, J. (1991). Marine regressions and the evolution of groundwater dwelling amphipods (Crustacea). *J. Biogeogr.* 18, 437–454.
3. Väinölä, R., Witt, J.D.S., Grabowski, M., Bradbury, J.H., Jazdzewski, K., and Sket, B. (2008). Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. *Hydrobiologia* 595, 241–255.
4. Stock, J.H. (1977). The taxonomy and zoogeography of hadziid Amphipoda, with emphasis on the West Indian taxa. *Stud. Fauna Curaçao Caribbean Isl.* 55, 1–130.
5. Stock, J.H. (1980). Regression mode evolution as exemplified by the genus *Pseudoniphargus* (Amphipoda). *Bijdr. Dierk.* 50, 104–144.
6. Holsinger, J.R. (1986). Zoogeographic patterns of North American subterranean amphipod crustaceans. In *Crustacean issues 4. Crustacean Biogeography*, R.H. Gore and K.L. Heck, eds. (Rotterdam, Boston: A.A. Balkema), pp. 85–106.
7. Holsinger, J.R. (1988). Trogllobites: The evolution of cave-dwelling organisms. *Am. Sci.* 76, 146–153.
8. Neiber, M.T., Hartke, T.R., Stemme, T., Bergmann, A., Rust, J., Iliffe, T.M., and Koenemann, S. (2011). Global biodiversity and phylogenetic evaluation of Remipedia (crustacea). *PLoS ONE* 6, e19627.
9. Page, T.J., Humphreys, W.F., and Hughes, J.M. (2008). Shrimps down under: Evolutionary relationships of subterranean crustaceans from Western Australia (Decapoda, Atyidae, *Stygocaris*). *PLoS One* 3, e1618.
10. Baratti, M., Filippelli, M., Nardi, F., and Messina, G. (2010). Molecular phylogenetic relationships among some stygobitic cirrolanid species (Crustacea, Isopoda). *Contrib. Zool.* 79, 57–67.
11. Holsinger, J.R. (1991). What can vicariance biogeographic models tell us about the distributional history of subterranean amphipods? *Hydrobiologia* 223, 43–45.
12. Boxshall, G.A. (1989). Colonization of inland marine caves by misophrioid copepods. *J. Zool. (Lond.)* 219, 521–526.
13. Sclater, J.G., Hellinger, S., and Tapscott, C. (1977). The paleobathymetry of the Atlantic Ocean from the Jurassic to the Present. *J. Geol.* 85, 509–552.
14. Jones, E.J.W., Cande, S.C., and Spathopoulos, F. (1995). Evolution of a major oceanographic pathway: the equatorial Atlantic. *J. Geol. Soc. London* 90, 199–213.
15. Boutin, C., and Messouli, M. (1988). *Longipodacrangonyx maroccanus* n. gen. n. sp., nouveau représentant du groupe *Metacrangonyx* dans les eaux souterraines du Maroc. *Crustaceana* 13 (suppl.), 256–271.
16. Messouli, M. (1988). Les Crustacés Amphipodes Souterrains du Groupe *Metacrangonyx*. Répartition, Systématique et Phylogénie. Ms Thesis. Université de Marrakech, Morocco, 234 pp.
17. Messouli, M. (1994). Evolution, Phylogénie et Biogéographie Historique des Metacrangonyctidae, Crustacés Amphipodes Stygobies du Nord de l'Afrique et des Régions Voisines. Ph.D Thesis. Université de Marrakech, Morocco, 312 pp.
18. Jaume, D., and Christenson, K. (2001). Amphi-Atlantic distribution of the subterranean amphipod family Metacrangonyctidae (Crustacea, Gammaridea). *Contrib. Zool.* 70, 99–125.
19. Bauzà-Ribot, M.M., Jaume, D., Juan, C., and Pons, J. (2009). The complete mitochondrial genome of the subterranean crustacean *Metacrangonyx longipes* (Amphipoda): a unique gene order and extremely short control region. *Mitochondrial DNA* 20, 88–99.
20. Meyer, M., Stenzel, U., Myles, S., Prüfer, K., and Hofreiter, M. (2007). Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res.* 35, e97.
21. Timmermans, M.J.T.N., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T.J., Pons, J., and Vogler, A.P. (2010). Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Res.* 38, e197.
22. Bauzà-Ribot, M.M., Jaume, D., Fornós, J.J., Juan, C., and Pons, J. (2011). Islands beneath islands: phylogeography of a groundwater amphipod crustacean in the Balearic archipelago. *BMC Evol. Biol.* 11, 221.
23. Bidegaray-Batista, L., and Arnedo, M.A. (2011). Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evol. Biol.* 11, 317.
24. Frizon de Lamotte, D., Zizi, M., Missenard, Y., Hafid, M., El Azzouzi, M., Maury, R.C., Charrière, A., Taki, Z., Benammi, M., and Michard, A. (2008). The Atlas System. In *Continental Evolution: The Geology of Morocco*, A. Michard, A. Chalouan, O. Saddiqi, and D. Frizon de Lamotte, eds. (Heidelberg: Springer), pp. 133–202.
25. Boutin, C. (1994). Phylogeny and biogeography of metacrangonyctid amphipods in North Africa. *Hydrobiologia* 287, 49–64.
26. Lefébure, T., Douady, C.J., Malard, F., and Gibert, J. (2007). Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Mol. Phylogenet. Evol.* 42, 676–686.
27. Smith, A.G., Smith, D.G., and Funnell, B.M. (1994). *Atlas of Mesozoic and Cenozoic Coastlines* (Cambridge: Cambridge University Press), pp. 103.
28. Iturralde-Vinent, M.A. (2006). Meso-Cenozoic Caribbean paleogeography: Implications for the historical biogeography of the region. *Int. Geol. Rev.* 48, 791–827.
29. Coello, J., Cantagrel, J.M., Hernan, F., Fuster, J.M., Ibarrola, E., Ancochea, E., Casquet, C., Jamond, C., Diaz De Teran, J.R., and Cendrero, A. (1992). Evolution of the eastern volcanic ridge of the Canary Islands based on new K-Ar data. *J. Volcanol. Geotherm. Res.* 53, 251–274.
30. Alvarez, W. (1972). Rotation of the Corsica-Sardinia microplate. *Nature* 235, 103–105.
31. Geldmacher, J., Hoernle, K., van den Bogaard, P., Zankl, G., and Garbe-Schönberg, D. (2001). Earlier history of the > 70-Ma-old Canary hotspot based on temporal and geochemical evolution of the Selvagens Archipelago and neighbouring seamounts in the eastern North Atlantic. *J. Volcanol. Geotherm. Res.* 111, 55–87.
32. Geldmacher, J., Hoernle, K., van den Bogaard, P., Duggen, S., and Werner, R. (2005). New 40K/39Ar age and geochemical data from seamounts in the Canary and Madeira volcanic provinces: Support for the mantle plume hypothesis. *Earth Planet. Sci. Lett.* 237, 85–101.
33. Fernández-Palacios, J.M., de Nascimento, L., Otto, R., Delgado, J.D., García-del-Rey, E., Arevalo, J.R., and Whittaker, R.J. (2011). A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *J. Biogeogr.* 38, 226–246.
34. Brower, A.V.Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci. USA* 91, 6491–6495.
35. Pons, J., Ribera, I., Bertranpetit, J., and Balke, M. (2010). Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Mol. Phylogenet. Evol.* 56, 796–807.
36. Kornobis, E., Pálsson, S., Kristjánsson, B.K., and Svarvarsson, J. (2010). Molecular evidence of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland. *Mol. Ecol.* 19, 2516–2530.
37. Craft, J.D., Russ, A.D., Yamamoto, M.N., Iwai, T.Y., Jr., Hau, S., Kahiapo, J., Chong, C.T., Ziegler-Chong, S., Muir, C., Fujita, Y., et al. (2008). Islands under islands: The phylogeography and evolution of *Halocaridina rubra* Holthuis, 1963 (Crustacean: Decapoda: Atyidae) in the Hawaiian archipelago. *Limnol. Oceanogr.* 53, 675–689.
38. Juan, C., Guzik, M.T., Jaume, D., and Cooper, S.J.B. (2010). Evolution in caves: Darwin's 'wrecks of ancient life' in the molecular era. *Mol. Ecol.* 19, 3865–3880.