Universitat de les Illes Balears

GENETIC LEGACY OF SEPHARDIC JEWS: PATERNAL AND MATERNAL LINEAGES OF CHUETA POPULATION

DOCTORAL THESIS
2017

Joana Francesca Ferragut Simonet

Universitat de les Illes Balears

DOCTORAL THESIS
2017

Doctoral Programme of Biotechnology, Genetics and Cell Biology

GENETIC LEGACY OF SEPHARDIC JEWS: PATERNAL AND MATERNAL LINEAGES OF CHUETA POPULATION

Joana Francesca Ferragut Simonet

Thesis Supervisor: Dra Antònia Picornell Rigo
Thesis Supervisor: Dr José A. Castro Ocón

Doctor by the Universitat de les Illes Balears

Universitat de les Illes Balears

Dra. Antònia Picornell Rigo and Dr. José A. Castro Ocón, professors of the Universitat de les Illes Balears,

DECLARE:

That the thesis titled GENETIC LEGACY OF SEPHARDIC JEWS: PATERNAL AND MATERNAL LINEAGES OF CHUETA POPULATION, presented by Joana Francesca Ferragut Simonet to obtain a doctoral degree, has been completed under our supervision and meets the requirements to opt for an European Doctoral degree mention.

For all intents and purposes, we hereby sign this document.

Palma, 26th June 2017

Dra. Antònia Picornell Rigo

Dr. José A. Castro Ocón

A la meva família

Agraïments

Tot i que semblava que no havia d'arribar mai, asseure's a escriure els agraïments és el moment d'adonar-se'n que el final d'aquesta etapa és a prop, i que toca donar les gràcies a tots aquells que d'una manera o una altra, heu fet d'aquesta tesi, una realitat. Tot i que no se si sabré expressar en paraules tot el que realment us voldria agrair a cada un de vosaltres, ho intentaré...

En primer lloc, he d'agrair infinitament als meus directors Dra. Antònia Picornell i Dr. José A. Castro per haver-me guiat, acompanyat, aconsellat i ensenyat tant en aquest camí. Antònia, digui el que digui...faré curt, segur! No només gràcies per la direcció de la tesi, sinó per creure en mi quan vaig entrar al despatx a tercer de carrera dient-te "vull dedicar-me a la Genètica", des d'aquell dia vas confiar en mi, i gràcies a això avui no només estic acabant una tesi en Genètica, sinó que m'has contagiat encara més l'entusiasme pel que estic fent. Sense deixar de banda, és clar, la feinada i les hores dedicades, també donar-te les gràcies per tots els moments que hem compartit fora de la Universitat: Sevilla, Cracòvia, Porto, Berlin, Israel...gràcies per deixar-me viure tot això al teu costat! Ets un gran exemple com a professional i com a persona, Gràcies per TOT! Pepe, pel teu suport dia a dia, pels ànims a qualsevol hora i per recordar-me com n'es d'important una actitud positiva. Per contagiar-nos les ganes de fer el que estam fent, i no cansar-te de recordar-nos que som afortunats. Per les hores de repassar taules infinites, dades, números i tots els detalls que sempre se'ns escapen i així i tot el cansament no fa que perdis el teu humor tan característic *(¿Cuál es el animal que está en la cola de la evolución?*). Als dos, us he d'agrair el vostre suport als moments durs, tant personals com professionals, i no haver-me deixat defallir en cap moment.

Aquest agraïment és extensiu a la Dra. Cori Ramon, per haver confiat en mi per fer feina al laboratori de Genètica, pels seus consells i suggeriments. Cori, gràcies per cuidar tant de tots nosaltres, preocupar-te i per fer sempre que tot sembli fàcil. Moltes gràcies. Fer la tesi al vostre grup ha estat un plaer.

La realització de la tesi també s'ha pogut dur a terme gràcies a l'ajuda econòmica del Govern Balear i els fons FEDER a través de les accions especials AAEE133/2009, AAEE009/2012, AAEE024/2014 i AAEE034/2015 i de la *Fundaçao para a Ciencia e a tecnología de Portugal (FCT)* al projecte PTDC/ATP-DEM/4545/2012.

A tots els membres de l'àrea de Genètica per haver-me contagiat les ganes i l'entusiasme per aquesta branca de la Biologia (els doctors Eduard Petitpierre, Carlos Juan, Joan Pons, Bàrbara Terrassa, Jose A. Jurado, Ana "Castillo", Marina Matas, Virginia Rodríguez, Kaoutar Bentayebi). També a tots els que m'heu acompanyat aquests anys en tantes hores de laboratori, anàlisis, i escriptura al seminari, amb vosaltres la Genètica ha estat encara més divertida!

Quan vaig arribar sols éreu na Marina i na Virginia, i de ben segur que es va haver acabat la tranquil·litat al seminari, gràcies per la paciència en els meus inicis al laboratori! En una de les meves partides...vaig tornar i ja estava el seminari ple! Tanta sort que vam aconseguir recol·locar-ho tot i afegir "la cinquena taula". ZIZAKA gràcies per comptar sempre amb sa tieta (tot i ser un trasto) com una més! Sabeu que he après i seguiré

aprenent MOLT de vosaltres, professional, i sobretot personalment. Jose, només te diré...gràcies per tot "*maestro yoda*"! Sergio...el gran descobriment del curs de Barcelona (i amb ell, un altre descobriment, Helena, perla valenciana, gràcies per ser la nostra "*mentor*"), ja ets una peça clau, ànims! Enanas, vosaltres si que heu donat vida a aquest seminari! Marta, gràcies per les xerradetes, vermuts, concerts, i per molta estona més! Iris, ets tan autèntica que és difícil no estimar-te (a tots els anteriors també, eh?), no canviïs brilli-brilli. Ana i Luz, gràcies per les xerradetes i els ànims, vosaltres sempre disposades!

A na Trinidad García, per la paciència i la cura que has tengut sempre a l'hora de passar fragments i seqüències pel seqüenciador. Però també pels dinars i cafès de cotorreig.

A gente de Porto, a o Professor Antonio Amorim muito obrigada por dar me a oportunidade de trabalhar em ou grupo de Genética Populacional do IPATIMUP e por compartilhar seu conhecimento científico e cultural (ainda não prove de fazer o Arroz de pato). A Luis, porque gracias a ti tuve la oportunidad de venir al IPATIMUP, porque me has cuidado mucho en Porto, y me has enseñado mucho en el terreno profesional, gracias también por dejarme ser parte de la familia en mis días en Porto. A Rui, por tu profissionalidade e tantas horas de *genemapper* compartilhadas (um, dois; um, um; dois dois...) e cuidar me tanto. Sofia Marques, muito obrigada por as horas de laboratório, pelas dicas, os pequenos almoços, os éclairs...a ti e as "meninas" por fazer me sentir como em casa. A professora Maria Joao Prata, Ines, Cíntia, Ana, Nadia e tudos los integrantes do grupo, muito obrigado! Porto já e a minha segunda cidade.

Donar les gràcies al GHEP-ISFG, per deixar-me formar part de la societat i aprendre tant d'un món fascinant com és la Genètica Forense. També he d'agrair a tots els professionals de la Genètica que a cada congrés, conferencia o curs, han compartit amb mi els seus coneixements i experiències, i fins i tot establert vincles per col·laboracions en els nostres estudis. Dra. Mercedes Aler, Dra. Lourdes Prieto, Dr. Walter Parson, Dr. Doron M Behar, Bennet Greenspan, etc. Thank you!

No puc deixar d'agrair als "historiadors": al Dr. Enric Porqueres per la seva ajuda amb les cites detallades d'esdeveniment històrics referents als Xuetes, i a Bernat Aguiló, no sols l'ajuda amb la bibliografia històrica, sinó les hores que ens ha dedicat, compartint amb nosaltres la seva feina en les genealogies, i sobretot, el seu ampli coneixement de la qüestió Xueta, fent que m'entusiasmés aprenent més de la història i les vivències d'aquesta comunitat.

Un agraïment infinit als membres del grup Memòria del Carrer, per la seva ajuda en la recollida de mostres, i a tots aquells Xuetes que voluntàriament han aportar sang o mucosa bucal, fent possible que aquesta tesi sigui una realitat, i pel seu gran interès en el treball. I no m'oblid de vosaltres, Luz i Carme, que m'heu acompanyat a voltar Mallorca a l'hora de treure mostres! Gràcies!

Tot i que és clar que la tesi l'he desenvolupada en l'àmbit professional, sense cap dubte no hagués estat possible sense el suport dels amics i la família. Sense ells el dia a dia no hagués estat el mateix.

AMICS, als de tota la vida, als que m'he anat trobant i us heu quedat, als "Biòlegs" i respectius. Als que per diversos motius ja no sou al meu dia a dia, però sí formeu part d'aquests anys i del que sóc. Als que ja us ho he dit mil vegades, i no me cansaré de fer-ho (*"tu trobes"?*). TOTS, que d'una manera o una altra, demanant-me què feia exactament, donant-me suport incondicional, anant de canyes o a passejar, però sempre preocupant-se i contents de saber que jo estava fent el que sempre havia volgut, GRÀCIES.

La família...que de ben segur no tenc les paraules i encara que n'escrigui alguna no seran suficients, però pot ser mai us ho he dit, i ara tenc l'oportunitat de fer-ho! Sense el vostre suport, ni hagués acabat la carrera, ni hagués recorregut món, ni hagués començat un doctorat...no cal dir-vos els motius, però el vostre recolzament ha fet possible que hagi complert els meus somnis. Tio Jaume, Madrina, tot i que la vida no us ho posat gens fàcil, sempre heu aconseguit anar endavant, i no tan sols això sinó que hi heu estat perquè jo, pogués prendre exemple i fes el mateix, se que no ho dic tot el que deuria però sou un pilar fonamental i un exemple a seguir. Padrino, Ventu, tampoc us ho faig saber amb paraules, però què faria jo sense vosaltres?...gràcies per fer-me sentir una més dels vostres tres fills! Kika, Pablo, gràcies també a vosaltres, primer per vigilar que no me desbaratàs massa en aquells dies per DDBÒ, i després per cuidar-me i fer-me madrina del vostre minitrastito! Padrina, tu també tens culpa de que sempre procuri anar pel món amb una rialla! Als cosins....Primo, gràcies per ser-hi sempre, sigui amb un cubata a la mà, quan tenc un problema amb el cotxe, o per fer la xerradeta! Joana, Antònia, sou les meves germanetes grans, gràcies per ser-hi i cuidar-me com a tal. I a les mini (Paula i Francina)...que han tornat l'alegria a la família i aconsegueixen que qualsevol moment, sigui una festa! GRÀCIES A TOTS.

PAPÀ...probablement tots els que me coneixen, saben meravelles de tu, de tot el que has fet per jo (per les dues) sempre, i en especial als darrers 10 anys, me sap greu no dir-t'ho a tu directament més sovint, però no hi ha RES, pel que no t'hagi d'estar agraïda.

I a tu peque...aquests darrers anys, crec que hem après ses dues a recolzar-nos, i agrair-nos el tenir-nos una a l'altra (ja era hora...hehe), però que sàpigues que aquesta tesi, també te l'he d'agrair a tu, perquè tot i ser sa petita, m'has donat un grapat de lliçons, gràcies per la teva capacitat de superació! *"Compta amb mi en els dies de lluita, si l'esperança et descuida. Als mals passos hi haurà uns braços, compta amb mi"*.

I per acabar, als que ja no hi sou, Jaumet, Abuelos, Padrí, i especialment i sobretot tu, MAMÀ (*Ja sols em queda protegir el què em vas deixar i creure...Que tot està per fer. Hem de caminar malgrat que avui ens faci mal, sobrevolar les pors i mirar endavant*), us tenc presents cada dia, i aquesta tesi l'he feta pensant que us sentiríeu orgullosos de mi, i que ho esteu de saber, que ho he aconseguit.

A TOTS, aquesta tesi també és un trosset vostra,

Joana Francesca

Summary

Chuetas are a group of descendants of the Jewish inhabitants of Majorca (Balearic Islands, Spain). Their historical origin dates back to the Sephardic communities that lived in Spain during the Diaspora. Despite their official conversion to Christianity (1391-1435) some of them maintained Crypto-Jewish practices. The descendants of those who were convicted in the inquisitorial processes due to these practices, in the 17th century, are the so-called Chuetas. The collective consciousness of their origin was preserved and historically they were discriminated against and isolated from the *old-Christian* Majorcan population until the middle of the twentieth century. The main characteristics that define belonging to the Chueta population are the fifteen surnames that are traditionally known in Majorca as Chuetas, and their high rate of endogamy.

The main points of this thesis on the genetics of Chuetas have been: a) to investigate whether cultural isolation has led to the impoverishment of genetic diversity in maternal and paternal lineages; b) to test the evidence of their supposed Sephardic origin; c) to estimate the extent of admixture with the host Majorcan population, and d) to evaluate whether there has been an asymmetrical sex-biased contribution from the parental populations to the Chuetas. In order to resolve these questions, different human genome regions have been studied in this work: autosomal markers (Indels); X-chromosome (STRs, Alu insertions, and Indels); mitochondrial DNA (D-loop, SNPs, and complete mitogenomes); and Y-chromosome (STRs and SNPs).

The results of all the markers studied show statistically significant differences between Chuetas and their host population, and also compared to other Jewish groups. Specific databases for populations with Jewish origin have been established for the 38 autosomal Indels and 53 X-chromosome markers, which is needful in forensic casework since they are differentiated populations. X-chromosome results suggest a sex-biased admixture process, with a genetic flow between Jewish communities mediated preferentially by males, and with preferential introgression from females of the host population. Mitochondrial DNA in Chuetas shows high values of diversity. Even though European haplogroups have been found in their maternal lineages, clear signatures of their Middle Eastern original are also present. The hallmark is the rare R0a2m, the modal haplogroup of Chuetas, which has a private mutation that defines a sub-branch not previously described in any other population. Y-chromosome results reveal a haplogroup composition very similar to the Sephardic Jews, with high frequencies of the Middle Eastern haplogroups J1 and J2 and a reduced presence of haplogroup R, the most common one in European populations. Besides, in the whole work, not only is the Sephardic influence clear, but also contact with North African and Ashkenazi Jews is detected in the Chueta population gene pool.

In conclusion, there does not exist an important reduction of genetic diversity either in mitochondrial DNA or Y-chromosome in Chuetas, as expected in a small-sized isolate population. Regarding lineages, they point towards a Middle Eastern ancestral signature along with a moderate degree of introgression from the host population, which seems to be higher in maternal than in paternal lines.

Resum

Els xuetes són els descendents d'un grup dels jueus que visqueren a Mallorca (Illes Balears, Espanya). Els seus orígens històrics es remunten a les comunitats sefardites assentades a Espanya durant la Diàspora. Tot i la conversió al cristianisme dels jueus mallorquins (1391-1435), hi hagué un grup que mantingué pràctiques criptojueves, i són els descendents dels condemnats per aquestes pràctiques, als processos inquisitorials del segle XVII, els coneguts com a xuetes. La memòria col·lectiva dels seus orígens es va conservar i varen estar discriminats i aïllats de la població "cristiana vella" de Mallorca fins a mitjan segle XX. Les característiques principals de la comunitat xueta són els quinze cognoms coneguts tradicionalment a Mallorca com a xuetes, i el seu alt grau d'endogàmia.

Els objectius principals d'aquesta tesi han estat: a) investigar si l'aïllament cultural de la població ha suposat un empobriment de la diversitat genètica en línies maternes i paternes; b) confirmar l'evidència del seu origen sefardita; c) estimar el grau de mescla amb la població hoste mallorquina; i d) avaluar si hi ha hagut una contribució asimètrica esbiaixada pel sexe de les poblacions parentals en els xuetes. Per resoldre aquestes qüestions, en aquest treball s'han estudiat diferents regions del genoma humà: marcadors autosòmics (*Indels*), cromosoma X (*STRs*, insercions *Alu* i *Indels*), ADN mitocondrial (Regió Control, *SNPs* i mitogenomes complets) i cromosoma Y (*STRs* i *SNPs*).

Els resultats, en tots els marcadors estudiats, mostren diferències significatives entre els xuetes i la resta dels mallorquins, i també amb altres poblacions jueves. S'han creat bases de dades específiques per poblacions d'origen jueu dels 38 *Indels* autosòmics i dels 53 marcadors de cromosoma X, necessàries en la pràctica forense, ja que són poblacions diferenciades. Els resultats de cromosoma X suggereixen un biaix influït pel sexe en el procés de mescla, amb un flux genètic entre comunitats jueves dut a terme principalment per mascles, i una introgressió preferencial de dones de la població hoste. Els resultats d'ADN mitocondrial en els xuetes presenten alts nivells de diversitat. Tot i que es troben haplogrups europeus en les línies maternes, també hi ha senyes clares del seu origen a l'Orient Mitjà. El tret més distintiu és que l'haplogrup modal en els xuetes és el rar llinatge R0a2m, que presenta una mutació específica que defineix una nova subbranca, no descrita prèviament en cap altra població. Els resultats de cromosoma Y mostren una composició d'haplogrups molt similar a la dels jueus sefardites, amb altes freqüències dels haplogroups J1 i J2 (propis de l'Orient Mitjà) i una falta d'R, el més comú en poblacions europees. A més, en aquest treball, no sols s'han trobat influencies sefardites en els xuetes, sinó que també s'ha detectat contacte amb jueus nord-africans i asquenasites.

Per tant, es pot concloure que no existeix a la població xueta una important reducció de la diversitat genètica ni a l'ADN mitocondrial ni al cromosoma Y, com caldria esperar en una població petita i aïllada. Pel que fa als llinatges observats, mostren senyes d'un origen a l'Orient Mitjà amb un grau moderat d'introgressió de la població hoste, que sembla més elevat a les línies maternes que a les paternes.

Resumen

Los Chuetas son descendientes de parte de los judíos que vivieron en Mallorca (Islas Baleares, España), y cuyos orígenes históricos se remontan a las comunidades sefardíes que se asentaron en España durante la Diáspora. A pesar de su conversión al cristianismo (1391-1435), un grupo mantuvo prácticas criptojudaicas. Los descendientes de los condenados por la Inquisición en el siglo XVII por éstas prácticas son los llamados Chuetas. La consciencia colectiva de sus orígenes se mantuvo, siendo discriminados y aislados por los "cristianos viejos" de Mallorca hasta mediados del siglo XX. Las características principales que distinguen a esta comunidad son los 15 apellidos conocidos tradicionalmente en Mallorca como Chuetas y su alto grado de endogamia.

Los principales objetivos de esta tesis han sido: a) investigar si el aislamiento cultural ha provocado un empobrecimiento de la diversidad genética en las líneas maternas y paternas; b) confirmar la evidencia de su supuesto origen sefardí; c) estimar el grado de mezcla con el resto de la población mallorquina, y d) evaluar si ha habido una contribución asimétrica sesgada por el sexo de las poblaciones parentales en los Chuetas. Para abordar estas cuestiones en este trabajo se han estudiado diferentes regiones del genoma humano: marcadores autosómicos (*Indels*), cromosoma X (*STRs*, inserciones *Alu* e *Indels*), ADN mitocondrial (Región Control, *SNPs* y mitogenomas completos), y cromosoma Y (*STRs* y *SNPs*).

Los resultados en todos los marcadores estudiados muestran diferencias significativas entre los Chuetas y el resto de los mallorquines, y también con otras poblaciones judías. Se han creado bases de datos específicas para poblaciones de origen judío para los 38 *Indels* autosómicos y para los 53 marcadores de cromosoma X, necesarias para la práctica forense, ya que son poblaciones diferenciadas. Los resultados del cromosoma X sugieren un sesgo influido por el sexo en el proceso de mezcla, con un flujo genético entre comunidades judías llevado a cabo principalmente por hombres, y una introgresión preferencial de mujeres de la población huésped. Los resultados del ADN mitocondrial en Chuetas presentan niveles altos de diversidad. Aunque se detectan haplogrupos europeos en las líneas maternas, hay señales claras de su origen en Oriente Medio. Su rasgo más característico es que presentan un haplogrupo modal tan poco frecuente como es el R0a2m, que además presenta una mutación específica que define una nueva sub-rama no descrita previamente en ninguna otra población. Los resultados del cromosoma Y muestran una composición de haplogrupos muy similar a la de los sefardíes, con altas frecuencias de los haplogrupos J1 y J2 (propios de Oriente Medio) y la falta de R, el más común en poblaciones europeas. Además, en este trabajo no solo se han encontrado influencias sefardíes en los Chuetas, sino también se han detectado contactos con judíos norteafricanos y asquenazis.

En resumen, se puede concluir que no existe una importante reducción de la diversidad genética ni en ADN mitocondrial ni en el cromosoma Y en los Chuetas, como sería de esperar en una población pequeña y aislada. En cuanto a los linajes detectados muestran señales de un origen en Oriente Medio, con un grado moderado de introgresión de la población huésped que parece mayor en las líneas maternas que en las paternas.

List of publications

1. **Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations.**
Ferragut JF; Bentayebi K; Castro JA; Ramon C and Picornell A
International Journal of Legal Medicine. 129(2): 253–255 (2015).

2. **Genetic diversity of 12 X-chromosomal short tandem repeats in Jewish populations.**
Ferragut JF; Castro JA; Ramon C and Picornell A
Forensic Science International: Genetics Supplement Series. 5: e327–e329 (2015).

3. **Founding mothers of Chueta population.**
Ferragut JF; Marques SL; Ramon C; Castro JA; Amorim A; Alvarez L and Picornell A
Forensic Science International: Genetics Supplement Series. 5: e492–e494 (2015).

4. **Genetic diversity of 38 insertion-deletion polymorphisms in Jewish populations.**
Ferragut JF; Pereira R; Castro JA; Ramon C; Nogueiro I; Amorim A and Picornell A
Forensic Science International: Genetics. 21: 1–4 (2016).

5. **Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alus and STRs.**
Ferragut JF; Bentayebi K; Pereira R; Castro JA; Amorim A; Ramon C and Picornell A
Forensic Science International: Genetics (Submitted).

6. **A GHEP-ISFG collaborative study on the genetic variation of 38 autosomal Indels for human identification in different continental populations.**
GHEP-ISFG collaborative exercise.
Forensic Science International: Genetics (Submitted).

7. **Maternal gene pool in Chuetas: Middle Eastern legacy and a novel sub-branching of the rare haplogroup R0a2m.** (In preparation)

8. **Y-chromosome lineages in Chuetas reveal the maintenance of their Middle Eastern genetic ancestry.** (In preparation)

List of abbreviations

AD	Anno Domini
AIM	Ancestry Informative Markers
AMOVA	Analysis of MOlecular VAriance
ASH	Ashkenazi Jews
BC	Before Christ
BCE	Before Christ Era
BI	Balearic Islands
bp	Base pair
CHU	Chuetas
CMH	Cohen Modal Haplotype
CR	Control Region
CR	Credible Region
CRS	Cambridge Reference Sequence
DC	Discrimination Capacity
ddNTP	Dideoxynucleotide
DLB	Denaturing Lysis Buffer
DNA	DeoxyriboNucleic Acid
ERDF	European Regional Development Fund
Exo-SAP	Exonuclease SAP
FCT	Foundation for Science and Technology
GD	Genetic Diversity
GEPH-ISFG	*Grupo de Habla Española y Portuguesa* of the ISFG
HD	Haplotype Diversity
Het	Heterozygosity
Hg	Haplogroup
HLA	Human Leukocyte Antigen
HMP	Haplotype Match Probability
Ht	Haplotype
HVRI	Hypervariable region I
HVRII	Hypervariable region II
HW	Hardy-Weinberg
HWE	Hardy-Weinberg Equilibrium
i3S	*Instituto de Investigação e Inovação em Saúde*
IBI	Ibiza
Indel	Insertion Deletion marker
IPATIMUP	Institute of Molecular Pathology and Immunology at the University of Porto
ISFG	International Society of Forensic Genetics
ISOGG	International Society of Genetic Genealogy
IUNICS	Institut Universitari d'Investigació en Ciències de la Salut
Kb	Kilobase
Kya	Thousand years ago
LD	Linkage Disequilibrium
LG	Linkage Group
LIZ	Lane Internal Size
MAJ	Majorca
Mb	Megabase

MCMC	Markov Chain Monte Carlo
MDS	MultiDimensional Scaling
MEC	Mean Exclusion Chance
MEJ	Middle Eastern Jews
MIN	Minorca
ML	Maximum Likelihood
MP	Match Probability
MPD	Mean Pairwise Differences
MRCA	Most Recent Common Ancestor
mtDNA	Mitochondrial DNA
N	Number
NA	Not Available
NA	Number of different Alleles
NAJ	North African Jews
N_e	Effective population size
Np	Nucleotide position
NRY	Non-Recombining Y
PAR	PseudoAutosomal Regions
PCR	Polymerase Chain Reaction
PD	Power of Discrimination
PE	Power of Exclusion
PhD	Philosophiæ doctor
PI	Paternity Index
PIC	Polymorphism Information Content
POPH	*Programa Operacional Potencial Humano*
rCRS	revised Cambridge Reference Sequence
RFLP	Restriction Fragment Length Polymorphism
RNA	RiboNucleic Acid
rRNA	ribosomal RNA
RSRS	Reconstructed Sapiens Reference Sequence
SAP	Shrimp Alkaline Phosphatase
SBE	Single-Base Extension
SDS	Sodium Dodecyl Sulfate
SEP	Sephardic Jews
SINE	Short Interspersed Nuclear Element
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
TBE	Tris-borate-EDTA (Ethylene Diamine Tetraacetic Acid)
TE	Tris-EDTA (Ethylene Diamine Tetraacetic Acid)
TMRCA	Time to the Most Recent Common Ancestor
TPI	Typical Paternity Index
tRNA	transfer RNA
UEP	Unique Event Polymorphism
UH	Unique Haplotype/Haplogroup
UV	Ultra Violet light
VAL	Valencia
W	Paternity Probability
YCC	Y-Chromosome Consortium

Index

Introduction

1. Introduction

1.1. Chueta population

The Chueta population is made up of descendants of a group of converted Jews from Majorca, who were convicted in the inquisitorial processes in the last quarter of the 17th century due to their Crypto-Jewish practices. The collective consciousness of their origin was preserved and historically they were stigmatized and segregated, so that until the first half of the 20th century, they practiced strict endogamy.

The word '*Chueta*' appeared documented for the first time in the inquisitorial processes in the 17th century, as an expression used by the prosecuted to refer to themselves (Porqueres, 2001). Its etymology is disputed and there are two hypotheses. The first and most accepted one, states that it comes from '*juetó*', diminutive of '*jueu*', Jew in Catalan (de Muntaner, 2002). The other claims that the word comes from a derivate of the word '*xulla*', bacon in Catalan, referring to the popular belief that Chuetas ate pork to demonstrate that they were not Jews anymore (Moore, 1976; Pons, 1984).

Chuetas have also been known as '*del Segell*' (from Segell), the name of a street on which many lived, or '*del carrer*' (from the street). In modern times, this name has been related to the '*carrer de l'Argenteria*' (silversmiths' street), a street in the neighbourhood where the majority of the Chueta lived, which has the name of one of their traditional occupations.

The main characteristics that define belonging to the Chueta population are the fifteen surnames that are traditionally known as Chuetas and their high rate of endogamy. Other features were the fact that in the past many Chueta families continued living in the same neighbourhood as their Jewish ancestors and that they mainly worked in a reduced number of specific jobs, related to jewelry and trade (Bestard, 1985).

Chuetas are, together with the Crypto-Jewish communities in Portugal (Nogueiro et al., 2015b), the only direct descendants of the original Sephardic population, and the characteristics of this population have kept them isolated from other Jewish and non-Jewish populations because, although for the last five centuries Chuetas have lived as Christians and, therefore, without religious barriers with the Majorcan population, social discrimination has acted as a cultural barrier to intermarriage.

1.1.1. History of the Chuetas

The first Jewish settlements in the Balearic Islands seem to date from the first century AD, after the destruction of Jerusalem by the Roman emperor, Titus; although the oldest remains found date back to the 4th or 5th century: some tombs with Hebrew inscriptions (Figure 1) and an encyclical from bishop Severus in 418 AD (Cortés, 1985). During the Arab domination (10th to 13th centuries) it is known that an important Jewish community lived in Medina Mayurqa (Palma) (Moore, 1976).

Figure 1. Hebraic inscriptions (Semuel, R. Haggay's son) in a tomb dated from 4th–5th century AD (Font, 2007).

The Christian conquest of the island by King James I in 1229 not only respected the existence of the Majorcan Jewish community but increased it, with Jews arriving from the Peninsula Mainland (Assis, 1992), and new aljamas appearing in some villages on the island. King James I ensured that the Jews were able to improve their quality of life and their businesses. The abilities of the Jews in terms of trade were essential to the king's idea of making Majorca a strategic point of trade in the Mediterranean Sea. At this time, Jews and Christians lived together for the economic and cultural development of the island, as the well-known Majorcan cartographic school reflects (Harwood, 2006; Llompart, 2011; Chacón, 2013), and the Jewish community was able to freely maintain its religion, traditions and organization.

In the middle of the 14th century, under the reign of Peter IV, the loss of Majorcan prosperity entailed the end of the good relationship between Jews and Christians (Font, 2007). In 1391, a rural revolution that focused in the beginning against the ruling class, ended up with an assault on the '*Call*', the so-called Jewish ghetto. Around 300 Jews died and there was an important loss of goods, properties and commercial archives. As a consequence, a number of Jews escaped to the nearest North African coast, and others submitted to baptism (Rozenberg, 2010). Even though the attack was not meant to be against the Jewry it was not an isolated case, because pogroms were taking place all over Spain at that time. The exodus and conversions continued from 1391 to 1435, due to pressure from the Roman Catholic Church and especially to some priests' preaching, for example that of Saint Vincent Ferrer. The last massive conversion occurred in Majorca in 1435, so officially there were no Jews in the island fifty-seven years before the official expulsion of the Jews from Spain (1492).

After the conversions, the so-called new Christians ('*conversos*') appeared, who mixed with the surrounding population, and as well as in many other cities in Spain, their track has been lost. Yet in Mallorca, a group of converted families, despite being legally

Christians, continued to live in the *Call* and kept up the relationships and occupations they had had before the conversion. They remained as a closed community and practised Jewish practices to a greater or lesser extent. This situation is known as Crypto-Judaism and was maintained until the end of the 17th century when the Inquisition managed to put an end to it.

The Inquisition started acting in Majorca in 1232, but its practices did not affect the Jews until 1488. The 'New Inquisition' (a tribunal newly created by the Catholic Kings as part of an effort to forge a nation state on the basis of religious uniformity) attacked the *conversos* and their descendants. Between 1488 and 1535, 806 *conversos* were condemned, 234 of them were '*reconciliats*' meaning they were readmitted in the Catholic Church, after undergoing certain punishments, such as seizure of properties and goods, expulsion from the neighbourhood, jail, or wearing the penitential garment known as the '*gramalleta*' or '*sambenet*' (Figure 2). The other 535 were '*relaxats*', that is, condemned to execution either in person, or as a sculpture (if they were fugitives) or burning the bones (if they were already dead) (Muntaner, 1986).

Figure 2. Francisco de Goya picture called *'Por mober la lengua de otro modo'* (for moving his tongue in a different way) which may mean praying in another language – possibly Hebrew – or perhaps expressing ideas contrary to official doctrine. In the image there is a person condemned by the inquisition dressed with a '*gramalleta*'. ©Museo Nacional del Prado.

After 1536 the inquisitorial tribunal remained more or less inactive regarding the *conversos* until 1673, even though there were signs of prohibited practices. In this period, although there was a reduction in the Crypto-Jewry, a small group persevered in their clandestine Jewish practises. They are essentially the people who would later be known as Chuetas. This was a period of strong economic growth and commercial influence for Majorcan *conversos*, who strongly focused their activity on trade. They created complex Mercantile Companies, participated actively in foreign trade, and dominated the market for insurance and retail commerce of imported products (Pons, 1988; Bibiloni, 1992; Porqueres, 2001).

In 1673 the Inquisition again started acting against the *conversos*. The confession of some servants of the converts led to the beginning of an inquisitorial process, whose consequences were that in 1677 around 250 residents of the *Call* were incarcerated, and all their possessions were confiscated. This procedure was called '*complicitat del 78*' (the

complicity of 1678), due to the mutual understanding in the testimony of the convicted and those in the street. This situation ended up with five *Autos de fe* in 1679, where most of the prosecuted were *reconciliats*. Imprisonment was not so hard for the convicts because, thanks to the corruption of the guards, the prisoners had a lot of liberties (Braunstein, 1936). They maintained some Jewish rites and traditions and the experience between 1677 and 1679 helped the group's cohesion. All in all it entailed a severe impoverishment of the community and the social pressure ended up leading to emigrations of small groups to other places in the Mediterranean basin.

In 1687, the confessions of a betrayer (Rafael Cortès, also known as 'Crazy head') and the abortive attempted escape of a group of *conversos*, triggered a massive arrest of the community. The trials lasted three years and concluded with four spectacular *Autos de fe* in 1691 known as the '*Cremadissa del 91*' (the burnings of 1691), where 86 *conversos* were condemned, 45 to the bonfire, three of them alive (Braunstein, 1936). Severe sentences were imposed on the *reconciliats*. These episodes definitely marked the end of Crypto-Judaism of Majorca, the effect of the loss of the leaders and the generalized fear after the mass burnings made it impossible to sustain the ancestral faith (Braunstein, 1936).

It is after these events, when in fact one can start talking about Chuetas (Moore, 1976) and, far from an assimilation of the community into Majorcan society, the hardest era of isolations started. Despite their submission to the Catholic faith, the memory of the heresy of the condemned was perpetuated among the Majorcan population and, by extension, also included their relatives or unrelated people bearing the same surnames. Some elements helped to maintain this collective memory: the first one was the publication of the book '*La fe triunfante*' in 1691 (Garau, 1931) and its consecutive reprints, written by a Jesuit who actively participated in the inquisitorial trials. This book explains the four *Autos de fe* in 1691 with extreme crudity and claims the need to perpetuate the memory of the 'infamy of the condemned' and extend it to their relatives. Another was the publication of the list of all the convicted. And another event that maintained the memory of the Crypto-Jews and their descendants alive was that the paintings of the people condemned in the last *Autos de fe*, wearing the clothes that they were forced to use, with their names (the family names today considered Chuetas), were exhibited in the cloister of Saint Dominic's Church until 1820, when a group of Chuetas assaulted and burned the church (Perdigó, 1946; Font, 1993).

The Chueta population was forced, by law, to live in the *Segell* district (the old *Call*), had important restrictions to access Majorcan ecclesiastical, university and military positions, and was professionally limited, due to the existence of statutes of '*Limpieza de sangre*' (purity of blood) in most of the guilds (Moore, 1976; Porqueres, 2001).

All these elements generated a community that, although it no longer contained Judaic religious elements, maintained a strong group cohesion and characteristics very similar to

other Jewish (or Crypto-Jewish) groups in the diaspora: inbreeding, system of internal cooperation and interdependence, consciousness of Jewishness, and external social hostility. Elements that, to varying degrees, meant they were perceived as still Jewish, or more accurately as Catholic Jews (Moore, 1976).

Notwithstanding the social and legal segregation, the Chuetas regained the leading economic role they had had before the inquisitorial trials, and this situation enabled them to fight actively for equal rights. One clear example was the allegation in defence of the rights of the Chuetas presented before the court of Charles III (1773). Although the king agreed to some of the requests of the Chuetas, by signing two Royal Decrees (1782 and 1785), the same attitudes of social discrimination, matrimonial endogamy, and traditional professions were kept. Moreover, segregation continued in education and ecclesiastic institutions (Riera and Melià, 1973; Riera and Porqueres, 1996).

In 1809 and 1823 there were two important assaults on the *Call* and similar incidents took place in some villages in the island where Chueta families had settled. Continuous discriminative situations took place until the 20th century. In 1936 the historian Braunstein highlighted the fact that around 300 families lived in the *Call* and their jobs were mainly as jewellers or traders (Figure 3). He also wrote that endogamy was still the main strategy of marriage in the community (Braunstein, 1936).

Figure 3. Family of Jaime Piña, in front of the jeweller's they owned in the *Call.* (Image from the Facebook page of FAM, Fotos Antiguas de Mallorca).

Anti-Chueta prejudice lessened with the arrival of tourism in the island in the first decades of the 20th century, along with economic development, which had started by the end of the previous century. The presence of outsiders living in the island (Spaniards or foreigners) to whom the status of the Chuetas meant nothing, marked a definite point of inflection in the history of this community.

Based on the genealogies constructed by the association '*Memòria del Carrer'* (http://www.memoriadelcarrer.com) from ecclesiastic documents and family documents, some charts were drawn up to confirm the endogamy over the centuries and the opening of the community from the second half of the 20th century. In Figure 4, by way of example we show the marriages we have found among the people called Aguiló.

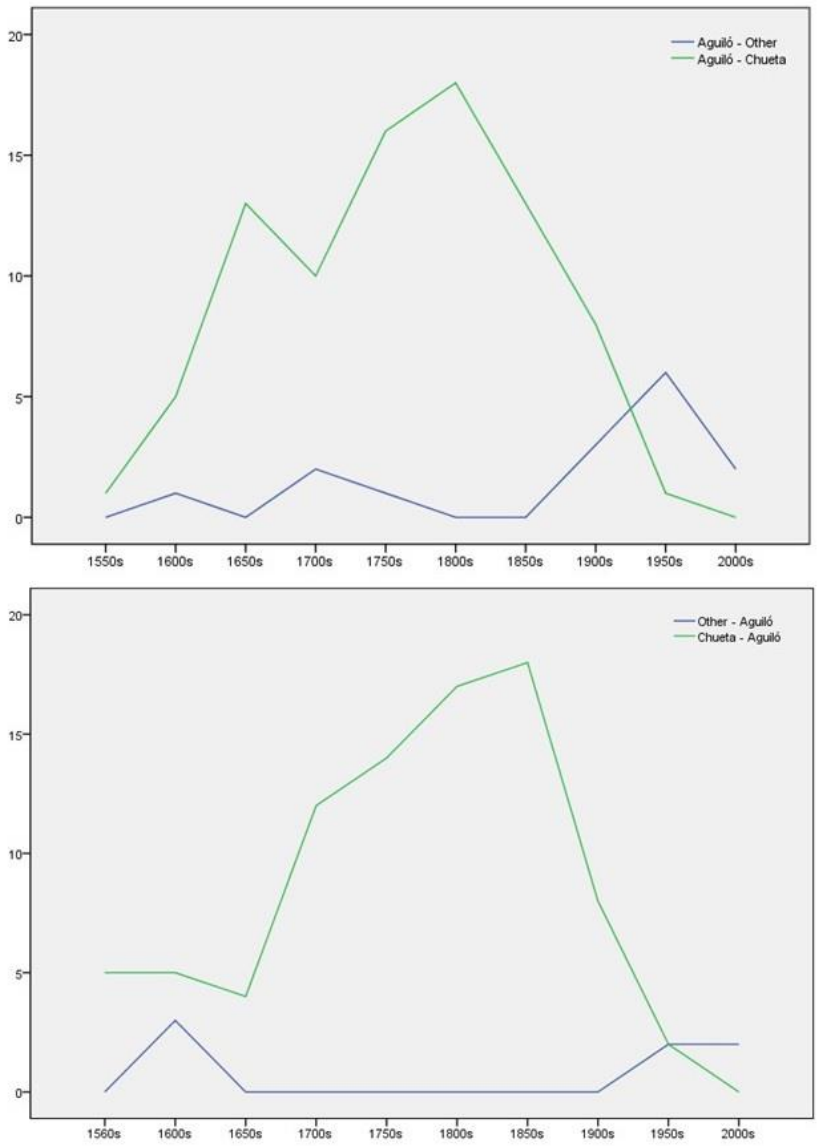


Figure 4. Marriages between people with the surname Aguiló and other Chueta people or non-Chueta (Other) from the 16th century to nowadays. The upper graph represents men's marriages, and the lower the women's.

1.1.2. Chueta surnames

Nowadays, the surnames known as Chuetas are the following: Aguiló, Bonnín, Cortès, Fortesa, Fuster, Martí, Miró, Picó, Pinya, Pomar, Segura, Tarongí, Valentí, Valleriola, and Valls (Forteza, 1972). The origin of the surnames is basically Catalan, Castilian and Italian and they only have a relationship with the Jewish identity in the island. Other surnames, with Hebrew etymology (Abraham, Daviu, Sansó, Salom, Maimó, Vidal, etc.) are not considered Chuetas despite being quite common in Majorca.

At any rate, it is worth noting that these 15 surnames came from a much more extensive set of surnames present in the converted community. In 1391 a list of Jewish people converted due to the increasingly violent assaults, contained 91 different surnames. Between 1481

and 1491, the number of surnames that appear in the registers of the 'New Inquisition' as condemned for maintaining Judaism is 180. Finally, between 1478 and 1536 there are 235 condemned carrying 112 different surnames. Taking into account the three lists of surnames, 238 different surnames could be identified from converted Jews and people condemned for Crypto-Judaism (Porqueres, 2011). Hence, it is important to highlight that to have a *converso* origin would not have been sufficient to be considered Chueta, the collective identification of families and surnames so considered would also have been necessary. Therefore, Chuetas are descendants of *conversos* but only a fraction of *conversos* descendants are considered Chuetas.

1.1.3. Majorcan Jews and the Jewish Diaspora

Jewish people describe themselves as 'people' since the definition of Jew is complex (Ostrer and Skorecki, 2013). According to religious law, a Jew is one whose mother is a Jew. Reality though, is that one is considered a Jew because of a number of factors nowadays. Jewishness can be determined by following the Jewish faith, or by descending from a Jewish family, belonging to a particular ethnic group (Jewish), or simply by the family's heritage and identification with the culture and history of the Jewish people (Levy-Coffman, 2005). Entry into Judaism through religious conversion is possible, but throughout history it has probably been a rare event.

To understand this complex collective identity, it is important to know the origin of contemporary Jews and, especially, the phenomenon that has shaped the Jewish people's history, namely the Diaspora. Historical evidence suggests common origins in the Middle East in the early Bronze Age (Shanks, 1988), followed by migrations leading to the establishment of communities of Jews worldwide, which have maintained continuous Jewish identity up to the present. The word Diaspora, from the Greek *diaspeirein*, means to disperse or scatter. Nowadays, the term 'Jewish Diaspora' is commonly defined as the dispersion of Israelites, Judahites, and later Jews, out of Israel, and their subsequent settlement in other parts of the world. The Babylonian and Roman conquests of Palestine led the Jewish people to migrate from the Levant to other Middle Eastern regions and the Mediterranean basin. Afterwards, other migratory movements throughout history led to the settlement of Jewish communities in different countries in Europe, America, Asia, and Africa.

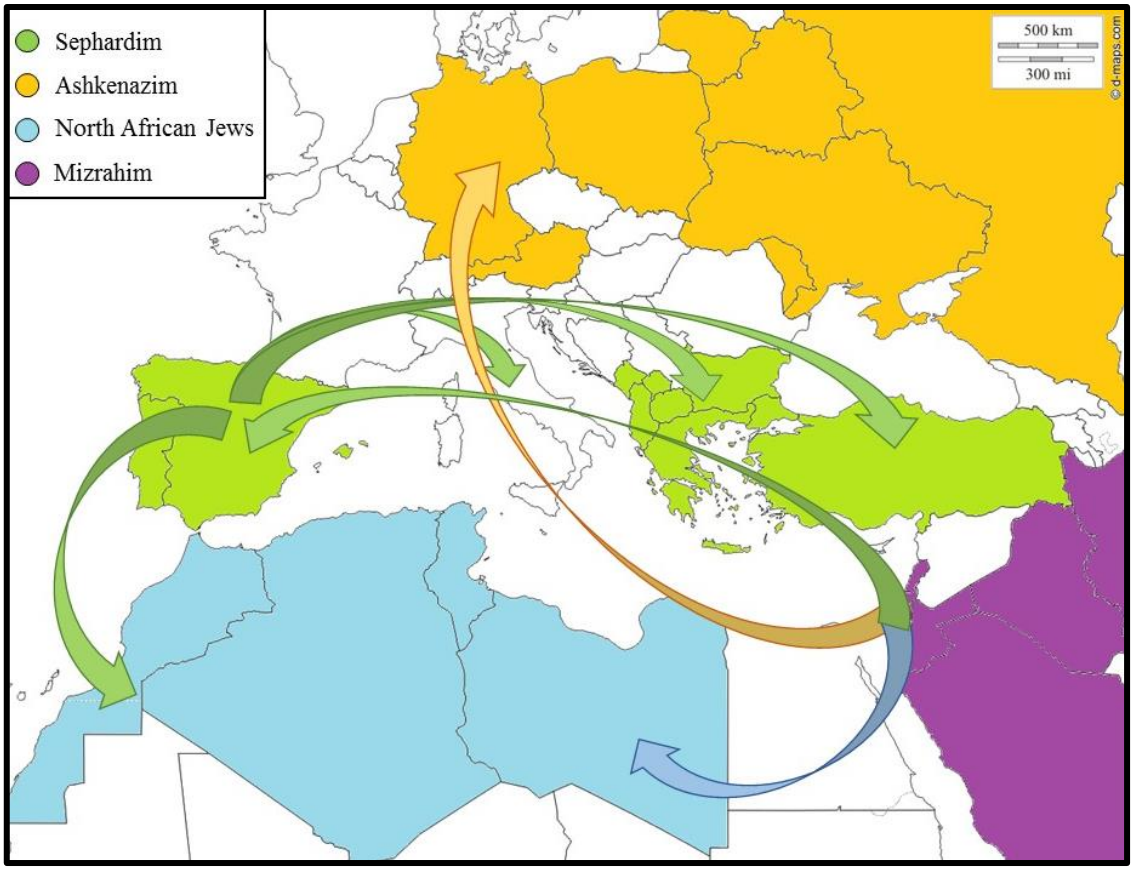


Figure 5. Map showing the four main groups of Jews.

Current Jews can be divided into the following four main groups on the basis of the long-term place of residence (Figure 5): Mizrahim, Ashkenazim, Sephardim, and North African Jews. The Mizrahim (also known as Middle Eastern or Oriental Jews) are those who stayed in actual Israel or Palestine or lived in Iraq, Iran, Central Asia or the Arabian Peninsula. The Ashkenazim (from the Hebrew word for German) moved north of the Alps. In the 12th and 13th centuries they were expelled from the countries in Western Europe and then settled in Poland and Lithuania. They developed their own language, Yiddish (similar to German with words derived from Hebrew and the Slavic languages). The Sephardic Jews (from the Hebrew word for Hispania, referring to the Iberian Peninsula) lived in Spain and Portugal up to the 15th century when the Inquisition in these countries forced them into exile with the Edicts of Expulsion. This group moved to North Africa, Italy, the Balkans, Turkey, Lebanon and the Americas. Sephardic Jews also maintained their own language, called 'Ladino' which is an ancient Spanish mixed with Hebrew words. The fourth group, the North African Jews is sometimes included in the Sephardim, but in fact comprises both Sephardim and Mizrahim, as there is evidence of Jewish communities in North Africa as early as the first centuries AD that were augmented as a consequence of the Spanish expulsion (Baron, 1937; Ben-Sasson et al., 1976; Goodman, 1979), therefore, due to their own history and composition, they deserve to be considered as a different group.

Apart from these four main groups, there are other Jewish communities such as Ethiopian Jews (Falashas) (Lucotte and Smets, 1999), Indian groups like Cochin or Bene Israel Jews (Chaubey et al., 2016), the Yemenite (Edholm and Samueloff, 1973), and Chinese Jews (Shapiro, 1984).

The Jewish Diaspora towards the Iberian Peninsula seems to date back to the Roman period, although the exact date of their arrival has not yet been unravelled. The oldest archaeological undoubted evidence can be traced to the 4th or 5th century AD (Iniesta et al., 2009; Graen, 2012), and there are also written documents indicating a relatively large number of Jews on the Iberian Peninsula already, such as the conclusions of the ecclesiastical council of Elvira, dating back to the 4th century AD (Dale, 1882).

The chronology of the first Jewish communities in the Balearic Islands seems to be similar to that of the mainland. The encyclical from the bishop Severus demonstrate the existence of large, prosperous Jewish communities in the islands of Minorca and Majorca at the beginning of the 5th century AD. In addition to these early Jewish settlers in the island, other population movements throughout history influenced the composition of the Majorcan Jewry, such as the arrival of Jews from Aragon and Catalonia after the Catalan conquest (13th century), from France and Portugal (14th century), and North Africa (Pons, 1984; Assis, 1996-1997; Pérez, 2005; Font, 2007).

1.2. Population Genetics

Population Genetics is a branch of Evolutionary Biology that deals with the study of genetic variation, at a molecular level, within and between populations, measuring distribution and changes in allele and genotype frequencies. Its aim is to infer populations' histories from genetic data and understand the evolutionary forces that have shaped the observed distribution of genetic variability. Its applications cover a variety of fields, such as Ecology, Conservation Genetics, Genetic Improvement and, particularly Human Population Genetics, Anthropology, Forensic Science, and Medicine.

Patterns of human genetic diversity depend on the complex interaction of variables such as the history of the populations, which includes the origin of the human groups, migrations, mating systems, physical and/or cultural barriers, and demographic fluctuations in the population size (founder effects, bottlenecks, etc.); and the evolutionary forces that affect gene frequencies such as mutation, gene flow, natural selection, and random genetic drift.

Mutation and recombination are the main sources of variation, while genetic drift and gene flow change gene frequencies upon which selection will act. Mutation is the only process generating new alleles by random changes anywhere in the genome and at a wide range of different rates. The evolutionary consequences of these changes depend on whether they occur in the germ line or in somatic tissues, and also on other evolutionary forces acting

upon the new variants by increasing or decreasing its frequency in a gene pool (Meier, 2010; Relethford, 2012). Recombination generates new combinations of alleles and enhances the ability of populations to adapt to their environments by combining advantageous alleles of different loci.

Gene flow (also known as admixture or migration) is the outcome of the genetic exchange between individuals from different populations. It can modify variation both within and between populations since, on the one hand, it can lead to the introduction of new alleles into a population from elsewhere and, therefore, genetic diversity within that population increases. On the other hand, interchanging alleles from one population to another reduces the difference or leads to a homogenization of allele frequencies between different populations over time (Meier and Raff, 2010; Relethford, 2012). The consequences of migration movements depend on the initial gene differences between the populations, on the proportion of individuals involved in the migration, and on their reproductive success (Jobling et al., 2014).

Genetic drift is the fluctuation in allelic frequencies between generations due to the stochastic process of sampling, since each generation represents a finite sample of the previous one. This is directly linked to the effective population size (N_e) since the smaller the population, the greater its effects. It is also more powerful in isolated populations (Meier, 2010). Long-term N_e is influenced by a series of factors such as variation in census population and reproductive success. Founder effects and bottlenecks are two important processes that shape diversity in many human populations; both involve reduction in population size. Non-random mating can also influence N_e. Population subdivision, inbreeding or endogamy by isolation, ethnicity or cultural traits have played an important role in the genetic make-up of many current human populations (Jobling et al., 2014). Genetic drift cannot only maintain or change the frequency of an allele in a population, but may eventually lead to its extinction or fixation. Due to the random nature of the process, it is impossible to predict the direction of the change or which alleles will survive, but the likelihood of extinction or fixation is related to the initial allele frequency and population size. Gene flow and genetic drift operate in opposite ways. While genetic drift makes populations more differentiated, gene flow makes them more similar (Relethford, 2012).

Lastly, natural selection, which is the differential reproduction of individuals of different genotypes in sequential generations, as defined by Darwin and elaborated by Fisher. Natural populations are composed of individuals with genotypic differences that provide them diverse capacities to survive and reproduce in different environments. The ability of an individual genotype to survive and reproduce is known as fitness. Different forms of selection, such as positive or negative selection, can increase or reduce the fitness of the carrier, and therefore favour the increase or decrease in frequency of one allele in the population, respectively. Purifying selection tends to lead to the disappearance of intrapopulation genetic diversity, while balancing selection tends towards the maintenance of diversity. At any rate, fitness is partly dependent on the environment, and therefore the

same genotype can be, depending on the conditions, unfavourable in one population and positive in another (Hedrick, 1985).

In summary, genetic diversity tends to be increased by mutation, recombination and gene flow, but decreased by random genetic drift. Selection, however, can act either way. In natural populations, there is a dynamic interaction among all these forces across generations that is strongly influenced by historical-demographic characteristics, to create the final genetic make-up of each particular population.

1.2.1. Genetic polymorphisms

More than 99.7% of the human genome is shared between individuals, so regions that differ need to be found in the remaining 0.3% of the genome (Butler, 2009). This genetic variation between individuals can be assessed using polymorphic markers. By definition, a genetic polymorphism is the occurrence in the same population of two or more alleles at one locus, each with appreciable frequency, where the minimum frequency is typically taken as 1% (Cavalli-Sforza and Bodmer, 1971).

Since the turn of the 20th century polymorphic markers have been used to assess human genetic diversity. Blood groups, serum proteins, erythrocyte enzymes, and HLA (currently known as ‘classical markers’) were the first markers defined, but in the last three decades, a number of molecular marker techniques have been developed. DNA markers provide an immensely popular tool for a variety of applications, due to their stability, cost-effectiveness, and ease of use (Grover and Sharma, 2016).

These DNA polymorphisms are found all over the genome. The ones appearing in the non-coding regions are considered to be neutral and are the most commonly used in the study of demographic human history, because they reveal population level effects such as migration, admixture, drift, and expansions (Kidd et al., 2004; Garrigan and Hammer, 2006; Rubicz and Crawford, 2007; Steiper, 2010).

In this work, three different types of polymorphisms have been performed: Short Tandem Repeats (STRs), Insertion deletion markers (Indels) (including Alu insertions), and Single Nucleotide Polymorphisms (SNPs).

1.2.1.1. SNPs

The simplest and smallest scale difference between two homologous DNA sequences is a base substitution, in which one base is exchanged for another. These differences are known as 'single nucleotide polymorphisms' if more than 1% of a population does not carry the same nucleotide at a specific position in the DNA sequence.

SNPs can arise throughout the nuclear and mitochondrial genomes elsewhere in coding, noncoding, or regulatory regions (Figure 6), although most SNPs reside within non-coding ones. There is an average of one SNP in each 1000 nucleotides, approximately, between two randomly chosen chromosomes of a population (Antonarakis, 2010). More than 97 million SNPs have been reported to date in the dbSNP database (Sherry et al., 2001).

Depending on the phenotypic effect, SNPs can be classified as synonymous or non-synonymous, if the alleles encode the same or different amino acid products, respectively; and silent or neutral if the SNP is located outside a coding region. Depending on the minor allele frequency in a population, a SNP can be considered common or rare (frequency >5% or frequency between 1 and 5%, respectively) (Antonarakis, 2010; Griffits et al., 2012).

Most SNPs have only two alternative alleles (ancestral and mutant) so, together with Indels, they are classified in the category of binary polymorphism. They represent quite rare events, occurring at very low rates, ~$2.5x10^{-8}$ in human history (Nachman and Crowell, 2000). This low mutation rate means that this class of mutation generally shows identity by descent, rather than identity by state (coincidental resemblance, sometimes called convergent evolution), meaning that the presence of the same base at a SNP in two independent genome copies usually implies that the base has been inherited from a common ancestor. Due to this stability, these markers generally allow the direction of the evolution to be established and, therefore, are very appropriate markers in phylogenetic studies (Underhill and Kivisild, 2007; Hughes and Rozen, 2012; Hallast et al., 2015; Pugach and Stoneking, 2015; Fregel et al., 2015) and population structure analysis (Liu et al., 2005; Henn et al., 2010; Haasl and Payseur, 2011). In forensic applications, SNPs can be more easily genotyped in degraded samples than other markers, because short amplicons can be used for Polymerase Chain Reaction (PCR) amplification, although the limited number of alleles per locus and the lack of available forensic kits make other markers, like STRs, more useful in forensic casework than SNPs (Canturk et al., 2014).

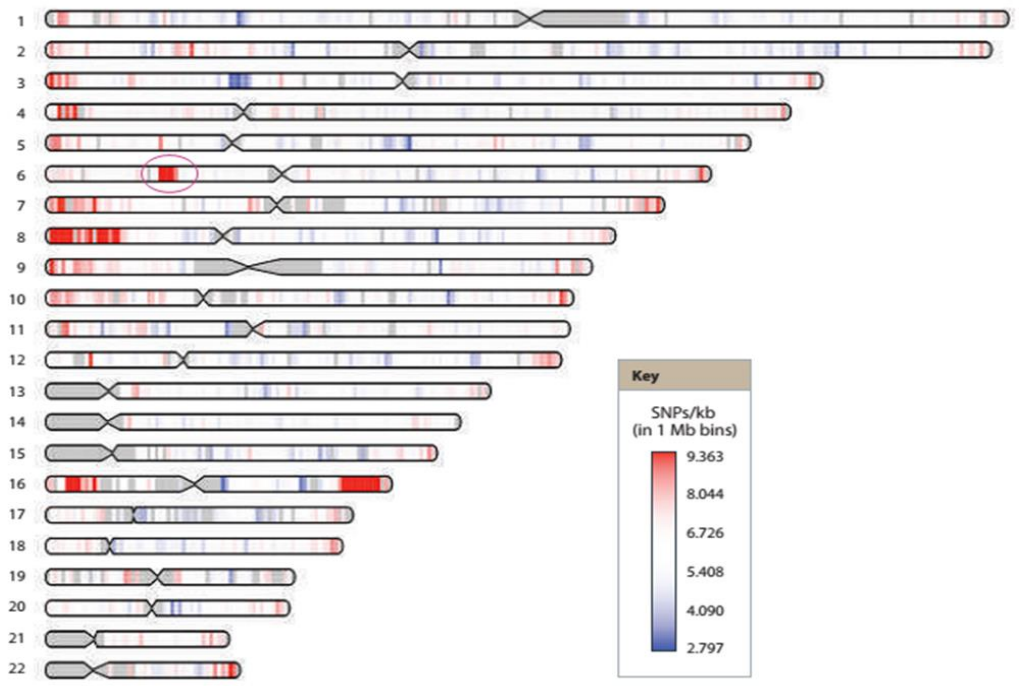


Figure 6. Genomewide SNP density. The circle indicates a SNP density peak around the HLA locus. Grey indicates regions where base calling is too unreliable to estimate SNP density (1000 Genomes Project Consortium (2010).

1.2.1.2. STRs

Short Tandem Repeats, also called microsatellites or Simple Sequence Repeats (SSRs), are arrays of short units, usually of 1 to 7 bp in length, tandemly repeated a variable number of times (typically between 10 and 30) (Antonarakis, 2010; Griffiths et al., 2012). These polymorphic markers are widespread throughout the genome including the 22 autosomal chromosomes and the X and Y sex chromosomes, and can be found in exons, introns, and regulatory regions, as well as non-coding regions (Oliveira et al., 2006). They occur on average every 10000 nucleotides, accounting for approximately 3% of the total human genome (Ellegren, 2004; Butler, 2011).

These markers can be named depending on the length of the repeat unit: mono-, di-, tri-, tetra-, penta-, hexa- and heptanucleotide for repeats between 1 and 7 nucleotides, respectively. STRs can also be classified into several categories based on the repeat pattern region: a) simple repeats, containing units of identical length and sequence, for instance $[ACTC]_n$, b) compound repeats, comprising two or more adjacent simple repeats, e.g. $[GTCT]_n$ $[TACG]_m$, and c) complex repeats, may contain several repeat blocks of variable unit length as well as variable intervening sequences, e.g. $[TCTA]_n$ $[TCTG]_n$ $\{[TCTA]_3TA[TCTA]_3TCA[TCTA]_2TCCATA\}$ $[TCTA]_n$ TA TCTA (Urquhart et al., 1994; Butler 2014).

STRs are extremely polymorphic, with vast numbers of alleles (often more than 20) at a particular microsatellite locus. This high degree of polymorphism is due to its relatively high mutation rate, in the order of 10^{-3} or 10^{-4} per STR per generation (Jobling et al., 2014). It is widely accepted that microsatellite mutation occurs as a result of DNA replication slippage, with the insertion or deletion of repeat units relative to the template strand (Ellegren, 2004).

Microsatellites satisfy all the requirements for a forensic and/or population genetic marker: they are robust, easy to amplify by PCR, and highly discriminatory, especially when analysing a large number of loci simultaneously (multiplexing). Moreover, there are a large number of STRs located in non-coding DNA that, therefore, are assumed to evolve neutrally (Ellegren, 2004; Goodwin et al., 2007; Griffits et al., 2012). Hence, STRs are currently the most commonly analysed genetic polymorphisms in forensic genetics and also in genome mapping and population genetic studies (Ellegren 2004; Oliveira et al., 2006; Sun et al., 2009; Steiper 2010; Kayser and de Knijff, 2011; Griffits et al., 2012).

At any rate, the underlying high mutation rates of these multi allelic loci represent some disadvantages in contrast to SNPs: a) alleles with the same size and sequence may not reflect identity by descent, but identity by state; and b) ancestral states cannot be determined by reference to great ape DNAs (Jobling et al., 2014).

1.2.1.3. Indels

Indels are length polymorphisms based on insertion or deletion of one or more nucleotides. Indels, ranging from 1 to 10000 bp, are highly abundant in humans and cause a great deal of variation, representing approximately 18% of all sequence polymorphisms in humans (Mullaney et al., 2010).

They are found in all the chromosomes, with normalized averages of one Indel for every 5.1-13.2 Kb of genomic DNA (Mills et al., 2006). Over the past decade, several million Indels have been discovered in human populations and personal genomes (Mullaney et al., 2010; 1000 Genomes Project Consortium, 2015). They can be classified in five categories: Indels of single-base pairs (29.1%), repeat expansions (29.5%) monomeric or multi-base of 2-15 bp repeat units, transposon insertions (including Alu insertions) (0.59%) and Indels containing random DNA sequences (40.8%). Over 36% map to functionally important sites within human genes and, thus, are likely to influence human traits and diseases (Weber et al., 2002; Mills et al., 2006; Mullaney et al., 2010).

Although some Indels have multiple alleles (multiallelic), most of them have just two alleles (diallelic). In the present work we deal only with diallelic ones. The size difference between both alleles can vary a lot, depending on the type of Indel. The largest group of Indels is composed of those with allele-length differences of relatively few nucleotides.

Other types can differ by hundreds or thousands of nucleotides (Weber et al., 2002; Jurka, 2004; Mills et al., 2006).

Diallelic Indels, especially short length ones, are being increasingly used in human identification, forensic genetics, and population genetic studies (Weber et al., 2002; Bastos-Rodrigues et al., 2006; Pereira et al., 2009; Pimenta and Pena, 2010; Li et al., 2011; Zidkova et al., 2013; Cardoso et al., 2017; Liu et al., 2017), due to the fact that these markers combine advantages from both SNPs and STRs, such as their widespread distribution throughout the genome; their origin from a single mutation event which occurs at a low frequency leading to their subsequent stability (low likelihood of having recurrent mutations); the significant differences in allele frequency between distinct geographical groups, enabling their use as ancestry informative markers (AIMs) (Pereira et al., 2012b; Phillips, 2015; Santos et al., 2015); and the ease of genotyping of small Indels giving rise to their suitability for large-scale multiplexing and automation.

Alu elements are abundant in all primate species and comprise more than 10% of the human genome (Cordaux and Batzer, 2009), with 1.2 million of copies. They belong to a class of retroelements termed SINEs (Short Interspersed Nuclear Elements) and are non-autonomous, depending on the enzymatic machinery of autonomous elements (Jurka, 2004). Alu sequences are approximately 300 bp long, formed from two divergent dimers, ancestrally derived from the 7SL RNA gene, separated by a short A-rich region (Figure 7).

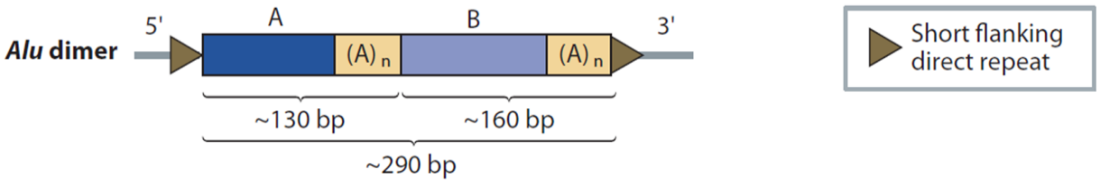


Figure 7. Structural features of full-length Alu elements. The two monomers within the Alu dimer differ in length because of a 32-bp insertion in the B monomer. Lengths are approximate because of variation in the length of the poly(A) tails [$(A)_n$] (Jobling et al., 2014).

Alu elements have been inserted into the human genome over the last 65 million years. Their evolutionary history led to two broad categories, to fixed Alu elements (monomorphic in the entire human population and presumably evolutionarily older), and polymorphic Alu elements (only in a subset of humans, due to the result of more recent retrotranscription events than the African radiation of humans) (Qian et al., 2015). Since there are no specific mechanisms for removal of Alu elements, their evolution is dominated by the accumulation of Alu insertions that accumulate sequence variation over time (Deininger, 2011).

The Alu insertion polymorphism has been used in human evolutionary studies because they are markers of identical descent with known ancestral states (Batzer and Deininger, 2002; Gómez-Pérez et al., 2007; Salem et al., 2014; Chinniah et al., 2016).

1.2.2. Human genome regions: Inheritance and features

When making inferences about human demographic history based on DNA data, it is very important to consider the processes that differentially affect the various genome regions that can be genotyped. It is important for example, to be aware of the location and inheritance mode, as well as the effect of evolutionary forces (such as recombination and mutation rates) and the effective population size on them (Garrigan and Hammer, 2006).

Human nuclear DNA is organised into autosomal and sex-chromosomes (X-chromosome and Y-chromosome). Apart from this nuclear DNA, the mitochondria have their own genome, known as mitochondrial DNA (mtDNA). Autosomes, sex chromosomes, and mitochondrial DNA have their particularities in terms of inheritance and how they are affected by diversity shaping factors (Table 1). Therefore, each of the distinct genomic material might provide differential information and will be more or less suitable to answer particular population genetics questions (Jobling et al., 2014).

Table 1. Comparison between features of different genomic compartments (Garrigan and Hammer, 2006).

	Genomic compartment			
Feature	**Autosomes**	**X-chromosomes**	**Y-chromosomes (NRY)**	**mtDNA**
Location	Nuclear	Nuclear	Nuclear	Cytoplasmic
Inheritance	Bi-parental	Bi-parental	Uni-parental	Uni-parental
Ploidy	Diploid	Haploid-diploid	Haploid	Haploid
Relative N_e	4	3	1	1
Recombination rate	Variable	Variable	Zero	Zero
Mutation rate	Low/Moderate*	Low	Low/High*	High/Very High*

* Schaffner (2004)

1.2.2.1. Autosomes

The 22 pairs of autosomal chromosomes contain 98% of human DNA and, as such, human genetic diversity mainly occurs in this genomic compartment (Jobling et al., 2014). As one chromosome of each pair is inherited from each parent, the polymorphisms found in the autosomes are biparentally transmitted in a typical Mendelian mode of inheritance. Recombination allows reshuffling of all the alleles at different loci when transmitted down in each generation. This inheritance mode permits new combinations of genetic material in each generation and provides information of both parents' ancestries (Rubicz and Crawford, 2007), unlike what occurs in uniparental markers. The study of many independent inherited autosomal loci is a powerful tool to unravel the multiple genetic histories in the past of a population, whereas non-recombinant markers only reflect a picture of the maternal or paternal genealogical past (Schiffels and Durbin, 2014).

Information contained in autosomes can be considered neutral when many loci are genotyped, due to recombination. This force tends to decouple the genealogical histories of sites along chromosomes (that is, neutral sites and those under selection), in contrast with the haploid regions, each behaving evolutionarily as a single locus. In these compartments selection acting on any particular site affects patterns of polymorphism on the entire mtDNA and Non-Recombining Y (NRY) (Garrigan and Hammer, 2006).

The presence of two copies of the autosomal loci in both males and females leads to an effective population size of autosomes that is 4-fold those of the Y chromosome and mtDNA, and 4/3 that of the X chromosome. Taking into account that for a neutrally evolving locus the expected TMRCA (time to the most recent common ancestor) is a function of the N_e, the larger N_e of the autosomes implies that autosomal loci are expected to have deeper ancestry than the haploid regions, allowing inferences of evolutionary processes that took place in more ancient times of evolutionary human history. Autosomal loci are also expected to result in lower levels of differentiation between human populations, and possibly greater effects of natural selection (Garrigan and Hammer, 2006).

Autosomal markers have been widely used for reconstructing human evolutionary history, for inferring structure and admixture proportions of human groups, and also in Forensic Genetics. In kinship testing, the use of the traditional autosomal markers only provides information on relationships spanning from one or two generations (Nothnagel et al., 2010), but the simultaneously study of a greater number markers can reduce the limitations related to testing complex pedigrees (Egeland and Sheehan, 2008; Skare et al., 2009).

1.2.2.2. X-chromosome

The sex chromosomes, X and Y, are thought to have evolved from an ancient pair of autosomes, about 300 million years ago (Hughes and Page, 2015). During their evolutionary process, they have accumulated differential mutations, leading to two chromosomes with very distinct structures and functions (Lahn and Page, 1999; Strachan and Read, 2006). Both chromosomes have two short homologous regions at the telomeres, known as pseudoautosomal regions (PAR) 1 and 2, which are autosomal-like and allow them to pair in meiosis. Whilst the X-chromosome do recombine in females, in males the Y-chromosome only recombines with the X in the pseudoautosomal regions.

The X-chromosome spans approximately 155 Mb in humans (~5% of the genome), with a low gene density compared with the rest of the genome, but a relative wealth in genes involved in reproductive and cognitive functions (Ross et al., 2005). The different ploidy in males and females leads to the need to balance doses of X-genes, thus in the somatic cells of a female, one of the two X-chromosomes is largely silenced (Carrel and Willard, 2005; Huynh and Lee, 2005).

The only X-chromosome present in males (inherited from the mother) is not subjected to recombination in most of its extension, and therefore the haplotype located in the non-recombining region is transmitted identically to all daughters, unless mutation occurs. On the other hand, in females the two X chromosomes undergo meiotic recombination (Figure 8). Since only 2/3 of X-chromosomes recombine in each generation, the measured recombination rate for the X-chromosome is, in fact, almost exactly two-thirds of the genome average (Schaffner, 2004). Moreover, the specific mode of inheritance of the X-chromosome yields that its effective population size is 3/4 that of an autosome, when there are equal numbers of breeding males and females, and random variation in offspring.

Another property of the X-chromosome is its lower mutation rate with respect to autosomes, result of the higher mutation rate in males than in females (Wilson Sayres and Makova, 2011), and the fact that the X-chromosome spends only one-third of its time in males (Johnson and Lachance, 2012).

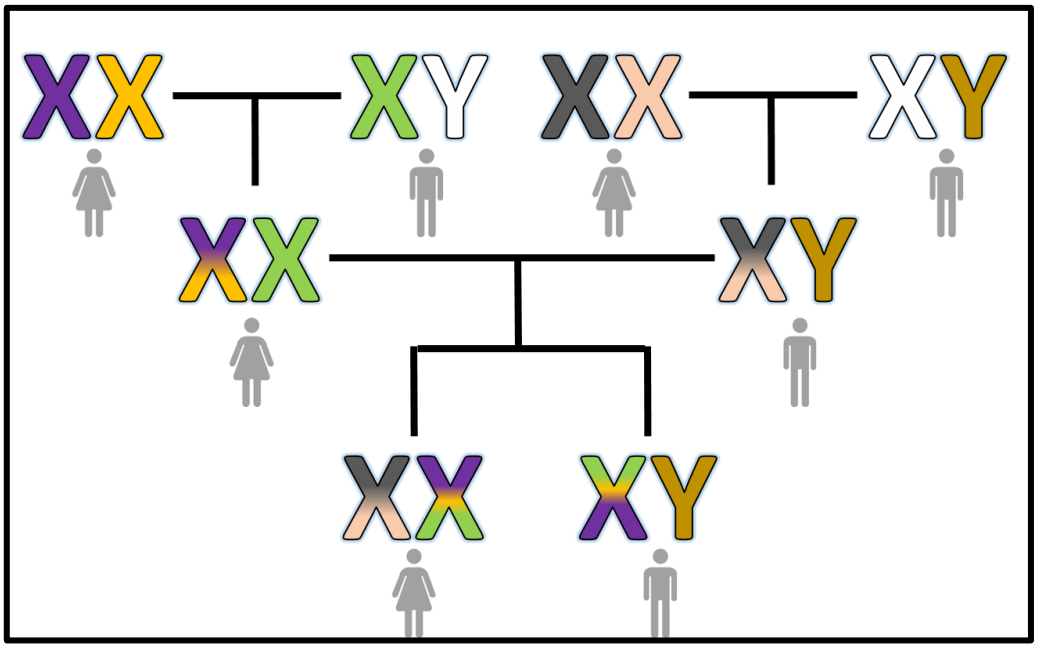


Figure 8. Different mode of inheritance of X-chromosome in males and females. Recombination can be observed in female transmission.

In short, the X-chromosome has a lower recombination rate, a lower mutation rate, and a smaller effective population size. All these features make the study of markers located on it an important tool in the fields of Anthropology and Population Genetics and, in recent years, the interest of study of this chromosome has notably increased. Compared with other genome regions, the X-chromosome suffers recombination, like the autosomes, and, equally to mtDNA and the Y-chromosome, it has a sex-biased mode of inheritance allowing direct haplotyping in males. Since males present only one copy of the X-chromosome this makes it one of its unique features alongside the fact that it exhibits lower genetic diversity in comparison to autosomal chromosomes. Following the comparison with autosomes, the population structure is expected to be more prominent as a consequence of the reduced

effective population size, thus explaining the faster genetic drift found in the X-chromosome compared to autosomes. This can explain why populations tend to differ more in their X-chromosomes than in their autosome variation. Since only two-thirds of chromosomes (female chromosomes) recombine in each generation, linkage disequilibrium is also stronger. Hence, the size of regions with a single genetic history is expected to be larger than in autosomes (Schaffner, 2004). Moreover, evolutionary forces can operate differentially in males and females due to social practices, which can lead to divergent patterns of genetic variation between autosomes, X-chromosome, Y-chromosome, and mitochondrial DNA (Bustamante and Ramachandran, 2009). A comparison of the genetic diversity present in a population for autosomes and for the X-chromosome can help reveal demographic histories involving sex-biased migration and breeding patterns (Labuda et al., 2010; Pereira et al., 2015).

The X-chromosome has also recently become an extremely useful tool in the field of Forensic Genetics; on the one hand as there is only a single copy in males, typing X-polymorphism multiplexes (STRs or SNPs) immediately yields high forensically efficient haplotypes. On the other hand, due to their particular inheritance, the X-chromosome can efficiently complement the analysis of other markers and may solve cases that would otherwise remain unsolved. Any investigated relationship situation where at least one female is involved may benefit from the use of X-markers, which can be applied to cases of missing persons, incest, immigration, deficiency paternity, and other questioned relationships (Diegoli and Coble, 2015; Houck, 2015)

1.2.2.3. Y-chromosome

The Y-chromosome is the smaller of the sex chromosomes. Its size is ~57 Mb (2% of the genome, approximately) and, according to the Ensembl release 87 (December, 2016) (Aken et al., 2016), it contains 180 genes (coding and non-coding genes). As explained in the X-chromosome section, these two sexual chromosomes evolved differentially from an ancient pair. The process of divergence started when one of them (the current Y-chromosome) acquired a male sex-determining gene such as SRY, followed by a progressive repression of the recombination. Much of their common ancestral gene content has been lost from the Y-chromosome, and this gene decay has led to presage that the Y-chromosome is doomed to eventual disappearance (Lahn et al., 2001; Sun and Heitman, 2012). Gene loss is due in part to the accumulation of deleterious mutations, resulting in the lack of recombination with a non-mutant homolog (Muller's ratchet). In contrast to its paucity of genes, the Y-chromosome is enriched with many different types of repeats, including SINEs, endogenous retroviruses, and segmental duplications. These repeat sequences lead to an important susceptibility to rearrangements by non-allelic homologous recombination, explaining the unusually high levels of structural polymorphism observed in this chromosome (Jobling et al., 2014).

Two different portions can be recognised in the Y-chromosome; on the one hand, the Pseudo Autosomal Regions located in the terminal regions of the chromosome representing 5% of the Y-chromosome. These PAR regions allow the pairing with the X-chromosome in cell division, and exchange genetic material (recombination) during male meiosis. On the other hand, the so-called Non-Recombining Y, which is the remaining 95% of the chromosome.

NRY is the largest non-recombining block in the human genome and thus one of the most informative haplotyping systems, with applications in forensic and medical genetics, genealogical reconstruction, and evolutionary population studies, especially to investigate recent human evolution from a male perspective (Jobling and Tyler-Smith, 2003; Underhill and Kivisild, 2007; Kundu and Ghosh, 2015; Calafell and Larmuseau, 2016). Henceforth, when referring to the Y-chromosome in this work, we shall be alluding to the NRY region.

Since it is inherited directly from father to sons, all the males in a paternal lineage share an identical Y-chromosome. Mutation is the only force that leads to variation. Considering that the effective population size of the Y-chromosome is 1/4 compared to autosomal loci, and 1/3 compared with the X-chromosome, it is more susceptible to genetic drift, and this accelerate differentiation between groups and different populations (Jobling and Tyler-Smith, 2003; Chiaroni et al., 2009).

Studies to define paternal lineages started in the 80's (Casanova et al., 1985), and a great list of suitable markers have appeared since then (Underhill et al., 2000; Hammer et al., 2001; YCC, 2002; Hallast et al., 2015; etc.). By sequencing 1244 human Y-chromosomes from the 1000 Genomes project (Poznik et al., 2016) at least 65000 short variants (including SNPs, STRs and Indels) were found, and in the Ensembl release 87 (December, 2016) (Aken et al., 2016) 211595 were estimated.

The markers used to construct a phylogenetic tree of Y-chromosome are SNPs or as they were known before called as UEPs (Unique Event Polymorphism). Because due to its low mutation rate, ~10^{-8} mutations per generation (Jobling and Tyler-Smith, 2003; Butler, 2005), it is not expected to have occurred twice during human evolution time. Conventionally, Y-chromosomes identified by SNPs are called haplogroups, and those defined by microsatellites (STRs) are known as haplotypes. STRs have a faster mutation rate than Y-SNPs (average 3.84 x 10^{-4} mutations per generation, Willems et al., 2016). Therefore, combining both (SNPs and STRs) enables us to determine how the Y-chromosome of a specific haplogroup has diversified and then to be able to define and to date lineages (de Knijff, 2000; Jobling and Tyler-Smith, 2003; Underhill and Kivisild, 2007). During the last few decades, a wealth of studies based on these polymorphisms of the Y-chromosome have led to the construction of a well-established human Y-chromosome phylogeny, which is constantly being updated (Karafet et al., 2008; van Oven et al., 2014; ISSOG, 2017).

The expansion and distribution of the Y-chromosome through the planet (Figure 9) allows us to learn about human population migrations and colonizations. The most recent common ancestor (MRCA) of modern humans is estimated to have appeared 140-160 thousand years ago (Kya) in Africa (Cruciani et al., 2011b; Lippold et al., 2014). However the discovery of a more ancestral haplogroup, named A00, in an African American sample, would date the origins of the human Y-chromosome at about 338 Kya (Mendez et al., 2013). Ancient haplogroups found in Africa are split into branches A and B, which can only be found in Africa. D and E haplogroups are shared between African and non-African populations, while macrohaplogroup CF is only found outside Africa. A subsequent expansion of haplogroup E carried by the Bantu speaker farmers from West Africa could explain why currently most African Y haplogroups belong to haplogroup E and A and B are rare nowadays (Jobling and Tyler-Smith, 2003). The CF(xDE) haplogroup was the common ancestor of all people who migrated outside of Africa until recent times. The defining mutation occurred 3.1-5.5 Kya (ISSOG, 2017).

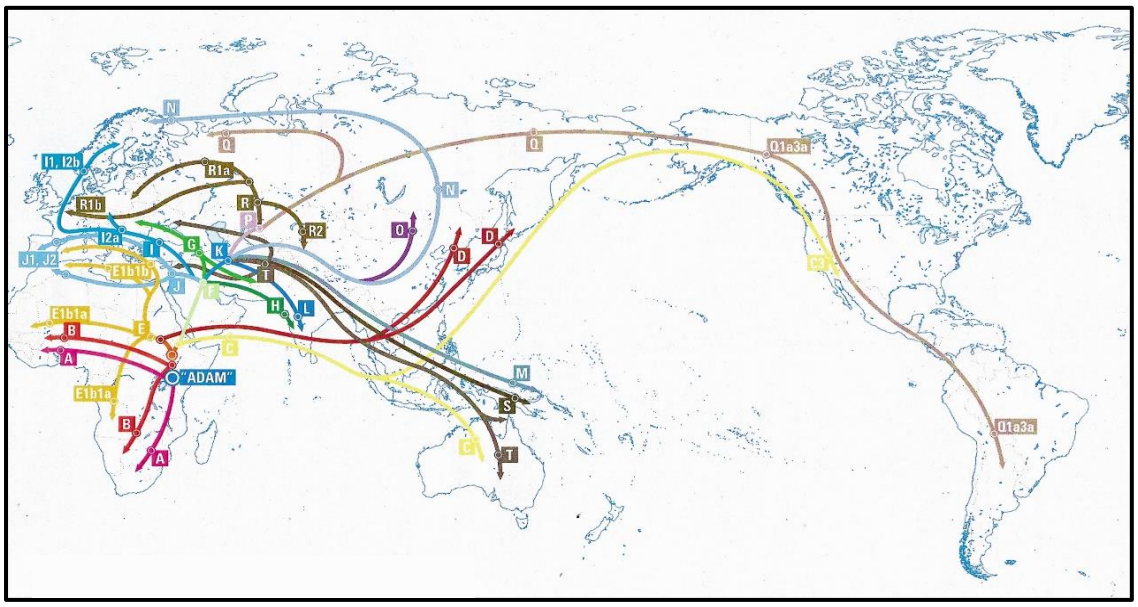


Figure 9. Major Y-chromosome haplogroup dispersion in the world, according to the Y-Chromosome Consortium (YCC) 2009 nomenclature (adapted from Family Tree).

Haplogrup F gave rise to haplogroups G, H, IJ and K (Figure 10). F and H are restricted to the Asian continent. The expansion of haplogroups I and J is found in Europe and the Middle East, respectively (Rootsi et al., 2004; Semino et al., 2004). Sub-lineages of haplogroups E, G and J are prevalent in the Mediterranean area (Batini et al., 2015). In East Asia, haplogroups N and O are mainly found, and to a lesser extent Q, which is the main haplogroup in America, especially its subhaplogroup Q1a3a. Haplogroup R derives from haplogroup P and has high frequencies in Europe. Due to this specific geographical distribution, the Y-chromosome is of great help to solve historical and anthropological questions regarding particular populations.

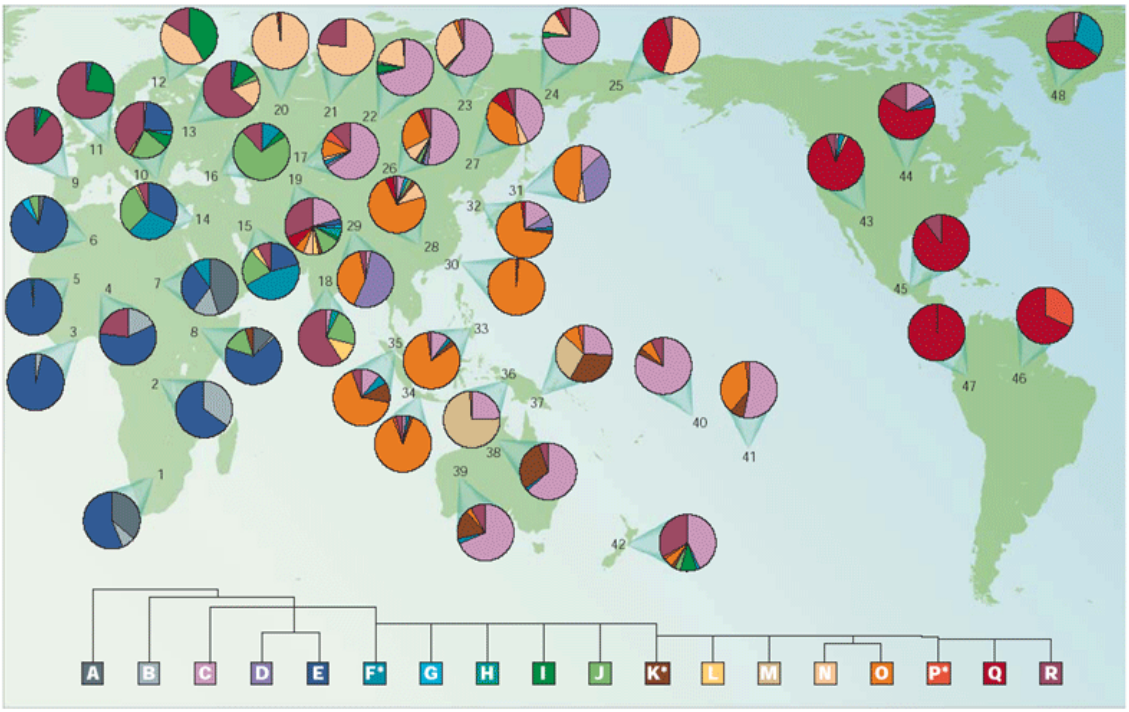


Figure 10. Frequency and phylogenetic tree of the 18 main Y-haplogroups worldwide identified by the YCC and indicated by the coloured sectors (Jobling and Tyler-Smith, 2003).

In the field of forensics, the Y-chromosome provides a powerful tool in different applications (Butler, 2009), including: a) male identification, especially in analysis of sexual assault evidence because the Y-chromosome is often the only detectable trace of the offender's DNA (Roewer, 2009; Purps et al., 2015), although it must be noted that all males who share the same paternal lineage have the same Y chromosome haplotype; b) conducting missing person's investigations, because patrilineal male relatives may be used for reference samples; c) performing deficient paternity testing; d) addressing historical questions or supplementing genealogical research. Surnames are usually paternally inherited, so Y-chromosome data can sometimes make links where a paper trail is limited; e) familial DNA searching or in the prediction of the surname within a forensic identification case (King and Jobling, 2009); and f) identification of the geographical origin of male lineages (Kayser and Ballantyne, 2014).

1.2.2.4. Mitochondrial DNA

The cytoplasmic location and energy-linked function of mitochondria derive from their origin as endosymbiotic prokaryotes (Margulis, 1981). Each mitochondrion contains many identical mitochondrial genomes that have a series of characteristics that make mtDNA especially interesting for forensic and phylogeographic genetic studies.

Mitochondrial DNA is a circular, double-stranded molecule with a 16569 base pair length (0.0005% of the size of the human genome). One strand (the heavy (or H) strand) is relatively rich in guanine bases, while the other (the light (or L) strand) is rich in cytosine. The mitochondrial genome encompasses two differentiated regions: the coding and the non-coding regions. As seen in Figure 11, the coding region represents the majority of human mtDNA and comprises 37 genes, 13 of which are essential components of the respiratory chain, 2 encoding ribosomal RNA (rRNA) (12S and 16S), and 22 transfer RNA (tRNA) (Kivisild, 2015). The non-coding region, better known as D-loop or control region (CR), ranges from position 16024 to 576. With a length of 1.1 Kb, it is the largest portion not directly associated with protein coding or RNA genes. The control region plays a major role in the regulation of transcription and replication of the molecule (Chinnery and Hudson, 2013).

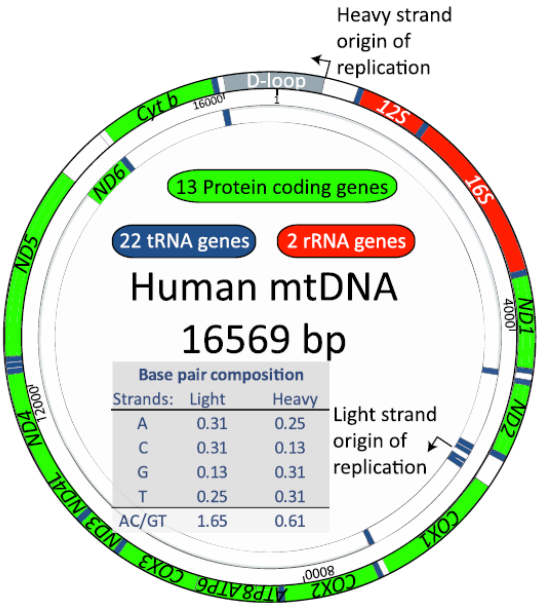


Figure 11. Functional map of mtDNA. Protein coding, rRNA and tRNA genes are shown in boxes distinguished by different colours (Kivisild, 2015).

The unique properties of human mtDNA are its high copy number by cell, exclusive maternal inheritance, lack of recombination, and high mutation rate.

Each cell contains several mitochondria and each has many copies of their genome, resulting in a large, variable amount of mtDNA genomes per cell, ranging from 100 to 10000 (Malyarchuk et al., 2002). This feature makes it easier to obtain mtDNA than nuclear DNA. Therefore, it is the molecule of choice for forensic cases with degraded samples, or for ancient DNA studies.

Mitochondrial DNA is exclusively maternally inherited. Although the mitochondria present in sperm are known to actually be able to enter the oocyte, mechanisms exist to eliminate paternal-derived mtDNA from the ovum (Sutovsky et al., 1999; Song et al., 2016). This sperm mitophagy may sometimes be defective, as suggested by the detection of an individual carrying a certain percentage of paternal mtDNA (Schwartz and Vissing, 2002), but this is likely to be an exceedingly rare phenomenon (Pakendorf and Stoneking, 2005). In accordance with the maternal inheritance pattern, the mtDNA effective population size is 1/4 relative to autosomes. For this reason, like NRY, mtDNA is more sensitive to demography events such as bottlenecks, genetic drift or founder events, due to its reduced N_e.

It is assumed that mtDNA does not undergo recombination. Mitochondria possess a functional recombinase, therefore recombination would be possible; however, leakage of paternal mtDNA is a very rare phenomenon, therefore any recombination events would result in mtDNAs that do not differ from the original maternal one (Pakendorf and Stoneking, 2005; Chinery, 2006). As a consequence, mtDNA is transmitted unaltered (except for mutation events) across generations and therefore enables stable female lineages to be defined back through time, thereby making it possible to trace the maternal ancestry of a population.

Mutation rate is on average much higher than in the autosomes, although the rates at which the mutations occur are different along the molecule and its different functional domains (Kivisild, 2015), with clearly higher rates in the CR than in the coding region. All the mitochondrial genomes in an individual are normally identical (homoplasmy); however, as a consequence of inefficient mtDNA repair and an oxidative environment, mutations in mtDNA are very common. When mutated and original molecules coexist in an individual, this is called heteroplasmy and can occur at different proportions among cell lines and different tissues (Stewart and Chinnery, 2015). Due to the different mutation rate between mtDNA positions, the identification of mutational hotspots can contribute significantly to distinguishing whether mtDNA molecules share nucleotide positions by descent or by state, which is essential in phylogenetic studies (Bandelt et al., 2006).

The first full mitochondrial genome sequence was determined in 1981 from human placenta of a European individual in the University of Cambridge, and subsequently became known as the Cambridge Reference Sequence (CRS) (Anderson et al., 1981). Currently, the nucleotide numbering of mtDNA sequences is based on a revised and corrected version of this, the rCRS (Andrews et al., 1999), although there are proposals to adopt a reconstructed ancestral sequence called RSRS (Reconstructed Sapiens Reference Sequence) instead (Behar et al., 2012).

The study of the mtDNA variation in a huge number of populations has allowed the construction of an updated comprehensive phylogenetic mtDNA tree. Sequential accumulation of mutations in different mtDNA molecules leads to the constitution of independent lineages, known as haplotypes. Groups of basal mutations (generally SNPs) shared for clusters of lineages define haplogroups, which represent the major branch points on the mitochondrial phylogenetic tree. Understanding the evolutionary path of female lineages has helped to draw a map of the main human migrations, from the origins in Africa to the subsequent spread throughout the world (Figura 12) (Maca-Meyer et al., 2001).

First studies on mtDNA proposed the most recent common ancestor of modern humans in Africa, 200,000 years before present (Cann et al., 1987). Although a large number of studies have focused on this question (Ingman et al., 2000; Maca-Meyer et al., 2001; Salas et al., 2002; Behar et al., 2008a; 2012; Barbieri et al., 2013; etc.), the exact place of the homeland of modern humans remains controversial (Rito et al., 2013; Cerezo et al., 2016). The most recent studies (Cerezo et al., 2016) suggest Southeast and East Africa rather than Central or South Africa as previously proposed (Rito et al., 2013). Regarding time, the last common ancestor of modern human mtDNAs possibly arose ~180 Kya, at a time of low population size. This 'mitochondrial Eve' (MRCA) split into haplogroup L0 and a second branch comprising L1'6 haplogroups, which include the L3 branch that migrated out of Africa 50-70 Kya (Soares et al., 2012; Kivisild, 2015). L3 haplogroup diversified in subclades M (India and Eastern Asia) and N (Caucasus through Levant) macrohaplogroups which will give rise to all the haplogroups spread in Europe, Asia, and later to America, as can be seen in Figure 12. All mtDNA haplogroups found nowadays in populations with European and Middle Eastern origin descend from the N branch, which split into a number of haplogroups, named H, I, J, K, T, U, V, W and X (except branch X2a which is found among Native Americans).

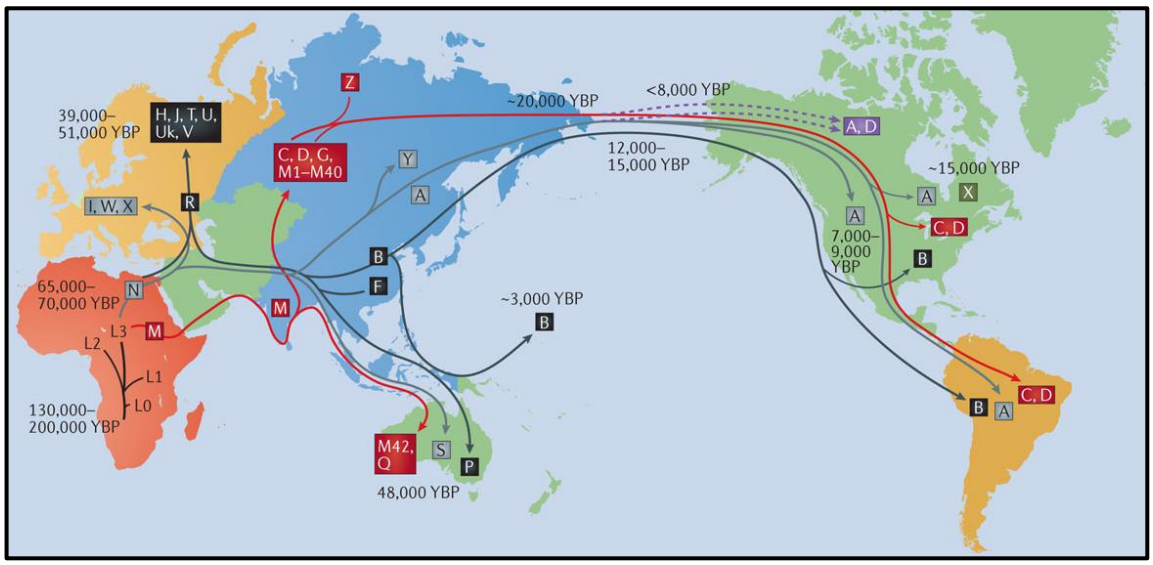


Figure 12. Map of mitochondrial DNA haplogroup migrations (adapted from Stewart and Chinnery, 2015).

In brief, mtDNA is an informative tool for the study of human evolutionary history and, together with the analysis of its counterpart NRY, provides an essential knowledge of sex-specific patterns in and between populations. In the field of Forensic Genetics it is especially useful where the amount of DNA is low or to solve questions involving potential maternal relatives. Although the fact that mtDNA profiles are relatively population specific must be taken into account when conclusions are drawn (Holland and Parsons, 1999).

1.3. Genetic studies in Jewish populations

The rise of modern Population Genetics in the second half of the twentieth century allowed geneticists to endeavour to define the origins and relatedness of Jewish people (Ostrer and Skorecki, 2013). Some of the questions put forward by researchers were: a) Do all contemporary Jews come from the ancient Israelites of the Middle East of three thousand years ago? b) Does DNA evidence show that European Jews are mainly comprised of European ancestors from the host populations? c) Were the Ashkenazi Jews descendants of the Khazars (an ancient tribe with roots in central Asia and Russia that converted to Judaism in the 8th century)? (Levy-Coffman 2005) and d) Can what must be a complex system of interrelations between Jewish groups and between Jews and their host peoples be unravelled?

Early studies were based on the so-called *classical markers* (blood groups, enzymes, serum markers, immunoglobulins and human leukocyte antigen), and provided the first evidence that the most Jewish diaspora groups originated in the Middle East and were more similar amongst themselves than compared with their corresponding host non-Jewish populations (e.g. Bonné-Tamir et al., 1978; Carmelli and Cavalli-Sforza, 1979; Karlin et al., 1979; Kobyliansky et al., 1982; Livshits et al., 1991; Nevo et al., 1996). Subsequent studies in DNA autosomal markers, based on STRs (Rosenberg et al., 2001; Picornell et al., 2002; Kopelman et al., 2009; Listman et al., 2010) also anchored the origin of different Jewish populations to the Middle East. However, the different studies inferred a variable component of Middle Eastern ancestry, with varying degrees of admixture and introgression from the corresponding Diaspora host populations (Ostrer and Skorecki, 2013).

More recent studies based on genome-wide SNP arrays have been performed on several populations with Jewish origin (e.g. Atzmon et al., 2010; Behar et al., 2010; 2013; Campbell et al., 2012; Waldman et al., 2016). The combined analysis of millions of polymorphism markers along the genome have led to greater precision in the clustering of different Jewish groups and in the estimate of the Middle Eastern, European, and African components in each group. These new analyses reflect that Jews are a mosaic of people (Levy-Coffman, 2005). In spite of the close genetic interrelatedness of many groups, modern Jews exhibit a diversity of genetic profiles. Each of today's Jewish populations is the result of the blending of Middle Eastern and host population (European, Asian or African) heritage to which its particular history has led through the centuries.

Genome-wide studies have also been carried out in an attempt to discover new genes for susceptibility to diseases in Jewish populations (Ostrer, 2001). It is well known that some genetic disorders are more prevalent in particular Jewish populations. This fact has been especially well-studied in Ashkenazi Jews (Rosner et al., 2009), where bottlenecks through their history (Behar et al., 2004b) could explain the relatively high frequency found in more than 20 recessive disorders, such as Gaucher or Tay-Sachs diseases (Jorde, 1992). The aim

in Clinical Genetics in the current genomic era is whole exome sequencing to address personalized medicine in the future (Ostrer, 2011).

In order to delve further into the origins of Jewish people and the demographic events that have contributed to the current make-up of the different Jewish groups settled in a vast geographic span after the Diaspora, haploid marker studies have been actively used. These research studies make it possible to focus attention on the paternal (Y-chromosome) and maternal (mtDNA) history of each population studied.

Concerning the Y-chromosome, early studies were based on Restriction Fragment Length Polymorphisms (RFLPs) (Lucotte and David, 1992; Lucotte et al., 1993; Santachiara Benerecetti et al., 1993) and searched for a common origin of modern Jewish male lineages. They also attempted to estimate the contribution of males from the host populations to the gene pool of the different Jewish groups. Subsequent studies (Hammer et al., 2000; Thomas et al., 2000; Nebel et al., 2001; Picornell et al., 2002; 2004) indicate that European, North African and Middle Eastern Jewish populations are more similar between them and other Middle Eastern populations, than between them and their host populations, suggesting, therefore, that most Jewish communities share an ancestral common Middle Eastern origin, and that they remained relatively isolated from neighbouring non-Jewish communities during the Diaspora.

Even though most of the studies focus on Ashkenazim (e.g. Behar et al., 2003; 2004a; Nebel et al., 2005) other Jewish communities have also been studied in the last decades, for example Middle Eastern, Sephardic Jews, Yemenite or Ethiopian Jews (e.g. Ritte et al., 1993; Nebel et al., 2001; Semino et al., 2004; Shen et al., 2004; Adams et al., 2008; Oefner et al., 2013). Furthermore, the male lineages of the Lemba, black Jews of Southern Africa (Spurdle and Jenkins 1996; Thomas et al., 2000; Soodyall, 2013) and Indian Jews (Chaubey et al., 2016) have been analysed. Although a specific Middle Eastern ancestry component is found in most of the studies, more distant geographical groups (such as Ethiopian, Lemba or Indian Jews) have a very important admixture with local populations in accordance with their history.

Other studies have concentrated on the Jewish priestly lineages, Cohanim and Levites. Cohanim (both Ashkenazi and Sephardic) reveal very high frequencies of a haplotype called CMH (Cohen Modal Haplotype) (Skorecki et al., 1997). The origin of CMH has been estimated at 3190±1090 years, supporting the hypothesis that the majority of contemporary Jewish priests descend from a limited number of paternal lineages with their origin in the Middle East before the dispersion of the Jewish people into separate communities (Hammer et al., 2009). Meanwhile, Levites show greater diversity, suggesting multiple origins in their paternal lines (Behar et al., 2003). This study claimed a founder effect for the haplogroup R1a-M17, and they suggested that it was an introgression by an Eastern European male or small group of males. Posterior studies performed with whole

sequences of Y-chromosomes (Rootsi et al., 2013) indicated that in fact there exists a founder effect by the R1a haplotype, but the Levite one comes from the Near East and not from Eastern Europe as thought at the beginning, since this haplotype is present in Hebrew populations pre-Diaspora. A recent study (Tofanelli et al., 2014) questions the previous results of Jewish Priestly lineages and indicates the need to look further into Y-chromosome typing to refine the phylogeny of this marker and ultimately clarify their role in the study of the Jewish groups' ancestry.

Regarding mitochondrial DNA, although the first studies were performed with RFLPs (e.g. Bonné-Tamir et al., 1986; Zoosmann-Diskin et al., 1991; Ritte et al., 1993), subsequent analyses were based on sequencing, first by means of fragments of the Control Region, later on expanded to the complete D-loop, and more recently to the whole mtDNA genome. Early studies in Jewish populations (Thomas et al., 2002; Picornell et al., 2006) indicated that each of the different Jewish communities is composed of descendants of a relatively small group of maternal founders, that the process was independent and different in each geographical area, and that inward gene flow from the host populations has probably been very limited on the female side.

Like in the Y-chromosome, many studies focus on Ashkenazi Jews (Behar et al., 2004b; 2006; Brandstätter et al., 2008; Costa et al., 2013; Tian et al., 2015). Mitochondrial DNA reveals a genetic bottleneck in the early history of the Ashkenazi Jewry. In fact, current variation in this Jewish group can be traced back to only 4 women, who would be the major founders, and several minor contributors. It is a matter of controversy where these 4 main contributors came from, some studies claim the Near East, while others suggest the Caucasus or Europe. A recent study (Tian et al., 2015) even suggests that there is a contribution to the Ashkenazi population that comes from the Far East through the Silk Road. Behar et al. (2008b) studied non-Ashkenazi communities to compare how these were founded. Contrary to the Ashkenazim, no evidence for narrow founder effects was found in these populations. For Indian and Ethiopian communities, an important local female contribution was detected, while mtDNAs in all other communities studied belong to a well-characterized West Eurasian maternal pool. In North African and Iberian Exile Jewish communities, a putative Iberian admixture has been shown. These results led them to conclude that there are striking differences in the demographic history of the widespread Jewish Diaspora communities.

Converted Jews have also been subject of study, either due to their contribution to the host population genetic pool (e.g. Carvajal-Carmona et al., 2000; Maca-Meyer et al., 2003; Gonçalves et al., 2005; Sutton et al., 2006; Adams et al., 2008; Santos et al., 2010; Velez et al., 2012; Bedford et al., 2013; Marques et al., 2016), or owing to their isolation and differentiation from their neighbours, such as the Portuguese Crypto-Jew communities in Belmonte and Bragança (Gerber et al., 2000; Behar et al., 2008b; Teixeira et al., 2011; Nogueiro et al., 2010; 2015a; 2015b) or the Chuetas in Majorca.

Regarding the subject of study in the present work, the Chueta population, it has previously been studied using classical markers (Picornell et al., 1990; 1991; 1992; 1994; 1997; Nevo

et al., 1996); autosomal STRs (Tomàs et al., 2000); HLA (Crespi et al., 2002; Cambra et al., 2009); and genetic illnesses (Buades et al., 1995; Domingo et al., 2000; Guix et al., 2002; Matas et al., 2006). The main points of this genetic research were: a) to investigate whether this group was a genetically differentiated population with respect to their neighbours; b) to test to what extent they have kept their genetic make-up, keeping evidence of their historical Middle Eastern origin; c) to estimate the extent of admixture with the host Majorcan population; and d) to investigate possible impoverishment of genetic diversity due to isolation and inbreeding.

In summary, the results of the studies conducted to date on the one hand demonstrate that Chuetas are a differentiated population that has kept a considerable proportion of its original genetic make-up. It was especially clear in some markers, where Chuetas showed polymorphic frequencies of alleles that are very rare in neighbour populations, but not in Middle Eastern populations. Considering their well-documented Jewish origin, the most probable reason for these findings is their ancestral genetic patrimony due to their Middle Eastern origin (Picornell et al., 1994; 2005; Nevo et al., 1996). On the other hand, a certain degree of introgression from and admixture with the host population was also detected. The extent of this admixture was estimated at approximately 50% using present-day Sephardim and Majorcans as parental populations (Tomàs et al., 2000). Lastly, despite being a small, isolated, endogamous group, Chuetas seem not to have suffered a significant loss of diversity, although the observed singular frequency in some mutations seems to point to genetic drift as the most probable cause (Matas et al., 2006).

Haploid markers, data from the X-chromosome and new autosomal approaches would contribute to produce a more detailed genetic landscape of Chuetas, the only current Spanish population whose ancestors can be traced to Sephardic Jewish populations, because of their peculiar history which has kept the memory of their Jewish origin for centuries, along with their genetic heritage.

Aims

2. Aims

The main goal of this PhD thesis was to better characterise the Chueta gene pool and to answer unsolved questions, especially regarding their maternal and paternal lineages, since previous studies in autosomal markers have demonstrated that they are a genetically differentiated population. In order to do this, we have studied in depth the Genetics of this community through new autosomal markers (Indels), X-chromosome, Y-chromosome, and mtDNA mitochondrial analyses.

To achieve this main purpose, specific aims were proposed:

- Characterization of a 38-plex panel of autosomal Indels in Chuetas, their host population (Majorca), and other populations with Jewish origin.

- To examine different X-markers (9 Alu insertion polymorphisms, 12 STRs, and 32 Indels) in these populations.

- To evaluate the usefulness of these markers in both Population and Forensic genetics, and establish specific frequency databases for forensic casework.

- To investigate the ancestry and demographic history of the maternal and paternal founding lineages of the Chueta population.

- To analyse whether cultural isolation has led to the impoverishment of genetic diversity in mtDNA and Y-chromosome lineages in Chuetas.

- Through a comparison with other populations in the literature, to attempt to estimate the amount of Sephardic and Majorcan contribution to the Chueta population

- To find out, based on data of X-chromosome, mtDNA, and Y-chromosome in Chuetas, their host population and Jewish populations with whom they share origin, whether there has been an asymmetrical sex-biased contribution from the Sephardic and Majorcan populations to the Chuetas' gene pool.

Materials and methods

3. Materials and methods

3.1. Samples

An analysis of the genetic structure of the Chueta population needs knowledge of its host population (Majorca-MAJ), and of the populations with which it shares origin (Jewish populations). Therefore, in markers where no data existed in the literature for these populations, these were also studied.

Jewish populations were categorized in four groups following classical criteria: Sephardic Jews (SEP) (those exiled from the Iberian Peninsula after the 1492 expulsion to Bulgaria and Turkey), North African Jews (NAJ) (Morocco, Libya and Tunisia), Middle Eastern Jews (MEJ) (Iraq and Iran), and Ashkenazi Jews (ASH) (Table 2).

DNA samples from unrelated individuals were obtained. The individuals in the Jewish populations belonged to the National Laboratory for the Genetics of Israeli populations at Tel-Aviv University, and most of the Chueta (CHU) and Majorca individuals to the collection of the Genetics Laboratory, University of the Balearic Islands. Additional blood and mouthwash samples from Chueta individuals were collected for this study. The criteria that were applied to the Chuetas to be included in the study were: a) self-identification as belonging to the community; b) having two Chueta surnames; and c) having at least three generations of maternal and paternal Chueta lineages. All participants signed the informed consent form and provided information concerning their parents' and grandparents' origins.

Table 2. Individuals studied in each marker genotyped in this work.

	Autosomal	**X-Chromosome**			**Y-Chromosome**		**Mitochondrial DNA**
	Indel	Alu	Indel	STR	STR	SNP	D-loop
Chuetas	136	96	72	87	100	100	104
Majorca	102	67	79	63	46	46	79
SEP	106	64	100	60	-	-	-
NAJ	60	52	55	54	-	-	-
MEJ	54	50	53	55	-	-	-
ASH	55	55	54	54	-	-	-
Total	513	384	413	373	146	146	183

3.2. DNA extraction

DNA extraction was performed from buccal swabs or bloodstains. The procedure followed was similar in both cases. Buccal swabs or bloodstain cuts (one mm fragments) were put in 500 µl of DLB buffer, and 50 µl of SDS 10% and five µl of proteinase K were added.

Then, the samples were placed in a bath at 56 °C for three hours. After incubation, 20 µl of NaCl 5M, 250 µl of phenol and 250 µl of chloroform-isoamylalcohol solution (24:1) were added and the suspension was centrifuged at 15777 g for ten minutes. The supernatant was mixed with 500 µl of chloroform-isoamylalcohol solution (24:1) and centrifuged at 15777 g for ten minutes. Finally, the upper phase was collected carefully and transferred into a new tube, and 1 ml of absolute ethanol was added to precipitate the DNA. The solution was mixed and stored at -20 °C. The day after, the samples were centrifuged for 30 minutes at 11337 g. Then all the liquid was eluted and the pellet dried completely. Finally, 30 µl of Tris-EDTA (TE) buffer (pH: 7.5) was added. In order to optimise the use of the samples, these were quantified with spectrophotometer NanoVue Plus (GE Healthcare) and diluted to 2 or 20 ng/µl concentrations.

3.3. Molecular analyses

3.3.1. Autosomal Indels

For the study of autosomal markers, 38 insertion-deletion polymorphisms described in Pereira et al. (2009) (Figure 13) were genotyped following the recommendations of Pereira and Gusmão (2012). Primers (whose sequences are confidential) were provided by Rui Pereira (IPATIMUP).

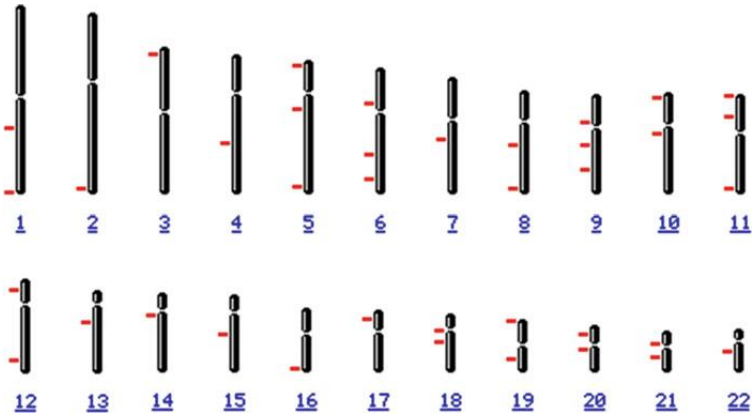


Figure 13. Location of the 38 Indels of the multiplex studied (Pereira and Gusmão, 2012).

Amplification of the 38 Indels was performed in a single multiplex PCR using 5 µl Qiagen Multiplex PCR kit (Qiagen), 1 µl primers, and 0.3–5 ng of genomic DNA in a final reaction volume of 10 µl.

Thermal cycling conditions were: initial incubation at 95 °C for 15 min; followed by 10 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 60 s; plus 20 cycles at 94 °C for 30 s, 58 °C for 90 s, and 72 °C for 60 s; and a final extension at 72 °C for 60 min.

PCR products were separated by capillary electrophoresis (see section 3.4).

3.3.2. X-chromosome Indels

The amplification of the 32 Indels (Figure 14) described in (Pereira et al., 2012a) was performed in a single multiplex PCR using 5 µl Qiagen Multiplex PCR kit (Qiagen), 1 µl primers (Table 3), and 0.3–5 ng of genomic DNA in a final reaction volume of 10 µl. Primer mix was provided by Rui Pereira (IPATIMUP).

Thermal cycling conditions were: initial incubation at 95 °C for 15 min; followed by 30 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 45 s; and a final extension at 72 °C for 60 min. PCR products were separated by capillary electrophoresis.

Figure 14. Idiogram of the X-chromosome showing the location of the 32 Indels studied (Pereira et al., 2012a).

Table 3. Sequence of the 32 pairs of primers of the X-Indels genotyped in this study. PET, NED, VIC and 6FAM are the dyes used to label primers (Applied Biosystems). Lowercase italic letters represent nucleotide tails added to the primers.

	Forward primers	Reverse primers
MID2612	gACCCACGGTGTTGAATTCAG	NED-CACAGCACCAGGAAAATAGC
MID3712	gtttAGTCTTGCTGCAATGTACCC	VIC-TTCAAGGGCAATGATGTTTG
MID357	gTTTTATAGACTGTGGCCCCC	PET-GTTAGTGGTTGGATTGCTCG
MID356	gtttCCAACTCCACGTGAGAAATG	PET-AGTCTGATGCAGTGGCAAAC
MID3703	VIC-AGCTTCCAAGTAGTTCTGCC	gTTTGGCTTACTTCCTCCTCC
MID3774	gAAGACGGGAATTGAGTCACC	NED-TTTTTGTGCACAGGCACTCC
MID3692	6FAM-ACATAAAAGCAAGCTTTGGC	gtttcttCCCGGTGTGTGAACTTTTTC
MID3716	6FAM-AAAGGGAGCATCTACTCCAG	gtttctAGGGCAATCCAGAATTGGAC
MID3690	GGGCACCATATTAGGCATGT	VIC-CCCACCATCTAACCCATTTC
MID3719	gTTCTTTCTCATCTGGCACCC	VIC-CTATGAAGCCTATAGATTGG
MID2089	VIC-AATCCATTCTGGAATAAGATGTCA	gtttcttTCCACTCTCAGGGATTCCTTT
MID2692	gtttcttCAAGTTCATATGGTGTCTTGG	PET-TGCATTACACAGAGCAACTC
MID3701	gtttctAGTTGGAGATGCAATGAAGC	NED-AGAGACAGGTGAATTGAGGC
MID198	6FAM-CAGGCACAGGAGAGGAAGAG	gTCCACCCCTAGTTAAACAGC
MID1736	VIC-GTGAAAGGTGAGCTTGTCTG	gtttctAGGCCTTTTTGGTTAACTGG
MID3730	6FAM-AGGATCCTGACTAAGATAGC	GAATCTCTGGAAACACTTGG
MID1511	gCTGCCTGGGATTTTTCCTTT	PET-CAGGGGAGAACACCCACTAA
MID3740	gtttctACTTGCTTTGCTTTTCCCTC	6FAM-GTACAACTGCAAGGAACRAGc
MID3732	VIC-CAGAGTCATCTATTCCCCAG	gtttcttCACCCATGTGGTTTCATTTC
MID3727	gtttctGGTGGAATTCTTTTGCATGTG	PET-TTTTGGGAAGCACACTCACC
MID3754	NED-TTTCACCAAGGACTTGAAAGG	gttCAGCTCACACTAGGGCCTTC
MID3722	VIC-TGGCCCTTCTGAGTTCAAAC	gCAGTGTAATAAGGTGGGAGC
MID1361	6FAM-TCAGTCCTTAACAAGGGAGC	gtttGTCATTGTGAAGGCTACCTG
MID243	PET-TACAGTTGGCTGCTTTTCCC	gtttcttATACGAAGATCTGTGGGAAC
MID2637	gtttctTATGTGTCAAAATGGGAGGC	6FAM-AATCCTCAAATCACAGTGGC
MID111	GAGGCAGGGAAATCAGTTAG	NED-TTGATTCCAGCTTTCCCTTT
MID3736	6FAM-GGGTTAGGAGCCCCTGCT	gtttcttGGATGTATGACACACAACGC

MID3753	GCTACACCAATGGACAGATG	PET-TGTGGTGTGCATGATTTGC
MID1839	gttGATATCCCATAACGCCCATTT	NED-TCCTTTTGCTACGCAGACCT
MID3760	gCAAGGTTCGTACTCATTTAG	NED-AACCTAGTTCACAACCCCTG
MID329	gTCTCAAAACCTTCCCATGGC	6FAM-AGAAGTTAGAGGGTGTCTGG
MID2652	PET-GCTGCTCTTTGCTTTAATTTC	gTATGGTAGGCACTGTGCTAA

3.3.3. X-chromosome STRs

Twelve X-chromosome STR markers (DXS10148, DXS10135, DXS8378, DXS7132, DXS10079, DXS10074, DXS10103, HPRTB, DXS10101, DXS10146, DXS10134, and DXS7423) were amplified using the Investigator Argus X-12 kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations.

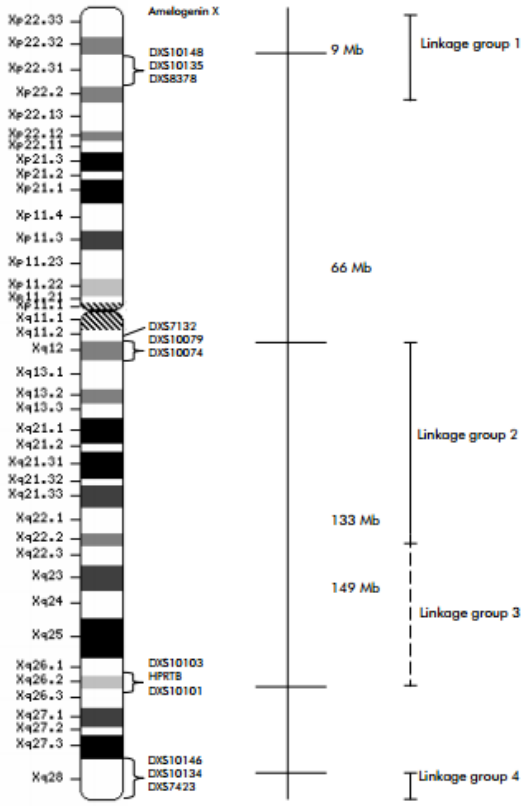


Figure 15. Twelve X-STRs of the Investigator Argus kit 12, and their position in the X-chromosome (Investigator Argus X-12 Handbook 06/2013).

As shown in Figure 15, the twelve markers are grouped by linkage in four groups. Linkage group (LG) 1 is comprised of DXS10148-DXS10135-DXS8378 markers, LG2 contains DXS7131-DXS10079-DXS10074 markers, LG3 is made up of DXS10103-HPRTB-DXS10101 markers, and finally DXS10146-DXS10134-DXS7423 markers compose LG4.

PCR was performed in a multiplex reaction using 3.95 µl reaction mix A, 0.625 µl primer mix, 0.15 µl Multi Taq2 DNA Polymerase, and 0.3–5 ng of genomic DNA in a final reaction volume of 7 µl. Positive controls were included in each PCR run. PCR amplification was carried out in five steps: initial incubation at 94 ºC for 4 min followed by 5 cycles: 30 s at 96 ºC, 120 s at 63 ºC, and 75 s at 72 ºC; 25 more cycles: 30 s at 94 ºC, 120 s at 60 ºC, and 75 s at 72 ºC; and a final extension at 68 ºC for 60 min. PCR products were separated by capillary electrophoresis.

3.3.4. X-chromosome Alu insertions

Nine Alu insertion polymorphisms (Ya5DP62, Yb8DP49, Yd3JX437, Yb8NBC634, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4 and Ya5DP13) described in Callinan et al. (2003) were analysed (Figure 16).

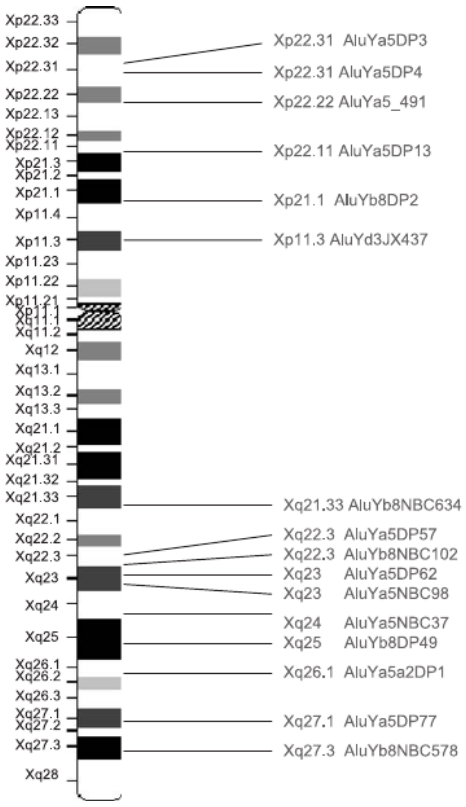


Figure 16. Idiogram of human sex chromosome-specific Alu insertion polymorphisms (Callinan et al., 2003).

Amplification of the 9 Alu polymorphisms was performed in a single PCR using 5 µl Qiagen Multiplex PCR kit (Qiagen), 1 µl primers, and 0.3–5 ng of genomic DNA in a final reaction volume of 10 µl. A negative control was included in each PCR run.

Thermal cycling conditions were: initial incubation at 95 °C for 15 min; followed by 30 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 45 s; and a final extension at 72 °C for 60 min.

Correct amplification checking and allele reading was carried out by agarose gel electrophoresis. Agarose gels were prepared at 2% concentration with TBE buffer and 0.5 µg/ml final concentration of ethidium bromide solution. After polymerization, 1 µl of each PCR product was mixed with 4 µl of loading buffer (bromophenol blue dye and glycerol) and loaded into the wells of the gel. Visualization was performed directly by UV fluorescence.

3.3.5. Y-chromosome STRs

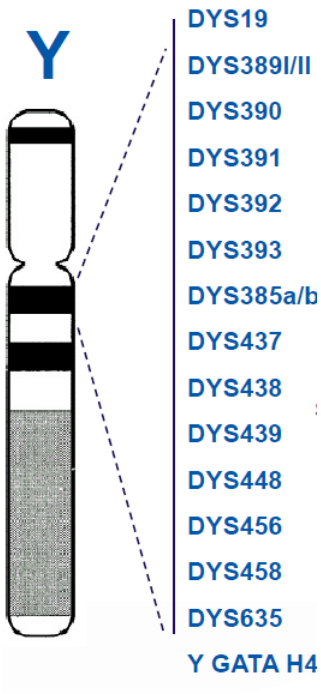


Figure 17. Y-chromosome markers included in the Y-filer kit (Applied Biosystems, Foster City, CA, USA).

Seventeen Y-chromosome STR markers (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a, DY385b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA_H4, DYS437, DYS438 and DYS448x) were amplified using the Y-filer kit (Applied Biosystems, Foster City, CA, USA) (Figure 17) following the manufacturer's recommendations.

Amplification of the Y-STRs was performed in a single multiplex PCR using 1.84 µl AmpFlSTR Yfiler PCR Reaction Mix, 1 µl AmpFlSTR Yfiler Primer Set, 0.16 µl AmpliTaq Gold® DNA Polymerase, and 0.3–5 ng of genomic DNA in a final reaction volume of 5 µl.

PCR products were separated by capillary electrophoresis. Once the haplotype of each sample was obtained, the online tool Whit Athey's Haplogroup predictor (Athey, 2013) was used in order to tentatively assign a haplogroup and so decide which set of SNPs was better to genotype in each sample so as to finally achieve its exact haplogroup.

3.3.6. Y-chromosome SNPs

Thirty-eight SNPs were typed to define the major male lineages. Thirty-one of them were genotyped using SNaPshot™ kit (Applied Biosystems) technique in five multiplexes. Multiplex 1: M13, $SRY_{1532.1}$, M213, M9, Tat, 92R7, M173, $SRY_{1532.2}$, P25 and M70; multiplex 2: M201, M170, M26, 12f2a, M62 and M172 (Figure 18); and multiplex 3: M78, M81 and M123, were described by Brion et al. (2005) and modified by Gomes et al. (2010). Besides, multiplex R: L23, U106, M153, M167, U152 and M529, described in Marques et al. (2016) and multiplex Q: M242, P36.2, M346 M3, M19, M194 and M199, described by Roewer et al. (2013) were typed in the samples belonging to the R and Q haplogroups.

SNPs M1 and M269 were genotyped with conventional PCR followed by agarose gel electrophoresis; while S116, M17 and M18 were genotyped by Sanger sequencing (explained extensively in section 3.7), with the primers in Table 4. DYS458.2 was used to determine the J1 chromosomes (Myres et al., 2007).

Table 4. Primers used to amplify SNPs M1, M269, S116, M17, and M18.

SNP	Forward primers	Reverse primers
M1	CAGGGGAAGATAAAGAAATA	ACTGCTAAAAGGGGATGGAT
M269	CATGCCTAGCCTCATTCCTC	CTGGATGGTCACGATCTCCT
S116	TCAGTCAGGGCAAATCTGAA	GGTGGAGTTGGGGCTAAAGT
M17/18	CCCCAGTTTGTGGTTGCTGGTTGTTA	AGCTGACCACAAACTGATGTAGA

Amplification of the multiplexed SNPs was performed in multiplex PCRs using 5 µl Qiagen Multiplex PCR kit (Qiagen), 1 µl primers, and 0.3–5 ng of genomic DNA in a final reaction volume of 10 µl.

PCR conditions for the amplification of the multiplexes were: initial incubation at 95 °C for 15 min; followed by 35 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 45 s; and a final extension at 72 °C for 10 min. Then agarose gel electrophoresis was carried out to ensure the amplification was performed correctly.

Purification of the samples was performed by mixing 1 µl PCR product and 0.80 µl ExoSAP-IT reagent. Then two incubations at 35 °C for 15 min followed by 85 °C for 15 min.

Afterwards, SNaPshot reaction was performed by adding 1 µl SNaPshot mix provided with the kit, and 1 µl SBE (single-base extension) primers in a final reaction volume of 5 µl to the purification tubes. Conditions for the SNaPshot reaction were: 25 cycles at 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s. Then, a final purification was conducted by adding 1 µl SAP to each tube and incubating at 37 °C for 60 min and at 85 °C for 15 min.

The SNaPshot chemistry is based on the dideoxy single-base extension of an unlabelled oligonucleotide primer (or primers). Each primer binds to a complementary template in the presence of fluorescently labelled ddNTPs and AmpliTaq® DNA polymerase. The polymerase extends the primer by one nucleotide, adding a single ddNTP to its 3´ end (Figure 19).

SNaPshot products were separated by capillary electrophoresis.

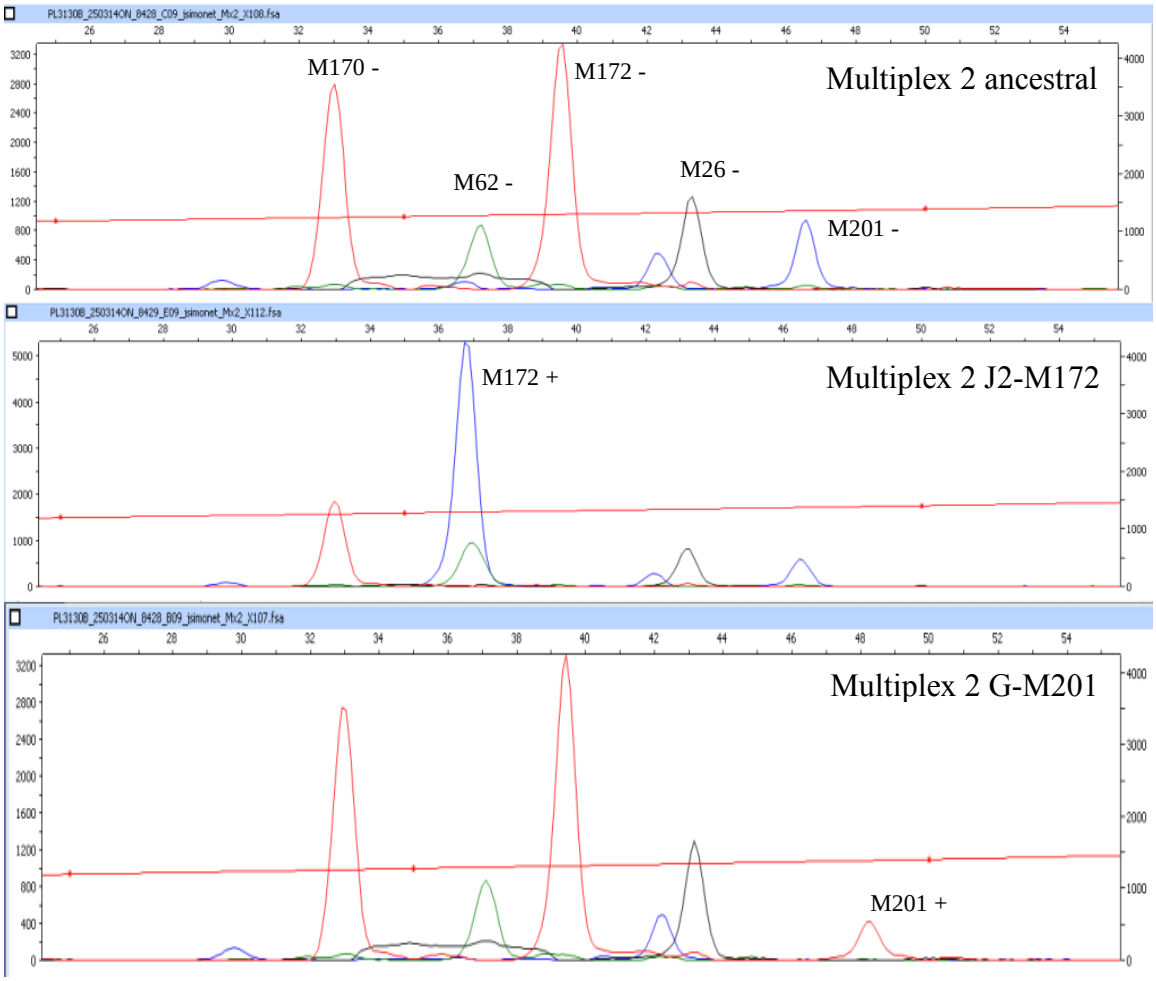


Figure 18. Multiplex 2 results showing an ancestral sample, a J2 M-172 sample, and a G-M201 sample.

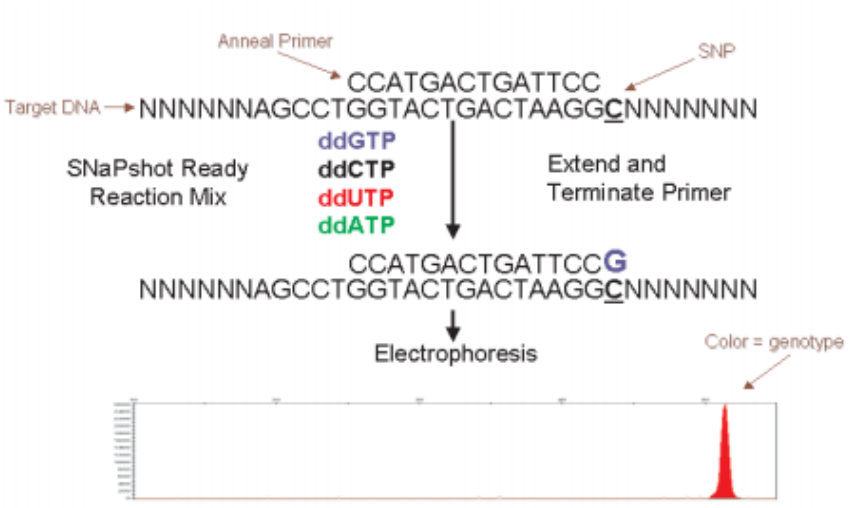


Figure 19. SNaPshot technique amplification, based on SBE primers minisequencing (ABI PRISM® SNaPshot™ Multiplex kit user manual).

3.3.7. Mitochondrial DNA sequencing

Sequences of mitochondrial DNA were obtained by conventional PCR followed by Sanger sequencing technique. For all samples, the amplification of the D-loop was performed, for 17 selected samples, all the mitochondrial genome was sequenced.

Mitochondrial DNA amplification was performed using 5 µl Qiagen Multiplex PCR kit (Qiagen), 0.5 µl forward and reverse primers, and 20–25 ng of genomic DNA in a final reaction volume of 10 µl. A negative control was included in each PCR run.

PCR conditions for the amplification were: initial incubation at 95 °C for 15 min; followed by 35 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 45 s; and a final extension at 72 °C for 10 min. Then agarose gel electrophoresis was carried out to ensure the amplification was performed correctly.

Samples amplified correctly were purified with the MBS® Spin PCRapace (Invitek) following the manufacturer's procedure. After quantification of purified samples, the sequencing reaction was performed for the two DNA strands (forward and reverse) separately using 0.6 µl BigDye® Terminator v.3.1 cycle sequencing premix (Applied Biosystems), 1.2 µl buffer, 1 µl primer, and 1 µl purified sample in a final reaction volume of 10 µl.

PCR conditions for the sequencing reaction were: initial incubation at 96 °C for 2 min; followed by 35 cycles at 96 °C for 15 s, 50 °C for 9 s, and 60 °C for 120 s; and a final extension at 60 °C for 10 min.

In order to purify the samples, conventional precipitation with ethanol was performed. First, samples were stored at -20 °C for 15 min with absolute ethanol (100%) and sodium acetate 3M. Then samples were centrifuged for 30 min at 15777 g to precipitate. After removing the supernatant, pellets were washed with 70% ethanol, followed by another centrifuging step of 10 min. Samples were dried in a vacuum pump and, finally, diluted in mili-Q water for capillary electrophoresis in the genetic analyser.

Once the haplotype had been obtained for each sample, haplogroups were classified following the updated mtDNA phylogeny, PhyloTree, mtDNA tree Build 17 (http://www.phylotree.org/) using HaploGrep2 tool (van Oven, 2015; Weissensteiner et al. 2016), and assigned haplotypes were validated by EMPOP (EMP00672) (Parson and Dür, 2007; http://empop.org/) curators.

3.3.7.1. D-loop

The entire D-loop (non-codifying control region) was amplified (position 16024 to 576). Amplifications were conducted in two fragments called F1 and F2 whose primers are specified in Table 5. Amplification and sequencing was carried out using the same primers.

Table 5. Primers used to amplify the D-loop region of mtDNA.

		Forward primers		Reverse primers
D-loopF1	L15997	CACCATTAGCACCCAAAGCT (Alonso et al., 2003)	H016	CCCGTGAGTGGTTAATAGGGT (Parson and Bandelt, 2007)
D-loopF2	L16555	CCCACACGTTCCCCTTAAAT (Marques et al., 2015)	H639	GGGTGATGTGAGCCCGTCTA (Parson and Bandelt, 2007)

3.3.7.2. Complete mtDNA genome

Attempting to go further into the classification of female lineages, complete mitogenomes were sequenced in those cases that could be relevant for the study. Primers described by Ramos et al. (2009) were used to redesign a strategy of amplification in order to attain shorter, easily amplifiable, fragments. Nineteen overlapping fragments were amplified (Table 6) and finally 31 inner fragments were sequenced by Sanger technique to obtain the whole genome.

Table 6. Primer pairs from Ramos et al. (2009; 2011) used to amplify mtDNA complete genomes.

Fragment	Primer pair	Fragment	Primer pair
MTT1A	14898for/15826rev	MTT5B	7713for/9220rev
MTT1B	15416for/151rev	MTT6A	8910for/10154rev
MTT2A	16488for/1159rev	MTT6B	9874for/10648rev
MTT2B	909for/1677rev	MTT7A	10360for/11673rev
MTT3A	1404for/2801rev	MTT7B	11461for/12226rev
MTT3B	2646for/3947rev	MTT8A	11977for/13297rev
MTT4A	3734for/5017rev	MTT8B	12988for/13830rev
MTT4B	4896for/6154rev	MTT9A	13477for/14838rev
MTT4C	5995for/6739rev	MTT9B	14440for/15349rev
MTT5A	6511for/8000rev		

3.3.8. Mitochondrial DNA SNPs

SNPs were genotyped by sequencing in those cases that could be relevant to better classify female lineages.

In the samples belonging to haplogroups R0a, K1a1b1a, T1a, and T2c1d, key mutations found in the mitogenomes amplified (Table 7) were genotyped, by sequencing the fragment containing the position of interest.

Table 7. Key mutations studied in the samples belonging to haplogroups R0a, K1a1b1a, T1a, and T2c1d.

Haplogroup	Key mutations
R0a	4767G, 13858G and 15734A
K1a1b1a	8029T
T1a	6656T and 10116G
T2c1d	11914A, 12363T, 1306T, 14544A, 3027C
	8475T, 8911C, 8980A and 15569T

In the case of the H haplogroup samples, a set of 51 SNPs arranged in 3 multiplexes designed by Alvarez-Iglesias et al. (2009) were typed using SNaPshot technique. SNPs typed were: 709, 750, 2581, 3010, 3796, 6253, 6296, 6365, 6776, 7337, 10810, 12858, 12957, 13708, 13759, 14365, and 14470A for multiplex 1; 951, 3915, 3936, 3992, 4310, 4336, 4727, 4745, 4769, 4793, 7028, 7645, 8569, 8473, 8592, 8598, 8602, 9066, 9150, 10044, 10394, 13404, 8271T, and 961G for multiplex 2; and 1438, 2259, 8994, 10166, 10211, 11140, 14869, 14872, 15833, and 13101C for multiplex 3.

3.4. Capillary electrophoresis and genotype analyses

For all the markers where capillary electrophoresis was used, analyses were performed in an ABI 3130/3130XL DNA Analyser.

PCR products were prepared for capillary electrophoresis by adding 1 µl of each amplified product to 12 µl Hi-Di™ Formamide (Applied Biosystems) and 0.5 µl of internal size standard (LIZ). For Indels and Y-STRs, GeneScan™ 500 LIZ® (Applied Biosystems) was used. SNaPshot products were analysed with GeneScan™ 120 LIZ® (Applied Biosystems), whereas for X-STR products 550 BTO LIZ (Qiagen) was used.

Results were visualised and analysed with the following programs: GeneMapper v.3.2 software (Applied Biosystems) for STRs and Indels; Peak Scanner software (Applied

Biosystems) for SNaPshot results; and BioEdit (Hall, 1999) and Geneious (Biomatters, Ltd) software for DNA sequencing.

3.5. Statistical analyses

3.5.1. Intra-population variability

3.5.1.1. Allele frequencies

Allele frequencies are the relative frequency of a particular allele of a specific locus in a determined population. In population genetics, allele frequencies are used to describe the amount of variation at a particular locus in a population.

For Indel, Alu, and STR markers, allele frequencies were calculated using Arlequin v.3.5 software (Excoffier and Lischer, 2010). In X-chromosome markers, allele frequencies of males and females were calculated separately, and then total frequencies were estimated using the following formula for each allele in each marker:

$$p_i = \frac{(2 * female\ frequeny) + male\ frequency}{3}$$

3.5.1.2. Hardy-Weinberg Equilibrium (HWE)

Once we know the allele frequencies of a population, the proportions of the genotypes in the succeeding generation by combining gametes at random can be predicted through the postulate of the Hardy-Weinberg principle.

In our studies, the HWE and p-values were calculated using Arlequin v.3.5 software (Excoffier and Lischer, 2010). In X-chromosome markers the calculations were made by taking into account only the female data.

In statistical significance testing, the p-value measures how the observation compares with the expectation. The null hypothesis can be rejected when the p-value is less than significance level α, which in our case is 0.05. When the null hypothesis is rejected it can be said that the results are statistically significant.

3.5.1.3. Diversity parameters

3.5.1.3.1. Gene diversity (GD)

This is equivalent to the expected heterozygosity for diploid data (X-chromosome and autosomal markers). It is defined as the probability that two randomly chosen haplotypes are different in the population (Nei, 1987).

$$Gene\ diversity\ = 1 - \sum_{i}^{n} (p_i)^2$$

where p_i is the allele frequency of each allele in the sample.

This was performed using only female data in X-chromosome markers and all the samples for Autosomal Indels.

For multiallelic markers, diversity parameters were estimated using Arlequin v.3.5 (Excoffier and Lischer, 2010).

3.5.1.3.2. Haplotype diversity (HD)

Haplotype diversity is a measure of the uniqueness of a particular haplotype in a given population. It is defined as the probability that two randomly chosen haplotypes are different in the population. This parameter is equivalent to gene diversity in haploid markers. It was calculated using Arlequin v.3.5 (Excoffier and Lischer, 2010).

Haplotype diversity is computed as Nei and Tajima (1981):

$$HD\ = \frac{N}{N-1}\ (1 - \sum_{i}^{n} (x_i)^2)$$

Here, x_i is the (relative) haplotype frequency of each haplotype in the sample, and N is the sample size.

This was calculated for male samples in X-chromosome markers. It was also calculated in the X-STRs for the four different linkage groups (LGs) included in the Investigator Argus X-12 kit, and for X-Indels for the markers that were shown to be linked in the linkage disequilibrium analysis. It was also performed for all the samples studied for Y-chromosome STRs, and mtDNA.

3.5.1.3.3. Mitochondrial DNA diversity

In mtDNA analyses additional parameters were calculated: a) K (number of different haplotypes); b) S (number of polymorphic sites); c) π (nucleotide diversity), which is the probability that two randomly chosen homologous (nucleotide or RFLP) sites are different, which is equivalent to gene diversity at nucleotide level for DNA data; and d) Theta (θ).

Theta is a fundamental parameter of molecular evolution that encapsulates the expected level of genetic diversity in a randomly mating, constant-sized population not subject to selection when an equilibrium is reached between genetic drift and mutation. It is defined as:

$$\theta = 2nN_e\mu,$$

where n is the number of heritable copies of the locus per individual (0.5 in the case of mtDNA), N_e is the effective population size, and μ is the mutation rate per nucleotide (or per sequence) and per generation (Nei, 1987; Tajima, 1993).

There are different ways to estimate θ from sequence data; depending on the parameter used the estimator is called θ_s (using the number of polymorphic sites), θ_K (using the number of different haplotypes), etc. The Theta estimator based on the number of different lineages (θ_K) (Tamura and Nei, 1993), which is based on the relationship between sample size and the number of distinct lineages, is more sensitive to the effects of lineage sorting during recent demographic history.

3.5.1.4. Neutrality tests

There are many methods to detect selection, which typically calculate a statistic that compares a feature of the observed diversity to that expected under neutral evolution.

In this work Tajima's D test of selective neutrality (Tajima, 1989; 1993) – which compares the number of segregating sites per site with nucleotide diversity – was estimated in mtDNA analysis. This test compares two estimators of the population parameter. Under the infinite-site model, both estimators should estimate the same quantity, but differences can arise under selection, population non-stationarity, or heterogeneity of mutation rates among sites. The test statistic D is defined as:

$$D = \frac{\hat{\theta}\pi - \hat{\theta}s}{\sqrt{Var\left(\hat{\theta}\pi - \hat{\theta}s\right)}}$$

3.5.1.5. Linkage disequilibrium (LD)

Linkage disequilibrium is an estimate of recombination at a population level. LD measures whether specific alleles at different loci are correlated with one another more or less often than would be expected by chance (Jobling et al., 2014). LD is influenced by many factors, including evolutionary forces and also population characteristics (mating system, population substructure…). Thus, the pattern of LD is a powerful tool to understand past evolutionary and demographic events in human history or in the history of a particular population.

Exact test of LD was calculated for haplotypic data (Y-chromosome, and in males in X-chromosome markers) using Arlequin v.3.5 software (Excoffier and Lischer, 2010) performing 1000000 steps in Markov chains and 1000 steps of demorization.

3.5.1.6. Forensic parameters

To test how efficient the recombinant markers used in this work (STR, Alu and Indel) are for forensic purposes, a series of parameters was calculated.

Polymorphism information content (PIC) (Botstein et al., 1980) and expected heterozygosity (formula equivalent to gene diversity) are devised for more general purposes and are valid for both autosomal and X-chromosome markers.

$$PIC = 1 - \left(\sum_{f=1}^{n} p_i^2\right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

An online tool in http://www.chrx-str.org was used to perform the calculations of X-chromosome markers.

Mean exclusion chance (MEC) was introduced by Krüger et al. (1968) (MECKRÜ) for autosomal markers typed in trios involving mother, child, and putative father. This parameter is not suitable for X-chromosome markers except for deficiency cases in which the paternal grandmother is investigated instead of the alleged father.

$$MECKR\ddot{U} = \sum_i f_i^3 (1 - f_i)^2 + f_i (1 - f_i)^3 + \sum_{i<j} f_i f_j \left(f_i + f_j\right) \left(1 - f_i - f_j\right)^2$$

Kishida et al. (1997) devised a MECKIS for X-chromosome markers that covers trios including a daughter. If MECKRÜ is compared to MECKIS, the latter is considerably

larger. This highlights the fact that in trios involving a daughter, X-chromosome markers are more efficient than autosomal markers.

$$MECKIS = \sum_i f_i^3(1 - f_i) + f_i(1 - f_i)^2 + \sum_{i<j} f_i f_j (f_i + f_j)(1 - f_i - f_j)$$

$$MECD_{trio} = 1 - \sum_i f_i^2 + \sum_i f_i^4 - \left(\sum_{i<j} f_i^2\right)^2$$

$$MECD_{duo} = 1 - 2\sum_i f_i^2 + \sum_i f_i^3$$

Finally, Desmarais et al. (1998) introduced formulae for the mean exclusion chance of X-chromosome markers in trios involving daughters ($MECD_{trio}$) and in father/daughter duos lacking maternal genotype information ($MECD_{duo}$). $MECD_{trio}$ is equivalent to MECKIS whilst $MECD_{duo}$ is also appropriate for maternity testing of mother/son duos.

Power of discrimination in Females (PD_{female}) and power of discrimination in males (PD_{male}) are parameters suitable to assess the power of markers for forensic identification purposes in males and females, respectively.

$$PD_{female} = 1 - 2\left(\sum_i f_i^2\right)^2 + \sum_i f_i^4$$

$$PD_{male} = 1 - \sum_i f_i^2$$

Here, f_i (f_j) are population frequencies of the i^{th} (j^{th}) marker alleles.

For autosomal Indels, PowerStats formulae (Brenner and Morris, 1989; Jones, 1972) were used to calculate forensic parameters. For the match probability (MP) and power of discrimination, the formulae are the following:

$$Match\ probability = \sum_{i=a}^{n} G_i^2$$

$$Power\ of\ discrimination = 1 - \sum_{i=a}^{n} G_i^2$$

where G_1 is the fraction of samples with genotype "i".

To determine the power of exclusion (PE) and typical paternity index (TPI or PI) the following formulae were used:

$$Power\ of\ exclusion = h^2(1 - 2 * h * H^2)$$

$$Typical\ paternity\ index = \frac{(H + h)}{2H}$$

where h defines the number of heterozygotes and H is the number of homozygotes. These formulae are valid for both autosomal and X-chromosome markers.

To test forensic efficiency in Y-STRs, discrimination capacity (DC) was calculated as the percentage of different haplotypes and haplotype match probability (HMP) as 1-haplotype diversity.

3.5.2. Genetic structure and Inter-population variability

3.5.2.1. *F*ST statistic

F_{ST} measures the apportionment of genetic variation between subpopulations; in other words, it compares the genetic diversity found within subpopulations (the "S" of the subscript) to the genetic diversity of the total population (the "T" of the subscript). It can also be regarded as measuring the proportion of genetic diversity due to allele frequency differences between subpopulations. F_{ST} can be estimated from genetic diversity data by a variety of methods, most commonly:

$$F_{ST} = \frac{(H_T - H_S)}{H_T}$$

where H_T is the expected heterozygosity of the entire population, and H_S is the mean expected heterozygosity across subpopulations. F_{ST} measures were performed using Arlequin v.3.5 software (Excoffier and Lischer, 2010).

Jobling et al. (2014) indicated a qualitative guideline that could be used to interpret F_{ST} values (Wright, 1920; 1969) (Figure 20).

SEWALL WRIGHT'S QUALITATIVE GUIDELINES FOR INTERPRETING F_{ST}	
F_{ST} values	Level of genetic differentiation
Less than 0.05	little
Between 0.05 and 0.15	moderate
Between 0.15 and 0.25	great
Greater than 0.25	very great

Figure 20. Guideline of F_{ST} values and its interpretation by Wright (Jobling et al., 2014).

3.5.2.2. Analysis of molecular variance (AMOVA)

Analysis of molecular variance estimates genetic structure indices using information on the allele content of haplotypes, as well as their frequencies. The significance of the covariance components associated with the different possible levels of genetic structure (within individuals, within populations, within groups of populations, among groups) is tested using non-parametric permutation procedures (Excoffier et al., 1992).

AMOVA was performed in Arlequin v.3.5 software (Excoffier and Lischer, 2010) for STR, Alu, and Indel markers.

3.5.2.3. Structure analysis

For autosomal Indels, population structure was further assessed with STRUCTURE 2.3.4 software using default settings (admixture model; correlated allele frequencies) (Pritchard et al., 2000; Falush et al., 2003). All runs included a burn-in period of 50,000 iterations followed by 100,000 MCMC (Markov chain Monte Carlo) repetitions (K = 1-9), and were repeated ten times each in order to test the consistency of the results. K was tested from 1 to 9 because there were 9 different populations: four Jewish populations, 2 Jewish descendant populations (Bragança Jews and Chuetas), and 3 host populations (Portugal, South East of Spain, and Majorca).

In the case of X-chromosome markers, linkage model and correlated allele frequencies were used. All runs included a burn-in period of 50,000 iterations followed by 100,000 MCMC repetitions (K = 1-6), and were repeated ten times each in order to test the consistency of the results. In this case, K was tested from 1 to 6 because the 53 markers were genotyped in four Jewish populations, Chuetas, and their host population, Majorca.

3.5.2.4. Genetic Distances

For use as a genetic distance, F_{ST} can be formulated to compare two populations (known as pairwise F_{ST}), and can be defined as:

$$F_{ST} = \frac{(V_P)}{p(1-p)}$$

where p and Vp are the mean and variance of gene frequencies between the two populations, respectively (Jobling et al., 2014).

Pairwise F_{ST} genetic distances were calculated with Arlequin v.3.5 software (Excoffier and Lischer, 2010) whenever the genotypes of the populations to be compared were available. In those cases when only the allele frequencies were available, pairwise F_{ST} genetic distances were obtained from POPTREE2 software (Takezaki et al., 2010).

3.5.2.5. Multidimensional Scaling (MDS) Plot

Using the matrix of F_{ST} genetic distances, a multidimensional scaling plot was obtained with SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA). MDS is used to determine an n-dimensional space and corresponding coordinates for a set of objects, strictly using a matrix of pairwise dissimilarities between objects; in our case, pairwise F_{ST} genetic distances.

After computations, the software gives an MDS plot, and Stress and R-squared measures. R-squared represents the level of variance in the data which is explained by the hypothesized n-dimensional configuration (Giguère, 2006), and the Stress is a badness-of-fit measure: the higher the stress score, the worse the fit of the plot (Sturrock and Rocha, 2000). To ensure our Stress result was defining a greater structure than would be obtained from a random dataset, the tables suggested by Sturrock and Rocha (2000) were consulted.

3.5.2.6. Admixture estimation

Admixture rates we estimated by means of Admix 2.0 software to better explain the current composition of the Y-chromosome lineages in Chuetas, taking the Majorcan (also studied in this work) and Sephardic populations (Adams et al., 2008) as parental populations.

Admix 2.0 is a coalescence based method (Bertorelle and Excoffier, 1998; Dupanloup and Bertorelle, 2001) which takes mutation and sampling error into account. We created molecular divergence matrices using data from literature, and following software guidelines. The bootstrap procedure was set to 1000000 repetitions.

3.5.3. Phylogenetic Network

Phylogenetic networks were generated with NETWORK 4.6.1.1 (Fluxus Technology Ldt) using the Median Joining Method (Bandelt et al., 1999). Network plot shows the relationship between haplotypes based on the minimal number of mutations between them. These plots were represented for Y-chromosome STRs using the haplotypes, and for mtDNA sequences using a Rohel format input file obtained from DnaSP 5.10.01 software (Librado and Rozas, 2009).

Samples from other populations used to compare with ours were collected from the bibliography and Genebank database (Benson et al., 2013). In the case of mtDNA, different mutations were given different weights following Bandelt et al. (2006) recommendations, in order not to underestimate less frequent mutations in the presence of the most common ones. For Y-STR, the weights were calculated depending on the variance of each locus, giving more weight to those markers with less variance.

Results

Chapter 1: Autosomal markers

1. A GHEP-ISFG collaborative study on the genetic variation of 38 autosomal Indels for human identification in different continental populations

GHEP-ISFG collaborative exercise

Forensic Science International: Genetics (Submitted)

2. Genetic diversity of 38 insertion-deletion polymorphisms in Jewish populations

Ferragut JF; Pereira R; Castro JA; Ramon C; Nogueiro I; Amorim A and Picornell A

Forensic Science International: Genetics. 21: 1–4 (2016)

Introduction

The reconstruction of human evolutionary history is based on uniparental (or non-recombinant) markers, such as Y-chromosome (NRY) and mtDNA, and recombinant ones (autosomes and X-chromosome), as each of them have particular features, as stated in the introduccion section. Hence, the whole information obtained by studying all of them may properly reflect the effect of different demographic events such as founder effects, migrations, colonizations and expansions, and also evolutionary forces, like mutation, genetic drift and selection in the genetic diversity pattern of a modern population (Balaresque et al., 2007; Jobling et al., 2014). Thus, even though the main aim of this PhD thesis was to characterize the paternal and maternal lineages of the Chueta population, it is of crucial importance to also have information regarding autosomal markers, which will supply information that is unavailable from uniparental markers and which will therefore provide additional tools to enable the complete reconstruction of the demographic history of the Chuetas.

Although there are previous studies based on autosomal markers in this population – classical markers (Picornell, 1992) and autosomal STR (Tomàs et al., 2000) – studying these new autosomal markers was considered interesting because Indels have different evolutionary characteristics from STRs or classical markers. For this reason, genotyping this set of 38 polymorphisms in Chuetas allows us to go further into and refine their genetic portrait. Moreover, Indels have an increasing use in forensic casework, for their proven success in genotyping degraded samples.

This first chapter contains the study of autosomal markers, specifically 38 Indels described by Pereira et al. (2009). It is divided into two sections: the first one is the characterization of the Majorcan population, which is included in a collaborative worldwide exercise proposed by the GEPH-ISFG (*Grupo de Habla Española y Portuguesa* of the International Society of Forensic Genetics) (paper submitted), and the second section is the study of this set of markers in 4 Jewish populations (Sephardim, Ashkenazim, North African and, Middle Eastern) and two Crypto-Jewish populations (Bragança and Chuetas) (Ferragut et al., 2016). A total of 568 samples were genotyped taking into account both sections.

Therefore, this work provides a database of this set of markers in populations with Jewish origin and in the insular population of Majorca, which is useful in Population and Forensic Genetics, because it is important to highlight that a specific database must be used in these populations, and especially in Chuetas or Bragança Crypto-Jews, due to the genetic heritage of their singular history.

A GHEP-ISFG collaborative study on the genetic variation of 38 autosomal Indels for human identification in different continental populations

(Submitted)

1. Background

In 2012, the GHEP-ISFG group proposed a collaborative exercise to all the members of the Society. The exercise consisted of two phases. First, the laboratories that wanted to participate had to correctly genotype three control samples. Once the control samples were genotyped successfully, participants were given the opportunity to type their own populations of study. The condition to participate was that a number of samples of at least 100 individuals was needed.

At the end of the deadline proposed by the organizers, the study contained around 5839 samples of 54 population groups from 21 different countries in Africa, America, East Asia, Europe, and the Middle East.

Our laboratory participated by studying the Majorcan population. These results are presented here.

2. Material and Methods

A hundred and two unrelated individuals from Majorca, were typed using the PCR multiplex protocol originally described in Pereira et al. (2009). These samples belonged to the collection of the Genetics Laboratory, University of the Balearic Islands. Amplification products were separated by capillary electrophoresis in a 3130 Genetic Analyser (Applied Biosystems) and fragment sizes and allele calls were determined automatically using GeneMapper v3.2 ID software (Applied Biosystems).

Allele frequencies and the Hardy-Weinberg equilibrium analysis, were calculated using Arlequin v.3.5 software (Excoffier and Lischer, 2010). Gene Diversity was calculated as $1-\sum_{i}^{m} (p_i)^2$. Statistical parameters of forensic interest: MP (matching probability), PD (power of discrimination), PIC (polymorphisms information content), PE (power of exclusion), and TPI (typical paternity index) were computed using the Powerstats v1.2 spreadsheet (Tereba, 1999).

Majorcan results were compared with those from 18 European, Middle Eastern, and African populations (from the same collaborative exercise) by means of F_{ST} pairwise

distances (Arlequin v. 3.5 software) that were represented on a multidimensional scaling (MDS) plot using SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA).

3. Results and discussion

Allele frequencies and Gene diversity are shown in Table 1. No deviations from the Hardy-Weinberg equilibrium were observed in the population studied after Bonferroni's correction for multiple tests (p>0.00132), showing that no significant levels of substructure were found. Expected heterozygosities ranged from 0.2509 (rs1160956 and rs36040336) to 0.4996 (rs1610919 and rs1160886). Linkage disequilibrium test showed no significant gametic association between all 38 markers after Bonferroni's correction.

Forensic parameters of interest were calculated for each Indel (Table 2). rs2307700 (B04) showed the highest values of power of discrimination, while the lowest values in PD and power of exclusion in trios were observed in rs36040336 (R02). Combined PD and combined PE were 0.999999999999995 and 99.55464, respectively. These results are in accordance with other studies of Spanish populations (Saiz et al., 2014).

The MDS plot performed for comparison between populations studied for the same set of Indel markers can be seen in Figure 1. European and African populations cluster separately. The first group includes European and Middle Eastern populations and it is not possible to differentiate substructures, indicating that although this set of markers can separate efficiently between populations from different continents, as shown in Pereira et al. (2009), it does not appear to be a suitable tool for studying the genetic sub-structure in Europe. In the African cluster, we can distinguish three groups, Sub-Saharan Africa (Sudan and Cape Verde islands), equatorial Africa (Uganda) and Afro-tropical populations (Angola and South African Zulus).

Taking into account the genetic distances matrix (Table 3), and as can also be observed in the MDS, the African populations are more distant between each other (F_{ST} values ranging between 0.007-0.038) than are the European ones (F_{ST} values <0.008). These results are in accordance with other studies (González-Pérez et al., 2007; Soundararajan et al., 2016), which indicate slight differences in autosomal markers between Mediterranean and European populations. Even though Iraq clusters with the European group, the pairwise distances with these populations are greater than the distances between the European populations with each other. Additional populations should be studied to correctly position Middle Eastern populations.

Our population of study, within the European cluster, presented the greatest distances with Portuguese and Sicilian populations. Since Sicily is also a Mediterranean island, it was not expected to be one of the most distant from Majorca. At any rate from its particular history

and heterogeneity (Piazza et al., 2000) Sicily has always been controversially situated in genetic studies. Whilst some authors claim the differentiation of the Sicilian population from the Western Mediterranean basin (Calo et al., 2003; Robino et al., 2006), others do not (Scozzari et al., 2001; Francalacci et al., 2003; Ghiani et al., 2004).

4. Conclusions

This study provides a useful database of the 38 Indel multiplex (Pereira et al., 2009) for the Majorcan population for forensic purposes, particularly in situations where conventional STRs may result in low paternity indexes. However, this set of markers does not appear to be appropriate for molecular anthropology studies in subcontinental terms, namely in the European region.

Table 1. Short allele frequencies and Gene Diversity of the 38 autosomal Indels studied in the Majorcan population.

Loci code	rs number	short allele frequency	Gene diversity	Loci code	rs number	short allele frequency	Gene diversity
B01	rs34541393	0.3873	0.4746	Y01	rs3051300	0.4461	0.4942
B02	rs16624	0.7892	0.3327	Y02	rs10629077	0.2549	0.3799
B03	rs2307689	0.2892	0.4111	Y03	rs10688868	0.3039	0.4231
B04	rs35769550	0.4020	0.4808	Y04	rs2067208	0.2353	0.3599
B05	rs2307700	0.5539	0.4942	Y05	rs2307579	0.5196	0.4992
B06	rs140809	0.3431	0.4508	Y06	rs2308020	0.7451	0.3799
B07	rs3047269	0.4069	0.4826	Y07	rs3080855	0.2745	0.3983
B08	rs33972805	0.4118	0.4844	Y08	rs1610919	0.4853	0.4996
B09	rs33917182	0.6520	0.4538	Y09	rs2307839	0.2500	0.3750
B10	rs16402	0.2990	0.4192	R01	rs2308137	0.3088	0.4269
G01	rs1610871	0.6128	0.4746	R02	rs36040336	0.8529	0.2509
G02	rs2067238	0.6569	0.4508	R03	rs1160886	0.4853	0.4996
G03	rs2067294	0.2892	0.4111	R04	rs2308026	0.3284	0.4411
G04	rs2307710	0.3628	0.4623	R05	rs2307526	0.4118	0.4844
G05	rs2308242	0.2402	0.3650	R06	rs34811743	0.6275	0.4675
G06	rs2307580	0.4559	0.4961	R07	rs2308189	0.3824	0.4723
G07	rs1160956	0.8529	0.2509	R08	rs5895447	0.2745	0.3983
G08	rs34511541	0.4706	0.4983	R09	rs2308171	0.2255	0.3493
G09	rs2307978	0.1667	0.2778	R10	rs35605984	0.3873	0.4746

Table 2. Forensic parameters for the 38 autosomal Indels in the Majorcan population.

Loci code	rs number	MP	PD	PIC	PE	TPI
B01	rs34541393	0.3910	0.6090	0.3600	0.1710	0.9600
B02	rs16624	0.4960	0.5040	0.2800	0.0690	0.7300
B03	rs2307689	0.4270	0.5730	0.3300	0.1210	0.8500
B04	rs35769550	0.3580	0.6420	0.3700	0.1090	0.8200
B05	rs2307700	0.3430	0.6570	0.3700	0.1040	0.8100
B06	rs140809	0.3910	0.6090	0.3500	0.1280	0.8600
B07	rs3047269	0.4360	0.5640	0.3700	0.2550	1.1600
B08	rs33972805	0.3960	0.6040	0.3700	0.1960	1.0200
B09	rs33917182	0.3910	0.6090	0.3500	0.1340	0.8800
B10	rs16402	0.4080	0.5920	0.3400	0.0930	0.7800
G01	rs1610871	0.4000	0.6000	0.3600	0.1880	1.0000
G02	rs2067238	0.4050	0.5950	0.3500	0.1630	0.9400
G03	rs2067294	0.4340	0.5660	0.3300	0.1280	0.8600
G04	rs2307710	0.4030	0.5970	0.3600	0.1790	0.9800
G05	rs2308242	0.4790	0.5210	0.3000	0.1090	0.8200
G06	rs2307580	0.3650	0.6350	0.3700	0.1630	0.9400
G07	rs1160956	0.5710	0.4290	0.2300	0.0610	0.7100
G08	rs34511541	0.3820	0.6180	0.3700	0.1960	1.0200
G09	rs2307978	0.5490	0.4510	0.2400	0.0740	0.7400
Y01	rs3051300	0.3610	0.6390	0.3700	0.1480	0.9100
Y02	rs10629077	0.4560	0.5440	0.3100	0.0980	0.8000
Y03	rs10688868	0.4390	0.5610	0.3400	0.1710	0.9600
Y04	rs2067208	0.4760	0.5240	0.3000	0.0610	0.7100
Y05	rs2307579	0.3870	0.6130	0.3700	0.2050	1.0400
Y06	rs2308020	0.4520	0.5480	0.3100	0.1040	0.8100
Y07	rs3080855	0.4370	0.5630	0.3200	0.0980	0.8000
Y08	rs1610919	0.4040	0.5960	0.3700	0.2340	1.1100
Y09	rs2307839	0.4600	0.5400	0.3000	0.0650	0.7200
R01	rs2308137	0.4170	0.5830	0.3400	0.1410	0.8900
R02	rs36040336	0.5970	0.4030	0.2200	0.0400	0.6500
R03	rs1160886	0.3620	0.6380	0.3700	0.1630	0.9400
R04	rs2308026	0.3980	0.6020	0.3400	0.0980	0.8000
R05	rs2307526	0.3590	0.6410	0.3700	0.1280	0.8600
R06	rs34811743	0.3800	0.6200	0.3600	0.1280	0.8600
R07	rs2308189	0.3800	0.6200	0.3600	0.1340	0.8800
R08	rs5895447	0.4360	0.5640	0.3200	0.0880	0.7700
R09	rs2308171	0.4790	0.5210	0.2900	0.0740	0.7400
R10	rs35605984	0.3910	0.6090	0.3600	0.1710	0.9600

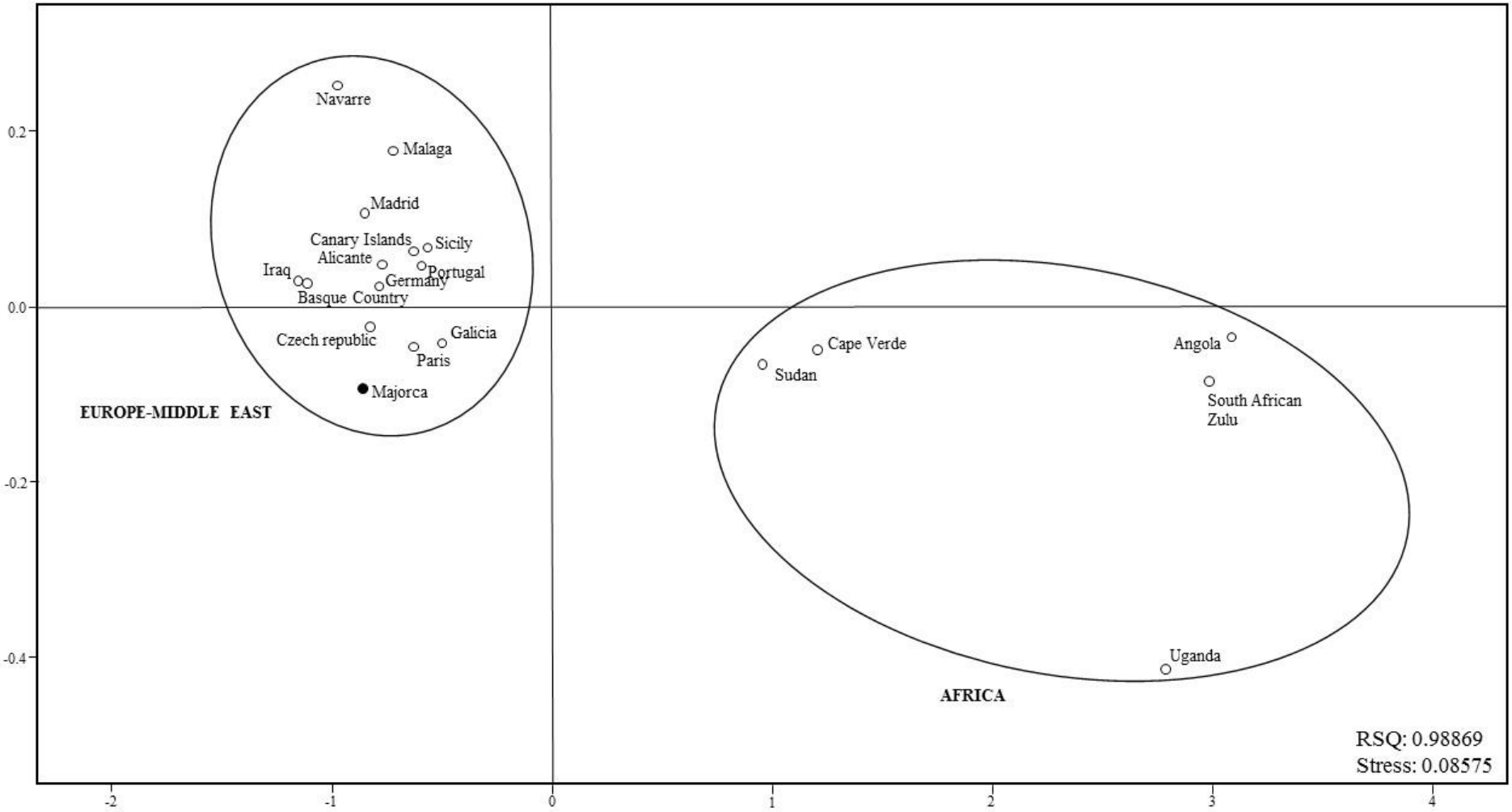


Figure 1. Multidimensional Scaling plot of the pairwise genetic distances between European and African populations. Majorcan population is labelled in black.

Table 3. Genetic distances matrix between European and African populations analysed with the 38 autosomal Indels. Darker colours correspond to greater genetic distances.

	Majorca	Madrid	Galicia	Basque Country	Alicante	Navarre	Malaga	Canary Islands	Portugal	Paris	Sicily	Germany	Czech Republic	Iraq	Sudan	Cape Verde	Uganda	Angola	South African Zulu
Majorca	0.00000																		
Madrid	0.00250	0.00000																	
Galicia	0.00150	0.00200	0.00000																
Basque Country	0.00450	0.00410	0.00660	0.00000															
Alicante	0.00130	0.00080	0.00100	0.00490	0.00000														
Navarre	0.00340	0.00420	0.00620	0.00390	0.00150	0.00000													
Malaga	0.00420	0.00140	0.00180	0.00620	0.00005	0.00370	0.00000												
Canary Islands	0.00150	0.00110	0.00120	0.00470	0.00070	0.00220	0.00190	0.00000											
Portugal	0.00800	0.00240	0.00140	0.00600	0.00000	0.00390	0.00320	0.00220	0.00000										
Paris	0.00190	0.00220	0.00040	0.00390	0.00020	0.00590	0.00360	0.00200	0.00000	0.00000									
Sicily	0.00590	0.00310	0.00460	0.00740	0.00280	0.00690	0.00550	0.00110	0.00370	0.00350	0.00000								
Germany	0.00000	0.00000	0.00060	0.00540	0.00000	0.00260	0.00290	0.00000	0.00000	0.00020	0.00310	0.00000							
Czech Republic	0.00170	0.00170	0.00310	0.00590	0.00020	0.00580	0.00350	0.00160	0.00090	0.00010	0.00380	0.00000	0.00000						
Iraq	0.01200	0.00700	0.00970	0.01060	0.00090	0.01330	0.01000	0.00470	0.00960	0.00660	0.00200	0.00820	0.00780	0.00000					
Sudan	0.04530	0.04680	0.04030	0.05270	0.04280	0.05510	0.04560	0.04020	0.04130	0.03930	0.03880	0.04460	0.04510	0.04920	0.00000				
Cape Verde	0.05580	0.05620	0.04820	0.06420	0.05400	0.05940	0.05170	0.04830	0.04930	0.05010	0.04980	0.05400	0.05630	0.06350	0.01190	0.00000			
Uganda	0.10580	0.10940	0.09500	0.11550	0.10620	0.11590	0.10670	0.10030	0.10020	0.09940	0.10250	0.10580	0.10650	0.11750	0.02410	0.02890	0.00000		
Angola	0.11750	0.11550	0.10590	0.12500	0.11390	0.11810	0.11050	0.10720	0.10740	0.11000	0.11010	0.11370	0.11590	0.12710	0.03760	0.01790	0.02220	0.00000	
South African Zulu	0.11470	0.11330	0.10270	0.12080	0.11090	0.11590	0.10990	0.10370	0.10580	0.10690	0.10550	0.11190	0.11280	0.12220	0.03510	0.02100	0.01850	0.00690	0.00000

Genetic diversity of 38 insertion-deletion polymorphisms in Jewish populations

Ferragut JF[1]; Pereira R[2,3]; Castro JA[1]; Ramon C[1]; Nogueiro I[2,3]; Amorim A[2,3,4] and Picornell A[1*]

[1] Institut Universitari d'Investigació en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Carretera de Valldemossa, km 7.5, 07122, Palma de Mallorca, Illes Balears, Spain

[2] i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[3] IPATIMUP - Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal

[4] Faculdade de Ciências da Universidade do Porto, Porto, Portugal

Key words: Indels, Jews, Ashkenazim, Sephardim, Mizrahim, Chuetas, Bragança

*Corresponding author: Dr. Antònia Picornell. Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Cra. Valldemossa, km 7.5, 07122-Palma de Mallorca, Illes Balears, Spain. Tel: +34-971-172050. Fax: +34-971-173184. E-mail: apicornell@uib.es

Abstract

Population genetic data of 38 non-coding bi-allelic autosomal Indels are reported for 466 individuals, representing six populations with Jewish ancestry (Ashkenazim, Mizrahim, Sephardim, North African, Chuetas and Bragança crypto-Jews). Intra-population diversity and forensic parameters values showed that this set of Indels was highly informative for forensic applications in the Jewish populations studied. Genetic distance analysis demonstrated that this set of markers efficiently separates populations from different continents, but does not seem effective for molecular anthropology studies in Mediterranean region. Finally, it is important to highlight that although the genetic distances between Jewish populations were small, significant differences were observed for Chuetas and Bragança Jews, and therefore, specific databases must be used for these populations.

1. Introduction

Indels are length polymorphisms created by small insertions or deletions of one or more nucleotides in the genome. These markers have a series of special characteristics that justify its increasing interest as a tool for a wide range of purposes, including population genetic studies, ancestry affiliation, and forensics. Indels combine many of the desirable characteristics of both SNPs and STRs, such as a widely spread distribution throughout the genome (Weber et al., 2002; 1000 Genomes Project Consortium, 2010; Mills et al., 2011); origin from a single rare mutation event and so unlikely to present recurrent mutations (Nachman and Crowell, 2000); amenable to PCR design of short amplicons, improving the chances of amplification success of degraded samples (Pereira et al., 2009; Romanini et al., 2012); significant differences in allele frequencies between populations, providing good results in population differentiation studies, and the ease of analysis by multiplex PCR and capillary electrophoresis (Weber et al., 2002; Yang et al., 2005; Pereira et al., 2009). In 2009, Pereira et al. described a new multiplex for human identification combining 38 small non-coding biallelic autosomal Indels into a single multiplex, with proven success for genotyping of degraded samples.

Jews can be traced back to populations occupying a small geographic area, in the Middle East, several thousand years ago and have maintained continuous cultural and religious traditions despite a series of Diasporas. Contemporary Jews comprise several communities that can be classified according to the location where each community developed. Among others, these include Middle Eastern Jews ("Mizrahim") (Iran and Iraq), who have always resided in the Near East; the Askenazim who lived in communities of central and eastern Europe; the Sephardim (from Sepharad, Hebrew word for Hispania) who, after their expulsion from the Iberian Peninsula in the late 15th century, lived in other Mediterranean

countries (especially Bulgaria and Turkey); and the North African Jews, comprising both Sephardim and Mizrahim (Baron, 1937; Ben-Sasson et al., 1976; Goodman, 1979).

Chuetas (Majorca, Spain) and the Crypto-Jewish communities in Portugal (especially in Belmonte and Bragança district) are the only current Iberian populations whose ancestors can be traced to the original Sephardic Jewish populations, because of their peculiar history that kept the memory of their Jewish origin through centuries, and their inbreeding, which has prevented their gradual assimilation into the general population, as it happened with most of converted Iberian Jews (Laub and Laub, 1987; Martins, 2006).

Geneticists have studied Jewish populations since the turn of the 20th century in an attempt to unravel what must be a complex system of interrelationships among Jewish communities and their non-Jewish neighbours. These studies have provided evidence for shared Middle Eastern ancestry among major Jewish Diaspora groups and variable degrees of admixture with local populations (Patai and Patai, 1975; Bonné-Tamir and Adam, 1992; Hammer et al., 2000; Thomas et al., 2002; Atzmon et al., 2010; Behar et al., 2010; Ostrer and Skorecki, 2013). Regarding the Jewish descendants in Iberia, genetic studies have shown that both Chuetas and Bragança Jews present a significant persistence of a Jewish heritage as well as signs of introgression from their host non-Jewish populations (Picornell et al., 1997; Crespí et al., 2002; Nogueiro et al., 2015b).

In the present study we aimed to characterize the diversity of the 38 Indel markers (Pereira et al., 2009) in populations with Jewish ancestry, and compare their distribution in Jewish and non-Jewish populations. Finally, we evaluate the usefulness of the Indel set both in forensic casework and population genetics, especially in Chuetas and in Bragança crypto-Jewish populations, due to the genetic heritage of their singular history.

2. Materials and methods

2.1. DNA Samples

DNA samples from 466 unrelated individuals with known Jewish ancestry were obtained after informed consent: 136 Chueta individuals (Majorca, Spain) belonged to the collection of the Genetics Laboratory, University of the Balearic Islands; 55 Jews from Bragança (NE Portugal) collected by the Institute of Pathology and Molecular Immunology, University of Porto; and 275 individuals of the National Laboratory for the genetics of Israeli populations at Tel-Aviv University. Following classical criteria, these samples were categorized into four groups: Ashkenazi (55), Middle Eastern (28 Iranian and 26 Iraqi), North African (34 Moroccan, 13 Libyan and 13 Tunisian), and Sephardic (62 Turkish and 44 Bulgarian).

2.2. Indel genotyping

The 38 Indels were genotyped using the PCR multiplex protocol originally described in (Pereira et al., 2009). Amplification products were separated by capillary electrophoresis in a 3130 Genetic Analyzer (Applied Biosystems) and fragment sizes and allele calls were determined automatically using GeneMapper v3.2 ID software (Applied Biosystems). Typing quality and allele designation were warranted by the analysis of control samples of known genotype (GHEP-ISFG collaborative exercise; details available at *www.gep-isfg.org/en/working-commissions/collaborative-exercise-indels-2012.html*).

2.3. Data analysis

Allele frequencies, Hardy-Weinberg equilibrium analysis, AMOVA and populations pairwise genetic distances (F_{ST}) were calculated using the Arlequin v.3.5 software (Excoffier and Lischer, 2010). Gene Diversity was calculated as $1-\sum_{i}^{m} (p_i)^2$. Statistical parameters of forensic interest: MP (matching probability), PD (power of discrimination), PIC (polymorphisms information content), PE (power of exclusion) and TPI (typical paternity index) were computed using Powerstats v1.2 spreadsheet (Tereba, 1999).

To examine the relationship of the populations under study and with other published population data (Pereira et al., 2009; Manta et al., 2012; Saiz et al., 2014), F_{ST} genetic distances were represented in a multidimensional scaling (MDS) plot using SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA). Population structure was further assessed with STRUCTURE 2.3.4 software using default settings (admixture model; correlated allele frequencies) (Pritchard et al., 2000; Falush et al., 2003). All runs included a burn-in period of 50,000 iterations followed by 100,000 MCMC repetitions (K = 1-9), and were repeated ten times each in order to test the consistency of the results.

3. Results and discussion

3.1. Intra-population variability

A total of 932 chromosomes were analyzed in this study. Genotypic data and allele frequencies for each Indel marker and population (Sephardic, North African, Middle Eastern and Ashkenazi Jews, Chuetas and Jews from Bragança) are shown in Supplementary Tables S1 (doi:10.1016/j.fsigen.2015.11.003) and S2. No deviations from Hardy-Weinberg equilibrium were observed in the populations studied after Bonferroni's correction for multiple tests ($p > 0.00132$), showing that no significant levels of substructure were found.

High genetic diversities (GD) were observed in all populations analysed (Supplementary Table S3), with mean values ranging between 0.4148–0.4336 (Table 1). The highest values were found for rs2307579 (Y05) and rs34511541 (G08) and expected heterozygosities ranged between 0.4861–0.4989 and between 0.4650–0.5000, respectively, while the lowest heterozygosities were observed for rs2307839 (Y09) in Ashkenazi (0.1503) and rs2308171 (R09) in North African Jews (0.1528). These GD values were in the range found in other studies (Pereira et al., 2009; Manta et al., 2012; Saiz et al., 2014), although the markers showing the highest and the lowest variability differed between populations.

Table 1. Diversity and Forensic Parameters for the set of 38 Indels studied in the 6 studied populations with Jewish ancestry.

	Obs Het	Exp Het	Combined MP	Combined PD	Combined PE
Sephardic Jews	0.4322 ± 0.0768	0.4244 ± 0.0738	1.9581E+14	0.9999999999999950	99.72%
North African Jews	0.4204 ± 0.0995	0.4190 ± 0.0831	1.5530E+14	0.9999999999999940	99.68%
Middle Eastern Jews	0.4096 ± 0.0871	0.4148 ± 0.0743	1.2510E+14	0.9999999999999920	99.50%
Ashkenazi Jews	0.4412 ± 0.1050	0.4187 ± 0.0799	6.3431E+13	0.9999999999999840	99.85%
Chuetas	0.4338 ± 0.0641	0.4336 ± 0.0555	3.3360E+14	0.9999999999999970	99.70%
Bragança Jews	0.4387 ± 0.0844	0.4335 ± 0.0574	2.0471E+14	0.9999999999999950	99.79%

Obs Het: Observed Heterozygosity; Exp Het: Expected Heterozygosity; MP: Matching probability; PD: Power of discrimination; PE: Power of Exclusion

Linkage disequilibrium (LD) test showed no significant gametic association between all 38 markers after Bonferroni correction, with the exception of the pair rs16624-rs2067238 in Ashkenazi population ($p \leq 10^{-5}$), in accordance with the reported founder effect and genetic drift of this population, indicating an early bottleneck, probably corresponding to initial migrations of ancestral Ashkenazim to Europe (Behar et al., 2004b). Still, it must be said that many generations have elapsed after the founder bottleneck and therefore this LD value may be a spurious result, and since the two Indels are located at different chromosomes, this association is rapidly vanished, provided random mating is established.

3.2. Forensic efficiency

Forensic parameters of interest were calculated for each Indel marker and population (Supplementary Table S4). As expected from the levels of diversity, rs34511541 (G08) showed the highest values of power of discrimination (PD) in different populations, while the lowest values in PD and power of exclusion in trios (PE) were observed in rs2308171 (R09), except in Ashkenazi and Bragança Jews.

The combined matching probability (MP) (Table 1) ranged from a value of 1 in $6.34x10^{+13}$ (in Ashkenazim) to 1 in $3.34x10^{+14}$ (in Chuetas). Considering the essentially biallelic nature of small Indels, a high combined PE was also obtained in all the populations (>0.995 in all cases). Although values differed slightly between populations, the set of loci was highly

informative in all the studied populations with Jewish ancestry, providing a suitable short amplicon marker set to complement STRs in forensic casework.

3.3. Application of this set of Indels to a challenging paternity investigation

A paternity case involving mother, an alleged father and one girl was investigated. In the initial conventional STR analysis, mismatches were observed between the alleged father and the child in two STRs (D2S1338 and D5S818) resulting in a paternity index (PI) of 0.827 (Table 2), if the possibility of mutation was considered. The alleged father matched perfectly when 12 additional X-STRs were typed, in order to investigate the possibility that he was not the father but a close relative (Pinto et al., 2013), and the same happened when subsequently typed for the set of 38 autosomal Indels (Supplementary Table S5). Joining the information from STRs and Indels resulted in a combined PI value of 10360 substantially contributing to solve the case.

Table 2. Paternity indices (PI) and probabilities (W) based on the results obtained with STRs, Indels, and with both type of markers in a paternity case with two paternal mutations.

	STRs *AmpF/STR Identifiler* (Applied Biosystems) *Powerplex® System* (Promega)	**INDELS** (38 Indel-plex)	**STRs + INDELs**
PI	0.827	12530.759	10360.431
W	45.259%	99.992%	99.990%

3.4. Inter-population variability

AMOVA analysis of the six Jewish populations studied showed a low but statistically significant value (F_{ST} = 0.00871; $p < 10^{-5}$). Interestingly, pairwise comparisons (Supplementary Table S6) evidenced the existence of a sub-set composed of the Jewish populations (Sephardic, North Africa, Mizrahim and Ashkenazi) with non-significant genetic distances, whereas the converted Jewish populations (Bragança Jews and Chuetas) presented significant differences with all other populations tested in this study. However, hierarchical AMOVA showed lack of homogeneity within these sub-sets (F_{SC} = 0.00250; p = 0.01075).

A comparison was also performed between the populations in this work and other populations previously studied for the same set of Indel markers (Portugal, Angola, Mozambique, Macau and Taiwan (Pereira et al., 2009), South East Spain (Saiz et al., 2014), Brazil (Rio de Janeiro and Terenas) (Manta et al., 2012), and Majorca (Chueta's host population) (GHEP-ISFG collaborative exercise *www.gep-isfg.org/en/working-*

commissions/collaborative-exercise-indels-2012.html). MDS results (stress value <0.199) showed greater structure than would be obtained from a random dataset (Sturrock and Rocha, 2000). The MDS plot (Figure 1) indicated that European, African and Asian populations are distantly positioned. Terenas (with major Amerindian ancestry living in Mato Grosso do Sul, Center-West Brazil) was closer to the Asians, while Rio de Janeiro was placed between Europe and Africa, as expected due to their origin. Jewish and South European populations (Portugal, Majorca and South East Spain) clustered together, indicating that this set of markers can separate efficiently between populations from different continents, but does not appear to be a suitable tool for studying the genetic sub-structure in Mediterranean context.

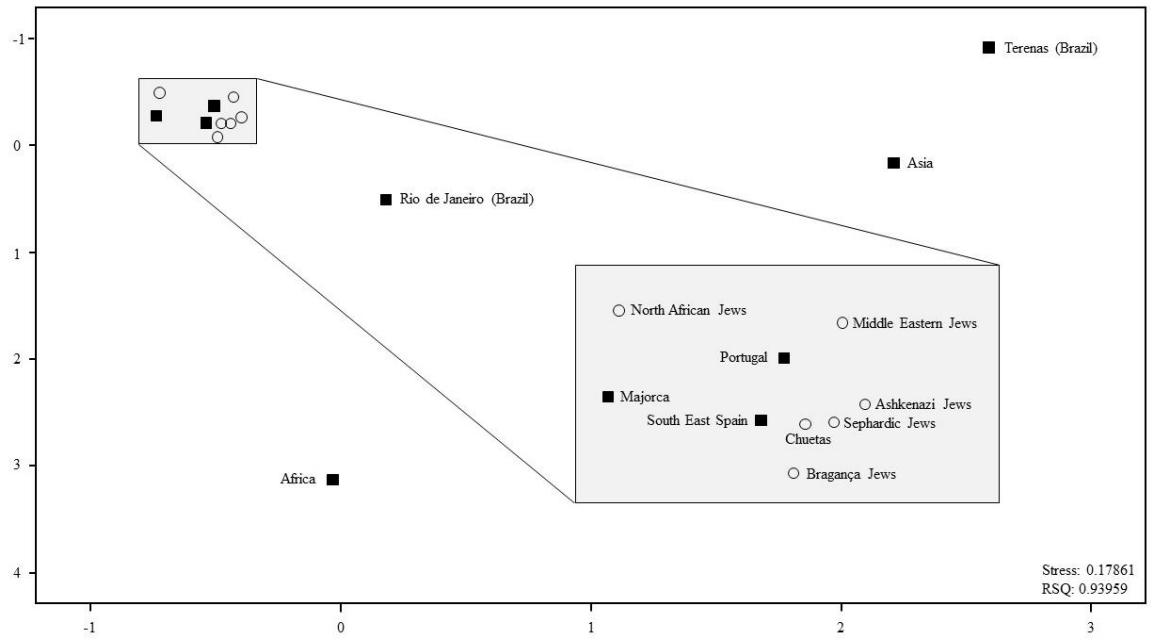


Figure 1. Multidimensional Scaling (MDS) plot of the pairwise genetic distances between the 6 populations with Jewish ancestry (showed as circles) in this study and South European populations (Portugal, South East Spain and Majorca), Brazilian populations (Rio Janeiro and Terenas), Asian and African populations. All non-Jewish populations are shown as black squares.

Within the Mediterranean region, taking into account the genetic distances matrix (Supplementary Table S6), we can observe two groups without significant differences within them; one is the aforementioned Jewish group, and the other includes the South European non-Jewish populations (Portugal, Majorca and South East Spain). North African Jews showed no significant distances with any population of any group. Regarding the converted Jewish populations, Chuetas presented significant differences with all other populations, including both the Jews and their host population, Majorca; but Braganza Jews were not significantly different to Portugal, their host population.

A Bayesian clustering analysis was performed with the same samples in STRUCTURE software in order to deeper scrutinize the substructure of the Mediterranean region for this set of 38 neutral, slow evolving, autosomal markers. The results consistently revealed

worse ln P(D) values for K>1 (Supplementary Figure 1), failing to detect distinct clusters in the data set and supporting that any possible genetic structure in the populations here analyzed could not be discerned with the level of resolution afforded by these 38 loci.

4. Conclusions

This study provides a useful database of the 38 Indel multiplex (Pereira et al., 2009) for populations with Jewish ancestry including two Sephardic isolates, the Chuetas and the Crypto-Jews from Bragança. The results obtained demonstrated also their usefulness for general forensic purposes, especially in situations where conventional STRs may result in low paternity indices. However, this set of markers does not seem adequate for molecular anthropology studies in subcontinental level, namely in the Mediterranean region.

It is important to highlight that although the genetic distances between Jewish populations were low, significant differences were observed for both Chuetas and Bragança Jews with the others. Therefore, specific databases for these Jewish populations must be used in the forensic field to correctly weigh the value of the evidence based on these Indel markers.

This study follows the guidelines for publication of population data proposed by the journal (Carracedo et al., 2013).

Acknowledgments

This work was partially supported by grant AAEE24/2014 from the Direcció General de R + D + I (Comunitat Autònoma de les Illes Balears) and European Regional Development Fund (ERDF). RP (SFRH/BPD/81986/2011) and IN (SFRH/BD/73336/2010) are recipients of grants awarded by the Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Social Fund (Human Potential Thematic Operational Programme –POPH).

Supplementary Table 2. Short Allele frequencies of the 38 Indels in Sephardic Jews, North African Jews, Middle Eastern Jews, Ashkenazi Jews, Chuetas and Bragança Jews populations.

Loci code	rs number	Sephardic Jews	North African Jews	Middle Eastern Jews	Ashkenazi Jews	Chuetas	Bragança Jews
B01	rs34541393	0.3585	0.4167	0.2778	0.3182	0.4191	0.3091
B02	rs16624	0.7736	0.8500	0.7407	0.8000	0.6949	0.7636
B03	rs2307689	0.3679	0.3667	0.3704	0.3546	0.4228	0.3364
B04	rs35769550	0.3679	0.4000	0.3611	0.4909	0.3971	0.4546
B05	rs2307700	0.6038	0.5583	0.5648	0.5636	0.5772	0.5463
B06	rs140809	0.2311	0.2667	0.2037	0.3455	0.2022	0.2182
B07	rs3047269	0.4717	0.4750	0.4352	0.4364	0.5662	0.4636
B08	rs33972805	0.4104	0.5167	0.3796	0.4273	0.5184	0.4727
B09	rs33917182	0.5896	0.5833	0.6019	0.5818	0.5110	0.6364
B10	rs16402	0.2453	0.3583	0.2407	0.2000	0.2353	0.3818
G01	rs1610871	0.4670	0.5583	0.5093	0.4818	0.5699	0.6364
G02	rs2067238	0.5381	0.6000	0.4434	0.5727	0.6912	0.6455
G03	rs2067294	0.3443	0.3167	0.3056	0.3455	0.3235	0.4000
G04	rs2307710	0.4717	0.4750	0.3611	0.3727	0.2831	0.3182
G05	rs2308242	0.2028	0.2250	0.2870	0.2000	0.2427	0.2364
G06	rs2307580	0.5425	0.5417	0.6574	0.5364	0.4228	0.4636
G07	rs1160956	0.8208	0.8000	0.7685	0.7727	0.7978	0.7818
G08	rs34511541	0.5000	0.4833	0.5278	0.3909	0.3677	0.5278
G09	rs2307978	0.2783	0.2833	0.3704	0.3182	0.2684	0.2364
Y01	rs3051300	0.4670	0.4310	0.4815	0.4091	0.5000	0.3727
Y02	rs10629077	0.2048	0.3083	0.2222	0.2091	0.2390	0.2818
Y03	rs10688868	0.3632	0.3750	0.3426	0.4546	0.3713	0.2500
Y04	rs2067208	0.1887	0.1917	0.1852	0.2182	0.3713	0.3727
Y05	rs2307579	0.4764	0.4750	0.4167	0.4455	0.4743	0.4636
Y06	rs2308020	0.6840	0.7881	0.7130	0.6546	0.6765	0.6909
Y07	rs3080855	0.2972	0.2833	0.2037	0.2636	0.4044	0.3636
Y08	rs1610919	0.5000	0.5750	0.6019	0.5000	0.5846	0.6182
Y09	rs2307839	0.1651	0.1583	0.2130	0.0818	0.2463	0.1909
R01	rs2308137	0.3396	0.3250	0.3519	0.4091	0.2831	0.4630
R02	rs36040336	0.8066	0.8000	0.7963	0.8182	0.7279	0.8091
R03	rs1160886	0.3821	0.4167	0.4167	0.4000	0.4118	0.3364
R04	rs2308026	0.3915	0.3583	0.3889	0.3364	0.3088	0.2546
R05	rs2307526	0.3491	0.3500	0.4352	0.4818	0.2868	0.3727
R06	rs34811743	0.6793	0.6667	0.8241	0.7091	0.6838	0.6364
R07	rs2308189	0.4481	0.4417	0.3519	0.4546	0.2831	0.4000
R08	rs5895447	0.3821	0.4250	0.3241	0.2182	0.3897	0.3182
R09	rs2308171	0.1557	0.0833	0.1019	0.2182	0.1875	0.2546
R10	rs35605984	0.4151	0.2917	0.3241	0.3000	0.3824	0.4727

Supplementary Table 3. Gene diversity values of the 38 Indels studied in Sephardic Jews, North African Jews, Middle Eastern Jews, Ashkenazi Jews, Chuetas and Bragança Jews populations.

Loci code	rs number	Sephardic Jews	North African Jews	Middle Eastern Jews	Ashkenazi Jews	Chuetas	Bragança Jews	Mean ± standard deviation
B01	rs34541393	0.4599	0.4861	0.4012	0.4339	0.4869	0.4271	0.4481 ± 0.0314
B02	rs16624	0.3503	0.2550	0.3841	0.3200	0.4241	0.3610	0.3448 ± 0.0527
B03	rs2307689	0.4651	0.4644	0.4664	0.4577	0.4881	0.4464	0.4645 ± 0.0125
B04	rs35769550	0.4651	0.4800	0.4614	0.4998	0.4788	0.4959	0.4800 ± 0.0142
B05	rs2307700	0.4785	0.4932	0.4916	0.4919	0.4881	0.4957	0.4898 ± 0.0056
B06	rs140809	0.3554	0.3911	0.3244	0.4522	0.3226	0.3412	0.3619 ± 0.0454
B07	rs3047269	0.4984	0.4988	0.4916	0.4919	0.4912	0.4974	0.4949 ± 0.0033
B08	rs33972805	0.4839	0.4994	0.4710	0.4894	0.4993	0.4985	0.4902 ± 0.0104
B09	rs33917182	0.4839	0.4861	0.4793	0.4866	0.4998	0.4628	0.4830 ± 0.0110
B10	rs16402	0.3702	0.4599	0.3656	0.3200	0.3599	0.4721	0.3875 ± 0.0554
G01	rs1610871	0.4978	0.4932	0.4998	0.4993	0.4902	0.4628	0.4904 ± 0.0129
G02	rs2067238	0.4971	0.4800	0.4936	0.4894	0.4269	0.4577	0.4734 ± 0.0248
G03	rs2067294	0.4515	0.4328	0.4244	0.4522	0.4377	0.4800	0.4461 ± 0.0180
G04	rs2307710	0.4984	0.4988	0.4614	0.4676	0.4059	0.4339	0.4598 ± 0.0333
G05	rs2308242	0.3234	0.3488	0.4093	0.3200	0.3675	0.3610	0.3538 ± 0.0300
G06	rs2307580	0.4964	0.4965	0.4504	0.4974	0.4881	0.4974	0.4874 ± 0.0170
G07	rs1160956	0.2942	0.3200	0.3558	0.3512	0.3226	0.3412	0.3302 ± 0.0211
G08	rs34511541	0.5000	0.4994	0.4985	0.4762	0.4650	0.4985	0.4894 ± 0.0138
G09	rs2307978	0.4017	0.4061	0.4664	0.4339	0.3927	0.3610	0.4090 ± 0.0330
Y01	rs3051300	0.4978	0.4905	0.4993	0.4835	0.5000	0.4676	0.4896 ± 0.0115
Y02	rs10629077	0.3257	0.4265	0.3457	0.3307	0.3637	0.4048	0.3643 ± 0.0375
Y03	rs10688868	0.4626	0.4688	0.4504	0.4959	0.4669	0.3750	0.4516 ± 0.0376
Y04	rs2067208	0.3062	0.3099	0.3018	0.3412	0.4669	0.4676	0.3588 ± 0.0730
Y05	rs2307579	0.4989	0.4988	0.4861	0.4940	0.4987	0.4974	0.4956 ± 0.0046
Y06	rs2308020	0.4323	0.3340	0.4093	0.4522	0.4377	0.4271	0.4135 ± 0.0386
Y07	rs3080855	0.4177	0.4061	0.3244	0.3883	0.4817	0.4628	0.4102 ± 0.0512
Y08	rs1610919	0.5000	0.4888	0.4793	0.5000	0.4857	0.4721	0.4875 ± 0.0102
Y09	rs2307839	0.2757	0.2665	0.3352	0.1503	0.3713	0.3089	0.2741 ± 0.0697
R01	rs2308137	0.4486	0.4388	0.4561	0.4835	0.4059	0.4973	0.4540 ± 0.0298
R02	rs36040336	0.3120	0.3200	0.3244	0.2975	0.3961	0.3089	0.3250 ± 0.0323
R03	rs1160886	0.4722	0.4861	0.4861	0.4800	0.4844	0.4464	0.4757 ± 0.0140
R04	rs2308026	0.4765	0.4599	0.4753	0.4464	0.4269	0.3795	0.4427 ± 0.0335
R05	rs2307526	0.4544	0.4550	0.4916	0.4993	0.4091	0.4676	0.4619 ± 0.0295
R06	rs34811743	0.4357	0.4444	0.2900	0.4126	0.4324	0.4628	0.4083 ± 0.0570
R07	rs2308189	0.4946	0.4932	0.4561	0.4959	0.4059	0.4800	0.4698 ± 0.0322
R08	rs5895447	0.4722	0.4888	0.4381	0.3412	0.4757	0.4339	0.4386 ± 0.0491
R09	rs2308171	0.2629	0.1528	0.1830	0.3412	0.3047	0.3795	0.2573 ± 0.0813
R10	rs35605984	0.4856	0.4132	0.4381	0.4200	0.4723	0.4985	0.4534 ± 0.0326

Supplementary Table 4. Forensic Parameters of the 38 Indels in the 6 Jewish populations studied.

	Loci	Sephardic Jews					North African Jews					Middle Eastern Jews					Ashkenazi Jews					Chuetas					Bragança Jews				
		MP	PD	PIC	PE	TPI	MP	PD	PIC	PE	TPI	MP	PD	PIC	PE	TPI	MP	PD	PIC	PE	TPI	MP	PD	PIC	PE	TPI	MP	PD	PIC	PE	TPI
B01	**rs34541393**	0.4199	0.5801	0.3542	0.1959	1.0192	0.3539	0.6461	0.3680	0.1139	0.8333	0.4321	0.5679	0.3207	0.0782	0.7500	0.4215	0.5785	0.3398	0.1507	0.9167	0.3958	0.6042	0.3684	0.2007	1.0303	0.4063	0.5937	0.3359	0.0754	0.7432
B02	**rs16624**	0.4931	0.5069	0.2889	0.1234	0.8548	0.5850	0.4150	0.2225	0.0507	0.6818	0.4492	0.5508	0.3103	0.0782	0.7500	0.5134	0.4866	0.2688	0.0754	0.7432	0.4244	0.5756	0.3341	0.1359	0.8831	0.4883	0.5117	0.2958	0.1376	0.8871
B03	**rs2307689**	0.4053	0.5947	0.3569	0.1794	0.9815	0.4106	0.5894	0.3566	0.1875	1.0000	0.4184	0.5816	0.3576	0.2042	1.0385	0.4321	0.5679	0.3529	0.2125	1.0577	0.3907	0.6093	0.3690	0.1940	1.0149	0.3977	0.6023	0.3468	0.1254	0.8594
B04	**rs35769550**	0.4147	0.5853	0.3569	0.1959	1.0192	0.3800	0.6200	0.3648	0.1600	0.9375	0.4342	0.5658	0.3550	0.2220	1.0800	0.3626	0.6374	0.3749	0.1647	0.9483	0.3771	0.6229	0.3642	0.1517	0.9189	0.3666	0.6334	0.3729	0.1647	0.9483
B05	**rs2307700**	0.4361	0.5639	0.3640	0.2521	1.1522	0.4106	0.5894	0.3716	0.2351	1.1111	0.3669	0.6331	0.3708	0.1571	0.9310	0.4089	0.5911	0.3709	0.2305	1.1000	0.3706	0.6294	0.3690	0.1573	0.9315	0.4534	0.5466	0.3728	0.3044	1.2857
B06	**rs140809**	0.4783	0.5217	0.2923	0.0858	0.7681	0.4489	0.5511	0.3146	0.0507	0.6818	0.5089	0.4911	0.2718	0.0782	0.7500	0.3970	0.6030	0.3500	0.1376	0.8871	0.5109	0.4891	0.2706	0.0841	0.7640	0.5015	0.4985	0.2830	0.0465	0.6707
B07	**rs3047269**	0.3564	0.6436	0.3742	0.1495	0.9138	0.3850	0.6150	0.3744	0.2025	1.0345	0.3669	0.6331	0.3708	0.1571	0.9310	0.3574	0.6426	0.3709	0.1376	0.8871	0.3646	0.6354	0.3706	0.1517	0.9189	0.3580	0.6420	0.3737	0.1507	0.9167
B08	**rs33972805**	0.3911	0.6089	0.3668	0.1875	1.0000	0.3356	0.6644	0.3747	0.0949	0.7895	0.3875	0.6125	0.3601	0.1571	0.9310	0.4506	0.5494	0.3697	0.2909	1.2500	0.3721	0.6279	0.3747	0.1811	0.9855	0.4023	0.5977	0.3743	0.2305	1.1000
B09	**rs33917182**	0.3822	0.6178	0.3668	0.1715	0.9636	0.3889	0.6111	0.3680	0.1875	1.0000	0.4163	0.5837	0.3644	0.2220	1.0800	0.3481	0.6519	0.3682	0.0933	0.7857	0.3589	0.6411	0.3749	0.1573	0.9315	0.4169	0.5831	0.3557	0.1956	1.0185
B10	**rs16402**	0.4640	0.5360	0.3017	0.0906	0.7794	0.3739	0.6261	0.3541	0.0863	0.7692	0.4760	0.5240	0.2987	0.1185	0.8438	0.5134	0.4866	0.2688	0.0754	0.7432	0.4735	0.5265	0.2951	0.0805	0.7556	0.3904	0.6096	0.3606	0.1647	0.9483
G01	**rs1610871**	0.3434	0.6566	0.3739	0.1174	0.8413	0.3606	0.6394	0.3716	0.1474	0.9091	0.3957	0.6043	0.3749	0.2220	1.0800	0.4136	0.5864	0.3747	0.2495	1.1458	0.3745	0.6255	0.3701	0.1689	0.9577	0.3997	0.6003	0.3557	0.1647	0.9483
G02	**rs2067238**	0.3558	0.6442	0.3735	0.1456	0.9052	0.4350	0.5650	0.3648	0.2528	1.1538	0.3457	0.6543	0.3718	0.1116	0.8281	0.4003	0.5997	0.3697	0.2125	1.0577	0.4347	0.5653	0.3358	0.1630	0.9444	0.4129	0.5871	0.3529	0.1797	0.9821
G03	**rs2067294**	0.4000	0.6000	0.3496	0.1426	0.8983	0.4022	0.5978	0.3391	0.0949	0.7895	0.4218	0.5782	0.3343	0.1305	0.8710	0.3970	0.6030	0.3500	0.1376	0.8871	0.4181	0.5819	0.3419	0.1517	0.9189	0.3825	0.6175	0.3648	0.1647	0.9483
G04	**rs2307710**	0.3637	0.6363	0.3742	0.1639	0.9464	0.4050	0.5950	0.3744	0.2351	1.1111	0.4342	0.5658	0.3550	0.2220	1.0800	0.4221	0.5779	0.3583	0.2125	1.0577	0.4426	0.5574	0.3235	0.1359	0.8831	0.4030	0.5970	0.3398	0.1032	0.8088
G05	**rs2308242**	0.5117	0.4883	0.2711	0.0956	0.7910	0.4850	0.5150	0.2879	0.0863	0.7692	0.4369	0.5631	0.3255	0.1305	0.8710	0.5147	0.4853	0.2688	0.0933	0.7857	0.4664	0.5336	0.3000	0.0878	0.7727	0.4724	0.5276	0.2958	0.0754	0.7432
G06	**rs2307580**	0.3494	0.6506	0.3732	0.1295	0.8689	0.3572	0.6428	0.3733	0.1474	0.9091	0.3957	0.6043	0.3490	0.1305	0.8710	0.4155	0.5845	0.3737	0.2495	1.1458	0.3604	0.6396	0.3690	0.1359	0.8831	0.3580	0.6420	0.3737	0.1507	0.9167
G07	**rs1160956**	0.5429	0.4571	0.2509	0.0568	0.6974	0.5200	0.4800	0.2688	0.0507	0.6818	0.4781	0.5219	0.2925	0.0698	0.7297	0.4830	0.5170	0.2896	0.0673	0.7237	0.5111	0.4889	0.2706	0.0704	0.7312	0.4922	0.5078	0.2830	0.0754	0.7432
G08	**rs34511541**	0.3434	0.6566	0.3750	0.1234	0.8548	0.3356	0.6644	0.3747	0.0949	0.7895	0.3971	0.6029	0.3742	0.2220	1.0800	0.4136	0.5864	0.3628	0.2125	1.0577	0.3858	0.6142	0.3569	0.1410	0.8947	0.3395	0.6605	0.3742	0.1073	0.8182
G09	**rs2307978**	0.4334	0.5666	0.3210	0.0956	0.7910	0.4339	0.5661	0.3236	0.1139	0.8333	0.3855	0.6145	0.3576	0.1433	0.9000	0.4215	0.5785	0.3398	0.1507	0.9167	0.4454	0.5546	0.3156	0.1078	0.8193	0.4724	0.5276	0.2958	0.0754	0.7432
Y01	**rs3051300**	0.3772	0.6228	0.3739	0.1875	1.0000	0.3936	0.6064	0.3702	0.2030	1.0357	0.3422	0.6578	0.3747	0.1185	0.8438	0.3607	0.6393	0.3666	0.1254	0.8594	0.3616	0.6384	0.3750	0.1630	0.9444	0.4221	0.5779	0.3583	0.2125	1.0577
Y02	**rs10629077**	0.5126	0.4874	0.2726	0.0542	0.6908	0.4172	0.5828	0.3356	0.1244	0.8571	0.4877	0.5123	0.2859	0.0782	0.7500	0.5061	0.4939	0.2760	0.1032	0.8088	0.4722	0.5278	0.2976	0.0994	0.8000	0.4321	0.5679	0.3229	0.1032	0.8088
Y03	**rs10688868**	0.3786	0.6214	0.3556	0.1174	0.8413	0.4150	0.5850	0.3589	0.2025	1.0345	0.4081	0.5919	0.3490	0.1571	0.9310	0.3666	0.6334	0.3729	0.1647	0.9483	0.3864	0.6136	0.3579	0.1463	0.9067	0.4712	0.5288	0.3047	0.1305	0.8710
Y04	**rs2067208**	0.5274	0.4726	0.2593	0.0724	0.7361	0.5239	0.4761	0.2619	0.0706	0.7317	0.5316	0.4684	0.2562	0.0782	0.7500	0.4949	0.5051	0.2830	0.0599	0.7051	0.3816	0.6184	0.3579	0.1359	0.8831	0.4221	0.5779	0.3583	0.2125	1.0577
Y05	**rs2307579**	0.3971	0.6029	0.3744	0.2227	1.0816	0.3550	0.6450	0.3744	0.1474	0.9091	0.3724	0.6276	0.3680	0.1571	0.9310	0.3501	0.6499	0.3720	0.1254	0.8594	0.3881	0.6119	0.3743	0.2075	1.0462	0.3468	0.6532	0.3737	0.1254	0.8594
Y06	**rs2308020**	0.4135	0.5865	0.3389	0.1295	0.8689	0.5001	0.4999	0.2782	0.0893	0.7763	0.4246	0.5754	0.3255	0.0872	0.7714	0.4102	0.5898	0.3500	0.1647	0.9483	0.3962	0.6038	0.3419	0.0878	0.7727	0.4076	0.5924	0.3359	0.0933	0.7857
Y07	**rs3080855**	0.4484	0.5516	0.3305	0.1715	0.9636	0.4339	0.5661	0.3236	0.1139	0.8333	0.5110	0.4890	0.2718	0.0620	0.7105	0.4453	0.5547	0.3129	0.0840	0.7639	0.3741	0.6259	0.3657	0.1517	0.9189	0.3997	0.6003	0.3557	0.1647	0.9483
Y08	**rs1610919**	0.3621	0.6379	0.3750	0.1639	0.9464	0.3650	0.6350	0.3693	0.1474	0.9091	0.3669	0.6331	0.3644	0.1305	0.8710	0.3706	0.6294	0.3750	0.1797	0.9821	0.3931	0.6069	0.3677	0.1940	1.0149	0.4076	0.5924	0.3606	0.1956	1.0185
Y09	**rs2307839**	0.5584	0.4416	0.2377	0.0683	0.7260	0.5672	0.4328	0.2310	0.0706	0.7317	0.5027	0.4973	0.2790	0.0548	0.6923	0.7263	0.2737	0.1390	0.0206	0.5978	0.4646	0.5354	0.3024	0.0994	0.8000	0.5246	0.4754	0.2612	0.0840	0.7639
R01	**rs2308137**	0.4219	0.5781	0.3480	0.1794	0.9815	0.4050	0.5950	0.3425	0.1244	0.8571	0.3772	0.6228	0.3521	0.0782	0.7500	0.4876	0.5124	0.3666	0.3368	1.3750	0.4385	0.5615	0.3235	0.1260	0.8608	0.3546	0.6454	0.3736	0.1433	0.9000
R02	**rs36040336**	0.5214	0.4786	0.2633	0.0767	0.7465	0.5150	0.4850	0.2688	0.0635	0.7143	0.5089	0.4911	0.2718	0.0782	0.7500	0.5385	0.4615	0.2533	0.0599	0.7051	0.4409	0.5591	0.3176	0.1035	0.8095	0.5299	0.4701	0.2612	0.0529	0.6875
R03	**rs1160886**	0.4028	0.5972	0.3607	0.1875	1.0000	0.3739	0.6261	0.3680	0.1600	0.9375	0.3724	0.6276	0.3680	0.1571	0.9310	0.3693	0.6307	0.3648	0.1376	0.8871	0.3982	0.6018	0.3671	0.2007	1.0303	0.3977	0.6023	0.3468	0.1254	0.8594
R04	**rs2308026**	0.3896	0.6104	0.3630	0.1715	0.9636	0.4239	0.5761	0.3541	0.2025	1.0345	0.3765	0.6235	0.3623	0.1433	0.9000	0.4089	0.5911	0.3468	0.1507	0.9167	0.4157	0.5843	0.3358	0.1212	0.8500	0.4605	0.5395	0.3075	0.1139	0.8333
R05	**rs2307526**	0.3848	0.6152	0.3512	0.1116	0.8281	0.4200	0.5800	0.3515	0.1875	1.0000	0.3669	0.6331	0.3708	0.1571	0.9310	0.3904	0.6096	0.3747	0.2125	1.0577	0.4373	0.5627	0.3254	0.1308	0.8718	0.4030	0.5970	0.3583	0.1797	0.9821
R06	**rs34811743**	0.4128	0.5872	0.3408	0.1360	0.8833	0.4306	0.5694	0.3457	0.1875	1.0000	0.5480	0.4520	0.2479	0.0548	0.6923	0.4274	0.5726	0.3275	0.1139	0.8333	0.4139	0.5861	0.3389	0.1308	0.8718	0.4380	0.5620	0.3557	0.2305	1.1000
R07	**rs2308189**	0.4014	0.5986	0.3723	0.2227	1.0816	0.3606	0.6394	0.3716	0.1474	0.9091	0.4287	0.5713	0.3521	0.2042	1.0385	0.4592	0.5408	0.3729	0.3133	1.3095	0.4426	0.5574	0.3235	0.1359	0.8831	0.4208	0.5792	0.3648	0.2305	1.1000
R08	**rs5895447**	0.3861	0.6139	0.3607	0.1566	0.9298	0.4383	0.5617	0.3693	0.2714	1.2000	0.3957	0.6043	0.3421	0.0698	0.7297	0.4949	0.5051	0.2830	0.0599	0.7051	0.4153	0.5847	0.3625	0.2146	1.0625	0.4559	0.5441	0.3398	0.2125	1.0577
R09	**rs2308171**	0.5730	0.4270	0.2283	0.0605	0.7067	0.7406	0.2594	0.1411	0.0142	0.5769	0.6756	0.3244	0.1662	0.0308	0.6279	0.4936	0.5064	0.2830	0.0933	0.7857	0.5410	0.4590	0.2583	0.0425	0.6602	0.4552	0.5448	0.3075	0.0933	0.7857
R10	**rs35605984**	0.3943	0.6057	0.3677	0.1959	1.0192	0.4306	0.5694	0.3278	0.1244	0.8571	0.4081	0.5919	0.3421	0.1305	0.8710	0.4698	0.5302	0.3318	0.2125	1.0577	0.4104	0.5896	0.3608	0.2007	1.0303	0.4023	0.5977	0.3743	0.2305	1.1000

MP: Matching probability; PD: Power od discriminations; PIC: Polymorphism Information Content; PE: Power of Exclusion; TPI: Typical Paternity Index.

Supplementary Table 5. Alleles presented for autosomal STRs and Indels and X-chromosome STRs by the trio.

Autosomal STR	Alleged Father	Daughter	Mother
D8S1179	13-14	11-14	11-14
D21S11	30-32.2	30	30-32.2
D7S820	11-13	12-13	12-13
CSF1PO	11	11-12	11-12
D3S1358	16	16-17	16-17
TH01	9.3	9.3	8-9.3
D13S317	12	12	12-13
D16S539	11-13	9-11	9
D2S1338	17-22	**18**-25	23-25
D19S433	13	13-15	13-15
VWA	17-18	16-17	16-18
TPOX	11	9-11	8-9
D18S51	13-16	13-14	14-15
D5S818	11-13	11-**12**	11
FGA	21-22	21	21-23
Penta E	5-10	10-11	10-11
Penta D	10-12	10-11	10-11

X-STR	Alleged Father	Daughter	Mother
DXS10103	18	17-18	17-20
DXS8378	10	10-11	11-12
DXS7132	15	13-15	13-14
DXS10134	37	36-37	36-37
DXS10074	14	14-19	17-19
DXS10101	32	31-32	29.2-31
DXS10135	21	21-25	25-26
DXS7423	15	15	15
DXS10146	30	27-30	27-46.2
DXS10079	20	20	19-20
HPRTB	12	12-14	13-14
DXS10148	26.1	26.1	26.1-27.1

Indels	Alleged Father	Daughter	Mother	Indels	Alleged Father	Daughter	Mother
B01	22	12	11	**Y01**	12	12	12
B02	11	11	11	**Y02**	22	12	12
B03	12	12	22	**Y03**	12	12	12
B04	11	12	12	**Y04**	22	22	22
B05	11	12	12	**Y05**	11	11	12
B06	12	12	22	**Y06**	12	11	11
B07	12	11	11	**Y07**	22	12	12
B08	12	11	11	**Y08**	11	11	12
B09	11	12	22	**Y09**	22	12	12
B10	22	22	22	**R01**	22	12	12
G01	11	12	22	**R02**	11	11	11
G02	22	12	11	**R03**	11	11	11
G03	22	12	12	**R04**	12	12	12
G04	22	22	12	**R05**	22	12	11
G05	12	22	22	**R06**	12	11	11
G06	12	22	12	**R07**	11	11	12
G07	12	12	11	**R08**	12	12	22
G08	12	12	11	**R09**	22	22	22
G09	12	22	22	**R10**	22	22	22

Mismatches in autosomal STR markers are labelled in bold.

Supplemetary Table 6. Genetic distances matrix between the populations in this study and Iberian populations (Portugal and south East Spain), Chuetas host population (Majorca), Brazilian populations (Rio Janeiro and Terenas), Asian and African populations.

	Sephardic Jews	**North African Jews**	**Ashkenazi Jews**	**Middle Eastern Jews**	**Bragança Jews**	**Chuetas**	**Majorca (Spain)[1]**	**South East Spain[2]**	**Portugal[3]**	**Rio Janeiro (Brazil)[4]**	**Asia[3]**	**Terenas (Brazil)[4]**	**Africa[3]**
Sephardic Jews	0												
North African Jews	**0**	0											
Ashkenazi Jews	**0.00259**	**0.00498**	0										
Middle Eastern Jews	**0.00212**	**0.00472**	**0.00439**	0									
Bragança Jews	0.00842	0.00714	0.00942	0.01367	0								
Chuetas	0.0112	0.01004	0.01509	0.01714	0.00835	0							
Majorca	0.00588	**0.00456**	0.00637	0.01205	0.00497	0.01188	0						
South East Spain	0.00307	**0.00249**	0.00702	0.01009	0.00522	0.00586	**0.00157**	0					
Portugal	**0**	**0**	**0.0021**	0.006	**0**	0.00325	**0**	**0**	0				
Rio Janeiro	0.01309	0.01453	0.01351	0.01639	0.0101	0.01557	0.01533	0.01313	0.00418	0			
Asia	0.07783	0.08834	0.0709	0.07665	0.08979	0.08226	0.08393	0.08275	0.08862	0.062	0		
Terenas	0.09512	0.10289	0.09741	0.09528	0.08938	0.09383	0.10614	0.09656	0.09056	0.06601	0.09143	0	
Africa	0.10134	0.11246	0.10343	0.10934	0.09564	0.10088	0.10607	0.10161	0.10564	0.04784	0.12031	0.13971	0

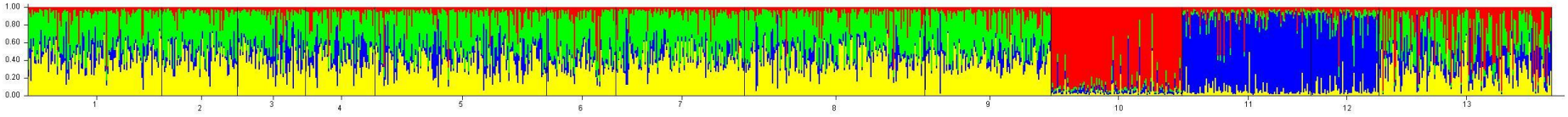
[1] Unpublished data; [2] Saiz et al. (2014); [3] Pereira et al. (2009); [4] Manta et al. (2012).
Non-significant distances (p>0.00385) are labelled in bold. Darker colours correspond to higher genetic distances.

Additional information not included in the publication

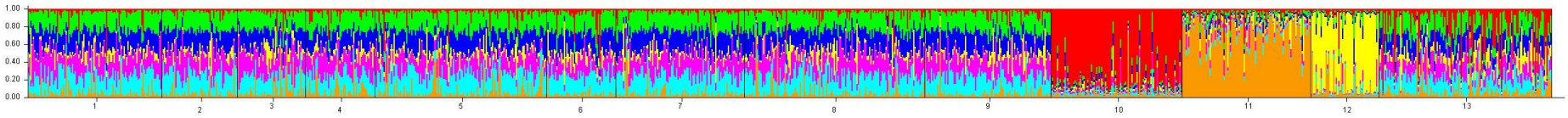
Supplementary Figure 1a. Mean values of ln P(D) from K=1 to K=9 for the 6 Jewish populations studied (Sephardic Jews, North African Jews, Middle Eastern Jews, Ashkenazi Jews, Chuetas and Bragança Jews).

K	Ln P(D)	Var [Ln P(D)]
1	-24433.93	18.97
2	-25073.31	1381.39
3	-25509.92	2314.4
4	-26072.43	3483.98
5	-26650.01	4729.99
6	-26219.53	3875.16
7	-27361.11	6250.36
8	-27792.25	7157.95
9	-29455.56	10523.74

Supplementary Figure 1b. STRUCTURE Plots for 13 worldwide populations. K=4 showed the high value of ln P(D). In K=7 Terenas population (Native American) appears as a separated cluster



K=4



K=7

1: Sephardic Jews; 2: North African Jews;3: Middle Eastern Jews; 4: Ashkenazi Jews; 5:Chuetas; 6:Braganza Jews; 7: Majorca (Spain); 8:South East Spain; 9: Portugal; 10: African; 11: Asian; 12: Terenas and 13: Rio de Janeiro (Brasil).

Chapter 2: X-Chromosome

1. Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations

Ferragut JF; Bentayebi K; Castro JA; Ramon C and Picornell A

International Journal of Legal Medicine. 129(2): 253–255 (2015)

2. Genetic diversity of 12 X-chromosomal short tandem repeats in Jewish populations

Ferragut JF; Castro JA; Ramon C and Picornell A

Forensic Science International: Genetics Supplement Series. 5: e327–e329 (2015)

3. Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alus and STRs

Ferragut JF; Bentayebi K; Pereira R; Castro JA; Amorim A; Ramon C and Picornell A

Forensic Science International: Genetics (submitted)

Introduction

The second chapter is devoted to the X-chromosome, since it is increasingly important in Forensic and Populations Genetics. On the one hand, the results of the X-chromosome combined with those of Y-chromosome and mtDNA will enable us to assess the paternal and maternal contribution to the genetic pool of Chuetas from the original Sephardic population and from the people among whom they lived (the gentile Majorcan population). On the other hand, as has been explained in the introduction, X-chromosome markers play an increasing role in Forensic Genetics, especially to solve complex kinship cases.

The first set of markers we studied was the 12 X-STRs of Investigator Argus X-12 Kit (Qiagen) due to its growing use in Forensic Genetics in the last years. Since it had not been studied before, first of all we characterized the host population of the Chuetas, Majorca and its neighbouring populations (Minorca, Ibiza and Valencia) (Ferragut et al., 2015a). A total of 255 samples were genotyped in this paper.

Subsequently, the same STR kit was used to genotype the Chueta population and 4 Jewish populations. In addition, in order to go further into the study of the X-chromosome, two other sets of markers (X-Alu and X-Indels) were also studied in these populations and Majorca (Ferragut et al., 2015b) (Ferragut et al., submitted). This work comprised 53 X-chromosome markers and a total of 500 samples genotyped. With this study, we aim not only to characterize and compare these six populations, but also to contrast the behaviour of these different X-chromosome markers in terms of Forensic and Population Genetics.

Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations

Ferragut JF [a]; Bentayebi K[a,b]; Castro JA[a]; Ramon C [a] and Picornell A[a]*

[a]Institut Universitari d'Investigació en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Carretera de Valldemossa, km 7.5, 07122, Palma de Mallorca, Illes Balears, Spain.

[b]Biochemistry and Immunology laboratory-Forensic Unit, Mohammed V Agdal University, Faculty of Sciences, Avenue Ibn Batouta Agdal, BP 1014 Rabat, Morocco

Abstract

Haplotype and allele frequencies of 12 X-STRs included in the Investigator Argus X-12 kit are reported for 255 individuals, representing four Western Mediterranean populations: Valencia (eastern mainland Spain) and the Balearic Islands (Majorca, Minorca and Ibiza). Ibiza shows the lowest intra-population variability and the highest level of linkage disequilibrium together with an important genetic distance with regard to the geographically close populations, which is consistent with the historical evidence for long-term demographic isolation and its different matrilineal background.

Key words: X-STRs, X-chromosome, Balearic Islands, Valencia, Western Mediterranean, Investigator Argus X-12 kit

*__Corresponding author:__ Dr. Antònia Picornell. Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Cra. Valldemossa, km 7.5, 07122-Palma de Mallorca, Illes Balears, Spain. Tel: +34-971-172050. Fax: +34-971-173184. E-mail: apicornell@uib.es

The Balearic archipelago (Majorca, Minorca and Ibiza islands, in the Western Mediterranean Sea) has been inhabited for 5500 years, and different people – especially the Romans (3rd century BC) and Catalans (early 13th century) – contributed to the genetic pool of the current population, but no remarkable additional contribution of foreign genes was received until recently because of immigration. Genetic studies (Rodríguez et al., 2009; Tomàs et al., 2006) show differences between the three islands and a remarkable differentiation of Ibiza compared to other Western Mediterranean populations, probably because of the Phoenician–Carthaginian origin of the first settlers and the effect of genetic drift. Few data regarding X-chromosome STRs have been reported in Western Mediterranean populations (Barbaro et al., 2012; Bentayebi et al., 2012) and there are no available data in the Balearic Islands. It is essential to give more insight into the genetic substructure in this area, focusing on known isolates such as the Ibiza population. In this work we studied the genetic diversity, geographic distribution and population structure of three insular populations from the Balearic Archipelago, and Valencia (eastern coast of the Iberian Peninsula), as a mainland Spanish reference.

Mouthwash samples were obtained from 255 unrelated individuals (160 males and 95 females), after informed consent. DNA was extracted by standard phenol-chloroform method. PCR amplification, capillary electrophoresis, data analysis and calculation of allele frequencies and additional relevant population and forensic statistical parameters were performed as previously described (Bentayebi et al., 2012). To examine the relationship of these populations with neighbouring populations, F_{ST} genetic distances, calculated using POPTREE2 (Takezaki et al., 2010), were performed to generate the multi-dimensional scaling (MDS) plot conducted with SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA).

Table 1. Distribution of X-STR haplotypes for the linkage groups in 160 males from Balearic Islands (n=121) and Valencian (n=39) populations.

Linkage Group	Number of Haplotypes			Unique Haplotypes			Most common Haplotype Frequency (%)			Haplotype diversity		
	B.I.	Val.	Total	B.I.	Val.	Total	B.I.	Val.	Total	B.I.	Val.	Total
1	102	39	135	88 (72.73%)	39 (100%)	116 (85.93%)	2.48	2.56	2.50	0.9967 ± 0.0016	1.000 ± 0.0058	0.9975 ± 0.0011
2	78	36	98	55 (45.45%)	33 (84.62%)	65 (66.33%)	5.79	5.13	5.00	0.9895 ± 0.0028	0.9960 ± 0.0069	0.9910 ± 0.0020
3	70	34	93	43 (35.54%)	30 (76.92%)	61 (65.59%)	5.79	7.69	5.00	0.9869 ± 0.0031	0.9919 ± 0.0082	0.9888 ± 0.0024
4	84	34	103	60 (49.59%)	30 (76.92%)	68 (66.02%)	3.31	7.69	3.13	0.9926 ± 0.0022	0.9919 ± 0.0082	0.9923 ± 0.0019

B.I. Balearic Islands; Val. Valencia.

Allele frequencies are shown in Supplementary Table 1. The highest variability was found in DXS10135, as described elsewhere (Tomàs et al., 2012), with 27 alleles and observed heterozygosities ranging between 0.864 and 0.950. The lowest heterozygosity (0.480) was observed for the DXS7423 system in Minorca. No deviations from HWE

were observed after Bonferroni correction. Typing of the 160 males from the Western Mediterranean region resulted in 160 different haplotypes when all 12 X-STRs were included. Linkage groups 1-4 revealed 135, 98, 93 and 103 haplotypes, respectively (Table 1). Among all the observed haplotypes, 96.4% showed frequencies <0.020 and the three most common haplotypes were observed in 8 individuals each, displaying a frequency of 0.050. In these populations, linkage group 1 proved to be the most polymorphic group and linkage group 3 the least, in accordance with other studies (Tomàs et al., 2012). The Valencia population showed greater haplotype diversity than Balearic Island populations, with a higher percentage of unique haplotypes in the four linkage groups. Supplementary Table 2 shows haplotype frequencies for each population. The Ibiza population showed the lowest haplotype diversity (HD) values – with a low percentage of unique haplotypes (between 46.3% and 56.1%) compared to the other populations studied (64.1% - 94.9%) – and the highest level of linkage disequilibrium (LD) in most of the loci pairs inside each linkage group, although after Bonferroni correction only one significant p-value was observed for DXS10148-DXS10135 pair of loci (p=0.000) (Supplementary Table 3). LD does not depend exclusively on the physical distance between loci, but may result from random genetic drift, founder effect, mutations, selection and population admixture or stratification (Chakravarti, 1999); therefore the greater LD found in Ibiza could result from the founder effect and genetic drift related to the demographic and historical features of this population (an isolated, consanguineous population with a reduced effective population size), supporting previous genetic studies (Tomàs et al., 2006). When Balearic Island populations were pooled, significant LD was observed inside each linkage group. Although no significant differences in allelic frequency were observed in the Balearic populations for the 12 X-STRs studied, different studies emphasize the importance of investigating these populations and attempt to describe their roots and genetic substructure (Rodríguez et al., 2009; López- Escribano et al., 2013); therefore, these LD values could result from the heterogeneous distribution of the haplotypes among the islands. Forensic parameters of interest were calculated for each X-STR and population (Supplementary Table 1). The combined power of discrimination (PD) in females ranged from 1 in 4.74E+14 (in Ibiza) to 1 in 1.80E+15 (in Mallorca and Valencia). Combined PD males values ranged from 1 in 5.33E+08 (in Ibiza) to 1 in 1.04E+09 (in Valencia). A high combined MEC was also obtained for father-daughter duos and father-girl-mother trios (>0.99999 in all cases). Although values differed slightly between populations, the set of loci in the Argus X-12 kit was highly informative in all the Western Mediterranean populations studied. A comparison between the Western Mediterranean population and 12 other European and African (mainly North African) populations with available data for the same set of X-STRS markers was performed. Supplementary Figure 1 shows a multi-dimensional scaling plot based on pairwise F_{ST} genetic distances. Along the X-axis, the distribution of the populations ranging from Northeast Africa to Northern Europe – with the Mediterranean populations showing an intermediate position – can be observed, as found in X-chromosome SNPs studies (Tomàs et al., 2008). However Majorca, Menorca and Valencia are closer to other European populations than to North Africans, in accordance with other studies that suggest the existence of a relative north-south gene flow barrier in

the Western part of the Mediterranean area (Comas et al., 2000; Rodríguez et al., 2009). Ibiza and Sahrawi populations show a remote position versus their geographic neighbouring populations; which may be due to historical and cultural background in the Sahrawi case. Regarding Ibiza, its differentiation versus other European populations was also evidenced by mtDNA analysis, but not by Y-chromosome markers, indicating a sex-biased contribution to the genetic pool of Ibiza (Tomàs et al., 2006). Indeed, as two thirds of the X-chromosomes descend from maternal origin, X-chromosome polymorphisms will mostly behave as matrilineal markers showing similar results to those obtained from mtDNA data. Therefore, the observed displaced plot of Ibiza based on X-STR results may reflect the differential matrilineal background of this population, probably accentuated by the effect of genetic drift. In conclusion, the results of the present study provide a useful X-STR database for the Western Mediterranean region. It is important to highlight that a local haplotype database must be used especially in Ibiza due to the matrilineal genetic features of this population.

Acknowledgments

This work was partially supported by grant AAEE133/09 from the Direcció General de R+D+I (Comunitat autònoma de les Illes Balears) and European Regional Development Fund (ERDF).

Supplementary Table 1. Allele frequencies of 12 X-STR in 350 chromosomes (Majorca (MAJ): 41 men and 22 women; Minorca (MIN): 39 men and 25 women; Ibiza (IBI): 41 males and 20 women; Valencia (VAL): 39 men and 28 women).

	DXS10079					DXS10101					DXS10103					DXS10134					DXS10146					DXS10148				
N	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95
Alelle	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL
13																														
13.3																														
14			0.017	0.005																										
15	0.023	0.013	0.008	0.015							0.015		0.025	0.013	0.021															
16	0.031	0.039	0.024	0.032							0.092	0.092	0.058	0.081	0.063															
17	0.07	0.079	0.074	0.074	0.114						0.062	0.109	0.025	0.067	0.097															
17.1																														
18	0.201	0.232	0.248	0.227	0.174						0.148	0.184	0.24	0.189	0.256											0.096	0.139	0.1	0.111	0.097
18.1																														
19	0.295	0.277	0.215	0.264	0.175						0.504	0.409	0.421	0.447	0.408											0.08	0.044	0.017	0.046	0.008
19.1																														
20	0.225	0.255	0.191	0.225	0.288						0.179	0.162	0.231	0.188	0.126												0.013		0.005	0.012
20.1																														
20.3																												0.017	0.005	
21	0.085	0.083	0.215	0.124	0.191							0.044		0.015	0.029													0.016	0.006	
21.1																														
21.2																														
22	0.054	0.013	0.008	0.025	0.049																						0.009		0.003	0.008
22.1																										0.032	0.035	0.033	0.032	0.024
23	0.016			0.006	0.009																					0.033	0.017	0.033	0.027	0.048
23.1																										0.08	0.03	0.049	0.052	0.038
24		0.009		0.003																	0.015	0.022	0.05	0.028	0.02		0.035	0.008	0.016	0.02
24.1																										0.128	0.198	0.198	0.176	0.237
24.2																														
25																					0.038	0.049	0.091	0.059	0.07	0.016			0.006	
25.1																										0.176	0.123	0.239	0.178	0.23
25.2										0.008																				
25.3																														
26																					0.079	0.07	0.132	0.093	0.114					
26.1																										0.167	0.154	0.108	0.144	0.109
26.2										0.024																				
27						0.015	0.009		0.008												0.161	0.175	0.099	0.146	0.143					
27.1																										0.04	0.088	0.141	0.088	0.08
27.2						0.016	0.062	0.058	0.046	0.082																				
28						0.07	0.067	0.083	0.074	0.036											0.225	0.158	0.148	0.178	0.182					
28.1																										0.128	0.075	0.033	0.079	0.041
28.2						0.133	0.119	0.124	0.125	0.041																				
29						0.038	0.035		0.025	0.026											0.132	0.214	0.017	0.17	0.172					
29.1							0.009		0.003	0.012																0.008	0.027		0.013	0.036
29.2						0.21	0.183	0.107	0.169	0.143																				
30						0.016	0.022	0.05	0.029	0.085											0.077	0.079	0.141	0.097	0.088					
30.1																												0.008	0.003	0.012
30.2						0.163	0.138	0.132	0.143	0.162																				
31						0.078	0.092	0.174	0.113	0.058											0.079	0.027	0.166	0.041	0.012					
31.1																										0.016	0.013		0.01	
31.2						0.124	0.154	0.116	0.132	0.18																				
32						0.039	0.052	0.024	0.039	0.053						0.015	0.026	0.008	0.016	0.032		0.008		0.003	0.012					
32.2						0.023	0.022	0.058	0.033	0.012																				

33						0.03	0.022	0.058	0.036	0.07						0.031	0.013	0.017	0.02	0.032											
33.2										0.008																					
34						0.03		0.016	0.015							0.092	0.08	0.066	0.081	0.099											
34.2																															
35						0.015	0.014		0.01							0.232	0.253	0.215	0.233	0.163											
35.2																0.015			0.005												
35.3																				0.021											
36																0.133	0.218	0.289	0.213	0.287											
36.2																	0.022		0.008	0.012		0.022		0.008							
37																0.185	0.159	0.132	0.161	0.141											
37.2																0.015	0.027		0.015												
37.3																	0.009		0.003	0.012											
38																0.041	0.061	0.116	0.070	0.097											
38.2																0.015			0.005												
38.3																0.015		0.016	0.011												
39																0.015	0.013	0.008	0.013	0.024											
39.2																					0.023	0.013	0.050	0.028	0.030						
39.3																0.038	0.013	0.008	0.020	0.038											
40																0.023	0.009		0.010												
40.2																					0.038	0.040		0.028	0.041						
40.3																0.040	0.009	0.058	0.034												
41.2																					0.040	0.052	0.024	0.040	0.024						
41.3																0.055	0.079		0.046	0.024											
42.2																					0.008	0.013	0.008	0.010	0.020						
42.3																0.032	0.009	0.017	0.018	0.009											
43.2																					0.008	0.040	0.008	0.020	0.012						
43.3																0.008		0.050	0.018	0.009											
44.2																					0.031			0.010	0.024						
44.3																															
45.2																					0.046	0.009	0.058	0.035	0.012						
46.2																						0.009		0.003							
47.2																							0.008	0.003	0.024						
Het	0.805	0.789	0.803	0.805	0.804	0.879	0.887	0.892	0.892	0.891	0.680	0.750	0.707	0.718	0.737	0.874	0.844	0.828	0.857	0.846	0.879	0.877	0.886	0.889	0.886	0.882	0.883	0.856	0.883	0.855	
Obs Het*	0.864	0.800	1.000	0.881	0.750	0.864	0.960	0.850	0.896	0.821	0.682	0.800	0.650	0.716	0.607	0.682	0.920	0.950	0.851	0.821	0.864	1.000	0.950	0.940	1.000	0.905	0.880	0.900	0.894	0.857	
HW p-value*	0.936	0.340	0.529	0.715	0.957	0.721	0.789	0.125	0.097	0.118	0.365	0.231	0.182	0.390	0.180	0.011	0.583	0.308	0.162	0.302	0.075	0.964	0.435	0.675	0.908	0.887	0.880	0.658	0.705	0.233	
PIC	0.779	0.758	0.774	0.778	0.776	0.868	0.877	0.882	0.882	0.882	0.643	0.718	0.659	0.680	0.701	0.863	0.827	0.808	0.843	0.831	0.868	0.865	0.875	0.879	0.875	0.871	0.872	0.840	0.872	0.840	
PD female	0.936	0.924	0.932	0.935	0.934	0.974	0.977	0.978	0.979	0.979	0.861	0.905	0.866	0.883	0.895	0.973	0.958	0.951	0.965	0.961	0.974	0.973	0.976	0.978	0.976	0.975	0.975	0.964	0.975	0.964	
PD male	0.805	0.789	0.803	0.805	0.804	0.879	0.887	0.892	0.892	0.891	0.680	0.750	0.707	0.718	0.737	0.874	0.844	0.828	0.857	0.846	0.879	0.877	0.886	0.889	0.886	0.882	0.883	0.856	0.883	0.855	
PE	0.609	0.579	0.605	0.609	0.607	0.753	0.769	0.779	0.779	0.778	0.397	0.510	0.439	0.456	0.489	0.743	0.683	0.652	0.709	0.688	0.753	0.749	0.766	0.773	0.766	0.759	0.760	0.706	0.761	0.705	
PI	2.566	2.374	2.539	2.569	2.554	4.133	4.427	4.617	4.621	4.597	1.560	2.003	1.708	1.771	1.905	3.972	3.207	2.909	3.503	3.256	4.138	4.062	4.376	4.505	4.367	4.240	4.257	3.460	4.276	3.451	
MEC Desmarais	0.779	0.758	0.774	0.778	0.776	0.868	0.877	0.882	0.882	0.882	0.643	0.718	0.659	0.680	0.701	0.863	0.827	0.808	0.843	0.831	0.868	0.865	0.875	0.879	0.875	0.871	0.872	0.840	0.872	0.840	
MEC Desmarais duo	0.657	0.630	0.649	0.655	0.651	0.777	0.790	0.797	0.798	0.797	0.497	0.582	0.515	0.539	0.563	0.771	0.720	0.695	0.742	0.726	0.778	0.774	0.787	0.794	0.788	0.780	0.782	0.738	0.783	0.739	

	DXS7132					DXS7423					DXS8378					DXS10074					DXS10135					HPRTB				
N	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95	85	89	81	265	95	72	89	81	265	95
Alelle	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL	MAJ	MIN	IBI	BI	VAL
7																0.015	0.043	0.041	0.033	0.068										
8																0.142	0.174	0.240	0.182	0.189										
9											0.047			0.015	0.041	0.031	0.013	0.025	0.023	0.026							0.009		0.003	
10											0.264	0.366	0.389	0.341	0.279											0.008	0.013		0.008	0.021
11	0.015	0.009		0.008	0.024			0.016	0.005		0.348	0.241	0.232	0.272	0.305											0.163	0.092	0.116	0.123	0.150
11.2													0.008	0.003																
12	0.130	0.126	0.249	0.163	0.076						0.333	0.349	0.290	0.323	0.349											0.342	0.321	0.339	0.334	0.363
13	0.248	0.245	0.272	0.253	0.298	0.084	0.084	0.066	0.078	0.100	0.008	0.044	0.065	0.040	0.026	0.008			0.003	0.009						0.301	0.373	0.371	0.349	0.245
13.3																														
14	0.375	0.349	0.339	0.359	0.397	0.310	0.330	0.339	0.328	0.382			0.016	0.006		0.008	0.022	0.041	0.023	0.012						0.147	0.140	0.124	0.136	0.180
15	0.216	0.232	0.140	0.198	0.187	0.427	0.429	0.389	0.415	0.344						0.109	0.110	0.033	0.087	0.168						0.039	0.043	0.050	0.044	0.041
16	0.016	0.039		0.019	0.009	0.126	0.118	0.132	0.125	0.165						0.192	0.128	0.091	0.138	0.135		0.022		0.008	0.009		0.009		0.003	
17					0.009	0.053	0.039	0.058	0.049	0.009						0.211	0.242	0.198	0.219	0.138		0.089		0.033	0.021					
17.1																						0.009		0.003						
17.2																				0.017										
18																0.192	0.202	0.240	0.210	0.189	0.070	0.022	0.041	0.045	0.043					
18.1																							0.008	0.003	0.009					
19																0.092	0.066	0.091	0.082	0.049	0.086	0.088	0.083	0.084	0.093					
19.1																					0.040	0.009		0.016						
20																					0.063	0.009	0.050	0.039	0.065					
20.1																					0.008			0.003	0.012					
21																					0.124	0.084	0.100	0.101	0.124					
21.1																						0.009		0.003	0.032					
21.2																														
22																					0.108	0.057	0.057	0.075	0.133					
22.1																					0.030	0.022	0.008	0.020						
23																						0.149	0.033	0.064	0.053					
23.1																						0.017		0.005	0.012					
24																					0.062	0.074	0.041	0.059	0.029					
24.1																														
24.2																														
25																					0.108	0.096	0.174	0.124	0.073					
25.1																														
25.3																														
26																					0.015	0.013	0.058	0.028	0.065					
26.1																														
26.2																														
27																					0.078	0.030	0.099	0.067	0.066					
27.1																														
27.2																														
28																					0.038	0.079	0.133	0.081	0.021					
28.1																					0.008			0.003						
28.2																														
29																					0.054	0.048	0.025	0.043	0.053					
29.1																						0.013		0.005						
29.2																														
30																					0.062	0.013	0.033	0.036	0.032					
30.1																														

30.2																															
31																					0.023	0.013	0.033	0.023	0.009						
31.2																															
32																					0.023	0.035	0.016	0.026	0.023						
32.2																															
33																							0.008	0.003	0.023						
33.2																															
Het	0.734	0.747	0.729	0.741	0.712	0.696	0.685	0.708	0.696	0.698	0.696	0.684	0.706	0.704	0.705	0.840	0.835	0.824	0.839	0.855	0.923	0.924	0.909	0.933	0.927	0.743	0.728	0.716	0.731	0.751
Obs Het*	0.591	0.760	0.700	0.687	0.857	0.727	0.480	0.850	0.672	0.679	0.636	0.640	0.650	0.642	0.643	0.955	0.840	0.800	0.866	0.821	0.864	0.920	0.950	0.910	0.929	0.727	0.840	0.750	0.776	0.857
HW p-value*	0.528	0.333	0.782	0.473	0.264	0.644	0.034	0.115	0.021	0.226	0.673	0.274	0.754	0.143	0.865	0.021	0.977	0.752	0.238	0.795	0.114	0.352	0.439	0.013	0.941	0.663	0.773	0.666	0.859	0.530
PIC	0.689	0.704	0.679	0.697	0.663	0.646	0.631	0.660	0.645	0.644	0.635	0.621	0.653	0.646	0.647	0.820	0.814	0.801	0.819	0.838	0.918	0.919	0.902	0.928	0.923	0.700	0.684	0.668	0.686	0.712
PD female	0.884	0.893	0.877	0.889	0.868	0.858	0.847	0.866	0.857	0.854	0.847	0.837	0.860	0.854	0.855	0.954	0.952	0.946	0.954	0.962	0.989	0.989	0.985	0.991	0.990	0.891	0.882	0.871	0.883	0.899
PD male	0.734	0.747	0.729	0.741	0.712	0.696	0.685	0.708	0.696	0.698	0.696	0.684	0.706	0.704	0.705	0.840	0.835	0.824	0.839	0.855	0.923	0.924	0.909	0.933	0.927	0.743	0.728	0.716	0.731	0.751
PE	0.483	0.504	0.475	0.494	0.447	0.422	0.405	0.441	0.422	0.426	0.422	0.404	0.438	0.434	0.436	0.675	0.666	0.644	0.674	0.704	0.843	0.844	0.813	0.862	0.852	0.497	0.472	0.454	0.478	0.512
PI	1.879	1.975	1.848	1.930	1.737	1.644	1.585	1.714	1.645	1.658	1.645	1.583	1.702	1.687	1.695	3.124	3.031	2.840	3.112	3.445	6.492	6.555	5.466	7.410	6.880	15.102	13.475	12.408	13.820	16.153
MEC Desmarais	0.689	0.704	0.679	0.697	0.663	0.646	0.631	0.660	0.645	0.644	0.635	0.621	0.653	0.646	0.647	0.820	0.814	0.801	0.819	0.838	0.918	0.919	0.902	0.928	0.923	0.700	0.684	0.668	0.686	0.712
MEC Desmarais duo	0.548	0.566	0.536	0.556	0.520	0.502	0.486	0.517	0.501	0.499	0.490	0.474	0.508	0.501	0.503	0.708	0.702	0.685	0.708	0.733	0.853	0.855	0.828	0.871	0.861	0.560	0.544	0.526	0.546	0.574

* Calculated using female data.

BI– pooled Balearic Islands; N– Number of X-chromosomes studied; Het– expected heterozygosity; Obs Het– observed heterozygosity; HW– Hardy-Weinberg; PIC – polymorphic information content; PDfemale – power of discrimination in women; PDmale – power of discrimination in men; PE- power of exclusion; PI- paternity Index; MEC Desmarais– mean exclusion chance in trios involving daughter; MEC Desmarais duo – mean exclusion chance in father/daughter or mother/son duos.

Supplementary Table 2. X-chromosome haplotypes of four linked STR trios in 121 males from Balearic Islands (41 from Majorca, 39 from Minorca and 41 from Ibiza) and 39 males from Valencia populations. More common haplotypes (5%) are labelled with *.

Linkage group 1												
DXS10148	**DXS10135**	**DXS8378**	**Majorca**	**Minorca**	**Ibiza**	**Valencia**	**DXS10148**	**DXS10135**	**DXS8378**	**Majorca**	**Minorca**	**Ibiza**	**Valencia**
18	**18**	**11**	1				**24.1**	**31**	**10**				1
18	**19.1**	**12**		1			**25**	**18**	**12**	1			
18	**20**	**10**	1				**25**	**22**	**11**	1			
18	**21**	**11**				1	**25.1**	**16**	**9**				1
18	**23**	**10**		1			**25.1**	**16**	**12**		1		
18	**23**	**11**		1			**25.1**	**18**	**11**				1
18	**23**	**12**		1			**25.1**	**18**	**12**				1
18	**23.1**	**10**		1			**25.1**	**19**	**10**	3			
18	**24**	**12**		1			**25.1**	**19**	**12**				1
18	**25**	**10**				1	**25.1**	**19.1**	**11**	1			
18	**27**	**10**				1	**25.1**	**19.1**	**12**	1			
18	**27**	**12**			1		**25.1**	**20**	**10**	1			1
18	**28**	**11**		1			**25.1**	**20**	**12**	1			
18	**33**	**11**			1		**25.1**	**21**	**12**		1		
19	**25**	**12**		2			**25.1**	**22**	**11**				1
19	**27**	**11**	1				**25.1**	**22**	**12**			1	
19	**27**	**12**	1				**25.1**	**22**	**13**			3	
19	**29**	**11**				1	**25.1**	**22.1**	**13**			1	
21	**25**	**13**			2		**25.1**	**23**	**11**		1	1	
22	**19**	**11**				1	**25.1**	**23**	**12**				1
22	**29**	**10**		1			**25.1**	**23.1**	**10**		1		
22.1	**19**	**12**		1			**25.1**	**24**	**11**			1	
22.1	**23**	**11**			1		**25.1**	**24**	**12**				1
22.1	**25**	**11**	1	2			**25.1**	**25**	**10**	1			
22.1	**26**	**12**			1		**25.1**	**25**	**11**			1	
22.1	**27**	**12**		1			**25.1**	**26**	**12**				1
22.1	**28**	**12**	1				**25.1**	**26**	**13**				1
23	**21**	**10**	1				**25.1**	**27**	**11**				1
23	**21.1**	**10**		1			**25.1**	**27**	**12**				1
23	**22**	**10**	2				**25.1**	**28**	**11**			1	
23	**24**	**11**		1			**25.1**	**28**	**12**		1		
23	**25**	**11.2**			1		**25.1**	**29**	**11**	1			
23	**25**	**12**			1		**25.1**	**30**	**11**				1
23	**27**	**11**	1				**25.1**	**30**	**14**			2	
23.1	**18**	**11**				1	**25.1**	**31**	**11**	1			
23.1	**18**	**12**			2		**26.1**	**17**	**12**				1
23.1	**20**	**12**	1	1			**26.1**	**18**	**10**				1
23.1	**20.1**	**12**	1				**26.1**	**19**	**10**		1		
23.1	**21**	**9**	1				**26.1**	**19**	**12**	1			1
23.1	**21.1**	**10**				1	**26.1**	**19.1**	**10**	1			
23.1	**22**	**11**		1			**26.1**	**21**	**10**	1			
23.1	**22**	**12**				1	**26.1**	**22**	**12**		1		
23.1	**30**	**11**	1				**26.1**	**24**	**13**		1		
24	**18.1**	**12**			1		**26.1**	**25**	**12**			1	
24	**22**	**12**				1	**26.1**	**27**	**13**		1		
24	**24**	**12**		1			**26.1**	**28.1**	**11**	1			
24.1	**17**	**12**		1			**26.1**	**29**	**10**			1	
24.1	**17.1**	**12**		1			**26.1**	**30**	**13**	1			
24.1	**18**	**10**				1	**26.1**	**32**	**12**			1	
24.1	**18**	**11**	1	1			**27.1**	**18**	**10**			1	
24.1	**18.1**	**12**				1	**27.1**	**18**	**12**			2	
24.1	**19**	**11**		1			**27.1**	**19**	**10**				1
24.1	**19**	**12**		1			**27.1**	**21**	**10**	1		2	
24.1	**21**	**13**				1	**27.1**	**21**	**12**		1		
24.1	**22**	**10**			1		**27.1**	**25**	**12**		1		
24.1	**22**	**11**	1				**27.1**	**28**	**11**		1		
24.1	**22.1**	**10**		1			**27.1**	**32**	**12**		1		
24.1	**24**	**9**	1				**28.1**	**20**	**9**				1
24.1	**24**	**13**				1	**28.1**	**23**	**11**		1		1
24.1	**25**	**10**			3	1	**28.1**	**24**	**12**	1			
24.1	**25**	**11**				1	**28.1**	**25**	**11**	1			
24.1	**25**	**12**	1				**28.1**	**27**	**10**			1	
24.1	**26**	**10**			1		**28.1**	**27**	**12**	1			
24.1	**26**	**12**			1		**28.1**	**32**	**10**			1	
24.1	**27**	**11**			1	1	**28.1**	**32**	**12**	1			
24.1	**27**	**12**			1	1	**29.1**	**19**	**12**	1			
24.1	**28**	**12**				1	**30.1**	**28**	**11**			1	
24.1	**29**	**11**				1							

Haplotype diversity ± standard desviation	**Majorca**	**Minorca**	**Ibiza**	**Valencia**
	0.9951 ± 0.0072	0.9973 ± 0.0066	0.9866 ± 0.0086	1.0000 ± 0.0058

Linkage group 2													
DXS7132	DXS10079	DXS10074	Majorca	Minorca	Ibiza	Valencia	DXS7132	DXS10079	DXS10074	Majorca	Minorca	Ibiza	Valencia
11	**24**	**16**		1			**14**	**18**	**15**				1
12	**17**	**19**		1			**14**	**18**	**17**	3			
12	**18**	**8**			2		**14**	**18**	**18**			1	
12	**18**	**14**			1		**14**	**19**	**8**	2		2	
12	**18**	**16**		1			**14**	**19**	**16**	1			
12	**19**	**8**	1				**14**	**19**	**17**	2		2	
12	**19**	**15**				1	**14**	**19**	**18**			2	
12	**19**	**16**		1	1		**14**	**19**	**19**				1
12	**19**	**17**	1	2			**14**	**20**	**8**	2		1	
12	**20**	**7**		1			**14**	**20**	**9**				1
12	**20**	**17**	1				**14**	**20**	**13**				1
12	**20**	**18**		1			**14**	**20**	**15**	1	1		1
12	**21**	**8**			1		**14**	**20**	**16**			1	
12	**21**	**18**			1		**14**	**20**	**17**	1	1		2
12	**21**	**19**				1	**14**	**20**	**19**			1	
13	**15**	**15**	1				**14**	**21**	**15**	1			
13	**15**	**18**			1		**14**	**21**	**18**	1		3	1
13	**16**	**16**	1				**14**	**22**	**17**	1			
13	**17**	**16**			1		**14**	**22**	**17.2**				1
13	**17**	**18**				1	**15**	**16**	**8**		1		
13	**18**	**8**		1	2		**15**	**16**	**17**			3	
13	**18**	**9**			1		**15**	**17**	**16**	1			
13	**18**	**15**				1	**15**	**17**	**17**			1	
13	**18**	**16**				1	**15**	**18**	**8**		1	1	
* **13**	**18**	**18**	1	4	2	1	**15**	**18**	**14**		1		
13	**18**	**19**	1	1	1	1	**15**	**18**	**16**	1			
13	**19**	**8**	1	2		1	**15**	**19**	**8**		2		
13	**19**	**15**				1	**15**	**19**	**15**				1
13	**19**	**16**	1	1		1	**15**	**19**	**17**	1			
13	**19**	**17**	1			1	**15**	**20**	**14**	1			
13	**19**	**18**			1		**15**	**20**	**15**		1		1
13	**20**	**7**		1		1	**15**	**20**	**16**			1	
13	**20**	**8**	1				**15**	**20**	**17**		1		
13	**20**	**17**		1		1	**15**	**20**	**17.2**				1
13	**20**	**18**	1			1	**15**	**21**	**15**	1			
13	**20**	**19**			1		**15**	**21**	**16**				2
13	**21**	**16**				1	**15**	**21**	**17**		1		
13	**21**	**17**			1	1	**15**	**21**	**18**		2		
13	**21**	**18**		1			**15**	**22**	**15**				1
13	**21**	**19**		1	2		**15**	**22**	**16**			1	1
13	**23**	**13**	1				**15**	**23**	**15**				1
14	**16**	**17**		1			**15**	**23**	**18**	1			
14	**16**	**18**	1				**16**	**16**	**8**		1		
14	**17**	**8**	1	2		1	**16**	**18**	**8**		1		
14	**17**	**9**				2	**16**	**19**	**8**				1
14	**17**	**15**	1			1	**16**	**19**	**17**		1		
14	**17**	**17**			1		**16**	**19**	**19**	1			
14	**18**	**7**			1		**16**	**20**	**15**	1			
14	**18**	**8**	2				**17**	**21**	**8**				1

Haplotype diversity ± standard desviation	**Majorca**	**Minorca**	**Ibiza**	**Valencia**
	0.9915 ± 0.0077	0.9852 ± 0.0104	0.9841 ± 0.0087	0.9960 ± 0.0069

Linkage group 3													
DXS10103	**HPRTB**	**DXS10101**	**Majorca**	**Minorca**	**Ibiza**	**Valencia**	**DXS10103**	**HPRTB**	**DXS10101**	**Majorca**	**Minorca**	**Ibiza**	**Valencia**
15	**12**	**33**			1		**19**	**11**	**28.2**	1			
15	**13**	**33**				1	**19**	**11**	**29.2**			1	2
16	**10**	**28**	1				* **19**	**11**	**30.2**	2	3	1	2
16	**12**	**25.2**				1	**19**	**11**	**31.2**	1			
16	**12**	**27.2**				1	**19**	**12**	**27.2**		1	2	
16	**12**	**28.2**		1			**19**	**12**	**28**	1		2	
16	**12**	**29.2**				1	**19**	**12**	**28.2**	1		1	
16	**12**	**30**	1				**19**	**12**	**29**	1			
16	**12**	**32**				1	**19**	**12**	**29.2**	2		1	1
16	**13**	**29**				1	* **19**	**12**	**30.2**	4	1	2	1
16	**13**	**29.2**		1			**19**	**12**	**31.2**	1	1		1
16	**13**	**33**				1	**19**	**13**	**27.2**	1			
16	**14**	**33**			1		**19**	**13**	**29**				1
16	**15**	**32**		1			**19**	**13**	**29.2**			3	
17	**12**	**31**				1	**19**	**13**	**30.2**		2		
17	**12**	**33**		1			**19**	**13**	**31**			1	1
17	**13**	**27**		1			**19**	**13**	**31.2**				1
17	**13**	**30**			1		**19**	**13**	**32**			1	
17	**13**	**31**	2	1		1	**19**	**13**	**32.2**	1			
17	**14**	**30**		1			**19**	**13**	**33**			1	
17	**14**	**31**		1		1	**19**	**13**	**34**			2	
18	**9**	**30.2**		1			**19**	**14**	**28.2**	1	2		
18	**11**	**28.2**	1				**19**	**14**	**29.2**		2		
18	**11**	**29.2**	1				**19**	**14**	**31**	1			
18	**11**	**32.2**			1		**19**	**14**	**31.2**				1
18	**12**	**27.2**				3	**19**	**14**	**33**				1
18	**12**	**28.2**			1		**19**	**15**	**28**	1			
18	**12**	**29.2**	1	1			**19**	**15**	**29.2**	1			
18	**12**	**30**				1	**19**	**15**	**32**				1
18	**12**	**30.2**		1	1		**20**	**11**	**28.2**				1
18	**12**	**33.2**				1	**20**	**12**	**28**			2	
18	**13**	**28.2**				1	**20**	**12**	**28.2**	1		1	
18	**13**	**29**		1			**20**	**12**	**29.2**	2	2	1	1
18	**13**	**29.2**			1		**20**	**12**	**30.2**				2
18	**13**	**30**				1	**20**	**12**	**31.2**	1			
18	**13**	**30.2**		1			**20**	**13**	**28.2**	2			
18	**13**	**31**	1		3		**20**	**13**	**29.2**	1	2		1
18	**13**	**32**	1	1			**20**	**13**	**30.2**		1	2	
18	**13**	**32.2**		1			**20**	**13**	**31.2**		2	2	
18	**13**	**33**				1	**20**	**14**	**29.2**	1			
18	**14**	**29.2**			2		**20**	**14**	**30.2**	1			
18	**14**	**31**		1			**20**	**16**	**29.1**		1		
18	**14**	**31.2**	1				**21**	**10**	**29.2**				1
18	**15**	**30**	1		1	1	**21**	**12**	**29.2**		1		
18	**15**	**31**			1		**21**	**12**	**31.2**		1		
18	**15**	**32**		1			**21**	**13**	**29**				1
19	**11**	**27.2**	1		1								

Haplotype diversity ± standard desviation	**Majorca**	**Minorca**	**Ibiza**	**Valencia**
	0.9866 ± 0.0095	0.9879 ± 0.0087	0.9829 ± 0.0088	0.9919 ± 0.0082

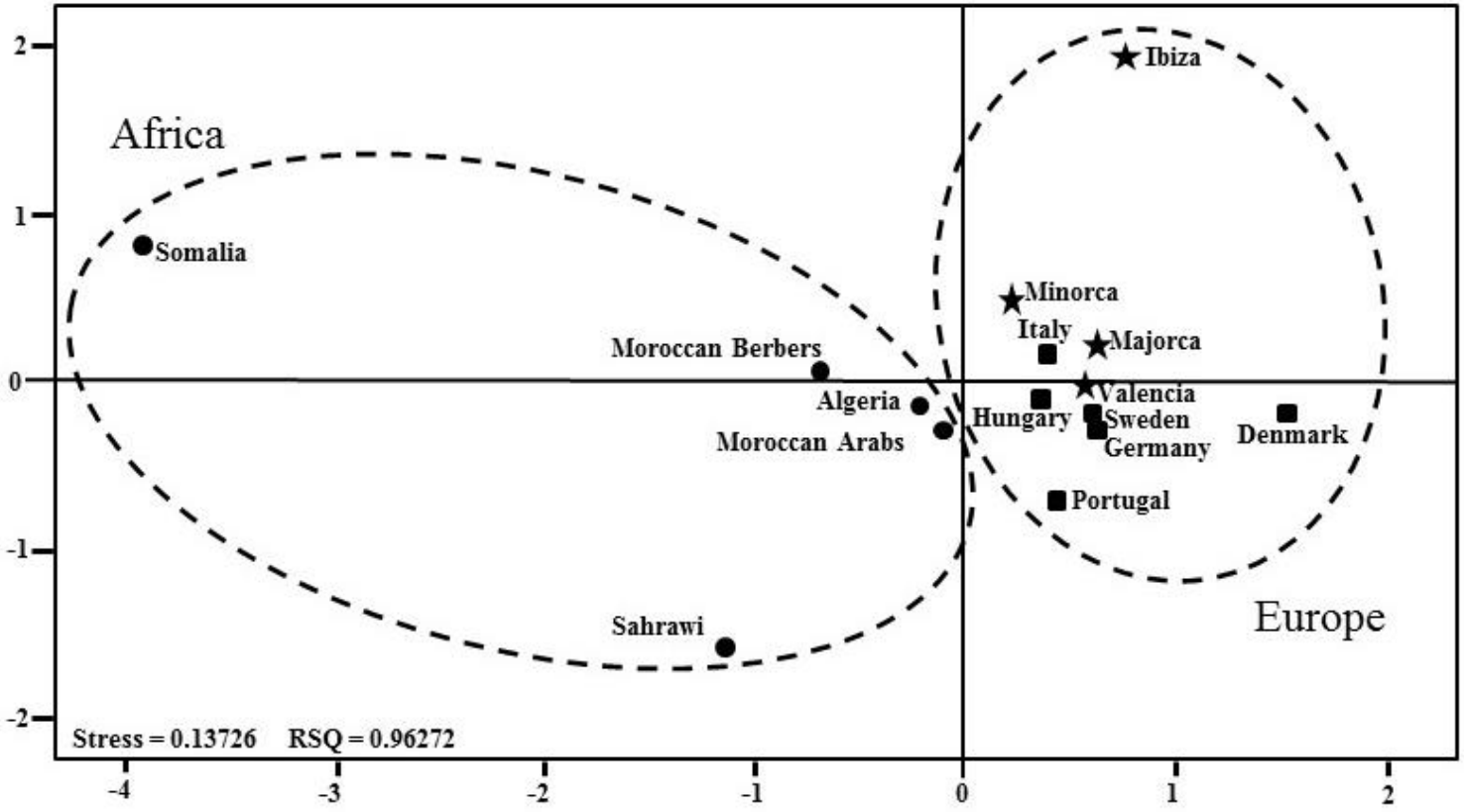
Linkage group 4													
DXS10146	**DXS10134**	**DXS7423**	**Mallorca**	**Menorca**	**Eivissa**	**València**	**DXS10146**	**DXS10134**	**DXS7423**	**Mallorca**	**Menorca**	**Eivissa**	**València**
24	**37**	**14**		1			**29**	**34**	**15**				1
24	**38**	**14**				1	**29**	**35**	**13**		1		
25	**34**	**15**				1	**29**	**35**	**14**	1			
25	**35**	**15**		1			**29**	**35**	**15**			2	
25	**36**	**13**			2	1	**29**	**35**	**16**	1			1
25	**36**	**14**	1				**29**	**35**	**17**		1		
25	**36**	**15**				2	**29**	**36**	**14**				2
25	**37**	**16**			1		**29**	**36**	**15**	1	2	1	3
26	**34**	**14**			1	1	**29**	**36.2**	**14**		1		
26	**35**	**14**	2	1		1	**29**	**37**	**14**	1			1
26	**35**	**15**				1	**29**	**37**	**16**	1			
26	**35**	**16**			1		**29**	**37**	**17**			1	
26	**36**	**15**	1				**29**	**37.3**	**16**		1		
26	**36**	**16**			2		**29**	**38**	**14**		1	2	
26	**36**	**17**		1			**29**	**38**	**16**		2		
26	**37**	**13**				1	**30**	**35**	**15**		1		1
26	**37**	**15**	1		1		**30**	**35**	**17**				1
26	**38**	**14**	1				**30**	**36**	**15**		1	1	
26	**39**	**14**			1		**30**	**36**	**16**			1	
26	**39.3**	**14**				1	**30**	**37**	**14**		1	1	
26	**40**	**14**	1				**30**	**38**	**16**	2			
27	**34**	**14**			2	1	**31**	**33**	**14**	1			
27	**34**	**15**				1	**31**	**34**	**16**	2			
27	**35**	**14**			1		**31**	**35**	**15**	1			
27	**35**	**15**		4		1	**31**	**36**	**15**	1			
27	**36**	**15**		2	1		**31**	**37**	**15**	1			
27	**36**	**16**		1	2		**32**	**40**	**16**		1		
27	**37**	**14**	1			1	**36.2**	**42.3**	**14**		1		
27	**38**	**13**	1				**39.2**	**39.3**	**15**				2
27	**38**	**14**		1		1	**39.2**	**40.3**	**14**	1			
27	**38**	**15**	1			1	**40.2**	**33**	**16**				1
27	**42.3**	**15**				1	**40.2**	**36**	**16**				1
28	**33**	**15**	1				**40.2**	**42.3**	**15**	1			
28	**34**	**14**				1	**41.2**	**36**	**15**		1		
28	**34**	**15**			1		**41.2**	**36**	**17**		1		
28	**35**	**14**			1		**41.2**	**38.3**	**15**			2	
28	**35**	**15**	1	2	1	1	**41.2**	**39.3**	**15**			1	
28	**35**	**16**	1				**41.2**	**40.3**	**15**	1			
28	**35.3**	**14**				1	**41.2**	**41.3**	**15**	1			
28	**36**	**14**			3	1	**41.2**	**41.3**	**16**		1		
28	**36**	**15**	1	1	2		**41.2**	**42.3**	**17**	1			
28	**36**	**16**	1				**42.2**	**35**	**15**				1
28	**37**	**13**		1			**42.2**	**36**	**14**	1		1	
28	**37**	**15**	1		1		**43.2**	**36**	**16**			1	
28	**37**	**16**				1	**43.2**	**39.3**	**14**	1			
28	**40.3**	**14**			1		**44.2**	**35**	**15**	1			
28	**40.3**	**15**	1	1			**44.2**	**41.3**	**15**	1			
28	**41.3**	**14**		1			**45.2**	**37**	**14**			1	
28	**41.3**	**15**	1				**45.2**	**41.3**	**14**		1		
28	**43.3**	**15**	1			1	**46.2**	**32**	**15**		1		
29	**32**	**14**				1	**47.2**	**32**	**15**			1	
29	**32**	**15**		2									

Haplotype diversity ± standard desviation	**Majorca**	**Minorca**	**Ibiza**	**Valencia**
	0.9963 ± 0.0063	0.9852 ± 0.0104	0.9866 ± 0.0081	0.9919 ± 0.0082

Supplementary Table 3. Linkage disequilibrium inside the 4 Linkage Groups (LG) in Majorca, Minorca, Ibiza and Valencia populations. In bold, significant p-values after Bonferroni correction.

pair locus	Majorca	Minorca	Ibiza	Balearic Islands (pooled)	Valencia
(DXS10148, DXS10135)	0.1890	0.1492	**0.0000**	**0.0000**	0.9338
(DXS10148, DXS8378)	0.5582	0.7408	0.0029	0.2295	0.8233
(DXS10135, DXS8378)	0.2457	0.6266	0.0049	0.1099	0.6107
(DXS7132, DXS10079)	0.9933	0.0863	0.0413	0.3657	0.1456
(DXS7132, DXS100747)	0.1752	0.4016	0.0955	0.0392	0.1567
(DXS10079, DXS10074)	0.2440	0.0069	0.0067	**0.0000**	0.6035
(DXS10103, HPRTB)	0.1558	0.2716	0.1435	0.0271	0.4451
(DXS10103, DXS10101)	0.2360	0.0113	0.0606	**0.0000**	0.0079
(HPRTB, DXS10101)	0.4058	0.1254	0.0048	**0.0000**	0.0504
(DXS10146, DXS10134)	0.3456	0.4500	0.4618	**0.0000**	0.0156
(DXS10146, DXS7423)	0.0248	0.3892	0.0401	0.1418	0.0965
(DXS10134, DXS7423)	0.7415	**0.0000**	0.7737	0.7527	0.6525

Supplementary Figure 1. Multidimensional scaling analysis (MDS) based on Reynolds genetic distances calculated between populations. African populations are represented by circular points, European populations by square dots and Balearic and Valencia populations are labelled with stars.



Moroccan Berbers, Moroccan Arabs and Sahrawi (Bentayebi K et al. 2011. Forensic Sci Int Genet 6: e48-49); Denmark and Somalia (Tomàs C et al. 2012. Int J Legal Med 126:121-128); Sweden (Tillmar AO. 2012. Forensic Sci Int Genet 6: e80-81); Portugal (Cainé L et al. 2013 Int J Legal Med 127:63-64); Hungary (Horváth G et al. 2012. Forensic Sci Int Genet 6: e46-47); Germany (Edelmann J et al.2012. Forensic Sci Int Genet 6:e24-34); Algeria (Bekada A et al. 2010. Int J Legal Med 124:287-294); Italy (Inturri S et al. 2011. Forensic Sci Int Genet 5:e152-154).

Genetic diversity of 12 X-chromosomal short tandem repeats in Jewish populations

J.F. Ferragut[1], J.A. Castro[1], C. Ramon[1] and A. Picornell[1]

[1]Institut Universitari d'Investigacions en Ciències de la Salut (IUNICS) and Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears. Palma de Mallorca, Spain

Corresponding author: Antonia Picornell (apicornell@uib.es). Cra. Valldemossa Km 7,5. 07122-Palma (Balearic Islands, SPAIN). Telephone number: +34-971172050. Fax number: +34-971173184.

Abstract: Haplotype and allele frequencies of 12 X-STRs included in the Investigator Argus X-12 kit were studied for 313 individuals, representing five populations with known Jewish ancestry. The high efficiency in forensic parameters in all the populations studied demonstrates that this set of markers provides a powerful tool for solving complex kinship cases in Jewish populations.

Keywords: X-STRs; X-Chromosome; Investigator Argus X-12 kit; Jews; Chuetas.

1. Introduction

X-Short Tandem Repeat (X-STR) markers have a series of special characteristics that justify their increasing interest in forensic practice, population genetics, and anthropology. Compared with autosomes, the X-chromosome has lower recombination and mutation rates and a smaller effective population size, resulting in faster genetic drift. Consequently, both linkage disequilibrium (LD) and population structure in the X-chromosome are expected to be stronger than in autosomes. X-STRs are particularly useful in paternity testing and kinship analysis, especially in deficient cases, such as grandmother-granddaughter, aunt-niece and cousins. Owing to the complexity of kinship investigation, a new approach of substituting single STRs for stable haplotypes of closely linked loci has been suggested.

Jews can be traced back to populations occupying a small geographic area, in the Middle East, several thousand years ago and have maintained continuous cultural and religious traditions despite a series of Diasporas. Contemporary Jews comprise several communities that can be classified according to the location where each community developed. Among others, these include Middle Eastern Jews (Mizrahim) (Iran and Iraq), who have always resided in the Middle East, dating from Babylonian or Persian communities in the fourth to sixth centuries BCE; the Askenazim – the vast majority of living Jews – who lived in communities in central and eastern Europe, but whose origins remain highly contested and enigmatic to this day; the Sephardim ("Spanish" in Hebrew) who, after their expulsion from the Iberian Peninsula in the late 15th century, lived in other Mediterranean countries (especially Bulgaria and Turkey), where they mixed with local Jewish communities formed during classical antiquity; and North African Jews, comprising both Sephardim and Mizrahim, as there exists evidence of Jewish communities in North Africa as early as the first centuries AD that were augmented as a consequence of the Spanish expulsion (Ben-Sasson et al., 1976). Chuetas are an isolated, inbred Spanish community, descending from Majorcan Sephardic Jews. Their peculiar history has kept the memory of their Jewish origin and has prevented their gradual assimilation into the general population (Laub and Laub, 1987).

The present study analyzed 474 X-chromosomes from five populations with Jewish ancestry, aiming to build an X-STR database, based on the markers included in the Investigator Argus X-12 kit (Qiagen, Hilden, Germany), for anthropological and forensic purposes.

2. Material and methods

DNA samples were collected from 313 unrelated individuals (152 men and 161 women), with known ancestors until at least the third generation, from four Jewish populations (Middle Eastern, Ashkenazi, Sephardic and North African) and Chuetas.

The 12 X-STR loci included in the Investigator Argus X-12 Kit were analyzed. PCR amplification for 12X-STR loci was performed using the Investigator Argus X-12 Kit (Qiagen) according to the manufacturer's instructions (Qiagen, 2010). For genetic typing, an ABIPrism 3130 DNA Genetic Analyzer along with GeneMapper ID3.2 software (Applied Biosystems, Foster City, CA) was used.

Forensic parameters were calculated on the Chromosome X homepage (http://www.chrx-str.org). Allele frequencies, exact test of the Hardy–Weinberg equilibrium for the female samples (HWE), pairwise exact test of linkage disequilibrium (LD), and haplotype diversity for the male samples, were carried out using the Arlequin v.3.5 software (Excoffier and Lischer, 2010).

3. Results and Discussion

We analyzed the 12 X-STR loci included in the Investigator Argus X-12 Kit in five populations with Jewish ancestry: Mizrahim, Ashkenazim, Sephardim, North African, and Chuetas, using 474 X-chromosomes.

The highest variability was found in DXS10146 and DXS10135 (29 and 28 alleles, respectively, and observed heterozygosities between 0.8077 and 0.9744), while DXS7423 was the least polymorphic one, with 6 alleles and observed heterozygosities between 0.5000 and 0.8000. No deviations from the Hardy–Weinberg equilibrium were observed after Bonferroni correction. New alleles were described in DXS10134, DXS10079, DXS10148 and DXS10135 markers (Table 1).

Table 1. Characteristics of news alleles found in this study.

Marker	General structure	New allele	Structure of new allele	Genbank accesion number
DXS10134	PrI_{22}-N_8-A_6-T-A_9- $(GAAA)_3$-GAGA-$(GAAA)_4$-AA-(GAAA)-GAGA-$(GAAA)_4$-GAGA-N_{18}-(GAAA)-GTAA-$(GAAA)_3$-AAA-$[(GAAA)_4$-$AAA]_{1-2}$-$(GAAA)_{12-20}$-N_5-$PrII_{19}$ (Edelmann et al., 2008)	44.2	PrI_{22}-N_8-A_6-T-A_9- $(GAAA)_3$-GAGA-$(GAAA)_4$-AA-(GAAA)-GAGA-**$(GAAA)_3$**-GAGA-N_{18}-(GAAA)-GTAA-$(GAAA)_3$-AAA-**$[(GAAA)_4$-$AAA]_3$**-$(GAAA)_{16}$-N_5-$PrII_{19}$	KC581793
DXS10079	$(AGAG)_3$-TGAA-AGAG-$(AGAA)_{10-21}$-AGAG-$(AGAA)_3$ (Hering et al., 2006)	20.1	$(AGAG)_3$-TGAA-AGAG-$(AGAA)_2$-AG**A**AA-$(AGAA)_{13}$-AGAG-$(AGAA)_3$	KC662329
DXS10148	$(GGAA)_4$-$(AAGA)_X$-$(AAAG)_Y$-N_8-$(AAGG)_2$ (Bekada et al., 2010)	22	$(GGAA)_4$-$(AAGA)_{10}$-$(AAAG)_4$-N_8-$(AAGG)_2$	KC662330
DXS10135	$(AAGA)_3$-GAAAG-$(GAAA)_X$-$(GGAA)_Y$-$(GAAA)_Z$ 3' flanking region AGAGAATAGAAAA(GAA/-) GAGA (Sumita and Whittle, 2009)	25.3	$(AAGA)_3$-GAAAG-$(GAAA)_{19}$-$(GGAA)_4$-(GAAA) 3' flanking region AGAGAA**[-----]**AAGAAGAGA	KC737839

Forensic parameters of interest were calculated for each X-STR and population. Overall values obtained for Power of Discrimination were high (>4.30E+08) in both females and males and Power of Exclusion ranged from 1.25E+05 (Mizrahim Jews) to 3.38E+05 (North African Jews). Although values differed slightly between populations, the set of loci in the Argus X-12 kit was highly informative in all the Jewish populations studied.

Table 2. Statistical parameters of haplotypes of 4 X-chromosomal STR trios in 152 Jewish men.

	Linkage group 1	**Linkage group 2**	**Linkage group 3**	**Linkage group 4**
Nº Haplotypes	123	95	93	108
Unique haplotypes	98 (79.80%)	62 (65.26%)	66 (70.96%)	77 (71.29%)
Haplotype Diversity	0.9970	0.9917	0.9881	0.9948
Discrimination Capacity	80.92%	62.50%	61.18%	71.05%
Match probability (%)	0.0030	0.0083	0.0119	0.0052
Frequency of the most common haplotype	0.0263	0.0461	0.0592	0.0263

Lg1: DXS10148-DXS10135-DXS8378; lg 2: DXS7132-DXS10079-DXS10074; lg 3: DXS10103-HPRTB-DXS10101; lg 4: DXS10146-DXS10134-DXS7423

Amongst the 152 males analyzed, the 4 X-STR trios of linkage group (lg) 1–4 revealed 123, 95, 93 and 108 haplotypes, respectively (Table 2). Most of them were only observed once, and the other haplotypes were shared by 2 to 9 men, displaying frequencies <0.059. Match probability ranged from 0.3% (lg 3) to 1.2% (lg 1).

4. Conclusion

In short, the forensic efficiency parameters of the twelve X-STRs investigated in this work demonstrates that this set of markers is highly discriminating and, therefore, provides a powerful tool for solving complex kinship cases in Jewish populations.

This study follows ISFG recommendations (Olaisen et al., 1998) and the guidelines for publication of population data proposed by the journal (Carracedo et al., 2013).

Acknowledgments

This work was partially supported by grant AAEE24/2014 from the Direcció General de R+D+I (Comunitat Autònoma de les Illes Balears) and European regional Development Fund (ERDF).

Conflict of interest statement

The authors declare no conflicts of interest.

Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs

(Submitted)

Ferragut JF[a], Bentayebi K[b], Pereira R[c,d], Castro JA[a], Amorim A[c,d,e], Ramon C[a], and Picornell A[a*]

[a]Institut Universitari d'Investigació en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Carretera de Valldemossa, km 7.5, 07122, Palma de Mallorca, Illes Balears, Spain

[b]Biochemistry and Immunology Laboratory-Forensic Unit, Mohammed V Agdal University, Faculty of Sciences, Avenue Ibn Batouta Agdal, BP 1014 Rabat, Morocco

[c]i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

[d]IPATIMUP - Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Rua Júlio Amaral de Carvalho, 45, 4200-135 Porto, Portugal

[e]Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal

Key words: X-chromosome; Alu insertions; X-STR; X-Indel; Jews; Chuetas; Majorca; Investigator Argus X-12 kit,

[*]**Corresponding author:** Dr. Antònia Picornell. Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Cra. Valldemossa, km 7.5, 07122-Palma de Mallorca, Illes Balears, Spain. Tel: +34-971-172050. Fax: +34-971-173184. E-mail: apicornell@uib.es

Abstract

Population genetic data for 53 X-chromosome markers (32 X-Indels, 9 X-Alu insertions and 12 X-STRs) are reported for five populations with Jewish ancestry (Sephardim, North African Jews, Middle Eastern Jews, Ashkenazim, and Chuetas) and Majorca, as the host population of Chuetas.

Genetic distances between these populations demonstrated significant differences, with the Chuetas as the most differentiated group, in accordance with the particular demographic history of this population. X-chromosome analysis and a comparison with autosomal data suggest a generally sex-biased demographic history in Jewish populations. Asymmetry was found between female and male effective population sizes both in the admixture processes between Jewish communities, and between them and their respective non-Jewish host populations.

Results further show that these X-linked markers are highly informative for forensic purposes, and highlight the need for specific databases for Jewish populations.

1. Introduction

Modern Jews comprise an aggregate of ethno-religious communities that can be traced back to a national and religious group originating several thousand years ago, and maintaining continuous cultural, historical, and religious traditions since that time. Historical evidence suggests a Middle Eastern origin, followed by a series of migrations leading to the establishment of communities of Jews worldwide, in what is termed the Jewish Diaspora. These communities can be classified on the basis of their main regions of residence: (i) the Sephardim who, after their expulsion from the Iberian Peninsula in the late 15th century, migrated to other Mediterranean countries where they mixed with local Jewish communities (Baron, 1937; Goodman, 1979); (ii) the North African Jews, of whom there is evidence in North Africa as early as the first centuries AD. These communities were augmented as a consequence of the Spanish expulsion (Baron, 1937; Chouraqui, 1968; Hirschberg, 1974); (iii) Middle Eastern Jews (Iran and Iraq), who originated from Babylonian or Persian communities in the fourth to sixth centuries BC (Rejwan, 1985; Levy, 1999); and (iv) the Ashkenazim, who have lived since the first millennium of the common era in central and Eastern Europe, but whose origins remain controversial to this day (Das et al., 2016; Flegontov et al., 2016).

Chuetas are a group of descendants of the Jewish population living in Majorca (Balearic Islands, Spain) who, despite their official conversion to Christianity (1391-1435), were discriminated against and isolated from the old-Christian Majorcan population until the middle of the 20th century. Chuetas, together with some Crypto-Jewish communities in Portugal (Martins, 2006), are the only current Iberian populations whose ancestors can be traced back to the original Sephardic Jewish populations, given their peculiar history that

kept the memory of their Jewish origin over the centuries. Unlike what happened with most of the converted Iberian Jews, their inbreeding has restricted their gradual assimilation into the general population (Laub and Laub, 1987).

Many historical and demographic events have shaped the genetic portrait of these groups, such as religious conversion, assimilation, bottlenecks, and intermarriage with different populations as a consequence of their various migrations. This complex demographic history imposes special challenges to better understanding the origins and genetic structure of these groups. For this reason, they have been the focus of genetic studies since the turn of the 20th century. These studies have provided evidence for shared Middle Eastern ancestry between major Jewish Diaspora groups, and variable degrees of admixture with local populations (e.g. Bonné-Tamir and Adam, 1992; Hammer et al., 2000; Thomas et al., 2002; Atzmon et al., 2010; Behar et al., 2010; Ostrer and Skorecki, 2013). Regarding Chuetas, genetic studies have shown that this population presents a significant persistence of Jewish heritage as well as signs of introgression from their non-Jewish host population (e.g. Picornell et al., 1997; 2005).

The choice of the X-chromosome comes from the many features it affords, making it a good source of information for population genetics and anthropology, and an important tool in forensic cases (Szibor et al., 2003; Diegoli, 2015). Compared with autosomes, the X-chromosome has a lower recombination rate, lower mutation rate, and a smaller effective population size, resulting in faster genetic drift. Consequently, both linkage disequilibrium (LD) and population structure in the X-chromosome are expected to be stronger than in the autosomes. On the other hand, since two thirds of X-chromosome history has been spent in females, X-chromosome polymorphisms mainly reflect the history of females (Schaffner, 2004). Finally, in kinship analysis, such as in cases of father-daughter, mother-son, grandmother-granddaughter, or putative sisters testing, X-chromosome markers are an extremely useful source of information (Szibor, 2007).

Population data on X-chromosome markers can be considered very scarce for Jewish groups (Zietkiewicz et al., 2003; Xiao et al., 2004; Lovell et al., 2005; Ferragut et al., 2015b). The present work focuses on the comprehensive analysis of 53 X-chromosomal markers of different types – 32 insertion-deletion polymorphisms (Indels), 9 Alu insertions, and 12 STRs – aiming to evaluate their usefulness in a forensic context and to contribute to refining knowledge regarding the complex system of interrelationships between Jewish communities and their non-Jewish neighbours.

2. Material and methods

2.1. DNA samples

DNA samples from 500 unrelated individuals (276 males and 224 females) were obtained after informed consent: 402 with known Jewish ancestry, and 98 from Majorcan individuals, included in the study as the host population of Chuetas.

Samples from Jewish populations were 281 individuals of the National Laboratory for the Genetics of Israeli Populations at Tel-Aviv University. Following classical criteria, these samples were categorized into four groups: Sephardic (65 Turkish and 44 Bulgarian), North African (35 Moroccan, 13 Libyan, and 12 Tunisian), Middle Eastern (30 Iranian and 27 Iraqi) and Ashkenazi (55). The 121 Chueta individuals and 98 individuals from Majorca (Balearic Islands, Spain) belonged to the collection of the Genetics Laboratory, University of the Balearic Islands.

2.2. Genetic markers and Genotyping

Samples were typed for three sets of X-chromosome genetic markers: (i) 32 X-Indels previously reported by Pereira et al. (2012a); (ii) a set of 9 X-chromosome Alu insertions (Ya5DP62, Yb8DP49, Yd3JX437, Yb8NBC634, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4 and Ya5DP13) described by Callinan et al. (2003); and (iii) 12 X-STRs included in the Investigator Argus X-12 kit (Qiagen GmbH, Hilden, Germany). X-STRs of Majorcan individuals have been published elsewhere (Ferragut et al., 2015a).

Amplification was performed according to the manufacturer's instructions (X-STRs) or using previously described protocols (X-Indels according to Pereira et al. (2012a), and X-Alu insertions as in González-Pérez et al. (2003)).

To genotype Alu sequences, 15 µl of the PCR products were run in 2% 1x TBE agarose gels containing ethidium bromide, and reaction products were directly visualized using ultraviolet fluorescence. For STRs and Indels, PCR products were separated by capillary electrophoresis on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) and analysed with GeneMapper ID v3.2 (Applied Biosystems).

2.3. Statistical analysis

Allele frequencies, exact test of Hardy-Weinberg equilibrium (HWE) for female samples, pairwise exact test of linkage disequilibrium (LD), and haplotype diversity (HD) for male samples were estimated using Arlequin v.3.5.1.2 software (Excoffier and Lischer, 2010). Statistical parameters of forensic interest were computed using Genoproof3 theory manual formulae (Qualitype GMbH, 2014) through the Forensic X-chromosome STR homepage (http://www.chrx-str.org).

In order to examine the relationship between the populations studied and with other published data, pairwise FST genetic distances were calculated with POPTREE2 software (Takezaki et al., 2010), while the Analysis of Molecular Variance (AMOVA) and corresponding non-differentiation p-values were assessed using Arlequin v3.5.1.2. For easier visualization of the observed genetic distances, a multidimensional scaling (MDS) plot of the pairwise FST matrix was represented using SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA).

3. Results and discussion

3.1. Genetic diversity

Allele frequency data for the different types of X-chromosome markers studied in five populations with Jewish ancestry (Sephardic, North African, Middle Eastern, Ashkenazi and Chuetas) and in Majorca are included in Supplementary Tables 1-3.

Values of average gene diversity across loci in the studied populations are summarized in Supplementary Tables 4-6. Most X-Indels showed diversities in the highest range of the possible values for biallelic markers (≤0.5), except MID2637, MID3753, MID3692 and MID3727, with average values below 0.25. Average gene diversity was similar in all populations, ranging between 0.368 and 0.384, in line with other studies (Pinto et al., 2015; Pereira et al. 2012c).

Table 1. Haplotype parameters of the 12 X-STRs studied in 149 males from the populations studied. Sephardic Jews: 21; North African Jews: 29; Middle Eastern Jews: 28; Ashkenazi Jews: 24 and Chuetas: 47.

		LG1	LG2	LG3	LG4	12 X-STR
Number of haplotypes	**SEP**	19	18	18	20	20
	NAJ	29	28	26	28	29
	MEJ	25	22	24	25	28
	ASH	22	21	21	21	24
	CHU	44	34	33	39	47
	Total	124	96	93	108	148
Frequency of unique haplotypes	**SEP**	0.895	0.833	0.889	0.950	0.950
	NAJ	1.000	0.964	0.885	0.964	1.000
	MEJ	0.960	0.818	0.833	0.880	1.000
	ASH	0.909	0.905	0.905	0.905	1.000
	CHU	0.932	0.765	0.818	0.872	1.000
	Total	0.823	0.656	0.699	0.722	0.993
Most common haplotype frequency	**SEP**	0.095	0.095	0.143	0.095	0.095
	NAJ	0.034	0.069	0.069	0.069	0.034
	MEJ	0.143	0.107	0.071	0.071	0.036
	ASH	0.083	0.125	0.125	0.125	0.042
	CHU	0.043	0.106	0.149	0.064	0.021
	Total	0.027	0.040	0.047	0.034	0.013

LG: Linkage Group; SEP: Sephardic Jews; NAJ: North African Jews; MEJ: Middle Eastern Jews; ASH: Ashkenazi Jews and CHU: Chuetas.

Five X-Alu insertions (Ya5DP62, Yb8DP49, Yd3JX437, Yb8NBC634 and Ya5DP13) were revealed to be polymorphic in all populations, whereas the others appeared as monomorphic in at least one studied population (Ya5DP4 insertion in Middle Eastern Jews, Ya5DP77 absence in Sephardic, and Ya5NBC491 and Yb8NBC578 absence in Ashkenazi Jews). Most Alu elements showed moderate to low diversity. Ya5DP62 displayed the highest heterozygosity (0.374) and Ya5NBC491 the lowest (0.063). Average gene diversity in the Jewish populations for this set of X-chromosome Alu insertions ranged from 0.156 to 0.171, in accordance with heterozygosity ranges found in other Mediterranean populations (Callinan et al., 2003; Athanasiadis et al., 2007; Bentayebi et al., 2012).

All X-STRs were highly polymorphic in all populations (average gene diversity ranged from 0.787 in Chuetas to 0.805 in North African Jews). Locus-by-locus analyses revealed that DXS10135 had the highest diversity (with 28 alleles and average heterozygosity of 0.926), whilst DXS10103 and DXS7423 were the least diverse markers (mean heterozygosity of 0.681), as described elsewhere (Ferragut et al., 2015a; Pinto et al., 2015; Pereira et al., 2015).

Table 2. Average Gene diversity and parameters of forensic efficiency over loci in the populations studied.

Markers	Population	N (men/women)	Het	MPD	PE*	MEC*	MEC* duo
X-Indels	**SEP**	56/44	0.368	12.012	3.914E+01	8.286E+04	7.153E+02
	NAJ	29/26	0.372	11.613	4.549E+01	9.284E+04	7.871E+02
	MEJ	28/25	0.384	12.664	5.148E+01	1.307E+05	9.665E+02
	ASH	24/30	0.384	12.664	5.424E+01	1.290E+05	9.726E+02
	CHU	48/24	0.372	11.276	4.021E+01	9.318E+04	7.661E+02
	MAJ	46/33	0.378	12.049	4.713E+01	1.111E+05	8.697E+02
X-Alus	**SEP**	25/39	0.156	1.567	1.306E+00	3.970E+00	2.118E+00
	NAJ	27/25	0.166	1.584	1.381E+00	4.322E+00	2.243E+00
	MEJ	25/25	0.165	1.267	1.408E+00	4.271E+00	2.243E+00
	ASH	25/30	0.141	1.187	1.320E+00	3.457E+00	1.986E+00
	CHU	52/44	0.171	1.747	1.337E+00	4.549E+00	2.278E+00
	MAJ	45/22	0.153	1.426	1.334E+00	3.834E+00	2.095E+00
X-STRs	**SEP**	21/39	0.802	9.729	2.826E+05	2.609E+08	1.237E+06
	NAJ	29/25	0.805	9.810	3.381E+05	3.223E+08	1.439E+06
	MEJ	28/26	0.788	9.550	1.246E+05	9.873E+07	5.453E+05
	ASH	24/30	0.798	9.656	2.332E+05	2.118E+08	1.032E+06
	CHU	47/40	0.787	9.428	1.277E+05	1.045E+08	5.749E+05
	MAJ†	24/39	0.803	9.722	2.522E+05	2.288E+08	1.071E+06
X-All	**SEP**	19/31	0.430	22.474	6.278E+07	1.430E+14	5.623E+09
	NAJ	23/23	0.435	23.399	9.396E+07	2.197E+14	7.696E+09
	MEJ	23/23	0.438	23.529	4.805E+07	9.898E+13	3.908E+09
	ASH	24/30	0.436	23.058	7.281E+07	1.580E+14	5.998E+09
	CHU	28/22	0.427	22.733	4.830E+07	6.299E+13	3.092E+09
	MAJ	24/11	0.436	23.275	1.847E+08	2.502E+14	9.694E+09

* the values indicated in the Table are 1 in X individuals. † data previously reported by Ferragut et al. (2015a). The number of individuals (N) studied for each set of markers varied according to the availability of the samples SEP: Sephardic Jews; NAJ: North African Jews; MEJ: Middle Eastern Jews; ASH: Ashkenazi Jews; CHU: Chuetas and MAJ: Majorca; HET: average gene diversity; MPD: mean number of pairwise differences; PE: power of exclusion; MEC: Mean exclusion chance in trios involving daughter and in duos father/daughter or mother/son (Desmarais et al.,1998. J Forensic Sci 43:1046–1049).

Typing of the 149 males from populations with Jewish ancestry resulted in 148 different haplotypes when all 12 X-STRs were included (Table 1); one haplotype was shared between two Sephardic individuals who are, to our knowledge, unrelated. Linkage groups (LG) 1–4 revealed 124, 96, 93, and 108 haplotypes, respectively (Supplementary Table 7). Out of all the observed haplotypes, 96% showed frequencies <0.030, and the most common haplotype was observed in seven Chueta individuals in LG3, displaying a frequency of 0.047. In these populations, LG1 proved to be the most polymorphic group and LG3 the least variable, similarly to other studies (e.g. Tomàs et al., 2012; Ferragut et al., 2015a). The Sephardic population had the lowest haplotype diversity value (0.995).

Overall, the results obtained with the 53 X-chromosomal markers showed high levels of diversity (Table 2), thereby revealing that no haplotype was shared between males either within or between the studied populations. Although there are no previous works based on these 53 markers, studies using the same set of STRs and Indels (Pereira et al., 2015; Pinto et al., 2015) have also reported high diversity levels.

Haplotype data for all the markers studied are available in Supplementary Table 8 (excel file not included).

3.2. Linkage disequilibrium analysis

Pairwise LD analysis for the 53 X-chromosome markers in male individuals from Jewish populations and Majorca revealed, after Bonferroni correction for multiple tests, significant associations between MID357-MID356 and MID3719-MID2089 pairs of Indels in Sephardic, Majorcan, and Chuetas; and between DXS10103-DXS10101 and DXS10103-HPRTB in STR pairs in Chuetas.

All these associations involve markers located in the same linkage group (for instance MID3690, MID3719, and MID2089 form a closely located cluster, spanning only along ~170 kb), and significant LD has also been reported for the same pairs of STRs or Indels in other studies (Freitas et al., 2010; Pereira et al., 2012c; Pinto et al., 2015; Edelmann et al., 2016). For these reasons, loci showing significant associations must be handled as haplotypes, especially for forensic purposes. For STRs, haplotype frequencies are shown in Supplementary Table 7; and for Indels, Supplementary Table 9 includes the frequencies observed for the two linkage disequilibrium blocks.

3.3. Forensic efficiency

Statistical parameters of forensic interest were calculated for each marker and population (Supplementary Tables 4-6). Table 2 shows gene diversity, mean number of pairwise differences, and some parameters of forensic interest for each set of markers and across the 53-markers set. As expected, considering the different nature of the polymorphisms, biallelic markers were less informative than STRs, especially Alu insertion polymorphisms.

Table 2 demonstrates that the use of this 53 X-loci set was highly efficient in all the populations studied, with values of combined PE >0.9999999, and MEC for duos and trios >0.999999999 in all cases. The combined power of discrimination (PD) in females and in males reached values greater than 1 in 9.007 E+15 in all populations. Therefore, this work provides a highly informative set of X-markers for forensic casework and for studies assessing genetic substructure in populations with Jewish ancestry.

Although values differed slightly between populations, it is interesting to note that Chuetas were the population with the lowest values in all forensic efficiency parameters, in accordance with the fact that Chuetas are a small, inbred population (Picornell et al., 2005).

3.4. Genetic structure of the Jewish populations

AMOVA analysis of the five Jewish populations studied was statistically significant (F_{ST} = 0.01431; p < 1 E-5). Genetic distances (Supplementary Table 10) revealed that Chuetas had the highest values compared with other Jewish populations, whereas the lowest values were observed between Sephardic and North African Jews. P-values were statistically significant in all pairwise comparisons, except between Sephardic and North African Jews. Hierarchical AMOVA confirmed the lack of differentiation between Sephardic and North African Jews (F_{SC} = 0.00252; p = 0.16520).

These results are in accordance with historical data. A considerably significant number of Sephardim were integrated into North African Jewish communities after the Spanish expulsion in the fifteenth century (Stillman, 1991) and, in fact, many studies consider North African Jews part of the current so-called Sephardic population (e.g. Ostrer, 2001; Levy-Coffman, 2005).

Genetic distance between Ashkenazi and Sephardic groups was at the limit of significance (F_{ST} = 0.00397; p = 0.02822). Although gene flow between both groups has been assumed to be restricted (Behar et al., 2010), a growing number of studies show evidence of contact between these two communities (Behar et al., 2006;2010;2013; Nogueiro et al., 2015a), possibly leading to a lower differentiation between these European Jews than between Ashkenazim and other Jewish communities from North Africa or the Middle East.

Regarding Chuetas, although Jewish in origin, genetic drift and admixture with the host population have probably shaped their genetic profile, giving as a result a population that is genetically differentiated from other Jewish groups (Picornell et al., 1997;2005; Ferragut et al., 2016) and also from their host population (F_{ST} = 0.01800; p < 1 E-5).

Interestingly, it was possible to compare genetic distances between the same Jewish populations based on X-chromosomal and autosomal data (Ferragut et al., 2016). Considering a similar number of Indel polymorphisms, pairwise F_{ST} values were approximately three times lower for the autosomes than for the X-chromosome (ratio ~0.3), except in Chuetas, where the ratio was considerably higher (0.5).

Considering that the effective population size of the X-chromosome is 3/4 that of the autosomes, the differences observed between Jewish populations were expected to be 3/4 in autosomes compared to the X-chromosome, if only random genetic drift was responsible for this differentiation. However, the observed ratio is much lower than expected, suggesting a gender-biased effective population size generated by the admixture process over time. Taking into account the history of Jewish communities in the Diaspora, this departure from the expected ratio between autosomal and X-chromosome differentiation could be explained by greater male mobility. This would have been related to the well-known commercial relations between different Jewish populations (e.g. Mea, 2007; Chacón, 2009; favouring genetic flow mediated preferentially by males. Previous studies with uniparental markers have also suggested greater differentiation between populations with Jewish origin in maternal lineages than in the paternal counterpart (Thomas et al., 2002; Picornell et al., 2006). However, this observed pattern may have resulted not only from asymmetrical sex-biased gene flow between Jewish communities and/or between them and their respective host populations, as suggested in Nogueiro (2015), but also from other mutually compatible events, such as gender-differentiated founder effects (Thomas et al., 2002).

3.5. Comparison with other populations

To evaluate the genetic relationship between Jewish populations and previously reported data on other populations (Supplementary Table 11), pairwise F_{ST} genetic distances were computed (Supplementary Table 12) and represented in an MDS plot (Figure 1). Since no other populations were studied with the full set of 53 X-chromosome markers, genetic distances were assessed separately for the three types of polymorphisms. Stress values in all MDS analyses showed greater structure than would be obtained from a random dataset (Sturrock and Rocha, 2000).

Clustering of the different populations in Figure 1 reveals the ability of the three sets of X-markers to clearly discriminate between continents, with European, Asian, and African populations distantly positioned. With the available data for these markers, Amerindian populations appeared closer to Asians, and North African groups were placed between sub-Saharan and European populations, as expected due to their origin. Within the European cluster, Mediterranean isolates like Ibiza, Sicily, and Calabrian populations showed the most distant positions, as in other genetic studies (Rodríguez et al., 2009).

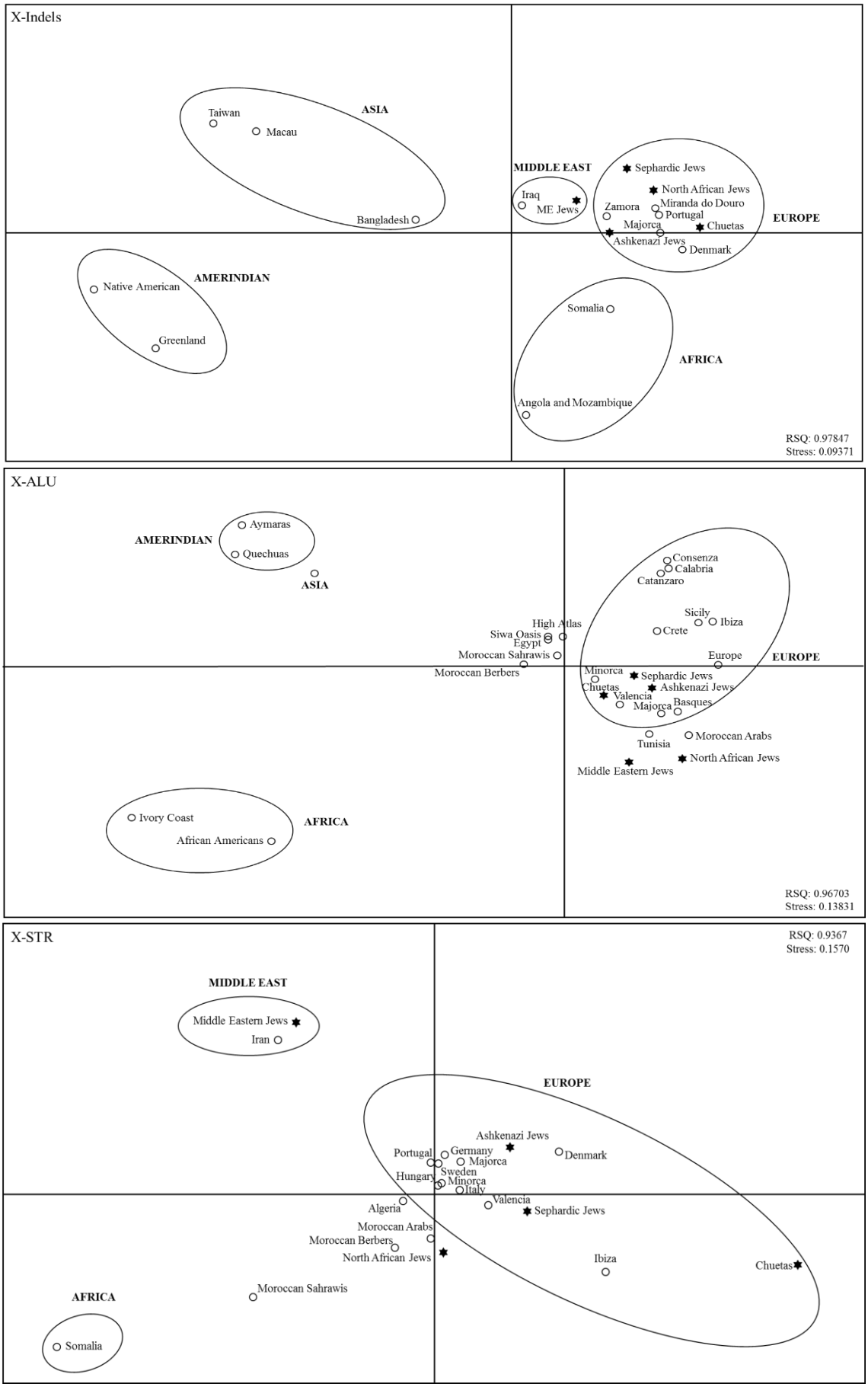


Figure 1. Multidimensional scaling analysis (MDS) based on F_{ST} genetic distances calculated for 32 X-chromosome Indels, 8 X-Alus and 12 X-STRs. Jewish populations are labelled with stars.

Regarding the Jewish populations, they do not form a well-defined cluster in any of the sets of markers in the MDS plots, but they do show up as relatively close to their host population or to other populations in the same geographical region, suggesting admixture with their neighbouring populations. These results mirror those obtained in uniparental markers such as mtDNA, which showed that Jewish pairwise comparisons have noticeably greater F_{ST} values than the Jewish-host comparisons, but not when the Y-chromosome was used (Thomas et al., 2002). Indeed, as two thirds of the X-chromosomes descend from maternal origin, X-chromosome polymorphisms will mostly behave as matrilineal markers, showing similar results to those obtained from mtDNA data. Therefore, the results obtained suggest a larger introgression from females in the respective host non-Jewish populations than from males. Nonetheless, it cannot be ruled out that other events in the complex demographic history of the different Jewish groups could also have contributed to the sex-biased genetic pattern observed.

4. Conclusions

This study provides a useful database of 53 X-chromosome markers for populations with Jewish ancestry, including the Chuetas – a converted Sephardic isolate – and their host population (Majorca). A comparison of differentiation patterns between autosomes and X-chromosome between Jewish groups and with non-Jewish populations suggests a sex-biased admixture process. Genetic flow between Jewish communities would have been mediated primarily by males, with preferential introgression from females in the respective non-Jewish host populations, although other scenarios explaining the genetic portrait of these populations with complex histories cannot be ruled out. The results obtained demonstrate that the use of this set of 53 X-markers is highly efficient for forensic purposes. Genetic distances show that the Jewish populations studied are a heterogeneous group. As such, specific databases for each Jewish community must be used in the forensic field to correctly weigh the value of the evidence based on these X-chromosome markers.

This study follows the guidelines for publication of population data proposed by the journal (Carracedo et al., 2013).

Acknowledgments

This work was partially supported by grant AAEE24/2014 from the Direcció General de R + D + I (Comunitat Autònoma de les Illes Balears) and European Regional Development Fund (ERDF). RP is a recipient of a grant (SFRH/BPD/81986/2011) awarded by the Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Social Fund (Human Potential Thematic Operational Programme— POPH). We are grateful to Maria Trinidad Garcia (from the Serveis Cientificotècnics of the Universitat de les Illes Balears) for her assistance and help with capillary electrophoresis procedures and Meryl Jones for the English language corrections. We would also like to sincerely thank all the Chueta people who volunteered to participate in this study.

Supplementary Table 1. Short Allele frequencies of 32 X-Indel markers in 595 chromosomes (Sephardic Jews: 144; North African Jews: 81; Middle Eastern Jews: 78; Ashkenazi Jews: 84; Chuetas: 96 and Majorcans: 112).

MID	rs number	Sephardic Jews	North African Jews	Middle Eastern Jews	Ashkenazi Jews	Chuetas	Majorca
MID3736	rs56162621	0.7062	0.7500	0.6143	0.6432	0.6528	0.6227
MID3730	rs3215490	0.2992	0.3531	0.2657	0.3167	0.3238	0.2681
MID1361	rs2307557	0.1807	0.2343	0.2819	0.1778	0.3194	0.1489
MID329	rs25553	0.2689	0.2126	0.2476	0.2861	0.1564	0.3063
MID3716	rs63344461	0.6472	0.5031	0.4229	0.6750	0.7153	0.6100
MID3692	rs67929163	0.1234	0.1755	0.0638	0.1306	0.1875	0.1880
MID2637	rs3053615	0.8631	0.9240	0.8443	0.7889	0.9306	0.9278
MID3740	−	0.3382	0.4584	0.3724	0.4389	0.4236	0.4076
MID198	rs16637	0.6926	0.5301	0.6262	0.4889	0.7014	0.5029
MID3703	rs59400186	0.3447	0.2599	0.2552	0.3639	0.3681	0.3799
MID3690	rs60283667	0.3615	0.4058	0.3233	0.3778	0.3819	0.3309
MID3722	−	0.3663	0.3457	0.4986	0.4222	0.4653	0.3076
MID3732	rs11277082	0.1174	0.1857	0.1157	0.2361	0.3065	0.2139
MID3712	rs55877732	0.0860	0.1499	0.2833	0.1639	0.2500	0.2312
MID1736	rs2307932	0.4897	0.4213	0.4867	0.4361	0.3819	0.5000
MID3719	rs3078486	0.2257	0.1702	0.2167	0.1000	0.1597	0.1994
MID2089	rs2308280	0.3377	0.3329	0.3205	0.1389	0.3542	0.3931
MID3774	rs5901519	0.1553	0.3112	0.4243	0.2167	0.3367	0.1634
MID3760	rs66676381	0.7998	0.7913	0.7671	0.7917	0.7566	0.7890
MID3701	rs4030406	0.3588	0.3727	0.4481	0.4111	0.6875	0.4234
MID2612	rs3048996	0.4724	0.5314	0.3248	0.3500	0.4097	0.4943
MID1839	rs2308035	0.8025	0.7259	0.7433	0.8972	0.9097	0.8426
MID3754	rs57608175	0.2137	0.0358	0.2405	0.3083	0.0833	0.2183
MID111	rs16397	0.5438	0.4969	0.5757	0.4556	0.7014	0.4783
MID2652	rs3080039	0.8420	0.7875	0.7938	0.7528	0.7431	0.7009
MID1511	rs2307707	0.7392	0.7995	0.7419	0.6556	0.7279	0.7195
MID2692	rs3047852	0.7468	0.8170	0.8990	0.6312	0.8171	0.6730
MID357	rs25581	0.4015	0.4213	0.3400	0.3417	0.2431	0.1965
MID356	rs25580	0.4803	0.4579	0.3667	0.4761	0.2558	0.2130
MID243	rs16680	0.7668	0.8515	0.8086	0.7720	0.7917	0.7587
MID3727	rs3050111	0.1250	0.1269	0.1500	0.1639	0.0981	0.1243
MID3753	rs72417152	0.0864	0.1627	0.1395	0.0778	0.2036	0.0722

Supplementary Table 2. Short Allele frequencies of 9 X-Alu markers in 569 chromosomes (Sephardic Jews: 103; North African Jews: 77; Middle Eastern Jews: 75; Ashkenazi Jews: 85; Chuetas: 140 and Majorcans: 89).

	Sephardic Jews	North African Jews	Middle Eastern Jews	Ashkenazi Jews	Chuetas	Majorca
Ya5DP62	0.1775	0.2198	0.3067	0.2311	0.2599	0.3313
Yb8DP49	0.2638	0.3501	0.2667	0.2911	0.1707	0.1727
Yd3JX437	0.8566	0.9620	0.9867	0.8822	0.9085	0.9098
Yb8NBC634	0.0230	0.0400	0.0667	0.0378	0.0510	0.0226
Ya5DP77	0.0000	0.0267	0.1600	0.0444	0.1014	0.0226
Ya5_491	0.0427	0.0923	0.0133	0.0000	0.0216	0.0300
Yb8NBC578	0.0652	0.0637	0.0667	0.0000	0.0355	0.0226
Ya5DP4	0.9781	0.9877	1.0000	0.9400	0.9848	0.9774
Ya5DP13	0.1026	0.1037	0.0533	0.0111	0.1725	0.1428

Suplementary Table 3. Allele frequencies of 12 X-STR in 470 chromosomes in the 5 Jewish populations studied (Sephardic Jews: 99; North African Jews: 79; Middle Eastern Jews: 80; Ashkenazi Jews: 84 and Chuetas: 127).

	DXS10079					DXS10101					DXS10103					DXS10134					DXS10146					DXS10148				
Alelle	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU
13.3																										0.0000	0.0133	0.0230	0.0000	0.0000
14	0.0000	0.0000	0.0000	0.0000	0.0083																					0.0085	0.0000	0.0000	0.0000	0.0000
15	0.0256	0.0230	0.0000	0.0111	0.0321						0.0000	0.0000	0.0000	0.0389	0.0072															
16	0.0244	0.0115	0.0115	0.0250	0.0000						0.0256	0.1067	0.0756	0.0806	0.0707						0.0000	0.0000	0.0000	0.0000	0.0083					
17	0.0659	0.0363	0.0615	0.0611	0.0842						0.1270	0.1030	0.0729	0.0694	0.0728											0.0000	0.0000	0.0115	0.0111	0.0000
18	0.1978	0.1393	0.2714	0.1750	0.2197						0.1734	0.1241	0.1715	0.1250	0.2515											0.1233	0.0897	0.0371	0.1250	0.1185
19	0.2808	0.3278	0.2126	0.3194	0.2805						0.4908	0.5251	0.5057	0.5000	0.4949						0.0000	0.0000	0.0000	0.0000	0.0083	0.0672	0.0248	0.0243	0.0361	0.0873
20	0.2454	0.2671	0.3532	0.2417	0.1105						0.1575	0.1163	0.1627	0.1611	0.0957											0.0000	0.0267	0.0000	0.0000	0.0000
20.1	0.0085	0.0000	0.0000	0.0000	0.0000																									
21	0.1087	0.1108	0.0641	0.1222	0.2197	0.0000	0.0000	0.0000	0.0111	0.0000	0.0256	0.0248	0.0000	0.0250	0.0072											0.0000	0.0000	0.0000	0.0111	0.0156
22	0.0342	0.0478	0.0128	0.0333	0.0296																					0.0000	0.0000	0.0128	0.0000	0.0072
22.1																										0.0256	0.0000	0.0345	0.0000	0.0167
23	0.0085	0.0133	0.0128	0.0111	0.0000																0.0000	0.0133	0.0000	0.0000	0.0000	0.0256	0.0115	0.0000	0.0333	0.0406
23.1																										0.0989	0.0860	0.0844	0.0778	0.0250
24	0.0000	0.0000	0.0000	0.0000	0.0083																0.0088	0.0133	0.0000	0.0345	0.0156	0.0171	0.0248	0.0256	0.0111	0.0000
24.1																										0.0085	0.0533	0.0385	0.0917	0.1446
24.2						0.0000	0.0115	0.0000	0.0000	0.0000																				
25						0.0246	0.0000	0.0000	0.0000	0.0000											0.0948	0.0915	0.0128	0.0254	0.0312	0.0000	0.0000	0.0115	0.0000	0.0000
25.1																										0.3150	0.2405	0.1768	0.3194	0.1819
26						0.0088	0.0000	0.0000	0.0111	0.0000											0.2231	0.0630	0.1114	0.0671	0.1924					
26.1																										0.1429	0.1545	0.2586	0.1417	0.1630
26.2						0.0088	0.0000	0.0500	0.0000	0.0000																				
27						0.0159	0.0267	0.0000	0.0333	0.0000											0.1387	0.1848	0.1229	0.1360	0.1558					
27.1																										0.0830	0.1145	0.1857	0.0694	0.1757
27.2						0.0493	0.0248	0.0486	0.0111	0.0392																0.0085	0.0000	0.0000	0.0000	0.0000
28						0.0159	0.0782	0.0128	0.0222	0.1213											0.1266	0.1375	0.1459	0.1566	0.0790					
28.1																										0.0342	0.1241	0.0628	0.0389	0.0000
28.2						0.0827	0.0115	0.1074	0.1472	0.1438																0.0000	0.0115	0.0000	0.0000	0.0000
29						0.0246	0.1067	0.0000	0.0250	0.0250											0.0739	0.1126	0.1472	0.0714	0.1830	0.0000	0.0133	0.0000	0.0000	0.0083
29.1																					0.0088	0.0000	0.0000	0.0000	0.0000	0.0244	0.0115	0.0128	0.0333	0.0083
29.2						0.1316	0.1223	0.2100	0.1944	0.1259																				
30						0.0844	0.0763	0.0486	0.0250	0.0083						0.0000	0.0139	0.0000	0.0000	0.0000	0.0175	0.0860	0.1384	0.1221	0.0217					
30.1																										0.0171	0.0000	0.0000	0.0000	0.0072
30.2						0.1387	0.0726	0.2012	0.1667	0.1638																				
31						0.0860	0.1508	0.0243	0.0861	0.0713						0.0000	0.0000	0.0000	0.0000	0.0329	0.0088	0.0115	0.0999	0.0484	0.0623					
31.2						0.1775	0.1945	0.0858	0.0806	0.2043																				
32						0.0422	0.0745	0.1242	0.0222	0.0463						0.0085	0.0460	0.0486	0.0111	0.0072	0.0000	0.0133	0.0000	0.0115	0.0000					
32.2						0.0739	0.0382	0.0115	0.0861	0.0142											0.0000	0.0000	0.0128	0.0000	0.0000					
33						0.0263	0.0000	0.0756	0.0444	0.0213						0.0427	0.1106	0.2281	0.0583	0.0329	0.0159	0.0000	0.0000	0.0000	0.0406					
33.2						0.0000	0.0000	0.0000	0.0222	0.0000																				
34						0.0088	0.0115	0.0000	0.0000	0.0154						0.1905	0.1547	0.1088	0.1028	0.1678										
34.2						0.0000	0.0000	0.0000	0.0111	0.0000																				
35																0.2320	0.2241	0.2060	0.1667	0.1585										
35.2																					0.0159	0.0000	0.0000	0.0115	0.0000					
36																0.1172	0.1667	0.1127	0.2278	0.2711	0.0000	0.0000	0.0128	0.0000	0.0000					
36.2																					0.0159	0.0000	0.0000	0.0000	0.0000					
36.3																0.0000	0.0345	0.0000	0.0111	0.0000										
37																0.1013	0.0972	0.1229	0.1500	0.0606										

37.2											0.0171	0.0000	0.0000	0.0111	0.0000					
37.3											0.0000	0.0115	0.0000	0.0000	0.0671					
38											0.1062	0.0000	0.0500	0.0667	0.0572					
38.2																0.0000	0.0133	0.0000	0.0000	0.0000
38.3											0.0171	0.0000	0.0000	0.0472	0.0243					
39											0.0403	0.0000	0.0000	0.0250	0.0243					
39.2																0.0422	0.0648	0.0115	0.0623	0.0239
39.3											0.0171	0.0000	0.0000	0.0361	0.0461					
40											0.0000	0.0230	0.0115	0.0000	0.0072					
40.2																0.0175	0.0115	0.0385	0.0230	0.0072
40.3											0.0256	0.0278	0.0628	0.0250	0.0000					
41.2																0.0334	0.0382	0.0000	0.0484	0.0167
41.3											0.0171	0.0508	0.0243	0.0000	0.0256					
42.2																0.0263	0.0267	0.0128	0.0345	0.0957
42.3											0.0427	0.0254	0.0000	0.0472	0.0171					
43.2																0.0652	0.0230	0.0243	0.0369	0.0083
43.3											0.0085	0.0000	0.0000	0.0000	0.0000					
44.2											0.0159	0.0139	0.0000	0.0000	0.0000	0.0510	0.0726	0.0716	0.0254	0.0500
44.3											0.0000	0.0000	0.0243	0.0000	0.0000					
45.2																0.0159	0.0000	0.0243	0.0460	0.0000
46.2																0.0000	0.0115	0.0128	0.0254	0.0000
47.2																0.0000	0.0000	0.0000	0.0139	0.0000
18-19	0.0000	0.0115	0.0000	0.0000	0.0071		0.0000	0.0000	0.0115	0.0000	0.0000									
20-21	0.0000	0.0115	0.0000	0.0000	0.0000															
30-31																0.0000	0.0115	0.0000	0.0000	0.0000
35-36											0.0000	0.0000	0.0000	0.0139	0.0000					

	DXS7132					DXS7423					DXS8378					DXS10074					DXS10135					HPRTB				
Alelle	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU	SEP	NAJ	MEJ	ASH	CHU
7																0.0244	0.0000	0.0243	0.0111	0.0142										
8											0.0085	0.0000	0.0000	0.0000	0.0083	0.1758	0.2253	0.1561	0.0861	0.0975										
9											0.0171	0.0133	0.0000	0.0639	0.0728	0.0647	0.0630	0.0371	0.0139	0.0463										
10											0.3248	0.3416	0.3585	0.3056	0.2754	0.0000	0.0000	0.0128	0.0111	0.0000						0.0342	0.0115	0.1512	0.0000	0.0071
11	0.0085	0.0000	0.0128	0.0000	0.0083						0.2991	0.3398	0.3625	0.4111	0.3348											0.1490	0.0860	0.1357	0.1111	0.1129
12	0.1184	0.0267	0.0959	0.1583	0.0873						0.3077	0.2938	0.2330	0.1944	0.2920											0.3260	0.2349	0.3559	0.3861	0.2527
13	0.1832	0.2768	0.2971	0.2306	0.3449	0.0342	0.0782	0.0000	0.0111	0.0642	0.0256	0.0115	0.0115	0.0250	0.0167	0.0244	0.0000	0.0000	0.0000	0.0000						0.3993	0.3986	0.1331	0.3306	0.4777
14	0.3858	0.4064	0.3842	0.3361	0.4554	0.3065	0.2538	0.4275	0.3444	0.2684	0.0171	0.0000	0.0345	0.0000	0.0000	0.0256	0.0248	0.0000	0.0139	0.0713						0.0659	0.1563	0.1742	0.1472	0.0963
15	0.2393	0.1871	0.1600	0.2083	0.0790	0.4029	0.3264	0.4226	0.4389	0.4752						0.0330	0.1260	0.0230	0.0806	0.1200	0.0000	0.0000	0.0000	0.0000	0.0071	0.0256	0.0993	0.0385	0.0250	0.0379
16	0.0647	0.0897	0.0500	0.0444	0.0250	0.2222	0.2290	0.1114	0.1472	0.1780						0.1758	0.1678	0.1857	0.2111	0.1163	0.0000	0.0000	0.0000	0.0000	0.0238	0.0000	0.0133	0.0115	0.0000	0.0154
17	0.0000	0.0133	0.0000	0.0222	0.0000	0.0342	0.1126	0.0385	0.0472	0.0142						0.2271	0.1582	0.2882	0.1944	0.2151	0.0085	0.0400	0.0000	0.0000	0.0225					
17.1																					0.0085	0.0000	0.0000	0.0000	0.0000					
18						0.0000	0.0000	0.0000	0.0111	0.0000						0.1258	0.1508	0.1574	0.1972	0.2493	0.0342	0.0115	0.0500	0.0000	0.0142					
18.1																					0.0085	0.0115	0.0000	0.0000	0.0000					
19																0.1074	0.0708	0.1026	0.1694	0.0225	0.0427	0.1030	0.0871	0.1250	0.0771					
19.1																					0.0000	0.0133	0.0385	0.0250	0.0142					
20																0.0159	0.0000	0.0128	0.0111	0.0475	0.0733	0.0745	0.1088	0.0361	0.0700					
20.1																					0.0342	0.0133	0.0243	0.0111	0.0571					
21																0.0000	0.0133	0.0000	0.0000	0.0000	0.0403	0.0612	0.0486	0.0694	0.0071					
21.1																					0.0000	0.0533	0.0128	0.0361	0.0000					
21.2																					0.0085	0.0000	0.0000	0.0000	0.0000					
22																					0.0830	0.0230	0.0871	0.0611	0.1117					
22.1																					0.0000	0.0382	0.0000	0.0111	0.0142					
23																					0.0769	0.0878	0.0615	0.0972	0.0500					
23.1																					0.0085	0.0133	0.0000	0.0000	0.0000					
24																					0.0745	0.0515	0.0371	0.0750	0.1046					
25																					0.1331	0.0860	0.0973	0.1056	0.1046					
25.3																					0.0000	0.0000	0.0000	0.0139	0.0000					
26																					0.0989	0.0400	0.0385	0.0528	0.1034					
27																					0.0745	0.0975	0.0959	0.0444	0.0321					
28																					0.1160	0.0533	0.0897	0.0750	0.0546					
29																					0.0342	0.0897	0.0473	0.1000	0.0617					
30																					0.0171	0.0115	0.0128	0.0500	0.0463					
31																					0.0000	0.0267	0.0500	0.0111	0.0083					
32																					0.0244	0.0000	0.0128	0.0000	0.0000					
33.1																					0.0000	0.0000	0.0000	0.0000	0.0154					

SEP: Sephardic Jews; NAJ: North African Jews; MEJ: Middle Eastern Jews; ASH: Ashkenazi Jews and CHU: Chuetas.

Supplementary Table 4. Forensic parameters of 32 X-Indel markers in 595 chromosomes (Sephardic Jews: 144; North African Jews: 81; Middle Eastern Jews: 78; Ashkenazi Jews: 84; Chuetas: 96 and Majorcans: 112).

Loci	Population	Satistical parameters			Power of discrimination		Mean paternity exclusion change		
		HET	PE	PI	PD female	PD male	MEC Krüger	MEC Desmarais	MEC Desmarais duo
MID3736	Sephardic Jews	0.4150	0.1233	0.8546	0.5716	0.4150	0.1644	0.3289	0.2075
	North African Jews	0.3750	0.0994	0.8000	0.5391	0.3750	0.1523	0.3047	0.1875
	Middle Eastern Jews	0.4739	0.1656	0.9503	0.6109	0.4739	0.1808	0.3616	0.2369
	Ashkenazi Jews	0.4590	0.1541	0.9242	0.6020	0.4590	0.1768	0.3537	0.2295
	Chuetas	0.4533	0.1498	0.9146	0.5984	0.4533	0.1753	0.3506	0.2267
	Majorca	0.4699	0.1625	0.9432	0.6086	0.4699	0.1797	0.3595	0.2349
MID3730	Sephardic Jews	0.4194	0.1261	0.8611	0.5749	0.4194	0.1657	0.3314	0.2097
	North African Jews	0.4568	0.1524	0.9205	0.6006	0.4568	0.1762	0.3525	0.2284
	Middle Eastern Jews	0.3902	0.1081	0.8200	0.5520	0.3902	0.1570	0.3141	0.1951
	Ashkenazi Jews	0.4328	0.1352	0.8815	0.5846	0.4328	0.1696	0.3391	0.2164
	Chuetas	0.4379	0.1387	0.8895	0.5882	0.4379	0.1710	0.3420	0.2190
	Majorca	0.3924	0.1094	0.8230	0.5539	0.3924	0.1577	0.3154	0.1962
MID1361	Sephardic Jews	0.2961	0.0619	0.7103	0.4607	0.2961	0.1261	0.2523	0.1480
	North African Jews	0.3588	0.0908	0.7798	0.5245	0.3588	0.1472	0.2944	0.1794
	Middle Eastern Jews	0.4049	0.1169	0.8401	0.5639	0.4049	0.1615	0.3229	0.2024
	Ashkenazi Jews	0.2924	0.0605	0.7066	0.4565	0.2924	0.1248	0.2496	0.1462
	Chuetas	0.4348	0.1365	0.8846	0.5860	0.4348	0.1701	0.3403	0.2174
	Majorca	0.2535	0.0461	0.6698	0.4106	0.2535	0.1107	0.2213	0.1267
MID329	Sephardic Jews	0.3932	0.1098	0.8240	0.5545	0.3932	0.1579	0.3159	0.1966
	North African Jews	0.3348	0.0789	0.7517	0.5015	0.3348	0.1394	0.2788	0.1674
	Middle Eastern Jews	0.3726	0.0981	0.7969	0.5369	0.3726	0.1516	0.3032	0.1863
	Ashkenazi Jews	0.4085	0.1192	0.8453	0.5667	0.4085	0.1625	0.3251	0.2042
	Chuetas	0.2639	0.0497	0.6792	0.4233	0.2639	0.1145	0.2291	0.1319
	Majorca	0.4250	0.1298	0.8695	0.5790	0.4250	0.1673	0.3347	0.2125
MID3716	Sephardic Jews	0.4567	0.1523	0.9202	0.6005	0.4567	0.1762	0.3524	0.2283
	North African Jews	0.5000	0.1875	1.0000	0.6250	0.5000	0.1875	0.3750	0.2500
	Middle Eastern Jews	0.4881	0.1773	0.9768	0.6188	0.4881	0.1845	0.3690	0.2441
	Ashkenazi Jews	0.4388	0.1393	0.8909	0.5887	0.4388	0.1713	0.3425	0.2194
	Chuetas	0.4073	0.1184	0.8436	0.5658	0.4073	0.1622	0.3243	0.2036
	Majorca	0.4758	0.1672	0.9538	0.6120	0.4758	0.1813	0.3626	0.2379
MID3692	Sephardic Jews	0.2163	0.0344	0.6380	0.3625	0.2163	0.0965	0.1929	0.1082
	North African Jews	0.2894	0.0593	0.7036	0.4532	0.2894	0.1238	0.2475	0.1447
	Middle Eastern Jews	0.1195	0.0116	0.5678	0.2175	0.1195	0.0562	0.1123	0.0597
	Ashkenazi Jews	0.2271	0.0376	0.6469	0.3768	0.2271	0.1007	0.2013	0.1135
	Chuetas	0.3047	0.0655	0.7191	0.4701	0.3047	0.1291	0.2583	0.1523
	Majorca	0.3053	0.0657	0.7197	0.4708	0.3053	0.1294	0.2587	0.1527

MID2637	**Sephardic Jews**	0.2363	0.0405	0.6547	0.3889	0.2363	0.1042	0.2084	0.1182
	North African Jews	0.1404	0.0156	0.5817	0.2513	0.1404	0.0653	0.1306	0.0702
	Middle Eastern Jews	0.2629	0.0494	0.6783	0.4221	0.2629	0.1142	0.2284	0.1315
	Ashkenazi Jews	0.3331	0.0781	0.7497	0.4997	0.3331	0.1388	0.2776	0.1665
	Chuetas	0.1292	0.0134	0.5742	0.2333	0.1292	0.0604	0.1208	0.0646
	Majorca	0.1340	0.0143	0.5773	0.2410	0.1340	0.0625	0.1250	0.0670
MID3740	**Sephardic Jews**	0.4476	0.1456	0.9052	0.5947	0.4476	0.1737	0.3475	0.2238
	North African Jews	0.4965	0.1845	0.9931	0.6233	0.4965	0.1866	0.3733	0.2483
	Middle Eastern Jews	0.4674	0.1606	0.9389	0.6071	0.4674	0.1791	0.3582	0.2337
	Ashkenazi Jews	0.4925	0.1811	0.9853	0.6212	0.4925	0.1856	0.3712	0.2463
	Chuetas	0.4883	0.1775	0.9772	0.6190	0.4883	0.1845	0.3691	0.2442
	Majorca	0.4829	0.1730	0.9670	0.6160	0.4829	0.1832	0.3663	0.2415
MID198	**Sephardic Jews**	0.4258	0.1304	0.8708	0.5796	0.4258	0.1676	0.3352	0.2129
	North African Jews	0.4982	0.1859	0.9964	0.6241	0.4982	0.1870	0.3741	0.2491
	Middle Eastern Jews	0.4681	0.1611	0.9401	0.6076	0.4681	0.1793	0.3586	0.2341
	Ashkenazi Jews	0.4998	0.1873	0.9995	0.6249	0.4998	0.1874	0.3749	0.2499
	Chuetas	0.4189	0.1258	0.8604	0.5746	0.4189	0.1656	0.3311	0.2094
	Majorca	0.5000	0.1875	1.0000	0.6250	0.5000	0.1875	0.3750	0.2500
MID3703	**Sephardic Jews**	0.4518	0.1487	0.9120	0.5974	0.4518	0.1749	0.3497	0.2259
	North African Jews	0.3847	0.1049	0.8126	0.5474	0.3847	0.1554	0.3107	0.1924
	Middle Eastern Jews	0.3801	0.1023	0.8066	0.5435	0.3801	0.1539	0.3079	0.1901
	Ashkenazi Jews	0.4630	0.1571	0.9310	0.6044	0.4630	0.1779	0.3558	0.2315
	Chuetas	0.4652	0.1588	0.9349	0.6058	0.4652	0.1785	0.3570	0.2326
	Majorca	0.4712	0.1635	0.9455	0.6093	0.4712	0.1801	0.3602	0.2356
MID3690	**Sephardic Jews**	0.4616	0.1561	0.9287	0.6036	0.4616	0.1775	0.3551	0.2308
	North African Jews	0.4823	0.1724	0.9657	0.6157	0.4823	0.1830	0.3660	0.2411
	Middle Eastern Jews	0.4376	0.1385	0.8890	0.5879	0.4376	0.1709	0.3418	0.2188
	Ashkenazi Jews	0.4701	0.1627	0.9436	0.6087	0.4701	0.1798	0.3596	0.2351
	Chuetas	0.4721	0.1642	0.9472	0.6099	0.4721	0.1803	0.3607	0.2361
	Majorca	0.4428	0.1422	0.8974	0.5915	0.4428	0.1724	0.3448	0.2214
MID3722	**Sephardic Jews**	0.4642	0.1581	0.9333	0.6052	0.4642	0.1782	0.3565	0.2321
	North African Jews	0.4524	0.1491	0.9130	0.5978	0.4524	0.1750	0.3501	0.2262
	Middle Eastern Jews	0.5000	0.1875	1.0000	0.6250	0.5000	0.1875	0.3750	0.2500
	Ashkenazi Jews	0.4879	0.1771	0.9764	0.6187	0.4879	0.1844	0.3689	0.2439
	Chuetas	0.4976	0.1854	0.9952	0.6238	0.4976	0.1869	0.3738	0.2488
	Majorca	0.4260	0.1305	0.8710	0.5798	0.4260	0.1676	0.3352	0.2130
MID3732	**Sephardic Jews**	0.2072	0.0318	0.6307	0.3501	0.2072	0.0929	0.1858	0.1036
	North African Jews	0.3024	0.0645	0.7168	0.4677	0.3024	0.1283	0.2567	0.1512
	Middle Eastern Jews	0.2046	0.0310	0.6286	0.3464	0.2046	0.0918	0.1837	0.1023
	Ashkenazi Jews	0.3607	0.0918	0.7821	0.5263	0.3607	0.1478	0.2957	0.1804
	Chuetas	0.4251	0.1299	0.8697	0.5791	0.4251	0.1674	0.3348	0.2126
	Majorca	0.3363	0.0796	0.7533	0.5029	0.3363	0.1399	0.2797	0.1681

MID3712	**Sephardic Jews**	0.1572	0.0192	0.5933	0.2773	0.1572	0.0724	0.1449	0.0786
	North African Jews	0.2549	0.0466	0.6710	0.4123	0.2549	0.1112	0.2224	0.1274
	Middle Eastern Jews	0.4061	0.1177	0.8419	0.5648	0.4061	0.1618	0.3236	0.2030
	Ashkenazi Jews	0.2741	0.0534	0.6888	0.4355	0.2741	0.1183	0.2365	0.1370
	Chuetas	0.3750	0.0994	0.8000	0.5391	0.3750	0.1523	0.3047	0.1875
	Majorca	0.3555	0.0891	0.7758	0.5214	0.3555	0.1462	0.2923	0.1777
MID1736	**Sephardic Jews**	0.4998	0.1873	0.9996	0.6249	0.4998	0.1874	0.3749	0.2499
	North African Jews	0.4876	0.1769	0.9758	0.6186	0.4876	0.1844	0.3687	0.2438
	Middle Eastern Jews	0.4996	0.1872	0.9993	0.6248	0.4996	0.1874	0.3748	0.2498
	Ashkenazi Jews	0.4918	0.1805	0.9839	0.6208	0.4918	0.1854	0.3709	0.2459
	Chuetas	0.4721	0.1642	0.9472	0.6099	0.4721	0.1803	0.3607	0.2361
	Majorca	0.5000	0.1875	1.0000	0.6250	0.5000	0.1875	0.3750	0.2500
MID3719	**Sephardic Jews**	0.3495	0.0860	0.7687	0.5158	0.3495	0.1442	0.2884	0.1748
	North African Jews	0.2825	0.0566	0.6968	0.4452	0.2825	0.1213	0.2426	0.1412
	Middle Eastern Jews	0.3395	0.0811	0.7570	0.5061	0.3395	0.1409	0.2819	0.1697
	Ashkenazi Jews	0.1800	0.0246	0.6098	0.3114	0.1800	0.0819	0.1638	0.0900
	Chuetas	0.2684	0.0513	0.6834	0.4287	0.2684	0.1162	0.2324	0.1342
	Majorca	0.3193	0.0718	0.7345	0.4857	0.3193	0.1342	0.2683	0.1596
MID2089	**Sephardic Jews**	0.4473	0.1454	0.9047	0.5945	0.4473	0.1736	0.3473	0.2237
	North African Jews	0.4442	0.1431	0.8995	0.5924	0.4442	0.1728	0.3455	0.2221
	Middle Eastern Jews	0.4356	0.1371	0.8858	0.5866	0.4356	0.1704	0.3407	0.2178
	Ashkenazi Jews	0.2392	0.0414	0.6572	0.3926	0.2392	0.1053	0.2106	0.1196
	Chuetas	0.4575	0.1529	0.9216	0.6010	0.4575	0.1764	0.3528	0.2287
	Majorca	0.4771	0.1683	0.9563	0.6128	0.4771	0.1817	0.3633	0.2386
MID3774	**Sephardic Jews**	0.2624	0.0492	0.6778	0.4215	0.2624	0.1140	0.2279	0.1312
	North African Jews	0.4287	0.1324	0.8752	0.5817	0.4287	0.1684	0.3368	0.2144
	Middle Eastern Jews	0.4885	0.1777	0.9776	0.6191	0.4885	0.1846	0.3692	0.2443
	Ashkenazi Jews	0.3395	0.0811	0.7570	0.5061	0.3395	0.1409	0.2819	0.1697
	Chuetas	0.4467	0.1449	0.9036	0.5941	0.4467	0.1735	0.3469	0.2233
	Majorca	0.2734	0.0532	0.6881	0.4347	0.2734	0.1180	0.2360	0.1367
MID3760	**Sephardic Jews**	0.3202	0.0722	0.7356	0.4866	0.3202	0.1345	0.2690	0.1601
	North African Jews	0.3303	0.0768	0.7466	0.4969	0.3303	0.1379	0.2757	0.1651
	Middle Eastern Jews	0.3573	0.0900	0.7780	0.5231	0.3573	0.1467	0.2935	0.1787
	Ashkenazi Jews	0.3298	0.0766	0.7461	0.4965	0.3298	0.1377	0.2754	0.1649
	Chuetas	0.3683	0.0958	0.7915	0.5331	0.3683	0.1502	0.3005	0.1842
	Majorca	0.3330	0.0780	0.7496	0.4996	0.3330	0.1388	0.2775	0.1665
MID3701	**Sephardic Jews**	0.4601	0.1549	0.9261	0.6027	0.4601	0.1771	0.3543	0.2301
	North African Jews	0.4676	0.1607	0.9391	0.6072	0.4676	0.1791	0.3583	0.2338
	Middle Eastern Jews	0.4946	0.1828	0.9893	0.6223	0.4946	0.1861	0.3723	0.2473
	Ashkenazi Jews	0.4842	0.1740	0.9694	0.6167	0.4842	0.1835	0.3670	0.2421
	Chuetas	0.4297	0.1330	0.8767	0.5824	0.4297	0.1687	0.3374	0.2148
	Majorca	0.4883	0.1774	0.9771	0.6189	0.4883	0.1845	0.3691	0.2441

MID2612	**Sephardic Jews**	0.4985	0.1862	0.9970	0.6242	0.4985	0.1871	0.3742	0.2492
	North African Jews	0.4980	0.1858	0.9961	0.6240	0.4980	0.1870	0.3740	0.2490
	Middle Eastern Jews	0.4386	0.1392	0.8906	0.5887	0.4386	0.1712	0.3424	0.2193
	Ashkenazi Jews	0.4550	0.1511	0.9174	0.5995	0.4550	0.1757	0.3515	0.2275
	Chuetas	0.4837	0.1736	0.9684	0.6164	0.4837	0.1834	0.3667	0.2418
	Majorca	0.4999	0.1874	0.9999	0.6250	0.4999	0.1875	0.3750	0.2500
MID1839	**Sephardic Jews**	0.3170	0.0708	0.7321	0.4833	0.3170	0.1334	0.2667	0.1585
	North African Jews	0.3979	0.1127	0.8305	0.5583	0.3979	0.1594	0.3188	0.1990
	Middle Eastern Jews	0.3816	0.1031	0.8086	0.5448	0.3816	0.1544	0.3088	0.1908
	Ashkenazi Jews	0.1845	0.0257	0.6131	0.3179	0.1845	0.0837	0.1675	0.0922
	Chuetas	0.1643	0.0208	0.5983	0.2881	0.1643	0.0754	0.1508	0.0821
	Majorca	0.2653	0.0502	0.6805	0.4250	0.2653	0.1150	0.2301	0.1326
MID3754	**Sephardic Jews**	0.3361	0.0795	0.7531	0.5027	0.3361	0.1398	0.2796	0.1680
	North African Jews	0.0690	0.0042	0.5371	0.1309	0.0690	0.0333	0.0667	0.0345
	Middle Eastern Jews	0.3653	0.0942	0.7878	0.5305	0.3653	0.1493	0.2986	0.1827
	Ashkenazi Jews	0.4265	0.1309	0.8718	0.5801	0.4265	0.1678	0.3356	0.2133
	Chuetas	0.1527	0.0182	0.5901	0.2705	0.1527	0.0705	0.1411	0.0764
	Majorca	0.3413	0.0820	0.7591	0.5079	0.3413	0.1415	0.2831	0.1706
MID111	**Sephardic Jews**	0.4962	0.1842	0.9924	0.6231	0.4962	0.1865	0.3731	0.2481
	North African Jews	0.5000	0.1875	1.0000	0.6250	0.5000	0.1875	0.3750	0.2500
	Middle Eastern Jews	0.4885	0.1777	0.9776	0.6191	0.4885	0.1846	0.3692	0.2443
	Ashkenazi Jews	0.4961	0.1841	0.9922	0.6230	0.4961	0.1865	0.3730	0.2480
	Chuetas	0.4189	0.1258	0.8604	0.5746	0.4189	0.1656	0.3311	0.2094
	Majorca	0.4991	0.1867	0.9981	0.6245	0.4991	0.1873	0.3745	0.2495
MID2652	**Sephardic Jews**	0.2661	0.0505	0.6813	0.4260	0.2661	0.1153	0.2307	0.1330
	North African Jews	0.3347	0.0788	0.7515	0.5014	0.3347	0.1393	0.2787	0.1673
	Middle Eastern Jews	0.3274	0.0754	0.7433	0.4940	0.3274	0.1369	0.2738	0.1637
	Ashkenazi Jews	0.3722	0.0979	0.7964	0.5366	0.3722	0.1515	0.3029	0.1861
	Chuetas	0.3818	0.1032	0.8088	0.5449	0.3818	0.1545	0.3089	0.1909
	Majorca	0.4193	0.1261	0.8610	0.5749	0.4193	0.1657	0.3314	0.2096
MID1511	**Sephardic Jews**	0.3856	0.1054	0.8138	0.5481	0.3856	0.1556	0.3112	0.1928
	North African Jews	0.3206	0.0724	0.7359	0.4870	0.3206	0.1346	0.2692	0.1603
	Middle Eastern Jews	0.3830	0.1039	0.8103	0.5459	0.3830	0.1548	0.3096	0.1915
	Ashkenazi Jews	0.4516	0.1485	0.9117	0.5973	0.4516	0.1748	0.3496	0.2258
	Chuetas	0.3961	0.1116	0.8280	0.5569	0.3961	0.1588	0.3177	0.1981
	Majorca	0.4036	0.1161	0.8384	0.5629	0.4036	0.1611	0.3222	0.2018
MID2692	**Sephardic Jews**	0.3782	0.1012	0.8041	0.5418	0.3782	0.1533	0.3067	0.1891
	North African Jews	0.2990	0.0631	0.7133	0.4639	0.2990	0.1272	0.2543	0.1495
	Middle Eastern Jews	0.1816	0.0250	0.6109	0.3137	0.1816	0.0826	0.1651	0.0908
	Ashkenazi Jews	0.4656	0.1591	0.9356	0.6060	0.4656	0.1786	0.3572	0.2328
	Chuetas	0.2989	0.0631	0.7132	0.4638	0.2989	0.1271	0.2542	0.1494
	Majorca	0.4401	0.1403	0.8931	0.5897	0.4401	0.1716	0.3433	0.2201

MID357	**Sephardic Jews**	0.4806	0.1711	0.9626	0.6147	0.4806	0.1826	0.3651	0.2403
	North African Jews	0.4876	0.1769	0.9758	0.6186	0.4876	0.1844	0.3687	0.2438
	Middle Eastern Jews	0.4488	0.1465	0.9071	0.5955	0.4488	0.1740	0.3481	0.2244
	Ashkenazi Jews	0.4499	0.1473	0.9089	0.5962	0.4499	0.1743	0.3487	0.2249
	Chuetas	0.3680	0.0956	0.7911	0.5329	0.3680	0.1501	0.3003	0.1840
	Majorca	0.3158	0.0702	0.7308	0.4820	0.3158	0.1330	0.2659	0.1579
MID356	**Sephardic Jews**	0.4992	0.1868	0.9985	0.6246	0.4992	0.1873	0.3746	0.2496
	North African Jews	0.4965	0.1844	0.9930	0.6232	0.4965	0.1866	0.3732	0.2482
	Middle Eastern Jews	0.4645	0.1583	0.9336	0.6053	0.4645	0.1783	0.3566	0.2322
	Ashkenazi Jews	0.4989	0.1865	0.9977	0.6244	0.4989	0.1872	0.3744	0.2494
	Chuetas	0.3807	0.1026	0.8074	0.5440	0.3807	0.1541	0.3083	0.1904
	Majorca	0.3353	0.0791	0.7522	0.5019	0.3353	0.1395	0.2791	0.1676
MID243	**Sephardic Jews**	0.3576	0.0902	0.7784	0.5234	0.3576	0.1468	0.2937	0.1788
	North African Jews	0.2529	0.0459	0.6692	0.4099	0.2529	0.1105	0.2209	0.1264
	Middle Eastern Jews	0.3095	0.0675	0.7241	0.4753	0.3095	0.1308	0.2616	0.1548
	Ashkenazi Jews	0.3520	0.0873	0.7716	0.5182	0.3520	0.1450	0.2901	0.1760
	Chuetas	0.3298	0.0766	0.7461	0.4965	0.3298	0.1377	0.2754	0.1649
	Majorca	0.3661	0.0946	0.7888	0.5312	0.3661	0.1496	0.2991	0.1831
MID3727	**Sephardic Jews**	0.2188	0.0351	0.6400	0.3657	0.2188	0.0974	0.1948	0.1094
	North African Jews	0.2216	0.0359	0.6423	0.3695	0.2216	0.0985	0.1970	0.1108
	Middle Eastern Jews	0.2550	0.0466	0.6711	0.4125	0.2550	0.1112	0.2225	0.1275
	Ashkenazi Jews	0.2741	0.0534	0.6888	0.4355	0.2741	0.1183	0.2365	0.1370
	Chuetas	0.1770	0.0238	0.6075	0.3069	0.1770	0.0806	0.1613	0.0885
	Majorca	0.2177	0.0348	0.6391	0.3643	0.2177	0.0970	0.1940	0.1089
MID3753	**Sephardic Jews**	0.1551	0.0187	0.5918	0.2740	0.1551	0.0715	0.1430	0.0775
	North African Jews	0.2725	0.0528	0.6872	0.4336	0.2725	0.1177	0.2353	0.1362
	Middle Eastern Jews	0.2401	0.0417	0.6580	0.3937	0.2401	0.1056	0.2113	0.1200
	Ashkenazi Jews	0.1435	0.0163	0.5838	0.2561	0.1435	0.0666	0.1332	0.0717
	Chuetas	0.3280	0.0757	0.7440	0.4946	0.3280	0.1371	0.2742	0.1640
	Majorca	0.1340	0.0143	0.5773	0.2410	0.1340	0.0625	0.1250	0.0670

	Combined PE	Combined PD female	Combined PD male	Combined MEC Desmarais
Sephardic Jews	1 in 3.91E+01	1 in 2.44E+10	1 in 3.67E+06	1 in 8.29E+04
North African Jews	1 in 4.55E+01	1 in 3.01E+10	1 in 4.87E+06	1 in 9.28E+04
Middle Eastern Jews	1 in 5.15E+01	1 in 6.43E+10	1 in 7.87E+06	1 in 1.31E+05
Ashkenazi Jews	1 in 5.42E+01	1 in 6.06E+10	1 in 8.25E+06	1 in 1.29E+05
Chuetas	1 in 4.02E+01	1 in 3.26E+10	1 in 4.27E+06	1 in 9.32E+04
Majorca	1 in 4.71E+01	1 in 4.52E+10	1 in 6.05E+06	1 in 1.11E+05

HET: heterozigosity; PE: power of exclusion; PI: paternity index; PD female: power of discrimination in females; PD male: power of discrimination in males; MEC Krüger: mean exclusion chance in trios involving mother child and putative father (Krüger et al., 1968. Dtsch Z Gerichtl Med 64:127–146); MEC Desmarais: mean exclusion chance in trios involving daughter; MEC Desmarais duo: mean exclusion chance in father/daughter or mother/son duos (Desmarais et al., 1998. J Forensic Sci 43:1046–1049).

Supplementary Table 5. Forensic parameters of 9 X-Alu markers in 569 chromosomes (Sephardic Jews: 103; North African Jews: 77; Middle Eastern Jews: 75; Ashkenazi Jews: 85; Chuetas: 140 and Majorcans: 89).

Loci	Population	Statistical parameters			Power of discrimination		Mean paternity exclusion change		
		HET	PE	PI	PD female	PD male	MEC Krüger	MEC Desmarais	MEC Desmarais duo
Ya5DP62	Sephardic Jews	**0.2920**	0.0603	0.7062	0.4561	0.2920	0.1247	0.2494	0.1460
	North African Jews	0.3430	0.0828	0.7610	0.5095	0.3430	0.1421	0.2842	0.1715
	Middle Eastern Jews	0.4253	0.1300	0.8700	0.5793	0.4253	0.1674	0.3348	0.2126
	Ashkenazi Jews	0.3554	0.0890	0.7757	0.5213	0.3554	0.1461	0.2922	0.1777
	Chuetas	0.3847	0.1049	0.8126	0.5474	0.3847	0.1554	0.3107	0.1924
	Majorca	0.4431	0.1424	0.8978	0.5917	0.4431	0.1725	0.3449	0.2215
Yb8DP49	Sephardic Jews	**0.3884**	0.1070	0.8176	0.5505	0.3884	0.1565	0.3130	0.1942
	North African Jews	0.4551	0.1511	0.9175	0.5995	0.4551	0.1758	0.3515	0.2275
	Middle Eastern Jews	0.3911	0.1086	0.8212	0.5528	0.3911	0.1573	0.3146	0.1956
	Ashkenazi Jews	0.4127	0.1218	0.8514	0.5699	0.4127	0.1638	0.3276	0.2064
	Chuetas	0.2831	0.0568	0.6975	0.4460	0.2831	0.1215	0.2430	0.1416
	Majorca	0.2857	0.0578	0.7000	0.4490	0.2857	0.1225	0.2449	0.1429
Yd3JX437	Sephardic Jews	**0.2457**	0.0435	0.6628	0.4008	0.2457	0.1077	0.2155	0.1228
	North African Jews	0.0731	0.0047	0.5394	0.1382	0.0731	0.0352	0.0704	0.0366
	Middle Eastern Jews	0.0262	0.0007	0.5135	0.0515	0.0262	0.0130	0.0259	0.0131
	Ashkenazi Jews	0.2078	0.0319	0.6312	0.3509	0.2078	0.0931	0.1862	0.1039
	Chuetas	0.1663	0.0213	0.5997	0.2911	0.1663	0.0762	0.1524	0.0831
	Majorca	0.1641	0.0208	0.5982	0.2878	0.1641	0.0753	0.1507	0.0821
Yb8NBC634	Sephardic Jews	0.0449	0.0019	0.5235	0.0869	0.0449	0.0220	0.0439	0.0225
	North African Jews	0.0768	0.0051	0.5416	0.1448	0.0768	0.0369	0.0739	0.0384
	Middle Eastern Jews	0.1245	0.0125	0.5711	0.2258	0.1245	0.0584	0.1168	0.0623
	Ashkenazi Jews	0.0727	0.0046	0.5392	0.1375	0.0727	0.0350	0.0701	0.0364
	Chuetas	0.0968	0.0079	0.5536	0.1795	0.0968	0.0461	0.0921	0.0484
	Majorca	0.0442	0.0018	0.5231	0.0854	0.0442	0.0216	0.0432	0.0221
Ya5DP77	Sephardic Jews	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
	North African Jews	0.0520	0.0024	0.5274	0.0999	0.0520	0.0253	0.0506	0.0260
	Middle Eastern Jews	0.2688	0.0515	0.6838	0.4292	0.2688	0.1163	0.2327	0.1344
	Ashkenazi Jews	0.0849	0.0062	0.5464	0.1589	0.0849	0.0406	0.0813	0.0424
	Chuetas	0.1822	0.0251	0.6114	0.3147	0.1822	0.0828	0.1656	0.0911
	Majorca	0.0442	0.0018	0.5231	0.0854	0.0442	0.0216	0.0432	0.0221
Ya5NBC491	Sephardic Jews	0.0818	0.0058	0.5445	0.1535	0.0818	0.0392	0.0784	0.0409
	North African Jews	0.1676	0.0216	0.6006	0.2930	0.1676	0.0768	0.1535	0.0838
	Middle Eastern Jews	0.0262	0.0007	0.5135	0.0515	0.0262	0.0130	0.0259	0.0131
	Ashkenazi Jews	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
	Chuetas	0.0423	0.0016	0.5221	0.0819	0.0423	0.0207	0.0414	0.0211
	Majorca	0.0582	0.0030	0.5309	0.1113	0.0582	0.0283	0.0565	0.0291

Yb8NBC578	**Sephardic Jews**	0.1219	0.0121	0.5694	0.2215	0.1219	0.0572	0.1145	0.0609
	North African Jews	0.1193	0.0116	0.5677	0.2172	0.1193	0.0561	0.1122	0.0596
	Middle Eastern Jews	0.1245	0.0125	0.5711	0.2258	0.1245	0.0584	0.1168	0.0623
	Ashkenazi Jews	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
	Chuetas	0.0685	0.0041	0.5368	0.1299	0.0685	0.0331	0.0661	0.0342
	Majorca	0.0442	0.0018	0.5231	0.0854	0.0442	0.0216	0.0432	0.0221
Ya5DP4	**Sephardic Jews**	0.0428	0.0017	0.5224	0.0829	0.0428	0.0210	0.0419	0.0214
	North African Jews	0.0243	0.0006	0.5125	0.0477	0.0243	0.0120	0.0240	0.0121
	Middle Eastern Jews	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
	Ashkenazi Jews	0.1128	0.0105	0.5636	0.2065	0.1128	0.0532	0.1064	0.0564
	Chuetas	0.0299	0.0008	0.5154	0.0585	0.0299	0.0147	0.0295	0.0150
	Majorca	0.0442	0.0018	0.5231	0.0854	0.0442	0.0216	0.0432	0.0221
Ya5DP13	**Sephardic Jews**	**0.1841**	0.0256	0.6129	0.3174	0.1841	0.0836	0.1672	0.0921
	North African Jews	0.1859	0.0260	0.6142	0.3200	0.1859	0.0843	0.1686	0.0929
	Middle Eastern Jews	0.1009	0.0085	0.5561	0.1866	0.1009	0.0479	0.0958	0.0505
	Ashkenazi Jews	0.0220	0.0005	0.5112	0.0432	0.0220	0.0109	0.0217	0.0110
	Chuetas	0.2855	0.0577	0.6998	0.4487	0.2855	0.1224	0.2447	0.1427
	Majorca	0.2448	0.0432	0.6621	0.3997	0.2448	0.1074	0.2148	0.1224

	Combined PE	Combined PD female	Combined PD male	Combined MEC Desmarais
Sephardic Jews	1 in 1.31E+00	1 in 1.81E+01	1 in 5.09E+00	1 in 3.97E+00
North African Jews	1 in 1.38E+00	1 in 2.14E+01	1 in 5.91E+00	1 in 4.32E+00
Middle Eastern Jews	1 in 1.41E+00	1 in 2.12E+01	1 in 5.98E+00	1 in 4.27E+00
Ashkenazi Jews	1 in 1.32E+00	1 in 1.36E+01	1 in 4.53E+00	1 in 3.46E+00
Chuetas	1 in 1.34E+00	1 in 2.41E+01	1 in 5.95E+00	1 in 4.55E+00
Majorca	1 in 1.33E+00	1 in 1.67E+01	1 in 5.07E+00	1 in 3.83E+00

HET: heterozigosity. In bold significant Hady-Weinberg equilibrium p-values; PE: power of exclusion; PI: paternity index; PD female: power of discrimination in females; PD male: power of discrimination in males; MEC Krüger: mean exclusion chance in trios involving mother child and putative father (Krüger et al., 1968. Dtsch Z Gerichtl Med 64:127–146); MEC Desmarais: mean exclusion chance in trios involving daughter; MEC Desmarais duo: mean exclusion chance in father/daughter or mother/son duos (Desmarais et al., 1998. J Forensic Sci 43:1046–1049).

Supplementary Table 6. Forensic parameters of 12 X-STR markers in 469 chromosomes (Sephardic Jews: 99; North African Jews: 79; Middle Eastern Jews: 80; Ashkenazi Jews: 84 and Chuetas: 127).

Loci	Population	Statistical parameters			Power of discrimination		Mean paternity exclusion change		
		HET	PE	PI	PD female	PD male	MEC Krüger	MEC Desmarais	MEC Desmarais duo
DXS10103	Sephardic Jews	0.6868	0.4081	1.5964	0.8636	0.6868	0.4617	0.6485	0.5030
	North African Jews	0.6727	0.3874	1.5278	0.8625	0.6727	0.4631	0.6423	0.4961
	Middle Eastern Jews	0.6772	0.3939	1.5491	0.8587	0.6772	0.4546	0.6401	0.4939
	Ashkenazi Jews	0.6950	0.4205	1.6392	0.8774	0.6950	0.4896	0.6654	0.5220
	Chuetas	0.6723	0.3867	1.5256	0.8493	0.6723	0.4410	0.6290	0.4833
DXS8378	Sephardic Jews	0.7091	0.4424	1.7185	0.8588	0.7091	0.4475	0.6525	0.5083
	North African Jews	0.6812	0.3998	1.5685	0.8312	0.6812	0.3982	0.6140	0.4669
	Middle Eastern Jews	0.6845	0.4046	1.5846	0.8376	0.6845	0.4118	0.6216	0.4753
	Ashkenazi Jews	0.6951	0.4207	1.6399	0.8528	0.6951	0.4400	0.6409	0.4959
	Chuetas	0.7212	0.4617	1.7931	0.8701	0.7212	0.4680	0.6690	0.5260
DXS7132	Sephardic Jews	0.7421	0.4963	1.9384	0.8937	0.7421	0.5163	0.7023	0.5633
	North African Jews	0.7143	0.4507	1.7500	0.8712	0.7143	0.4729	0.6671	0.5242
	Middle Eastern Jews	0.7267	0.4707	1.8292	0.8809	0.7267	0.4924	0.6823	0.5414
	Ashkenazi Jews	0.7629	0.5322	2.1092	0.9057	0.7629	0.5417	0.7249	0.5892
	Chuetas	0.6591	0.3679	1.4667	0.8248	0.6591	0.4031	0.6001	0.4548
DXS10134	Sephardic Jews	0.8671	0.7289	3.7627	0.9693	0.8671	0.7380	0.8541	0.7577
	North African Jews	0.8681	0.7308	3.7899	0.9693	0.8681	0.7381	0.8548	0.7583
	Middle Eastern Jews	0.8558	0.7063	3.4667	0.9635	0.8558	0.7137	0.8400	0.7372
	Ashkenazi Jews	0.8718	0.7383	3.9011	0.9713	0.8718	0.7466	0.8595	0.7654
	Chuetas	0.8553	0.7053	3.4544	0.9650	0.8553	0.7196	0.8412	0.7401
DXS10074	Sephardic Jews	0.8519	0.6986	3.3756	0.9611	0.8519	0.7045	0.8349	0.7299
	North African Jews	0.8477	0.6903	3.2823	0.9584	0.8477	0.6939	0.8293	0.7215
	Middle Eastern Jews	0.8200	0.6366	2.7772	0.9446	0.8200	0.6471	0.7970	0.6791
	Ashkenazi Jews	0.8354	0.6663	3.0376	0.9517	0.8354	0.6703	0.8141	0.7012
	Chuetas	0.8440	0.6830	3.2043	0.9578	0.8440	0.6922	0.8261	0.7182
DXS10101	Sephardic Jews	0.9175	0.8312	6.0587	0.9879	0.9175	0.7002	0.9121	0.8453
	North African Jews	0.8872	0.7693	4.4320	0.9770	0.8872	0.7735	0.8769	0.7900
	Middle Eastern Jews	0.8673	0.7292	3.7671	0.9688	0.8673	0.7358	0.8537	0.7565
	Ashkenazi Jews	0.8851	0.7652	4.3527	0.9764	0.8851	0.7709	0.8748	0.7874
	Chuetas	0.8699	0.7344	3.8426	0.9695	0.8699	0.7390	0.8563	0.7600
DXS10135	Sephardic Jews	0.9216	0.8397	6.3748	0.9885	0.9216	0.8411	0.9162	0.8506
	North African Jews	0.9323	0.8617	7.3851	0.9913	0.9323	0.8627	0.9282	0.8700
	Middle Eastern Jews	0.9275	0.8519	6.8963	0.9900	0.9275	0.8527	0.9228	0.8610
	Ashkenazi Jews	0.9229	0.8424	6.4870	0.9888	0.9229	0.8435	0.9177	0.8528
	Chuetas	0.9263	0.8494	6.7841	0.9898	0.9263	0.8507	0.9215	0.8591

DXS7423	**Sephardic Jews**	0.6920	0.4160	1.6235	0.8479	0.6920	0.4305	0.6348	0.4893
	North African Jews	0.7578	0.5232	2.0645	0.9011	0.7578	0.5313	0.7176	0.5807
	Middle Eastern Jews	0.6248	0.3217	1.3325	0.7838	0.6248	0.3451	0.5494	0.4046
	Ashkenazi Jews	0.6646	0.3757	1.4908	0.8267	0.6646	0.4020	0.6038	0.4579
	Chuetas	0.6661	0.3778	1.4976	0.8343	0.6661	0.4117	0.6119	0.4648
DXS10146	**Sephardic Jews**	0.8883	0.7716	4.4758	0.9783	0.8883	0.7797	0.8790	0.7942
	North African Jews	0.9011	0.7977	5.0566	0.9823	0.9011	0.8020	0.8932	0.8150
	Middle Eastern Jews	0.8917	0.7784	4.6149	0.9783	0.8917	0.7807	0.8817	0.7970
	Ashkenazi Jews	0.9152	0.8266	5.8968	0.9869	0.9152	0.8300	0.9093	0.8400
	Chuetas	0.8790	0.7527	4.1320	0.9739	0.8790	0.7592	0.8676	0.7769
DXS10079	**Sephardic Jews**	0.8031	0.6048	2.5391	0.9340	0.8031	0.6162	0.7758	0.6525
	North African Jews	0.7848	0.5712	2.3236	0.9246	0.7848	0.5919	0.7557	0.6282
	Middle Eastern Jews	0.7480	0.5064	1.9845	0.8961	0.7480	0.5231	0.7076	0.5702
	Ashkenazi Jews	0.7883	0.5775	2.3618	0.9254	0.7883	0.5924	0.7585	0.6308
	Chuetas	0.8034	0.6053	2.5431	0.9337	0.8034	0.6152	0.7758	0.6521
HPRTB	**Sephardic Jews**	0.7059	0.4375	1.7002	0.8643	0.7059	0.4635	0.6567	0.5138
	North African Jews	0.7439	0.4995	1.9527	0.8979	0.7439	0.5266	0.7074	0.5696
	Middle Eastern Jews	0.7825	0.5669	2.2984	0.9235	0.7825	0.5854	0.7533	0.6236
	Ashkenazi Jews	0.7070	0.4392	1.7064	0.8631	0.7070	0.4587	0.6559	0.5123
	Chuetas	0.6842	0.4042	1.5833	0.8569	0.6842	0.4523	0.6409	0.4959
DXS10148	**Sephardic Jews**	0.8401	0.6754	3.1267	0.9595	0.8401	0.6967	0.8252	0.7182
	North African Jews	0.8697	0.7340	3.8374	0.9705	0.8697	0.7368	0.8572	0.7619
	Middle Eastern Jews	0.8499	0.6947	3.3312	0.9617	0.8499	0.7072	0.8341	0.7301
	Ashkenazi Jews	0.8376	0.6706	3.0788	0.9584	0.8376	0.6927	0.8224	0.7143
	Chuetas	0.8639	0.7224	3.6728	0.9664	0.8639	0.7256	0.8488	0.7489

	Combined PE	Combined PD female	Combined PD male	Combined MEC Desmarais
Sephardic Jews	1 in 2.83E+05	1 in 2.25E+15	1 in 1.05E+09	1 in 2.61E+08
North African Jews	1 in 3.38E+05	1 in 3.00E+15	1 in 1.27E+09	1 in 3.22E+08
Middle Eastern Jews	1 in 1.25E+05	1 in 4.09E+14	1 in 4.30E+08	1 in 9.87E+07
Ashkenazi Jews	1 in 2.33E+05	1 in 1.80E+15	1 in 8.48E+08	1 in 2.12E+08
Chuetas	1 in 1.28E+05	1 in 4.50E+14	1 in 4.34E+08	1 in 1.04E+08

HET: heterozigosity; PE: power of exclusion; PI: paternity index; PD female: power of discrimination in females; PD male: power of discrimination in males; MEC Krüger: mean exclusion chance in trios involving mother child and putative father (Krüger et al., 1968. Dtsch Z Gerichtl Med 64:127–146); MEC Desmarais: mean exclusion chance in trios involving daughter; MEC Desmarais duo: mean exclusion chance in father/daughter or mother/son duos (Desmarais et al., 1998. J Forensic Sci 43:1046–1049).

Supplemenary Table 7. X-STR haplotypes in 149 males of the 5 Jewish populations studied (Sephardic Jews: 21; North African Jews: 29; Middle Eastern Jews: 28; Ashkenazi Jews: 24 and Chuetas: 47).

LG1			124 haplotypes													
DXS 10148	**DXS 10135**	**DXS 8378**	**SEP**	**NAJ**	**MEJ**	**ASH**	**CHU**	**TOTAL**	**DXS 10148**	**DXS 10135**	**DX S8378**	**SEP**	**NAJ**	**MEJ**	**ASH**	**CHU**	**TOTAL**
17	18	11			1			1	25.1	26	11	1					1
18	19	12					1	1	25.1	26	12	1				1	2
18	20	10		1				1	25.1	27	11		1				1
18	20	11	1					1	25.1	28	10	1					1
18	20	12					1	1	25.1	28	12	1					1
18	22	12	1					1	25.1	29	10					1	1
18	24	10					1	1	25.1	29	11		1				1
18	25	10	1			1		2	25.1	30	12		1				1
18	25	11		1				1	25.1	19.1	10					1	1
18	28	10					1	1	25.1	21.1	10				1		1
18	28	12					1	1	25.1	25.3	12				1		1
18	29	10					1	1	26.1	16	12					1	1
18	32	10	1					1	26.1	19	11		1	1			2
18	20.1	10			1			1	26.1	20	10			1			1
19	20	11			1			1	26.1	20	11				1		1
19	21	10		1				1	26.1	20	13		1	1			2
19	26	10				1	1	2	26.1	21	10	1					1
19	26	11	1				1	2	26.1	21	12					1	1
19	19.1	11					1	1	26.1	22	11					2	2
19	20.1	11					1	1	26.1	23	10				2		2
21	17	12					1	1	26.1	23	12			1			1
22	28	10					1	1	26.1	24	11					1	1
23	27	10					1	1	26.1	25	10			1		1	2
23	22.1	10		1				1	26.1	25	11			1	1		2
24	21	11		1				1	26.1	26	11				1		1
25	25	11			1			1	26.1	27	12		1				1
?	20	?					1	1	26.1	28	10				1		1
13.3	20	14			1			1	26.1	28	12	1					1
22.1	19	10			1			1	26.1	29	12			1		1	2
22.1	21	14			1			1	26.1	30	11				1	1	2
22.1	27	11			1			1	26.1	31	10			1			1
23.1	18	12		1				1	26.1	22.1	10					2	2
23.1	19	12		1				1	26.1	33.1	11					1	2
23.1	24	9				2		2	27.1	17	12					1	1
23.1	24	10		1				1	27.1	19	11					1	1
23.1	25	10	2	1				3	27.1	20	10					1	1
23.1	27	10	1					1	27.1	21	12			1			1
23.1	27	12			4			4	27.1	22	10			1			1
23.1	30	12				1		1	27.1	22	11	1					1
23.1	19.1	10				1		1	27.1	23	10		1	1			2
24.1	19	12					1	1	27.1	25	11			1			1
24.1	20	12					1	1	27.1	26	11					1	1
24.1	22	9					1	1	27.1	27	10		1				1
24.1	24	12					1	1	27.1	27	12	1					1
24.1	25	11					1	1	27.1	28	10				1		1
24.1	28	13				1		1	27.1	29	11					1	1
25.1	18	12					2	2	27.1	29	12			1			1
25.1	19	10					1	1	27.1	30	10					1	1
25.1	19	11				1		1	27.1	30	11					1	1
25.1	19	12					1	1	27.1	18.1	11		1				1
25.1	20	11	2					2	28.1	20	11		1				1
25.1	21	12		1				1	28.1	21	11				1		1
25.1	22	9					1	1	28.1	22	10		1				1
25.1	22	10			1	1		2	28.1	23	10		1				1
25.1	22	11		1		1		2	28.1	23	12		1				1
25.1	23	11				1		1	28.1	24	9				1		1
25.1	24	10			1			1	28.1	27	11		1				1
25.1	24	11	1					1	28.1	29	10			1			1
25.1	24	12	1					1	28.2	27	11		1				1
25.1	25	10		1			1	2	29.1	21	12	1					1
25.1	25	12		1				1	29.1	29	11		1				1
25.1	26	10				1		1	30.1	15	11					1	1

LG2			96 haplotypes													
DXS 7132	DXS 10079	DXS 10074	SEP	NAJ	MEJ	ASH	CHU	TOTAL	DXS 7132	DXS 10079	DXS 10074	SEP	NAJ	MEJ	ASH	CHU	TOTAL
12	17	18					1	1	14	18	9	1		1			2
12	18	16			1			1	14	18	16					2	2
12	18	18	1					1	14	18	18	1					1
12	19	18			1			1	14	19	7			1			1
12	19	19				2		2	14	19	9		1			2	3
12	20	8			2			2	14	19	16				1		1
12	20	18				1	1	2	14	19	17				3		3
12	21	16				1		1	14	19	18					3	3
12	21	17					1	1	14	19	19	1			1	1	3
12	21	19				1		1	14	20	8	1	1	1			3
12	18/19	18					1	1	14	20	13	1					1
13	15	16					1	1	14	20	14					1	1
13	16	15		1				1	14	20	15			1			1
13	17	8			1			1	14	20	16		1	3			4
13	17	16					1	1	14	20	17			1			1
13	17	19	1					1	14	20	18		1	1			2
13	17	20	1					1	14	21	8	1	2				3
13	18	7					1	1	14	21	17				1	3	4
13	18	8		1	1		1	3	14	21	20					2	2
13	18	14					1	1	14	22	8		1				1
13	18	17		1	1			2	14	22	15					1	1
13	18	18			1		5	6	14	22	17					1	1
13	19	9		1				1	15	16	7	1					1
13	19	14				1		1	15	16	16			1			1
13	19	15	1					1	15	17	8				1		1
13	19	16				1	1	2	15	17	9					1	1
13	19	17	1					1	15	17	18			1			1
13	19	19		1				1	15	18	8			1			1
13	20	14		1				1	15	18	16				1	1	2
13	20	15			1		1	2	15	18	19	1	1		1		3
13	20	17			3			3	15	19	8					1	1
13	20	18					1	1	15	19	9	2			1		3
13	20	19					1	1	15	19	16					1	1
13	21	14					1	1	15	19	17	2			1		3
13	21	15				1		1	15	19	18			2			2
13	21	17					2	2	15	20	15		1				1
13	21	18		1			2	3	15	20	17				1		1
13	22	7					1	1	15	20	18		1		1		2
13	22	19		1				1	15	21	8	1	1				2
13	18/19	18		1				1	15	21	16		1				1
14	15	18		1				1	15	22	19		1				1
14	15	19		1				1	15	20/21	15		1				1
14	16	18				1		1	16	18	8		1				1
14	17	8		1		1		2	16	18	16	2					2
14	17	15					1	1	16	18	18			1			1
14	17	16		1				1	16	19	16	1					1
14	17	17					1	1	16	20	15		1				1
14	18	8			1	1		2	?	17	15					1	1

LG3			93 haplotypes													
DXS 10103	HPRTB	DXS 10101	SEP	NAJ	MEJ	ASH	CHU	TOTAL	DXS 10103	HPRTB	DXS 10101	SEP	NAJ	MEJ	ASH	CHU	TOTAL
15	11	30.2				2		2	19	12	31		1				1
15	14	32					1	1	19	12	33					1	1
16	11	34					1	1	19	12	26.2			1			1
16	12	31				1		1	19	12	27.2					1	1
16	13	32			1			1	19	12	28.2		1	2			3
16	13	33					1	1	19	12	29.2		2	2	1	1	6
16	14	31					1	1	19	12	30.2		1			2	3
16	14	32.2					1	1	19	12	31.2		1				1
17	12	31		1	1			2	19	13	28					1	1
17	12	28.2			2			2	19	13	31		1				1
17	13	29	1					1	19	13	34		1				1
17	13	31		1			1	2	19	13	27.2	1				1	2
17	14	29				1		1	19	13	28.2					2	2
17	16	31					1	1	19	13	29.2			1	1		2
18	11	29.2			1			1	19	13	30.2		1		1	4	6
18	12	25	1					1	19	13	31.2	2	1			3	6
18	12	30	1		1			2	19	13	32.2	3			1		4
18	12	31	1					1	19	14	32		1				1
18	12	28.2				1		1	19	14	33			1			1
18	12	29.2				1		1	19	14	27.2			1			1
18	12	30.2		1			1	2	19	14	28.2	1					1
18	12	31.2					7	7	19	14	29.2			1	1		2
18	12	32.2				1	1	2	19	14	31.2			1			1
18	13	30			1			1	19	14	32.2				1		1
18	13	31		1				1	19	15	28					1	1
18	13	32	1				2	3	19	15	29.2					1	1
18	13	33					1	1	19	15	30		2				2
18	13	27.2			1			1	19	15	31				1		1
18	13	28.2					1	1	20	11	28.2	1					1
18	13	29.2		1				1	20	11	30.2	1					1
18	13	30.2			2		1	3	20	12	32		1				1
18	14	30				1		1	20	12	27.2	1					1
18	14	31				1		1	20	12	28.2					1	1
18	14	32		1				1	20	12	29.2					1	1
18	14	31.2	1					1	20	12	30.2				1	1	2
18	15	32.2		1				1	20	12	31.2			1	1		2
19	10	32			1			1	20	13	30	1					1
19	10	29.2		1				1	20	13	28.2				3		3
19	10	30.2			1			1	20	13	29.2				1	1	2
19	11	32			1			1	20	15	31.2		1				1
19	11	27.2		1				1	20	16	32.2			1			1
19	11	28.2	1		1	1	1	4	21	10	29.2					1	1
19	11	29.2		2	1			3	21	11	29.2				1		1
19	11	30.2		1				1	21	13	24.2		1				1
19	11	31.2	1					1	18/19	12	28.2			1			1
19	12	27	1					1	?	11	29.2					1	1
19	12	28	1	1			1	3									

LG4			108 haplotypes													
DXS 10146	**DXS 10134**	**DXS 7423**	**SEP**	**NAJ**	**MEJ**	**ASH**	**CHU**	**TOTAL**	**DXS 10146**	**DXS 10134**	**DXS 7423**	**SEP**	**NAJ**	**MEJ**	**ASH**	**CHU**	**TOTAL**
24	36	15					1	1	30	33	16		1				1
25	33	14				1		1	30	34	14		1		1		2
25	36	13					1	1	30	34	15		1				1
25	36	14	1	1				2	30	35	14			1			1
25	36	15					1	1	30	35	15		1	1			2
26	34	14			1		1	2	30	36	15				1		1
26	34	15	1					1	30	36	16					1	1
26	35	14		1	1	2	1	5	30	37	14					1	1
26	35	16	1					1	30	37	16					1	1
26	36	15				1		1	30	40.3	15				1		1
26	37	15		1		1		2	31	33	15			1			1
26	37	16					1	1	31	34	14					2	2
26	38	14			1			1	31	34	16					1	1
26	38	16	1					1	31	35	14			1			1
26	40	14					1	1	31	36	15					1	1
26	37.3	15					1	1	31	38	13		1				1
26	39.3	14					2	2	31	35/36	16				1		1
27	32	15			2		1	3	33	35	14	1					1
27	34	14			1		1	2	33	36	13					1	1
27	34	15	1				3	4	52	33	16		1				1
27	35	14		1	1	1		3	?	38.3	16					1	1
27	35	15				3		3	35.2	34	16	1					1
27	36	15					3	3	36.2	44.2	14	1					1
27	38	15					1	1	39.2	33	15			1			1
27	39	15					1	1	39.2	37	16					1	1
28	34	14		1				1	39.2	39	16	1					1
28	34	15	1	1				2	39.2	39.3	15				1		1
28	34	17		1				1	39.2	42.3	17				1		1
28	35	14	1					1	39.2	43.3	16		1				1
28	35	15			2			2	40.2	35	16					1	1
28	36	14			1	1		2	40.2	40.3	16		1				1
28	36	15		1			1	2	41.2	35	15				1		1
28	36	16		1			1	2	41.2	35	16		1				1
28	37	14			1		1	2	41.2	37	16	1					1
28	37	15				1		1	42.2	31	15					1	1
28	38	14	1					1	42.2	35	15					1	1
28	39	14				1		1	42.2	39.3	14					1	1
28	?	16					1	1	42.2	39.3	15					1	1
28	40.3	14			1			1	43.2	35	15	1			1		2
28	44.3	14			1			1	43.2	36	15		1				1
29	33	16					1	1	43.2	37	15		1				1
29	34	15		1				1	43.2	38	16	2					2
29	34	16					1	1	43.2	40	14			1			1
29	35	15	1	1	1			3	44.2	33	15		2				2
29	35	16			1			1	44.2	34	16			2			2
29	36	14	1				1	2	44.2	35	14	1					1
29	36	15					3	3	44.2	36	15			1	1		2
29	37	14		1	1			2	44.2	36	17		1				1
29	37	17					1	1	44.2	37	15			1			1
29	38	17					1	1	44.2	42.3	17		1				1
29	39	15	1					1	45.2	34	14	1		1			2
29	38.3	15				1		1	46.2	36	14				1		1
29	41.3	14			1			1	46.2	40.3	14		1				1
29	42.3	17		1				1	47.2	36	15				1		1

		LG1	LG2	LG3	LG4	12 X-STR
Haplotype diversity	**SEP**	0.9905 ± 0.0178	0.9857 ± 0.0189	0.9810 ± 0.0225	0.9952 ± 0.0165	0.9952 ± 0.0165
	NAJ	1.0000 ± 0.0091	0.9975 ± 0.0099	0.9926 ± 0.0111	0.9975 ± 0.0099	1.0000 ± 0.0091
	MEJ	0.9828 ± 0.0170	0.9754 ± 0.0163	0.9877 ± 0.0120	0.9901 ± 0.0116	1.0000 ± 0.0095
	ASH	0.9928 ± 0.0144	0.9855 ± 0.0179	0.9855 ± 0.0179	0.9855 ± 0.0179	1.0000 ± 0.0120
	CHU	0.9972 ± 0.0051	0.9806 ± 0.0099	0.9695 ± 0.0147	0.9898 ± 0.0072	1.0000 ± 0.0044
	Total	0.9973 ± 0.0012	0.9923 ± 0.0018	0.9896 ± 0.0024	0.9948 ± 0.0015	0.9998 ± 0.0008

LG: Linkage Group; SEP: Sephardic Jews; NAJ: North African Jews; MEJ: Middle Eastern Jews; ASH: Ashkenazi Jews and CHU: Chuetas.

Supplementary Table 9. Haplotype frequencies for MID357-MID356 and MID3690-MID3719-MID2089 in the populations studied.

	SEP	NAJ	MEJ	ASH	CHU	MAJ
MID357-MID356	(N=55)	(N=29)	(N=28)	(N=23)	(N=44)	(N=46)
SS	0.4909	0.3448	0.4643	0.3043	0.3182	0.1957
SL	0.0000	0.0345	0.0357	0.0000	0.0000	0.0000
LS	0.1091	0.0690	0.0357	0.0870	0.0682	0.0435
LL	0.4000	0.5517	0.4643	0.6087	0.6136	0.7609
MID3690-MID3719-MID2089	(N=56)	(N=29)	(N=28)	(N=24)	(N=48)	(N=46)
SSS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0217
SSL	0.0000	0.0345	0.0000	0.0000	0.0000	0.0000
SLS	0.0179	0.1379	0.1071	0.0000	0.1667	0.2174
SLL	0.3393	0.2759	0.1429	0.3333	0.2292	0.0870
LSS	0.2321	0.2069	0.2143	0.0833	0.1458	0.1522
LSL	0.0357	0.0000	0.0357	0.0833	0.0000	0.0000
LLS	0.0357	0.0000	0.0000	0.0000	0.0417	0.0000
LLL	0.3393	0.3448	0.5000	0.5000	0.4167	0.5217

N: number of males; SEP: Sephardic Jews; NAJ: North African Jews; MEJ: Middle Eastern Jews; ASH: Ashkenazi Jews; CHU: Chuetas and MAJ: Majorca.

Supplementary Table 10. F_{ST} genetic distances matrix between the Jewish populations in this study, based on 53 X-markers. P-values below diagonal. Significant values (p>0.05) are shown in bold.

	SEP	**NAJ**	**MEJ**	**ASH**	**CHU**
SEP	-	0.00267	0.01087	0.00397	0.02265
NAJ	0.10941	-	0.01340	0.01117	0.02256
MEJ	**0.00000**	**0.00000**	-	0.01324	0.02108
ASH	**0.02822**	**0.00010**	**0.00000**	-	0.02477
CHU	**0.00000**	**0.00000**	**0.00000**	**0.00000**	-

SEP: Sephardic Jews; NAJ: North African Jews; MEJ: Middle Eastern Jews; ASH: Ashkenazi Jews and CHU: Chuetas. Darker colours correspond to greater genetic distances.

Supplementary Table 11. Samples used in the population comparisons, number of chromosomes (N) studied and respective references

	Population	N	Reference
X-STR	**Jews**		
	Sephardic Jews	99	This study
	North African Jews	79	This study
	Middle Eastern Jews	80	This study
	Ashkenazi Jews	84	This study
	Chuetas	127	This study
	Morocco		
	Sahrawis	57	Bentayebi et al. (2012)
	Arabs	72	Bentayebi et al. (2012)
	Berber	64	Bentayebi et al. (2012)
	Algeria	316	Bekada et al. (2010)
	Somalia	275	Tomàs et al. (2012)
	Iran	375	Poulsen et al. (2015)
	Portugal	296	Cainé et al. (2013)
	Hungary	595	Horváth et al. (2012)
	Germany	1037	Edelmann et al. (2012)
	Sweden	652	Tillmar et al. (2012)
	Italy	160	Inturri et al. (2011)
	Denmark	212	Tomàs et al. (2012)
	Spain		
	Majorca	85	Ferragut et al. (2015a)
	Minorca	89	Ferragut et al. (2015a)
	Ibiza	81	Ferragut et al. (2015a)
	Valencia	95	Ferragut et al. (2015a)
X-Indels	**Jews**		
	Sephardic Jews	144	This study
	North African Jews	81	This study
	Middle Eastern Jews	78	This study
	Ashkenazi Jews	84	This study
	Chuetas	96	This study
	Greenland	129	Pereira et al. (2015)
	Native American	83	Ibarra et al. (2014)
	Bangladesh	56	Pereira et al. (2015)
	Macau	75	Pereira et al. (2012a)
	Taiwan	20	Pereira et al. (2015)
	Angola and Mozambique	116	Pereira et al. (2012a)
	Somalia	162	Pereira et al. (2012)
	Iraq	136	Pereira et al. (2011)
	Spain		
	Zamora	303	Pinto et al. (2015)
	Majorca	112	This study
	Portugal	324	Pereira et al. (2012a)
	Miranda do Douro	182	Pinto et al. (2015)
	Denmark	71	Pereira et al. (2015)
X-ALU	**Jews**		
	Sephardic Jews	103	This study
	North African Jews	77	This study
	Middle Eastern Jews	75	This study
	Ashkenazi Jews	85	This study
	Chuetas	140	This study
	Bolivia		
	Ayamara	152	Gayà-Vidal et al. (2010)
	Quechua	147	Gayà-Vidal et al. (2010)
	Asia	31	Callinan et al. (2003)
	Morocco		
	Sahrawis	49	Bentayebi et al. (2011)

Arabs	72	Bentayebi et al. (2011)
Berbers	67	Bentayebi et al. (2011)
High Atlas	151	Athanasiadis et al. (2007)
Siwa Oasis	143	Athanasiadis et al. (2007)
Egypt	33	Callinan et al. (2003)
Tunisia	168	Athanasiadis et al. (2007)
Ivory Coast	87	Athanasiadis et al. (2007)
African American	27	Callinan et al. (2003)
Europe	32	Callinan et al. (2003)
Italy		
Consenza	30	Bentayebi et al. (2011)
Catanzaro	61	Bentayebi et al. (2011)
Calabria	164	Bentayebi et al. (2011)
Sicily	84	Bentayebi et al. (2011)
Crete	121	Athanasiadis et al. (2007)
Spain		
Basque Country	138	Athanasiadis et al. (2007)
Valencia	60	Bentayebi et al. (2011)
Majorca	89	This study
Minorca	71	Bentayebi et al. (2011)
Ibiza	70	Bentayebi et al. (2011)

Supplemetary Table 12. F_{ST} distances matrices of the populations used in the MDS plots for X-Alu insertions, X-STR, and X-Indels. Darker colours correspond to greater genetic distances.

X-Alu

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1. Sephardic Jews	0.000																											
2. North African Jews	0.005	0.000																										
3. Middle Eastern Jews	0.033	0.016	0.000																									
4. Ashkenazi Jews	0.006	0.013	0.019	0.000																								
5. Chuetas	0.016	0.026	0.015	0.027	0.000																							
6. Majorca	0.018	0.026	0.020	0.023	0.000	0.000																						
7. Minorca	0.024	0.041	0.019	0.036	0.000	0.000	0.000																					
8. Ibiza	0.029	0.069	0.070	0.044	0.030	0.037	0.021	0.000																				
9. Valencia	0.000	0.000	0.032	0.018	0.012	0.013	0.023	0.044	0.000																			
10. Basques	0.012	0.017	0.021	0.000	0.033	0.019	0.039	0.055	0.021	0.000																		
11. European	0.013	0.033	0.026	0.000	0.018	0.006	0.015	0.000	0.030	0.001	0.000																	
12. Calabria	0.058	0.108	0.113	0.063	0.074	0.086	0.068	0.028	0.078	0.078	0.031	0.000																
13. Catanzaro	0.062	0.108	0.114	0.062	0.077	0.088	0.074	0.041	0.080	0.076	0.036	0.000	0.000															
14. Cosenza	0.051	0.108	0.112	0.068	0.068	0.086	0.057	0.000	0.073	0.088	0.029	0.010	0.027	0.000														
15. Sicily	0.033	0.070	0.079	0.042	0.052	0.061	0.048	0.010	0.050	0.050	0.007	0.001	0.002	0.009	0.000													
16. Crete	0.016	0.048	0.047	0.008	0.030	0.027	0.028	0.009	0.034	0.016	0.000	0.023	0.027	0.019	0.012	0.000												
17. Moroccan Arabs	0.001	0.000	0.014	0.012	0.013	0.007	0.023	0.047	0.000	0.010	0.013	0.091	0.093	0.090	0.054	0.033	0.000											
18. Moroccan Berbers	0.011	0.040	0.035	0.022	0.020	0.028	0.017	0.033	0.016	0.025	0.024	0.047	0.052	0.034	0.035	0.016	0.031	0.000										
19. Moroccan Sahrawis	0.002	0.030	0.030	0.011	0.019	0.022	0.015	0.019	0.011	0.013	0.006	0.040	0.047	0.022	0.024	0.003	0.020	0.000	0.000									
20. Tunisia	0.021	0.012	0.000	0.008	0.021	0.023	0.025	0.051	0.026	0.010	0.009	0.088	0.090	0.085	0.054	0.028	0.009	0.026	0.017	0.000								
21. High Atlas	0.013	0.046	0.050	0.017	0.039	0.053	0.040	0.033	0.028	0.032	0.027	0.041	0.047	0.024	0.031	0.014	0.043	0.000	0.000	0.034	0.000							
22. Siwa oasis	0.018	0.054	0.051	0.020	0.037	0.046	0.035	0.035	0.032	0.031	0.025	0.047	0.054	0.027	0.040	0.013	0.047	0.000	0.000	0.037	0.000	0.000						
23. Egypt	0.035	0.073	0.041	0.028	0.020	0.026	0.007	0.009	0.050	0.032	0.000	0.011	0.014	0.010	0.011	0.000	0.053	0.000	0.000	0.031	0.013	0.005	0.000					
24. Ivory Coast	0.273	0.281	0.240	0.275	0.246	0.276	0.245	0.340	0.258	0.285	0.327	0.333	0.327	0.340	0.334	0.302	0.290	0.196	0.231	0.264	0.242	0.230	0.238	0.000				
25. African Americans	0.215	0.209	0.192	0.227	0.200	0.225	0.201	0.273	0.192	0.229	0.260	0.258	0.250	0.274	0.252	0.244	0.216	0.155	0.181	0.204	0.202	0.197	0.190	0.073	0.000			
26. Aymaras	0.227	0.283	0.251	0.241	0.210	0.250	0.197	0.251	0.230	0.266	0.270	0.216	0.215	0.210	0.245	0.223	0.279	0.129	0.161	0.253	0.159	0.144	0.137	0.155	0.182	0.000		
27. Quechuas	0.225	0.278	0.245	0.235	0.207	0.241	0.194	0.254	0.227	0.257	0.266	0.224	0.222	0.219	0.251	0.221	0.274	0.127	0.159	0.249	0.161	0.142	0.136	0.133	0.173	0.000	0.000	
28.Asia	0.168	0.224	0.219	0.191	0.181	0.213	0.174	0.207	0.170	0.212	0.226	0.174	0.174	0.162	0.195	0.178	0.224	0.085	0.111	0.212	0.109	0.097	0.119	0.143	0.146	0.005	0.007	0.000

X-STR

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Sephardic Jews	0.000																				
2. North African Jews	0.002	0.000																			
3. Middle Eastern Jews	0.010	0.012	0.000																		
4. Ashkenazi Jews	0.001	0.004	0.006	0.000																	
5. Chuetas	0.010	0.010	0.024	0.012	0.000																
6. Majorca	0.002	0.002	0.004	0.000	0.011	0.000															
7. Minorca	0.005	0.003	0.010	0.003	0.009	0.000	0.000														
8. Ibiza	0.008	0.012	0.015	0.005	0.011	0.005	0.004	0.000													
9. Valencia	0.008	0.007	0.011	0.006	0.010	0.001	0.001	0.005	0.000												
10. Portugal	0.006	0.003	0.008	0.003	0.012	0.001	0.002	0.010	0.003	0.000											
11. Italy	0.004	0.006	0.008	0.001	0.011	0.000	0.000	0.006	0.003	0.001	0.000										
12. Germany	0.005	0.006	0.008	0.002	0.013	0.000	0.000	0.008	0.001	0.002	0.001	0.000									
13. Hungary	0.006	0.006	0.009	0.003	0.011	0.002	0.001	0.008	0.001	0.002	0.000	0.000	0.000								
14. Sweden	0.006	0.005	0.009	0.002	0.013	0.001	0.001	0.009	0.003	0.001	0.001	0.000	0.000	0.000							
15. Denmark	0.006	0.007	0.011	0.002	0.010	0.001	0.003	0.008	0.001	0.003	0.002	0.000	0.001	0.001	0.000						
16. Moroccan Arabs	0.005	0.003	0.012	0.005	0.011	0.003	0.004	0.007	0.002	0.004	0.003	0.004	0.002	0.004	0.005	0.000					
17. Moroccan Berbers	0.004	0.001	0.012	0.002	0.012	0.000	0.003	0.008	0.005	0.002	0.005	0.004	0.004	0.004	0.005	0.002	0.000				
18. Moroccan Sahrawis	0.011	0.009	0.017	0.012	0.020	0.009	0.009	0.014	0.006	0.006	0.008	0.008	0.008	0.009	0.011	0.002	0.005	0.000			
19. Algeria	0.005	0.004	0.008	0.004	0.011	0.001	0.001	0.008	0.003	0.003	0.002	0.003	0.003	0.004	0.004	0.001	0.004	0.005	0.000		
20. Iran	0.011	0.010	0.009	0.009	0.018	0.004	0.003	0.014	0.007	0.006	0.005	0.005	0.005	0.005	0.008	0.008	0.009	0.012	0.006	0.000	
21. Somalia	0.017	0.014	0.019	0.019	0.026	0.016	0.014	0.017	0.016	0.016	0.015	0.016	0.015	0.016	0.020	0.014	0.010	0.013	0.013	0.016	0.000

X-Indels

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Sephardic Jews	0.000																	
2. North African Jews	0.007	0.000																
3. Middle Eastern Jews	0.015	0.009	0.000															
4. Ashkenazi Jews	0.009	0.017	0.019	0.000														
5. Chuetas	0.031	0.025	0.020	0.031	0.000													
6. Majorca	0.013	0.016	0.021	0.011	0.023	0.000												
7. Zamora	0.001	0.007	0.014	0.002	0.016	0.004	0.000											
8. Miranda	0.005	0.007	0.020	0.004	0.019	0.005	0.000	0.000										
9. Portugal	0.010	0.010	0.021	0.010	0.018	0.002	0.003	0.004	0.000									
10. Denmark	0.013	0.016	0.025	0.003	0.027	0.002	0.004	0.004	0.001	0.000								
11. Iraq	0.011	0.015	0.010	0.011	0.026	0.016	0.008	0.011	0.011	0.015	0.000							
12. Somalia	0.073	0.061	0.057	0.065	0.064	0.052	0.056	0.066	0.052	0.051	0.052	0.000						
13. Angola-Mozambique	0.130	0.121	0.106	0.100	0.113	0.093	0.103	0.112	0.102	0.100	0.101	0.030	0.000					
14. Greenland	0.176	0.180	0.161	0.142	0.178	0.169	0.154	0.164	0.174	0.169	0.136	0.171	0.152	0.000				
15. Native American	0.181	0.184	0.152	0.162	0.199	0.191	0.168	0.189	0.188	0.198	0.141	0.164	0.146	0.098	0.000			
16. Taiwan	0.137	0.145	0.119	0.143	0.168	0.153	0.138	0.151	0.152	0.170	0.091	0.150	0.166	0.120	0.105	0.000		
17. Macau	0.117	0.137	0.115	0.120	0.156	0.139	0.121	0.136	0.134	0.149	0.078	0.136	0.160	0.121	0.092	0.004	0.000	
18. Bangladesh	0.071	0.077	0.057	0.065	0.102	0.075	0.070	0.085	0.070	0.075	0.036	0.061	0.092	0.125	0.095	0.072	0.053	0.000

Chapter 3: Mitochondrial DNA

1. Founding mothers of Chueta population

Ferragut JF; Marques SL; Ramon C; Castro JA; Amorim A; Alvarez L and Picornell A

Forensic Science International: Genetics Supplement Series. 5: e492–e494 (2015)

2. Maternal gene pool in Chuetas: Middle Eastern legacy and a novel sub-branching of the rare haplogroup R0a2m

In preparation

Introduction

The third chapter contains the genetic characterization of female lineages in the Chueta population based on mtDNA analyses.

Mitochondrial DNA has been widely studied to shed light on the origin and demographic history of human populations. Specifically, in Jewish populations, Ashkenazi maternal lineages have been well characterized (e.g. Behar et al., 2006; 2008b; Brandstätter et al., 2008; Costa et al., 2013), whereas Sephardic lineages studies are more limited (e.g. Behar et al., 2008b; Nogueiro et al., 2015a). These studies will enable a comparison with the Chueta results.

A preliminary study carried out only with mtDNA Hypervariable region I (HVRI) revealed that the Middle Eastern haplogroup pre-HV was the most prevalent (23%) in Chuetas (Picornell et al., 2005). In the present work, we studied the D-loop region of 104 individuals from the Chueta population, and 79 individuals from its host population (Majorca) in order to obtain a better picture of Chueta maternal lineages, and to scrutinize, on the one hand, the genetic footprints of Sephardim in this population and, on the other hand, the introgression of mtDNA lineages from their host population.

Moreover, to go further into the identification of different sub-haplogroups, and taking into account the complexity of haplogroup H classification, 51 SNPs were analysed to ensure the correct assignation of haplogroup H samples.

Once the Chuetas' mtDNA lineages were identified in terms of the D-loop region, the complete genome of 16 molecules, belonging to 10 different haplogroups, were sequenced to compare them with other published sequences, and to infer a more detailed origin of these haplogroups.

This chapter is divided into two sections: first, preliminary results in D-loop published in Ferragut et al. (2015c); and, second, a manuscript in preparation with full mtDNA results.

Founding mothers of Chueta population

Ferragut JF[1]; Marques SL[2]; Ramon C[1]; Castro JA[1]; Amorim A[2,3]; Alvarez L[2,3] and Picornell A[1]

[1]Institut Universitari d'Investigació en Ciències de la Salut (IUNICS) and Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Palma de Mallorca, Illes Balears, Spain
[2]Institute of Pathology and Molecular Immunology of the University of Porto (IPATIMUP) and Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal
[3]Faculty of Sciences of the University of Porto, Porto, Portugal

Corresponding author

Joana F. Ferragut (jf.ferragut@uib.es). Ctra Valldemossa Km 7.5 07122 Palma (Balearic Islands, SPAIN) Telephone number: +34 971173143 Fax number: +34 971173184

Keywords

mtDNA; Control Region; Jews; Chuetas; haplogroups

Abstract

In the present work, we studied 109 unrelated Chueta individuals aiming to characterize their maternal gene pool composition. The entire mtDNA control region (1121 bp) was sequenced. The highest frequencies were found for haplogroups R0a and H, indicating a remarkable signature of Middle Eastern ancestry along with some degree of European admixture. These data confirm that Chuetas have been able to maintain some ancestral genetic identity. The significant differences with their host population should therefore be taken into account in forensic casework.

1. Introduction

The presence of Jews in the Balearic Islands is archaeologically documented since the 5th century AD, but the Jewish communities that can trace a historical continuity with today's Chuetas date back to the Muslim period (10th-13th centuries). Although the Christian conquest of Majorca in 1229 guaranteed the survival of the Jewish population, social and religious pressures forced their conversion between 1391 and 1435. Many of these converted Jews were assimilated into the general population, but a good few families continued to live in the ghetto and keep a secret adherence to Judaism. This Crypto-Jewish community was persecuted by the Inquisition (15th-17th centuries) (Braunstein, 1936) and, although the last "Autos de Fe" in 1691 put an end to their hidden Jewish religious practices, these convicts and their descendants (called Chuetas) were socially isolated and discriminated by their Majorcan neighbours until the mid-twentieth century. Therefore, Chuetas were an inbred population with scarce intermarriage with the Majorcan host population (Laub and Laub, 1987). The family names known as Chuetas in Mallorca are Aguiló, Bonnin, Cortés, Fortesa, Fuster, Martí, Miró, Picó, Piña, Pomar, Segura, Tarongí, Valentí, Valleriola and Valls, which do not have any relationship with Judaism anywhere but in Majorca.

Genetic similarities and differences among Jewish populations and between Jews and their host people have been widely studied in an attempt to unravel what must be a complex system of interrelations. Chuetas are, together with the Crypto-Jewish communities in Portugal (from Bragança and Belmonte) (Nogueiro et al., 2015b), the only direct descendants of the original Sephardic population; hence they are an interesting population to be studied due to their particular history.

Monoparental genetic markers have been widely used to further the knowledge of factors that have shaped modern human population's structure. Moreover, monoparental lineage studies enable to estimate the maternal and paternal contributions to the genetic pool of a particular population.

The aim of the present study was to study the mitochondrial DNA (mtDNA) haplogroups in a sample of Chueta individuals and to compare with other Jewish populations and with their host population, in order to investigate the founding maternal lineages in this population.

2. Material and methods

In the present work, we analysed 109 non-related Chueta individuals, with ancestors in the community until at least their third generation. Buccal cells collected on cytology brushes were obtained, under informed consent. DNA was extracted by standard phenol-chloroform method. Two mitochondrial DNA (mtDNA) fragments were amplified using mtDNA-

specific primers (L15997, H016, L16555, H639) and sequenced with ABI prism 3130 in order to obtain the entire Control Region (CR) (16024-576). Sequences were aligned against the revised Cambridge Reference Sequence (rCRS) using Genious software version 7.1.3 (Kearse et al., 2012). Haplogroups were classified following the updated mtDNA phylogeny, PhyloTree, mtDNA tree Build 16 (http://www.phylotree.org/) using Haplogrep tool (Kloss-Brandstätter et al., 2011) and assigned haplotypes were validated by EMPOP (EMP00672) (http://empop.org/) curators. Diversity parameters were estimated using the ARLEQUIN package version 3.5 (Excoffier and Lischer, 2010).

3. Results and Discussion

In the 109 individuals studied 50 different haplotypes were identified. The haplotype and nucleotide diversities of the complete CR were 0.952 and 0.010, respectively. Comparing with other Iberian Jewish population, the genetic diversity values were just slightly lower than in Bragança Crypto-Jews (Teixeira et al., 2011) and much higher than in the Portuguese Jewish community from Belmonte (Table 1), where a very strong founder maternal effect has been previously described (Behar et al., 2008b).

Table 1. Diversity parameters of HVRI (16024-16365bp) and HVRII (72-300bp) of the Chueta population and other Iberian populations with Jewish ancestry.

Parameters of Diversity	Chuetas (N=109)	Bragança Jews (N= 57)	Belmonte Jews (N=30)
K (%)	46 (42.20%)	35 (61.40%)	2 (6.67%)
S	75	61	6
$\hat{H}$ ± SD	0.950 ± 0.013	0.967 ± 0.012	0.129 ± 0.115
π ± SD	0.013 ± 0.007	0.014 ± 0.008	0.001 ±0.001
θ_K [95%CI]	29.477 [19.852, 43.505]	37.616 [22.328, 63.866]	0.279 [0.065, 1.097]

N, sample size; K, number of different haplotypes; S, number of polymorphic sites; $\hat{H}$, gene diversity; π, nucleotide diversity; θ_K, theta estimator based on the number of different haplotypes.

High haplogroup (Hg) diversity was found in Chuetas (Figure 1), with 35 different haplogroups. The Middle Eastern haplogroup R0a was the most prevalent (19.3%), followed by the widespread European haplogroup H (16.5%).

In order to identify traits of Jewish ancestry, we took into account those Hgs with frequencies higher than 4% and significantly different frequencies to those found in the Majorcan host population (unpublished data). The haplogroups that stood out in the Chueta sample were: R0a, T1a, T2c1d, K1a1b1a, U1a1a and L3e2b, indicating a remarkable signature of Middle Eastern ancestry along with some degree of European and North African admixture. The haplogroup pattern in Chuetas pointed out that the most important Jewish putative founding lineage is R0a, found in other Jewish (especially North Africans) and Middle Eastern populations, like Druzes, Palestinian and Bedouins, but not in Portuguese Jewish populations.

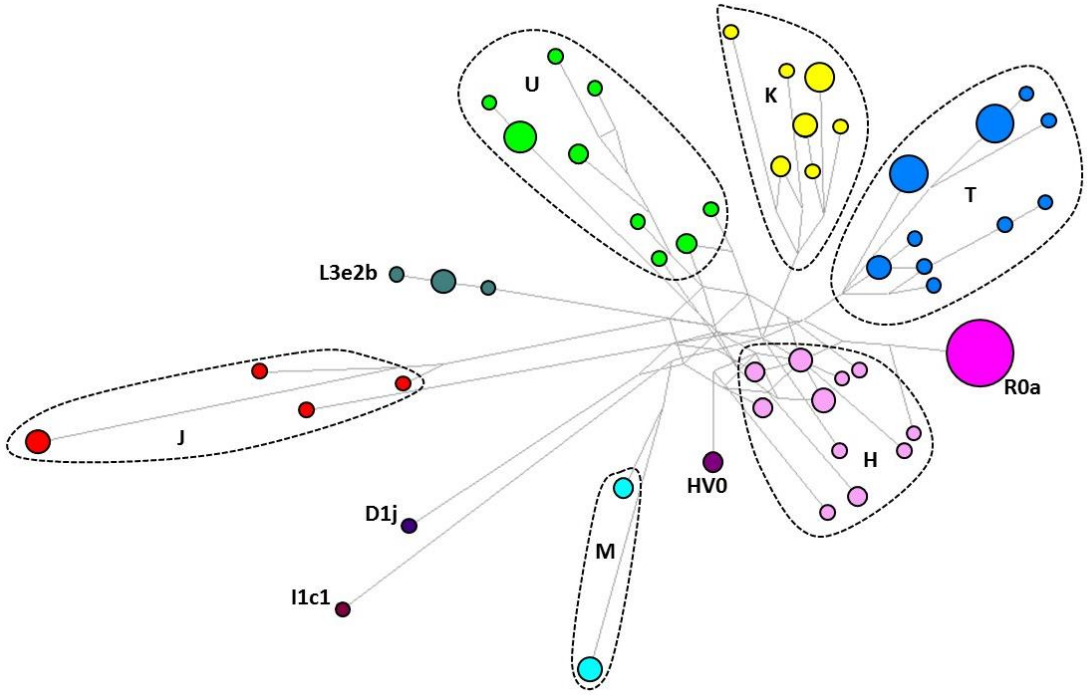


Figure 1. Network showing the haplogroup distribution of the 109 mtDNA Control Regions studied in Chuetas.

4. Conclusion

These data confirm that the Chuetas, due to their singular history, have kept not only the cultural memory of their Jewish origin through centuries but also a substantial degree of ancestral genetic signature. Also, some degree of the host admixture can be detected, as in other diaspora Jewish populations.

The significant differences observed between Chuetas and their host population should therefore be taken into account in forensic casework.

Acknowledments

This work was partially supported by grant AAEE24/2014 from the Direcció General de R+D+I (Comunitat Autònoma de les Illes Balears) and European regional Development Fund (ERDF). SLM is supported by FCT (grant PTDC/ATP-DEM/4545/2012) and financed by the European Social Funds (COMPETE-FEDER).

Conflict of interest statement

The authors declare no conflict of interest.

Maternal gene pool in Chuetas: Middle Eastern legacy and a novel sub-branching of the rare haplogroup R0a2m

(In preparation)

Key words: mitochondrial DNA, Chuetas, Middle Eastern, Haplotype, Haplogroup, Jewish.

1. Introduction

The first mitochondrial DNA (mtDNA) studies in Jewish populations (Thomas et al., 2002; Picornell et al., 2006; Brandstätter et al., 2008) revealed that most Jewish communities were founded by few women, that the process was independent and different in each geographical area, and that there were different rates of introgression from their corresponding host populations.

Most of the studies were mainly focused on the analysis of Ashkenazi Jews, a situation that has remained until today. As a major outcome, from the study of different parts of the mitochondrial genome, 4 women, together with several minor contributors, were identified as the founders of all the Ashkenazi Jews. However, there is some controversy regarding the identification of the origin of these 4 main contributors, with different studies claiming the Near East, while others suggest the Caucasus, Europe, or even the Far East through the Silk route (Behar et al., 2004b; 2006; Costa et al., 2013; Tian et al., 2015).

Behar et al. (2008b) widen the focus, analysing non-Ashkenazi communities, in order to investigate their demographic history, including how these were founded. Contrary to the Ashkenazim, they found no evidence for narrow founder effects for non-Ashkenazi populations: e.g. in Indian and Ethiopian communities an important local/host female contribution was found, while for North African and Iberian communities the contribution was admixed. These results led them to conclude that there are differences in the demographic history of the widespread communities resulting from the Jewish Diaspora in terms of maternal ancestries.

A specific group can be identified with its origin in Spain and Portugal: the Sephardic and Iberian Crypto-Jewish descendants – where the Crypto-Jewish phenomenon is defined as the secret adherence to Judaism while publicly professing another faith. Data from these populations still living in the Iberian Peninsula – which constitutes the original geographic source of Sephardic Jews – is very limited (Nogueiro et al., 2015b). The first study including Iberian Crypto-Jewish communities was performed by Behar et al. (2008b) on the Belmonte community, revealing a great founder effect in the community with extremely low diversity levels; unlike the analysis of the Bragança community, located in northeast Portugal, which found high levels of diversity, defining 5 haplogroups as founding lineages (Teixeira et al., 2011; Nogueiro et al., 2015a). For the descendants of Majorcan Jews, a

preliminary study was carried out (Picornell et al., 2005) using only the HVRI and a small number of samples, whose results also showed high rates of diversity in this community.

The presence of Jews in the Balearic Islands is archaeologically documented since the 5th century AD. Although the Christian conquest of Majorca in 1229 guaranteed the survival of the Jewish population, social and religious pressures forced their conversion between 1391 and 1435. Consequently, there were officially no more Jews in Majorca nearly 60 years before the Edict of Expulsion by the Catholic Kings in 1492. Many of these converted Jews were assimilated into the general population, but a good few families continued to live in the ghetto and keep a secret adherence to Judaism. This Crypto-Jewish community was persecuted by the Inquisition (15th-17th centuries) (Braunstein, 1936) and, although the last "*Autos de Fe*" in 1691 put an end to their hidden Jewish religious practices, these convicts and their descendants (called Chuetas) were socially isolated and discriminated by their Majorcan neighbours until the mid-twentieth century. Therefore, Chuetas were an inbred population with scarce intermarriage with the Majorcan host population (Laub and Laub, 1987; Porqueres, 2001).

In this study, we aim to evaluate the maternal genetic composition of the Majorcan Chueta population in order to infer its demographic history. Specifically, we attempt to identify and describe its founders and trace back their origins. Besides, we endeavour to determine the Majorcan host population's influence on the Chueta population gene pool.

2. Material and methods

2.1. Samples

In this study, 183 healthy non-maternally-related individuals were analysed: 104 from the Chueta community, and 79 from the Majorcan general population (with at least three generations born in Majorca). DNA was extracted by standard phenol-chloroform method, after obtaining the appropriate informed consent of all donors.

2.2. Mitochondrial DNA analysis

The mitochondrial DNA control region, or D-loop (16024-576), was amplified with two overlapping fragments using mtDNA-specific primers (L15997, H016 and L16555, H639) (Marques et al., 2015). The amplified product was purified with the MBS® Spin PCRapace (Invitek), and subsequent sequencing reactions were carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), following the manufacturer's procedures. Finally, products were run in an ABI prism 3130 analyser.

Sequences were assembled and compared to the revised Cambridge Reference Sequence (rCRS) (NC_012920) using Geneious software version 7.1.3 (Biomatters, Ltd).

Haplogroups were classified following the updated mtDNA phylogeny, PhyloTree, mtDNA tree Build 17 (http://www.phylotree.org/) using HaploGrep2 tool (van Oven, 2015; Weissensteiner et al., 2016), and assigned haplotypes were validated by EMPOP (EMP00672) (http://empop.org/) curators.

In order to obtain a fine grading sub-haplogroup classification in the samples identified as belonging to haplogroup H (18 in Chuetas and 32 in Majorca), a set of 51 SNPs located in the coding region were typed using a hierarchical strategy as described in Alvarez-Iglesias et al. (2009) and Marques et al. (2015) with the SNaPshot™ (Applied Biosystems) technique.

Moreover, the entire mtDNA molecule was obtained for the Chueta samples belonging to haplogroups showing differences with the host Majorcan population. Specifically, the mtDNA molecule of at least one of the samples belonging to haplogroups M5a1, M1a1, J2a1a1, T1a, T2c1d, U1a1a, K1a1b1a, K1a4a1a+195, and R0a+60.1T were fully sequenced. Amplification of the whole molecule was carried out by 19 overlapping fragments, and the sequencing strategy used 31 smaller fragments, primers, and specifications as described in Ramos et al. (2009; 2011). When Chueta specific mutations were identified in the complete mtDNA molecule for a particular haplogroup, the status of such nucleotide positions was interrogated in the rest of the samples belonging to the haplogroup.

2.3. Data analysis

Based on mtDNA diversity parameters such as the number of different haplotypes (K), number of polymorphic sites (S), haplotype diversity (H), nucleotide diversity (π), and theta estimator based on the number of different lineages (θk) were estimated using the ARLEQUIN package version 3.5 (Excoffier and Lischer, 2010) for the entire D-loop region (positions 16024 to 576) and in a shorter fragment (HVRI and HVRII; positions 16024 to 16365, and 72 to 300, respectively) for comparative purposes. ARLEQUIN was also used to assess the differences in haplogroup composition between Chuetas and other populations through F_{ST} test.

To attain statistical significance for frequencies of the putative Jewish founding lineages, we calculated Bayesian 0.90 credible region (90% CR) using SAMPLING software (Macaulay, personal communication). Furthermore, another criterion established by Behar et al. (2008b) was to consider haplogroups with a frequency greater than 5% and TMRCA previous to 2 Kya as founder lineages.

Median-joining networks were constructed for the complete mtDNA sequences, using other literature mitogenomes, with the Network software v 4.6.1.1 (Fluxus Technology Ltd.). For the construction of networks, Indels in position 309, 315, and 523-524 were not taken into

account. In order to adjust the weight of the mutations, the weights described in Bandelt et al. (2006) were used. Additionally, detailed maximum parsimony trees of complete sequences and complete sequences available in the public databases NCBI-GenBank, empop.org and mtDNA Community were also performed.

Coalescence times of R0a2m and R0a2m1 subclades were estimated using the ρ statistic (Forster et al., 1996) accompanied by a heuristic estimate of the standard error (σ) calculated from an estimate of the genealogy (Saillard et al., 2000). Calculations were made considering all substitutions except those at np 16519 and 16182C and 16183C, because they are considered hotspot mutations. Mutational distances were converted into years using the substitution rate for the entire mitogenome of about one mutation every 3,624 years, and correcting for purifying selection using the calculator provided by Soares et al. (2009).

In order to visualise the distribution of mtDNA haplogroup R0a frequencies in the Mediterranean geographic context, an isometric spatial frequency distribution map was constructed with the program Surfer 9 (Golden Software, http://www.goldensoftware.com/products/surfer).

2.4. Populations

Inter-population comparisons were performed using European, African, Middle Eastern, and Jewish populations (Supplementary Table 1). For phylogenetic analyses, all the samples found in mtDNA Community (Behar et al., 2012), PhyloTree (van Oven, 2015), and Genbank until December 2015 belonging to each haplogroup studied, were included.

3. Results and discussion

3.1. Haplotype diversity

Complete control region haplotypes, as well as the assigned haplogroups for all individuals of Majorcan and Chueta populations, are presented in Supplementary Tables 2 and 3. In the 104 samples from the Chueta population, 50 (48.08%) different haplotypes were identified; meanwhile, in the 79 Majorcan samples, 67 (84.81%) different haplotypes were found. Calculated diversity estimators are summarised in Table 1.

Theta k values (θk), as expected, are lower when obtained from HVRI and HVRII than when the complete D-loop is used. The longer the studied region, the more precise the information obtained concerning maternal lineage diversity, and the more accurate the estimate of the female effective population size. Estimated θk in Chuetas is much lower than in Majorcans, but in the same range as the ones in other non-Ashkenazi Jews. Therefore, the estimated number of putative female founders in Chuetas is similar to the one estimated in most of these Jewish groups.

Table 1. Diversity indices results calculated for the complete D-loop and for HVRI + HVRII fragment for inter-population comparison.

Population	N	K (% K)	S	HD ± SD	Π ± SD	θk (95% CI)
D-loop (16024-576)						
Chuetas	104	50 (48.08%)	110	0.950 ± 0.015	0.010 ± 0.005	37.216 (25.061;55.103)
Majorca	79	67 (84.81%)	120	0.995 ± 0.003	0.009 ± 0.005	205.807 (117.911;374.184)
Miranda do Douro[1]	121	60 (49.59%)	102	0.979 ± 0.005	0.007 ± 0.000	NA
Zamora[1]	214	144 (67.29%)	150	0.991 ± 0.003	0.008 ± 0.0047	NA
Egypt[2]	277	238 (85.92%)	228	0.999 ± NA	NA	NA
HVRI (16024-16365) and HVRII (72-300)						
Chuetas	104	46 (44.23%)	74	0.948 ± 0.014	0.013 ± 0.007	31.003 (20.769;46.042)
Bragança Jews[3]	57	35 (61.40%)	61	0.967 ± 0.012	0.014 ± 0.008	37.616 (22.328;63.866)
Belmonte Jews[4]	30	2 (6.67%)	6	0.129 ± 0.115	0.001 ± 0.001	0.279 (0.065;1.097)
Bulgarian Jews[4]	71	46 (64.79%)	70	0.982 ± 0.007	0.012 ± 0.007	55.477 (34.470;90.159)
Turkish Jews[4]	123	85 (69.11%)	109	0.985 ± 0.005	0.013 ± 0.007	120.394 (82.638;176.996)
Libyan Jews[4]	83	36 (43.37%)	63	0.922 ± 0.019	0.013 ± 0.007	23.631 (15.104;36.717)
Moroccan Jews[4]	148	80 (54.05%)	92	0.979 ± 0.005	0.012 ± 0.006	70.307 (50.484;97.941)
Tunisian Jews[4]	36	25 (69.44%)	43	0.971 ± 0.015	0.012 ± 0.007	34.939 (18.044;69.325)
Iranian Jews[4]	82	43 (52.44%)	76	0.971 ± 0.008	0.016 ± 0.008	35.821 (23.098;55.513)
Iraqi Jews[4]	134	48 (35.82%)	79	0.950 ± 0.009	0.016 ± 0.008	26.344 (18.151;37.920)

K: Number of different haplotypes; S: number of polymorphic sites; HD: haplotype diversity; Π: nucleotide diversity averaged over loci; θk: theta estimator based on the number of different lineages; NA: Not Available. (1. Mairal et al., 2013; 2. Saunier et al., 2009; 3. Nogueiro et al., 2015a; 4. Behar et al., 2008b)

3.2. Haplogroup composition

The Majorcan population had haplogroup H as the most common one, 39.24% of the samples, as was expected of a typical European population (e.g. Picornell et al., 2005; Brandestätter et al., 2007; Karachanak et al., 2012; Mairal et al., 2013). Haplogroup H together with haplogroups U, K, and HV (frequencies ranging from 12.66% to 13.92%) accounted for 78.48% of the total diversity. Other haplogroups found in this population were I, J, L, N, T, V, and X (Table 2).

In contrast with the Majorcan population, in the Chueta samples, haplogroup H only accounted for 17.31% of the total diversity. The modal haplogroup in this population (20.19%) was found to be the Middle Eastern haplogroup R0a+60.1T, followed by haplogroups T, K, U, and J (19.23%, 11.54%, 7.69%, and 5.77%, respectively). Together these 5 haplogroups represented 64.42% of all variation. The remaining lineages were observed at frequencies ranging from 0.96 to 4.81% (Table 2).

Both population gene pools are mainly composed of West Eurasian lineages, with the exception of the East Asian haplogroup M found in Chuetas (3.85%), and the African haplogroup L (4.81% and 2.53% in Chuetas and Majorcans, respectively).

Table 2. Haplogroup frequencies in Chueta and Majorcan populations.

Haplogroup	Chueta	Majorca	Haplogroup	Chueta	Majorca	Haplogroup	Chueta	Majorca
D1j	0.0096	-	K1a	0.0385	0.0380	T2b23	0.0192	-
H*	-	0.0633	K1a1b1a	0.0385	-	T2b5a1	0.0096	-
H1	0.0865	0.1772	K1a4a1a+195	0.0096	0.0127	T2c1d	0.0481	0.0127
H2a2a	0.0385	0.0253	K1b1a1+199	0.0096	-	U1a1a	0.0481	-
H3	0.0288	0.0380	K1b1+16093	-	0.0127	U2e1'2'3	-	0.0127
H4a1a	-	0.0253	K1c	0.0096	-	U2e1e	-	0.0127
H6a1	0.0096	0.0380	K2a5	-	0.0127	U2e2a2	-	0.0127
H11a	0.0096	0.0253	K2b1a1	0.0096	0.0506	U3	0.0192	-
HV0	-	0.0759	L2a1b+143	-	0.0127	U3a	0.0096	-
HV0 + 195	0.0192	0.0380	L3d1b2	-	0.0127	U4b3	-	0.0127
HV15	-	0.0127	L3e2b+152	0.0481	-	U5a2	0.0096	-
HV4a2a	-	0.0127	M1a1	0.0192	-	U5b1d2	0.0096	-
I	-	0.0127	M5a1	0.0192	-	U5b1f1a	0.0192	-
I1c1	0.0096	-	N1b1	-	0.0127	U5b2a2	-	0.0127
I2'3	-	0.0127	R0a+60.1T	0.2019	-	U5b2b1a1	-	0.0127
J1b1a1	-	0.0127	T	0.0288	0.0127	U5b2b3	-	0.0253
J1b1b	-	0.0127	T1a	0.0577	-	U5b3	0.0096	0.0253
J1c	-	0.0127	T1a1'3	0.0096	-	U6a	0.0096	-
J1c2o	0.0096	-	T2	0.0096	0.0127	V+16298	-	0.0127
J1d1	0.0096	-	T2a1b	-	0.0253	X2c	-	0.0127
J2a1a1	0.0288	-	T2b	0.0096	-	Unclassified	-	0.0127
J2b1a	0.0096	0.0127						

Since haplogroup H is difficult to classify according only to D-loop motifs (branch defining mutations) using tools such as HaploGrep (Marques et al., 2015), and taking into account that this haplogroup also has a high prevalence (10 to 30%) in the Middle East and the Caucasus (Roostalu et al., 2007), it was important to obtain a fine grade classification of its sub-branches (Supplementary Table 4). The sub-haplogroup H distribution in Chueta and Majorcan populations is presented in Figure 1. H1, H2, H3, H6a, and H11 haplogroups were present in both populations. However, whereas H1, H2, and H3 had higher percentages in Chuetas than in Majorcans; H6a and H11 had the opposite behaviour, with higher frequencies in Majorcan populations than in Chuetas. Finally, H* and H4a were only found in the Majorcan population. The overall H haplogroup frequency in Chuetas is similar to or even lower than other Jewish communities such as Ashkenazim, North African Jews, and Sephardic communities (Behar et al., 2004b; 2008b; Nogueiro et al., 2015a).

In Figure 1, two samples that were initially defined as N according to the D-loop region haplotype (X36 and X68), but reclassified within haplogroup H2a2a when the complete genome was sequenced, have been included.

All in all, the haplogroup composition indicates that Chuetas are statistically different from their host population (Majorca), and also different from other Jewish populations (F_{ST} p-values < 10^{-5}). With the SAMPLING analyses, and also with the Behar et al. (2008) criterion, two haplogroups showed up as the founder lineages in the Chueta population: R0a+60.1T (20.19%) and T1a (5.77%).

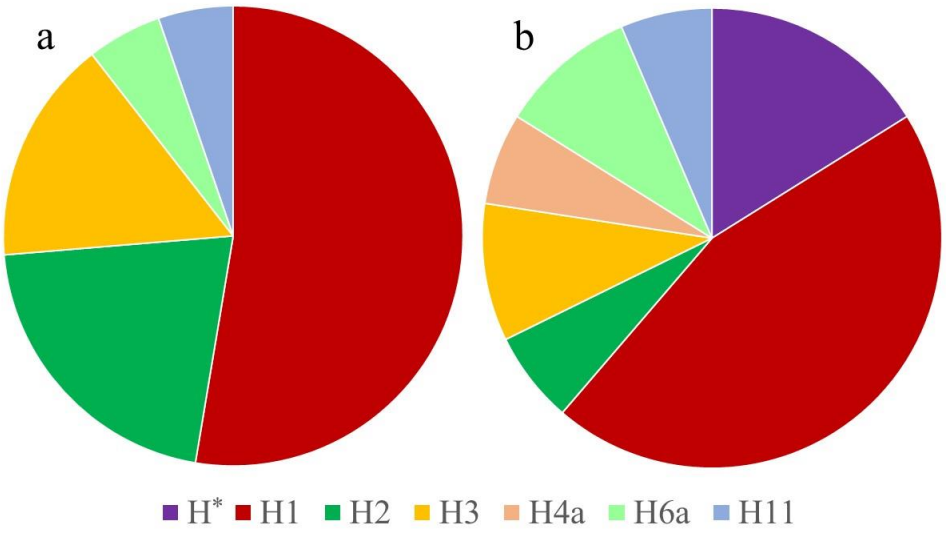


Figure 1. H sub-haplogroup distribution in Chueta (a) and Majorcan (b) populations.

A comparison of these founding lineages with those of other non-Ashkenazi Jewish populations confirmed the idea that each Jewish community was founded by different females, as previously described in the literature. For instance, while HV0b is found as a founder in Crypto-Jewish communities in Belmonte and Bragança (Behar et al., 2008b; Nogueiro et al., 2015a), in Chuetas and other Sephardic groups this haplogroup is absent or very uncommon; additionally, haplogroup K1a1b1a, which is a founder (8.5%) in the Iberian Exile Jewish community from Bulgaria, has a lower frequency than in other populations with Sephardic origin (3.85% in Chuetas and 0.81% in Turkey Jews) (Behar et al., 2008b). Therefore, two possible scenarios could explain these findings: either a lack of homogeneity in maternal lineages of the original Sephardic groups that settled in different areas of the Iberian Peninsula (and Balearic Islands), or genetic drift in the current populations, resulting in a lack of lineages that do not allow us to infer the original mtDNA composition of the Sephardic Jews that lived in Spain and Portugal from the surviving lines.

3.3. Phylogenetic comparison from complete genomes

In order to go further into the knowledge, origin, and distribution of the most interesting haplogroups, complete molecules of samples belonging to haplogroups M5a1, M1a1, J2a1a1, T1a, T2c1d, U1a1a, K1a1b1a, K1a4a1a+195, and R0a+60.1T were obtained. Mitogenome haplotypes are listed in Supplementary Table 5.

- Haplogroup M

All the non-African mtDNA lineages descended from two L3 subclades (M and N). The time when the split of the two macrohaplogroups was originated and what route was followed remain controversial (Kivisild, 2015). While some authors claim its origin can be

dated back 62 to 95 Kya (Fu et al., 2013) others give later times (40-70 Kya) (Macaulay et al., 2005; Soares et al., 2009). The most accepted route was following a Southern coastal route across Arabia and India to reach Australia shortly after, but Marrero et al. (2016) proposed that it could have happened along simultaneous routes: 1) the Northern route through the Levant to reach South Asia, the Philippines, and nearly Oceania; and 2) secondary expansions, northward through Asia to the Americas, and Southwest to North Africa and Europe.

Haplogroup M5 is exclusive to South Asia and typical of India. This haplogroup has also been related to Romani populations around Europe, due to their Indian origin (Mendizabal et al., 2011). Analysing the Chueta M5a1b genome together with those in the literature (Network analysis in Supplementary Figure 1), we were able to identify that it presents an identical haplotype to those previously described in Gómez-Carballa et al. (2013) in Spanish Romani samples, and defined as M5a1b1a1. The presence of this haplogroup in the Chueta population could be explained because, due to their difficulty marrying people outside the community, one marriage strategy was to attempt to marry outsiders to whom the status of Chueta meant nothing. Accordingly, there are some cases reported of marriages with wandering entertainers who arrived on the island, who used to be Romani nomadic people (Porqueres, personal communication).

Haplogroup M1 is predominantly African specific. Although it is present in North Africa, it is more common in East Africa, mainly in the Chad basin (Olivieri et al., 2006; Cerezo et al., 2016). Since haplogroup M1a has been related to Jewish populations (González et al., 2007) we analysed one of the Chueta M1a1 samples together with the complete genomes found in the literature (Network analysis in Supplementary Figure 2). No Jewish complete sequences clustered with the M1a1 Chueta sample. The Chueta sample shares the mutation 10506G with a Cretan (Greece) sample (Olivieri et al., 2006); however, this is not enough to conclude that they could pertain to a new not previously defined Mediterranean branch.

- Haplogroup JT

Haplogroup JT derives from the macrohaplogroup R which itself derives from N. The time of its appearance can be estimated ~58 Kya, before the settlement of the Fertile Crescent. It has been suggested that haplogroups J and T diverged during the settlement ~40 Kya and ~30 Kya respectively. Haplogroup J has higher frequencies in the Middle East and Arabia than in Europe (13-20% vs. 9%) while haplogroup T shows the opposite behaviour (10% in Europe and 8% in the Middle East) (Pala et al., 2012; Fernandes et al., 2015).

Two Chueta samples were classified as belonging to haplogroup J2a1a1. The complete molecule of one of them was sequenced and analysed together with the available data in the literature. According to the network this haplogroup is clearly of European origin

(Supplementary Figure 3) as was already proposed by Pala et al. (2012) and Fernandes et al. (2015). However, the Chueta sample does not cluster in any of the currently defined subhaplogroups. We could think of a case of introgression from the host population, but despite the fact that haplogroup J frequencies are similar (~5-6 %) in the Chueta and Majorcan populations, they do not reveal shared haplotypes. Therefore, it is more likely that the assimilation of this haplogroup into the Chueta community occurred before their arrival on the island.

T1a (5.77%) is considered one of the founders of the Chueta population, as indicated previously. This haplogroup originated in the Near East ~17 Kya, although most of its sub-branches seem to be European (Pala et al., 2012). When one complete molecule, of the six T1a samples detected, was analysed together with those in the literature, it remained as haplogroup T1a* without clustering in any of the sub-branches described (Network analysis in Supplementary Figure 4). The Chueta sample has two mutations downstream of the T1a phylogeny: 6656T and 10116G. To infer whether all of the T1a samples in Chuetas could be identical, we typed these two mutations in the five remaining samples. 10116G was not found in any of the samples, whereas 6656T was present in all of them. Another sample of the literature (JQ703693) from Lithuania has the 6656T mutation, so a new mtDNA branch could be proposed in the future when more entire molecules are sequenced. The fact that all the samples of the Chueta populations have this mutation enables us to suggest a founder effect, but one old enough to have had time to diverge (since only one sample has the 10116G mutation, opening the likelihood of more variation between all six samples). Complete genome sequencing and dating of the "new" branch would be needed to clarify its origin and diversity.

One Majorcan sample and 5 Chuetas were classified as T2c1d. Haplogroup T2c1 originated in the Near East, ~18.5 Kya, and later expanded to Europe in the Neolithic expansion (Pala et al., 2012). One of the T2c1d Chueta's complete genome and the Majorcan one, were analysed together in a phylogenetic network with the T2c1 samples from the literature (Supplementary Figure 5). Moreover, a phylogenetic tree was constructed using the T2c1d samples alone (Figure 2). Both samples displayed the 11914A! mutation which classified them in the T2c1d1 branch. Majorcan samples also had the 12363T mutation, characteristic of the T2c1d1a branch. This latest mutation was not present in the Chueta sample, instead, it presented a private mutation 8475T and another two mutations 13056T and 14544A shared with JQ704020 sample (from Galicia, Spain). The presence of 13056T and 14544A mutations was tested and confirmed in the remaining four Chueta samples, therefore a new sub-branch could be proposed. Moreover, exclusive Chueta mutation 8475T was tested in the other samples and was found to be present in all of them. Private Majorcan mutations (3027C, 8911C, 8980A and 15569T) were also tested but were absent in all Chueta samples.

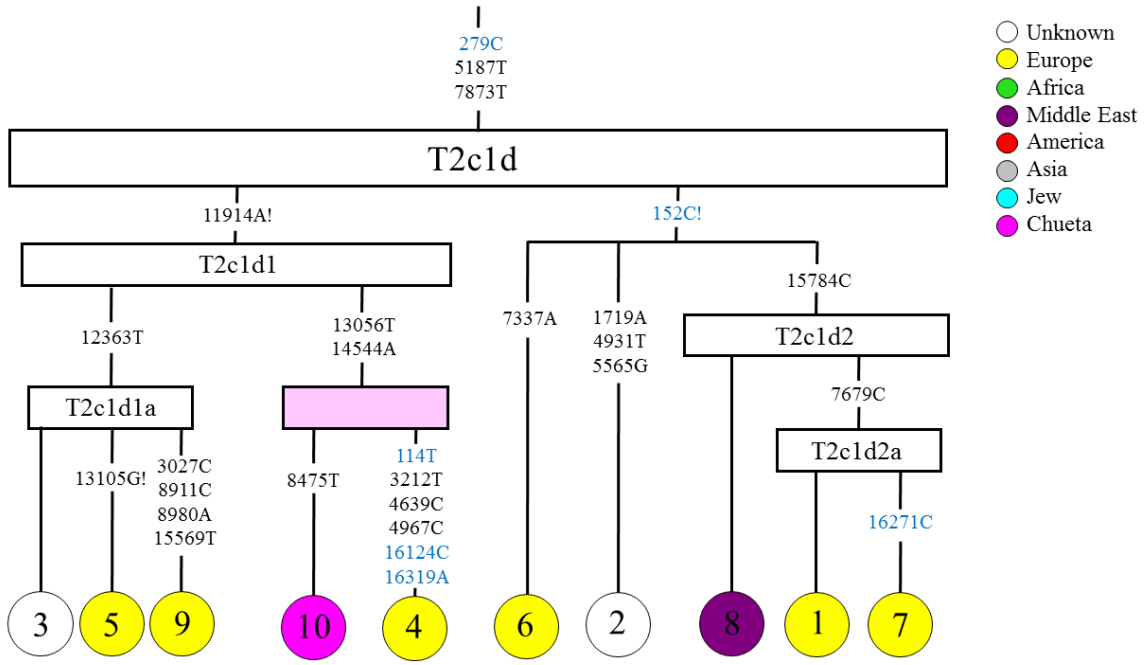


Figure 2. Phylogenetic tree of haplogroup T2c1d. A proposal of a new branch is coloured in pink. (1. DQ523629; 2. JF833037; 3. JN580589; 4. JQ704020; 5. JQ798092; 6. JQ798094; 7. JQ798095; 8. JQ798096; 9. M33 and 10. X139).

- Haplogroup U

Soares et al. (2010) estimated the origin of the West Eurasian haplogroup U (which is also a derivate of the macrohaplogroup R) at ~53 Kya. Haplogroup U is the second most frequent in modern Europeans and was predominant in pre-agricultural Europe (Davidovic et al., 2017). The U1a sub-haplogroup is dated at ~13-15 Kya and is present in Southwest and South Asia, the Caucasus, and Europe.

Five U1a1a1 samples were found in the Chueta population while no presence of the U1 sub-haplogroup was observed in the host population. When the complete genome of one of the samples was analysed together with other U1a1 complete genomes from the literature in a phylogenetic network (Supplementary Figure 6) we observed that sub-haplogroup U1a1a was made up of European samples, but there was also one Jewish and one Middle Eastern sample. To further explore relationships between U1a1a samples, a phylogenetic tree was built (Figure 3): four samples – a Middle Eastern sample (KC477757), a Jewish sample (EF556161), a European sample from Lucca, Italy (HQ615282), and a sample of unknown origin (HM156682) – clustered together with our Chueta sample, while the other European samples pertained to other sub-branches.

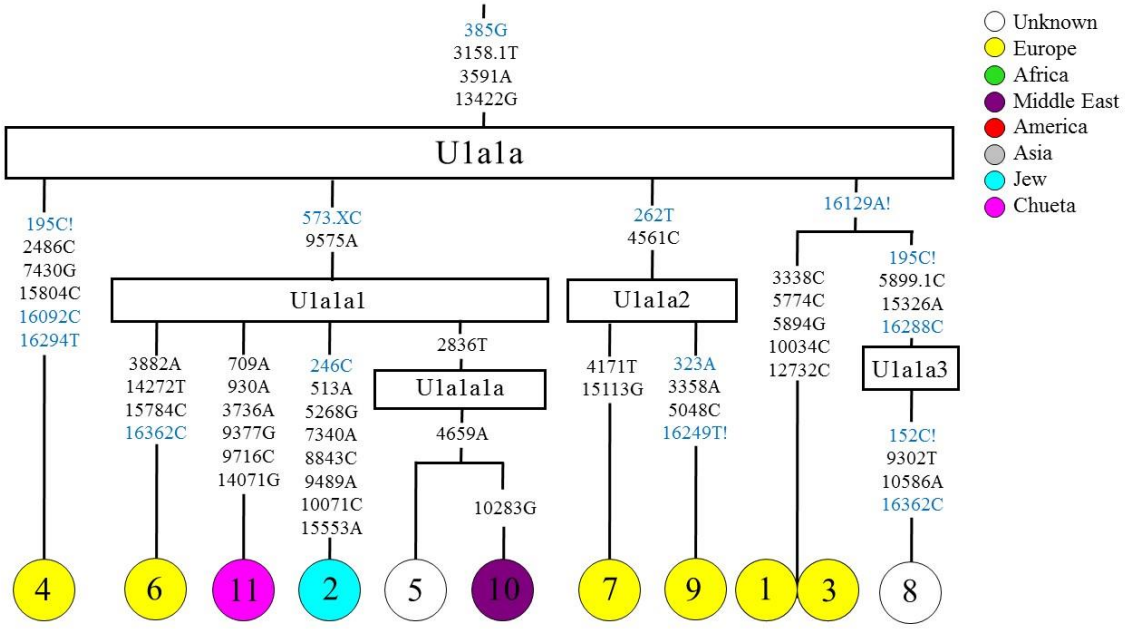


Figure 3. Phylogenetic tree of haplogroup U1a1a (1. AY882396; 2. EF556161; 3. EU597497; 4. GU218692; 5. HM156682; 6. HQ615882; 7. JQ703793; 8. JQ705601; 9. JX289842; 10. KC477757 and 11. X02).

- Haplogroup K

Haplogroup K is the main branch of U8 and its origin has been dated to ~36 Kya. The place of origin is still under discussion (Behar et al., 2006, Soares et al., 2010; Costa et al., 2013), but a Levantine origin seems the most likely. K splits into two branches (K1 and K2), and the main clade of K1, K1a, is dated to ~20 Kya.

The founder Ashkenazi lineage K1a1b1a (Behar et al., 2006), dated to ~4.4 Kya (Costa et al., 2013), also present in Sephardic communities (Behar et al., 2008b), was found in four Chueta samples. This haplogroup is absent in non-European Jews, which can be seen as evidence of its European origin (Costa et al., 2013). Due to its relation to the Jewish populations, and the absence of the haplogroup in the host population, the complete genome of two of the four samples was sequenced. Analysed together in a phylogenetic network with other available sequences, we observed that this haplogroup has a star-like shape. Both Chueta samples have identical haplotypes not shared with any other sample in the network (Supplementary Figure 7). Moreover, these samples showed a private mutation when compared with the basal haplotype, 8029T. This mutation was checked in the other two remaining K1a1b1a samples, and was found to be present in both. The presence of a private mutation in all the Chueta samples could suggest an identical haplotype of all the samples, probably resulting from a founder effect in this community.

Another K1a lineage, K1a4a1, which has a European Neolithic origin, arrived from the Near East ~8 Kya (Costa et al., 2013). Analysing the whole genome together with the available K1a4a1 sequences in the literature (Network analysis in Supplementary Figure 8)

we observe that this is a European haplogroup with presence also in Africa. Taking into account the network, a new cluster could be proposed for 4 samples: two European, one Caucasian, and one from the USA, sharing the 10398G mutation.

- Haplogroup R0a

Haplogroup R0a is a subclade of macrohaplogroup R. Due to the geographic distribution of its frequencies, its origin in terms of location and timescale has been under debate in the recent years (Abu-Amero et al., 2007; Černý et al., 2011; Gandini et al. 2016). As shown in Figure 4, this haplogroup is practically absent in Europe (frequencies ranging from 0 to 2% approximately), although some exceptions are found in a few populations, such as Capadoccia (Italy) (14.61%) (Messina et al., 2015), and Chuetas (20.19%). In Gandini et al. (2016) the proportion described of R0a in the Balearic Islands is ~5%, the highest in Europe according to their dataset, but the reason for these high values is the inclusion of previously preliminary Chueta results based only on HVRI (Picornell et al., 2005) as part of the autochthonous Balearic population.

The highest frequencies of haplogroup R0a are found in the Arabian Peninsula and the Horn of Africa, reaching values as high as ~25% in Soqotra island in Yemen (Černý et al., 2011). Frequencies in Jewish groups (Thomas et al., 2000; Picornell et al., 2006, Behar et al., 2008b; Černý et al., 2011) in general show similar frequencies to their host populations (Figure 4), with the exceptions of Iranian Jews who display a lower frequency, and the exceptional case of the Chuetas.

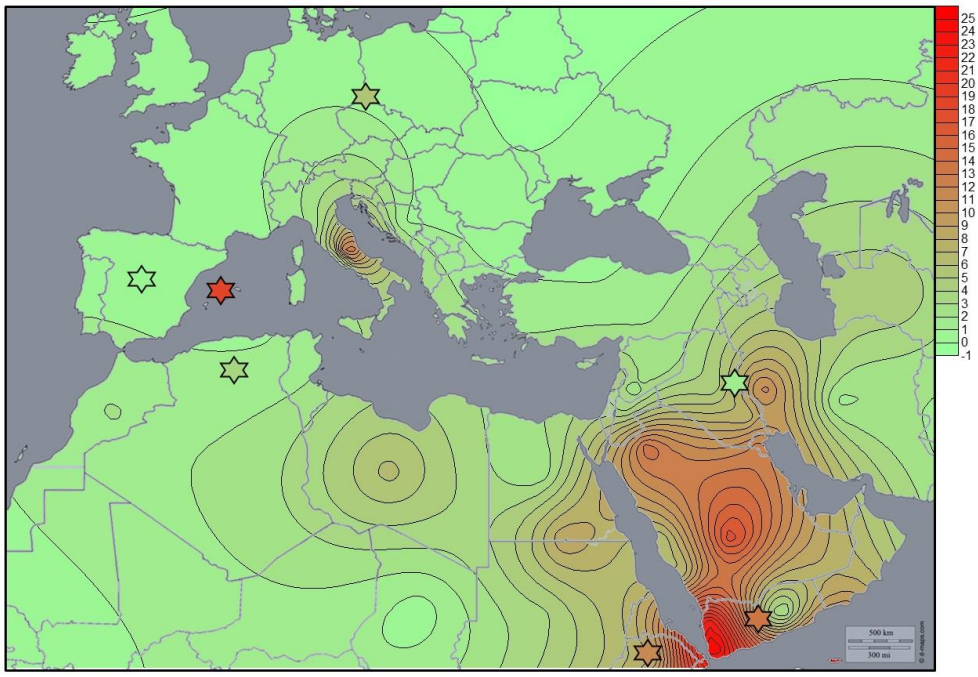


Figure 4. Isofrequency map of haplogroup R0a based on data from the literature (Supplementary Table 1) generated by Surfer v.8 (Golden Software Inc., Golden, Colorado). Jewish populations are indicated with a Star of David.

The first dating of the haplogroup (~19 Kya) suggested an Arabian origin (Abu-Amero et al., 2007). Later studies dated the haplogroup earlier, ~22.5 Kya (Černý et al., 2011) and ~30 Kya (Gandini et al., 2016). Both studies discuss whether the origin could be in the Horn of Africa or the Arabian Peninsula. Phylogeographic differences in the regional distribution of R0a throughout the Arabian Peninsula and East Africa, and the fact that the most ancient reservoir of R0a variation is found in Arabia, led the authors to conclude an Arabian origin of the haplogroup.

Two main branches characterise this haplogroup: R0a1 (~26 Kya) and R0a2'3 (~21 Kya). R0a1 dispersed from Arabia to Africa (especially North) with the Muslim conquest, and is also found in the Near East. R0a2 (~17 Kya) expanded much further across the Red Sea into the Horn of Africa (Gandini et al., 2016). The main branch of R0a1, R0a1a, can be identified by the 16355T transition. R0a2'3 is defined by the insertion 60.1T. Following this phylogenetic criterion, a map showing the sub-haplogroups of R0a – in the populations where it is present and the complete D-loop was studied – was drawn (Figure 5). R0a1a is more prevalent in the Middle East, whereas most of the Jewish populations where information was available showed the sub-haplogroup R0a2'3. Only the Yemenite Jews show both (Behar et al., 2008b).

Taking into account the high prevalence of this haplogroup in the Chueta population, it can be considered its main maternal founder. Therefore, it was important to delve into the phylogeny of our samples. To do so, for 5 out of the 21 R0a+60.1T samples, the complete genome was obtained, which classified the Chueta samples as R0a2m. In addition to the five Chueta sequences, nine R0a2m samples belonging to the Family tree initiative (www.familytreedna.com; Greenspan, personal communication) were added to the analyses. These nine samples belonged to eight Ashkenazi Jews and an Australian with known maternal Chueta ancestors. A phylogenetic network with other published R0a2 complete sequences was performed (Supplementary Figure 9). The R0a2 network reflects the prevalence of the haplogroup in the Middle Eastern and African populations (mainly from Arabia Saudi, Yemen, Ethiopia, and Sudan), while very few European samples (three of them Italian) have it. In the literature, just three samples cluster within the R0a2m branch: two Jewish samples from the Czech Republic and the Ukraine (JQ705916 and JQ705196), and another sample from Poland with unknown ethnicity (JQ703505). While most of the Jews cluster together in the R0a2m branch, three other Jews (from Tunisia, Yemen, and Ethiopia) appear together with non-Jewish samples from the same locations.

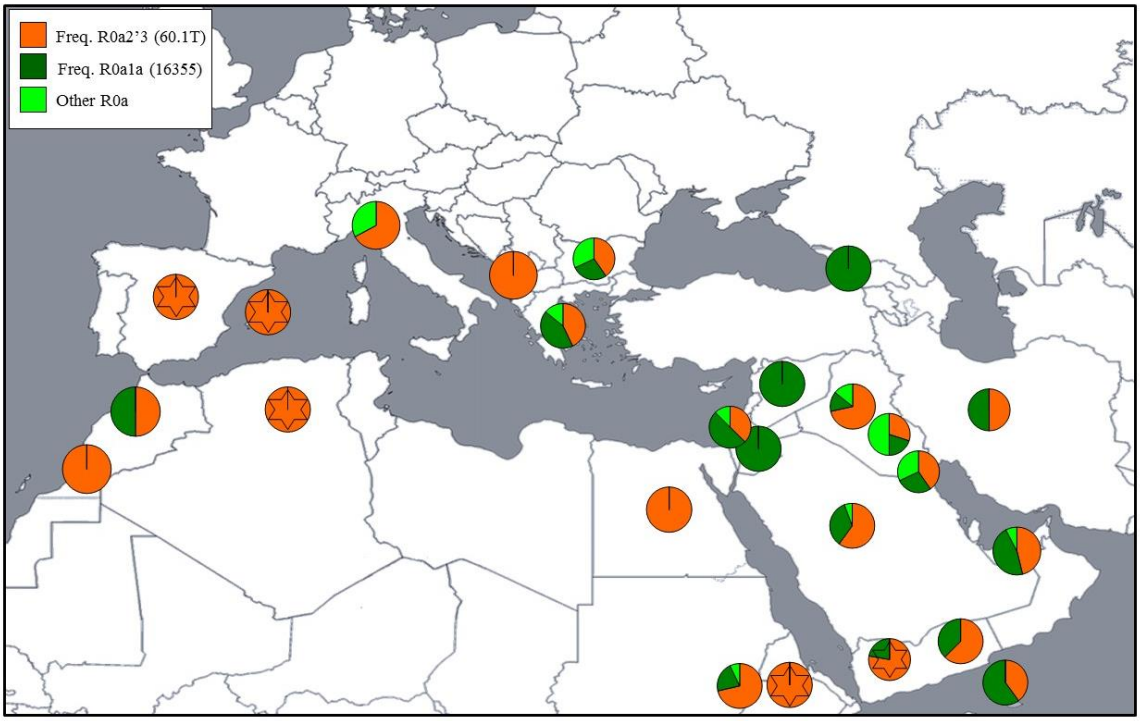


Figure 5. Distribution of R0a sub-haplogroups in data set where 16355 and 60.1 positions were analysed. Populations used are indicated in Supplementary Table 1. Jewish populations are indicated with a Star of David.

The R0a2m branch was dated by Gandini et al. (2016) to ~1.41 Kya (ML) and 1.29 Kya (ρ). Since this dating was performed with just three samples, we performed the dating by adding the new samples. With a total of 17 R0a2m samples the dating was established at ~1.36 [0-3.25] Kya (ρ), which is in accordance with the previously calculated values. This date coincides with the 7th century; so, taking into account that all the samples but one are from Jewish origin, this branch might possibly have originated in a post Diaspora Jewish community. With the data available, it seems logical to consider the Ashkenazi community, but we must be aware that few studies on complete mitochondrial genomes have been performed so far in non-Ashkenazi communities; thus, more R0a2m samples might well be found in other Jewish communities. Following historical evidence, this haplogroup is more likely to have arrived on the island through Sephardic or North African Jews than by Ashkenazim, but contact between Sephardic and Ashkenazi communities were not uncommon (Roth and Novella, 1979; Mea 2007) and introgression of other mtDNA lineages between both Jewish groups has previously been described (Nogueiro et al., 2015a). Bearing in mind the fact that the settlement of some German Jews in the Majorcan Jewry in the Middle Ages is documented (Pons, 1984), introgression from the Ashkenazi R0a2m lineage into Chuetas cannot be rejected.

A phylogenetic tree of the R0a2m sub-haplogroup was drawn (Figure 6). The haplogroup has six samples with the basal haplotype and the others differ in only one or two positions. All the Chueta samples (including the Family Tree sample with Chueta ancestry) share the 13858G mutation. Since six different samples have the variant, a new sub-branch named

R0a2m1 could be proposed. Five of the six samples show an identical haplotype (without taking into account 309 and 315 Indels and 16519 hotspot position) whereas one sample has an additional 15734A private mutation. For the 16 Chueta R0a+60.1T samples where the complete genome was not sequenced, these two new mutations were checked. The R0a2m1 defining position, 13858G, was present in all the samples, while none of them revealed the 15734A mutation. It is important to highlight that this rare newly described branch (R0a2m1) contains 22 Chueta samples, so it can be identified as an exclusively Chueta branch.

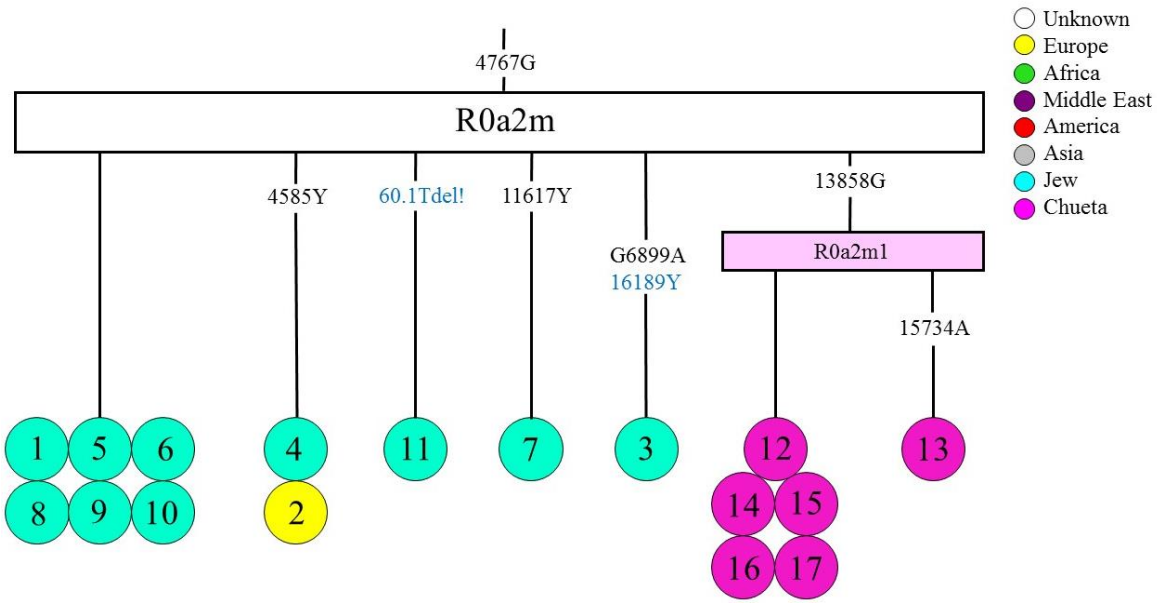


Figure 6: Phylogenetic tree of the R0a2m haplogroup. A proposal of a new branch is labelled in pink. (1. JQ705916; 2. JQ703503; 3. JQ705196; 4. Family tree 1 (FT1); 5. FT2; 6. FT3; 7. FT4; 8. FT5; 9. FT6; 10. FT7; 11. FT8; 12. FT9; 13. X24; 14. X70; 15. X127; 16. X132; 17. X142).

With the six complete molecules of the newly described R0a2m1 sub-branch, dating calculations were performed, obtaining a time of origin of ~0.43 [0-1.30] Kya (ρ). This value points towards a very recent origin of the branch, in the late 16th century. This time is an indicator that this new branch appeared on the island when the Crypto-Jew community was still present. In the last quarter of 17th century, with the last actions of the Inquisition, a bottleneck was produced because many of the members of the Crypto-Jewish community were prosecuted, resulting in condemns and emigrations. The people who survived these prosecutions and remained in Majorca are the ancestors of the current Chueta population, so this could explain how this specific branch became the most prevalent in our current population. Tajima's D results (-1.5557, p-value= 0.024) would support an expansion after the bottleneck mentioned before.

4. Conclusions

Gene diversity in maternal lineages in Chueta population was lower than in other populations, but higher than expected in a small, endogamous population. It remains to be explained what mating strategies were undertaken by this community to avoid the expected impoverishment of heterogeneity in their gene pool.

The Chuetas' haplogroup composition indicated a remarkable signature of Middle Eastern ancestry, with R0a and T1a being putative founding lineages; as well as the presence of other haplogroups found in Jewish/Middle Eastern populations (K1a1b1a and U1a1a1); and the low frequency of H, the most frequent haplogroup in Europe. These data confirm that the Chuetas have kept not only the cultural memory of their Jewish origin over centuries, but also a substantial degree of ancestral genetic signature.

Regarding the host population admixture, even though both populations (Chuetas and Majorcan) present quite similar percentages in haplogroups such as U or K, which could be a sign of introgression of the host population in the Chuetas' gene pool, the fact that only a few haplotypes (in the D-loop region) are shared raises the question of the amount of admixture, unsolved. The presence of haplogroup L (4.81%) in Chuetas also indicates some degree of North African introgression. In order to better define the amount and origin of introgression, a complete genome analysis of the host population and other Jewish communities would be enlightening.

The hallmark in the maternal gene pool in Chuetas is a new sub-branching of the rare haplogroup R0a2m (R0a2m1), originated very recently in this population, which has become their modal haplotype.

Supplementary Table 1. Populations used in the inter-population comparison, and their references.

Jews	Ashkenazi	Picornell et al. (2006) Int. J. Legal Med. 120: 271–281
	Chuetas	Present study
	Ethiopian Jews	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420; Behar et al. (2008) PLoS One 3: e2062
	Iranian Jews	Behar et al. (2008) PLoS One 3: e2062
	North African Jews	Picornell et al. (2006) Int. J. Legal Med. 120: 271–281; Behar et al. (2008) PLoS One 3: e2062
	Sephardic	Behar et al. (2008) PLoS One 3: e2062
	Yemen Jews	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420; Behar et al. (2008) PLoS One 3: e2062; Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
Europe	Balkans	Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231
	Bulgaria	Karachanak et al. (2012) Int. J. Legal Med. 126: 497–503
	Cappadocia (Italy)	Messina et al. (2015) Am. J. Hum. Biol. 27: 508–519
	France	Badro et al. (2013) PloS One 8(1): e54616
	Galicia (Spain)	Santos et al. (2014) Am. J. Hum. Biol. 26: 130–141
	Georgia	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420; Quintana-Murci et al. (2004) Am. J. Hum. Genet. 74: 827–845
	Greece	Badro et al. (2013) PloS One 8(1): e54616
	Majorca (Spain)	Present study
	Piglio (Italy)	Messina et al. (2015) Am. J. Hum. Biol. 27: 508–519
	Portugal Centre	Santos et al. (2014) Am. J. Hum. Biol. 26: 130–141
	Portugal North	Santos et al. (2014) Am. J. Hum. Biol. 26: 130–141
	Portugal south	Santos et al. (2014) Am. J. Hum. Biol. 26: 130–141
	Saracinesco (Italy)	Messina et al. (2015) Am. J. Hum. Biol. 27: 508–519
	Tuscan (Italy)	Achilli et al. (2007) Am. J. Hum. Genet. 80(4): 759-768
	Valencia (spain)	Santos et al. (2014) Am. J. Hum. Biol. 26: 130–141
Middle East	Amman	Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231
	Bedouin	Behar et al. (2008) PLoS One 3: e2062; Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Druze	Behar et al. (2008) PLoS One 3: e2062; Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231
	Hadram (Saudi Arabia)	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Hadramawt (Yemen)	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Hajja (Yemen)	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Iran	Derenko et al. (2013) PLoS 8(11): e80673
	Iraq	Al-Zaheri et al. (2011) BMC Evol. Biol. 11: 288; Badro et al. (2013) PloS One 8(1): e54616
	Israeli Arabs	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420
	Jordan	Badro et al. (2013) PloS One 8(1): e54616
	Kalash	Quintana-Murci et al. (2004) Am. J. Hum. Genet. 74: 827–845
	Kuwait	Scheible et al. (2011) Forensic Sci. Int. Genet. 5: e112–e113
	Lebanon	Badro et al. (2013) PloS one. 8(1): e54616
	Marsh Arabs	Al-Zaheri et al. (2011) BMC Evol. Biol. 11: 288
	Pakistan (Karachi)	Quintana-Murci et al. (2004) Am. J. Hum. Genet. 74: 827–845
	Palestine	Behar et al. (2008) PLoS One 3: e2062; Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231; Badro et al. (2013) PloS one. 8(1): e54616
	Pathans	Quintana-Murci et al. (2004) Am. J. Hum. Genet. 74: 827–845
	Persian (Iran)	Derenko et al. (2013) PLoS 8(11): e80673
	Qashqais (Iran)	Derenko et al. (2013) PLoS 8(11): e80673
	Saudi Arabia	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78; Badro et al. (2013) PloS One 8(1): e54616
	Sudan	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Tajikstan	Ovchinnikov et al. (2014) Legal Med. 16: 390–395
	Tihama (Saudi Arabia)	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Turkmenistan	Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231
	Turkey	Quintana-Murci et al. (2004) Am. J. Hum. Genet. 74: 827–845; Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231
	Yemen	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420; Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78; Badro et al. (2013) PloS One 8(1): e54616
	Zagros Mountain (Iran)	Quintana-Murci et al. (2004) Am. J. Hum. Genet. 74: 827–845
Africa	Chad	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Egypt	Badro et al. (2013) PloS One 8(1): e54616; Elmadawy et al. (2013) Legal Med. 15: 338–341
	Ethiopia	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420; Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78; Badro et al. (2013) PloS one. 8(1): e54616
	Kenya	Badro et al. (2013) PloS One 8(1): e54616
	Libya	Badro et al. (2013) PloS One 8(1): e54616
	Morocco	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78; Aboukhalid et al. (2013) Int. J. Legal Med. 127: 757–759; Badro et al. (2013) PloS One 8(1): e54616
	Nubia	Gonzalez et al. (2008) Ann. Hum. Biol. 35(2): 212-231
	Socotra (Yemen)	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Syria	Thomas et al. (2002) Am. J. Hum. Genet. 70: 1411–1420; Badro et al. (2013) PloS One 8(1): e54616
	Taizz (Yemen)	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78
	Tunisia	Černý et al. (2011) Mol. Biol. Evol. 28(1): 71-78

Supplementary Table 2. D-loop haplotypes and haplogroups of the 104 Chueta samples, classified according to HaploGrep2 (Build 17, PhyloTree).

Sample	Haplogroup	Haplotypes																	
X2	U1a1a1	16182C	16183C	16189C	16249C	73G	263G	285T	309.1C	309.2C	315.1C	385G	d523	d524	573.1C	573.2C	573.3C		
X3	H1a3	16051G	16162G	16519C	73G	263G	315.1C	534T											
X4	U5b1f1a	16192T	16270T	16319A	73G	150T	263G	315.1C	533G										
X10	H1bo	16189C	16519C	263G	267C	315.1C	485C												
X11	U5b3	16192T	16270T	16304C	16526A	73G	150A	228A	263G	309.1C	315.1C								
X12	H1+152	152C	263G	309.1C	309.2C	315.1C	466C												
X15	K1a1b1a	16223T	16224C	16234T	16311C	16519C	73G	114T	263G	315.1C	497T								
X16	K1a	16129A	16224C	16256T	16311C	16519C	73G	263G	315.1C	497T									
X17	H66a	16172C	16519C	263G	315.1C														
X19	H2a2a	16519C	263G	315.1C															
X21	T1a	16126C	16163G	16186T	16189C	16294T	16298C	16319A	16519C	73G	263G	309.1C	315.1C						
X22	H66a	16172C	16519C	263G	315.1C														
X23	M5a1	16129A	16223T	16291T	16298C	16519C	73G	263G	315.1C	489C	524.1A	524.2C							
X24	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X25	K1c	16224C	16311C	16519C	73G	146C	152C	263G	309.1C	315.1C	d498								
X26	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X27	H1+152	152C	263G	309.1C	309.2C	315.1C	466C												
X28	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X29	U3	16343G	73G	150T	263G	309.1C	315.1C												
X30	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X31	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X32	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X33	H2a2a	16519C	263G	309.1C	315.1C														
X34	T2b	16126C	16294T	16296T	16304C	16519C	73G	263G	309.1C	309.2C	315.1C								
X35	T2c1d	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	315.1C								
X36	N/H	16223T	16519C	263G	315.1C														
X37	J1d1	16069T	16126C	16193T	16300G	73G	152C	195C	263G	295T	309.1C	315.1C	462T	489C					
X38	H2a2a	16519C	263G	309.1C	315.1C	d523	d524												
X39	H1j8	16129A	16240G	16519C	152C	185A	263G	309.1C	315.1C										
X40	U6a	16092C	16172C	16219G	16278T	73G	263G	315.1C											
X42	T	16126C	16294T	16304C	16519C	73G	263G	309.1C	309.2C	315.1C									
X43	H11a2	16092C	16169T	16293G	16298C	16311C	263G	315.1C											
X45	T2c1d	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	309.2C	315.1C							
X46	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X47	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C									
X48	I1c1	16129A	16223T	16264T	16270T	16311C	16319A	16362C	16391A	16519C	73G	199C	204C	250C	263G	309.1C	315.1C	455.1T	573.1C
X49	M5a1	16129A	16223T	16291T	16298C	16519C	73G	263G	315.1C	489C	524.1A	524.2C							
X52	M1a1	16129A	16183C	16189C	16193.1C	16223T	16249C	16311C	16359C	16519C	73G	195C	263G	315.1C	489C				

X53	J2a1a1	16069T	16126C	16145A	16231C	16261T	73G	150T	152C	195C	203A	215G	263G	295T	315.1C	319C	489C	513A
X54	K1a	16129A	16224C	16256T	16311C	16519C	73G	263G	315.1C	497T								
X55	J2a1a1	16069T	16126C	16145A	16231C	16261T	73G	150T	152C	195C	203A	215G	263G	295T	315.1C	319C	489C	513A
X56	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X57	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X58	K1a1b1a	16223T	16224C	16234T	16311C	16519C	73G	114T	263G	309.1C	315.1C	497T						
X59	L3e2b+152	16172C	16183C	16189C	16320T	16519C	73G	150T	152C	195C	263G	315.1C						
X60	H1e1a6	16147T	16264T	16519C	150T	263G	315.1C											
X61	T	16126C	16294T	16304C	16519C	73G	263G	309.1C	315.1C									
X62	K1a	16129A	16224C	16256T	16311C	16519C	73G	263G	315.1C	497T								
X63	U1a1a1	16182C	16183C	16189C	16193.1C	16249C	73G	263G	285T	309.1C	309.2C	315.1C	385G	d523	d524	573.1C	573.2C	
X64	D1j	16223T	16242T	16311C	16325C	16362C	73G	152C	235G	263G	309.1C	315.1C	489C					
X65	U3	16343G	73G	150T	263G	315.1C												
X66	K1b1a1+199	16093C	16224C	16311C	16319A	16463G	16519C	73G	152C	199C	263G	309.1C	315.1C	524.1A	524.2C	524.3A	524.4C	
X67	U3a	16343G	16390A	16519C	73G	150T	263G	309.1C	309.2C	315.1C								
X68	N/H	16223T	16519C	263G	315.1C													
X69	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X70	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X71	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X104	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X105	L3e2b+152	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C					
X106	U5b1d2	16239T	16270T	73G	150T	263G	315.1C											
X107	K1a	16129A	16224C	16311C	16519C	73G	263G	315.1C	497T									
X108	T2c1d	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	315.1C							
X109	U1a1a1	16182C	16183C	16189C	16249C	73G	263G	285T	309.1C	309.2C	315.1C	385G	d523	d524	573.1C	573.2C		
X110	L3e2b+152	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C					
X111	T2b23	16126C	16147T	16294T	16296T	16297C	16304C	16519C	73G	263G	309.1C	309.2C	315.1C					
X112	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X113	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X115	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X117	M1a1	16129A	16183C	16189C	16223T	16249C	16311C	16359C	16519C	73G	195C	263G	315.1C	489C				
X118	K1a4a1	16129A	16224C	16256T	16311C	16519C	73G	263G	315.1C	497T								
X120	U1a1a1	16182C	16183C	16189C	16249C	73G	263G	285T	309.1C	309.2C	315.1C	385G	d523	d524	573.1C	573.2C		
X126	H1j8	16129A	16240G	16519C	152C	185A	263G	309.1C	315.1C									
X127	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X128	T2c1d	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	315.1C							
X129	T1a	16126C	16163G	16186T	16189C	16294T	16298C	16301T	16319A	16519C	73G	263G	309.1C	315.1C				
X130	T1a	16126C	16163G	16186T	16189C	16294T	16298C	16319A	16519C	73G	263G	309.1C	315.1C					
X131	HV0+195	16298C	72C	195C	263G	309.1C	309.2C	315.1C										
X132	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X133	K1a1b1a	16223T	16224C	16234T	16311C	16519C	73G	114T	263G	315.1C	497T							

X134	U5b1f1a	16192T	16270T	16319A	73G	150T	263G	315.1C	533G									
X135	T1a	16126C	16163G	16186T	16189C	16294T	16298C	16319A	16519C	73G	263G	309.1C	315.1C					
X136	H66a	16172C	16519C	263G	315.1C													
X137	J2b1a	16069T	16126C	16193T	16278T	73G	150T	152C	263G	295T	315.1C	489C	523d	524d				
X139	T2c1d	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	315.1C							
X140	T1a1'3	16126C	16163G	16186T	16189C	16294T	16519C	16527T	73G	152C	195C	263G	309.1C	315.1C	573.1C			
X141	K2b1a1a	16222T	16224C	16270T	16311C	16519C	73G	146C	195C	263G	315.1C							
X142	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								
X143	T2b23	16126C	16147T	16294T	16296T	16297C	16304C	73G	263G	309.1C	309.2C	315.1C						
X144	U5a2	16192T	16256T	16270T	16526A	73G	263G	309.1C	315.1C									
X145	L3e2b+152	16172C	16183C	16189C	16223T	16320T	16335G	16519C	73G	150T	152C	195C	263G	315.1C				
X146	H6	16126C	16362C	16482G	16519C	239C	263G	309.1C	315.1C									
X147	T1a	16126C	16163G	16186T	16189C	16294T	16298C	16319A	16519C	73G	263G	309.1C	315.1C					
X148	T2b5a1	16126C	16294T	16304C	16519C	73G	152C	263G	309.1C	309.2C	315.1C	573.1C						
X149	H1n6	16519C	263G	309.1C	315.1C	552A												
X150	J2a1a1	16069T	16126C	16145A	16231C	16261T	73G	150T	152C	195C	203A	215G	263G	295T	315.1C	319C	489C	513A
X151	T2	16093C	16126C	16294T	16296T	16519C	73G	263G	315.1C									
X153	J1c2o	16069T	16126C	16163G	16266T	16311C	16519C	73G	185A	188G	204C	228A	263G	295T	315.1C	462T	489C	
X154	T	16126C	16294T	16304C	16519C	73G	263G	309.1C	309.2C	315.1C								
X155	K1a1b1a	16223T	16224C	16234T	16311C	16519C	73G	114T	263G	315.1C	497T							
X156	T1a	16126C	16163G	16186T	16189C	16294T	16298C	16319A	16519C	73G	263G	309.1C	315.1C					
X157	U1a1a1	16182C	16183C	16189C	16249C	16545C	73G	263G	285T	309.1C	309.2C	315.1C	385G	523d	524d	573.1C	573.2C	573.3C
X199	HV0+195	16298C	72C	195C	263G	309.1C	315.1C											
X200	L3e2b+152	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C					
X201	R0a+60.1T	16126C	16362C	16519C	58C	60.1T	64T	263G	309.1C	315.1C								

Supplementary Table 3. D-loop haplotypes and haplogroups of the 79 Majorcan samples, classified according to HaploGrep2 (Build 17, PhyloTree).

Sample	Haplogroup	Haplotypes																
M01	HV0	16298C	72C	263G	295T	315.1C												
M02	H1+152	152C	263G	309.1C	315.1C	466C												
M03	T2	16126C	16240G	16296T	16519C	73G	146C	263G	315.1C									
M04	X2c	16093C	16169T	16183C	16189C	16223T	16255A	16278T	16519C	73G	153G	195C	225A	227G	263G	315.1C		
M05	H13a1a2a	16278T	16519C	263G	309.1C	315.1C												
M06	U5b2b1a1	16270T	16292T	16362C	16366T	73G	150T	263G	309.1C	315.1C								
M07	H1bo	16519C	263G	267C	315.1C	485C												
M08	K1a	16129A	16224C	16311C	16519C	73G	263G	315.1C	497T									
M09	K1a4a1a+195	16093C	16224C	16311C	16519C	73G	195C	263G	315.1C	497T	d523	d524						
M10	H1ag1a	16183C	16189C	16218T	16256T	16519C	263G	315.1C										
M11	U5b2a2	16189C	16192T	16270T	16398A	73G	263G	315.1C										
M13	U2e1e	16051G	16111T	16129C	16145A	16183C	16189C	16193.1C	16362C	16519C	73G	152C	217C	263G	309.1C	309.2C	315.1C	340T
M14	H1e5	16256T	16519C	263G	315.1C													
M15	H11a	16257T	16293G	16311C	93G	195C	263G	315.1C										
M17	K2b1a1	16224C	16270T	16311C	16519C	73G	146C	263G	315.1C									
M18	H1bv1	16362C	16519C	263G	315.1C													
M20	I2´3	16129A	16174T	16223T	16261T	16391A	16519C	73G	152C	199C	204C	207A	250C	263G	309.1C	309.2C	315.1C	
M21	H3z	16189C	16294T	16362C	16519C	263G	293C	309.1C	315.1C									
M22	H1+16189	16189C	16519C	56.1G	263G	315.1C												
M23	V+@16298	72C	263G	315.1C														
M24	H1bv1	16104T	16362C	16519C	152C	263G	315.1C											
M25	H1+152	16519C	152C	263G	309.1C	309.2C	315.1C											
M26	H1+16189	16189C	16519C	263G	309.1C	315.1C												
M27	HV0	16298C	72C	263G	309.1C	315.1C	d522	d523										
M28	H1m1	16519C	146C	263G	315.1C													
M29	U5b2b3	16224C	16270T	16519C	73G	150T	263G	309.1C	315.1C	517T								
M30	H3w	16248T	16304C	146C	263G	309.1C	309.2C	315.1C										
M31	J2b1a	16069T	16093C	16126C	16193T	16278T	73G	150T	152C	263G	295T	315.1C	489C					
M32	I	16129A	16223T	16262T	16391A	16519C	73G	199C	204C	250C	263G	315.1C						
M33	T2c1d	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	315.1C							
M34	K2b1a1	16224C	16270T	16311C	16519C	73G	146C	263G	315.1C									
M35	K2b1a1	16224C	16270T	16311C	16519C	73G	146C	263G	315.1C	560.1C	573.1C							
M37	HV0	16298C	64T	72C	263G	309.1C	315.1C	498.1C	d522	d523								
M38	K1a	16224C	16311C	16519C	73G	263G	315.1C	497T	523.1C	523.2A	523.3C	523.4A	523.5C	523.6A				
M39	HV0+195	16298C	72C	195C	263G	309.1C	315.1C	573.1C										
M40	H1+152	16261G	16519C	152C	263G	309.1C	315.1C											
M41	U2e2a2	16051G	16092C	16129C	16183C	16189C	16362C	16519C	16525G	73G	152C	217C	263G	315.1C	508G			
M42	H57	16519C	64T	93G	146C	263G	315.1C											
M43	U5b2b3	16224C	16270T	73G	150T	152C	263G	315.1C	517T									

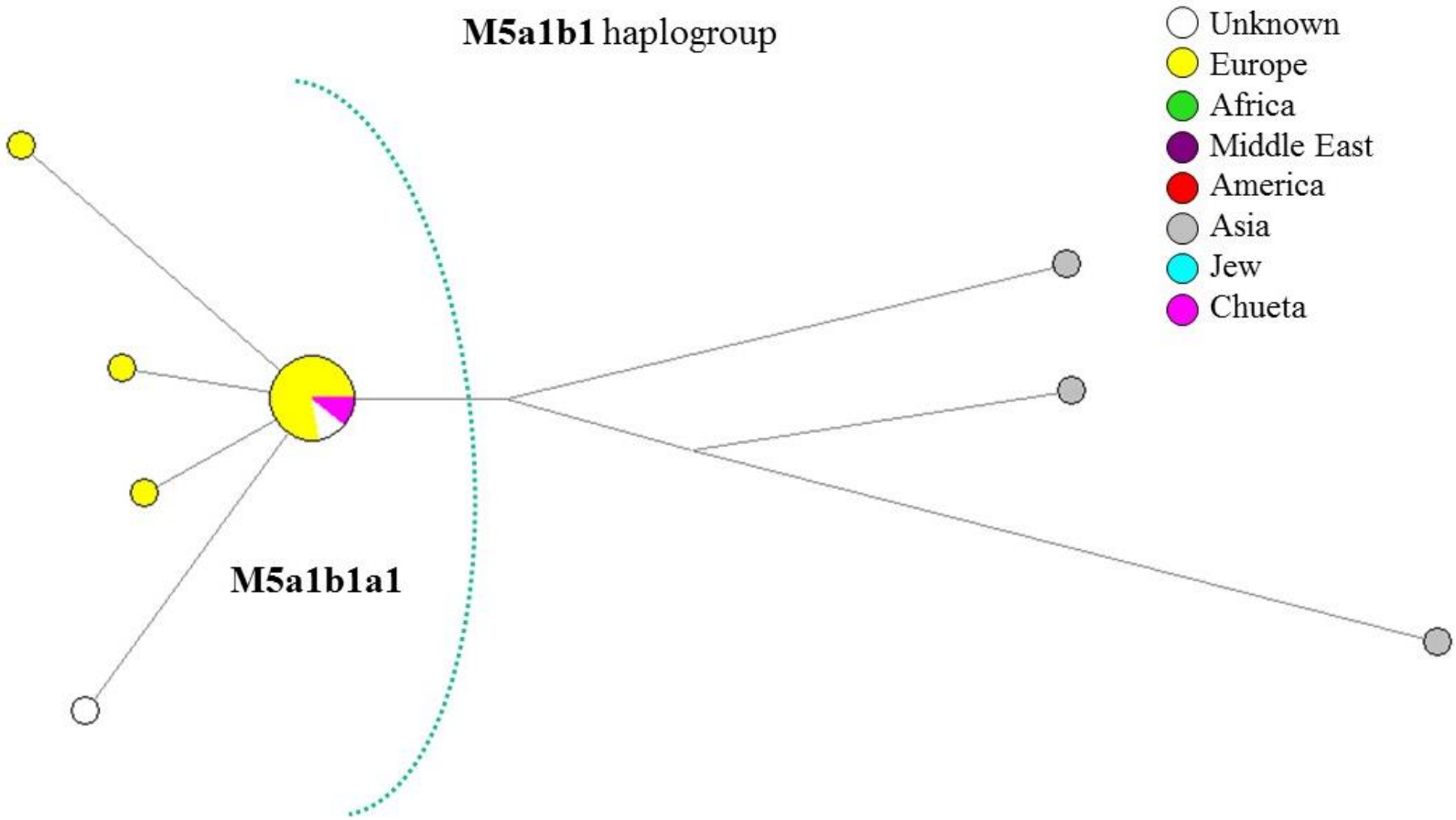
M44	J1b1b	16069T	16126C	16145A	16261T	16263C	16519C	73G	263G	271T	295T	309.1C	315.1C	462T	489C	d522	d523	
M45	HV0	16298C	16311C	72C	73G	263G	309.1C	309.2C	315.1C									
M46	H2a2a	16519C	263G	315.1C														
M47	J1c	16069T	16126C	73G	228A	263G	295T	309.1C	315.1C	462T	489C	537.1C						
M48	HV15	16129A	16234T	16311C	16519C	263G	309.1C	315.1C										
M49	U5b3	16270T	16304C	16399G	64T	73G	150T	204C	228A	263G	315.1C	573.1C						
M50	H24	16293G	309.1C	315.1C														
M51	K1a	16224C	16304C	16311C	16519C	73G	263G	315.1C	497T									
M52	K2a5	16224C	16311C	16519C	73G	146C	152C	263G	315.1C	324T								
M53	T2a1b	16126C	16218T	16294T	16296T	16324C	16519C	73G	263G	315.1C								
M54	N1b1	16145A	16176G	16187T	16223T	16311C	16390A	16519C	73G	152C	263G	315.1C						
M55	L3d1b2	16124C	16223T	16380.1C	16519C	73G	150T	152C	263G	309.1C	315.1C	d523	d524					
M56	T	16126C	16294T	16304C	16519C	73G	93G	199C	263G	315.1C								
M57	L2a1b +143	16189C	16192T	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C		
M58	H5s	16111T	16304C	16311C	16391A	16519C	263G	309.1C	309.2C	315.1C	456T							
M59	H3ak	16519C	143A	263G	309.1C	309.2C	315.1C											
M60	U4b3	16278T	16356C	16519C	73G	195C	215G	263G	309.1C	315.1C	499A	524.1A	524.2C					
M61	H11a2	16092C	16257T	16293G	16311C	93G	195C	263G	315.1C									
M62	K2b1a1a	16222T	16224C	16270T	16311C	16519C	73G	146C	195C	263G	315.1C							
M63	HV0	16298C	72C	263G	309.1C	315.1C												
M64	K1b1 +(16093)	16093C	16192T	16224C	16311C	16319A	16519C	73G	152C	263G	309.1C	315.1C	524.1A	524.2C				
M65	H1m1	16519C	146C	263G	315.1C													
M66	H1e +16129	16129A	16311C	152C	263G	309.1C	315.1C	d523	d524									
M67	HV0 +195	16298C	72C	195C	263G	309.1C	315.1C											
M68	H24	16293G	309.1C	315.1C														
M69	HV0	16298C	72C	263G	309.1C	315.1C												
M70	H1j8	16129A	16240G	16519C	152C	185A	263G	309.1C	315.1C									
M71	H6	16362C	16482G	16545C	239C	263G	309.1C	315.1C										
M72	HV0 +195	16298C	72C	195C	228A	263G	309.1C	315.1C										
M73	J1b1a1	16069T	16092C	16126C	16145A	16172C	16222T	16261T	73G	146C	242T	263G	295T	315.1C	462T	489C		
M74	H1e +16129	16129A	16519C	263G	315.1C													
M75	H1m1	16519C	146C	263G	315.1C													
M76	U2e1'2'3	16051G	16129C	16183C	16189C	16274A	16362C	16519C	73G	152C	217C	263G	309.1C	309.2C	315.1C	508G	523.1C	523.2A
M77	T2a1b	16092C	16126C	16294T	16296T	16324C	16519C	73G	263G	315.1C								
M78	H2a5a1	16291T	16519C	217C	263G	315.1C												
M79	H6	16362C	16482G	239C	263G	309.1C	315.1C											
M80	HV4a2a	16287T	16519C	263G	309.1C	315.1C												
M81	H6	16362C	16482G	239C	263G	315.1C												
M104	H1e +16129	16129A	16311C	152C	263G	309.1C	315.1C	d523	d524									
M198	U5b3	16192T	16270T	16304C	16526A	73G	150T	228A	263G	309.1C	315.1C							

Supplementary Table 4. Multiplex strategy used for each sample with respective coding region mutations found Haplogroup: classification obtained after coding region sub-typing. Haplogroup names following Phylotree (Build 17).

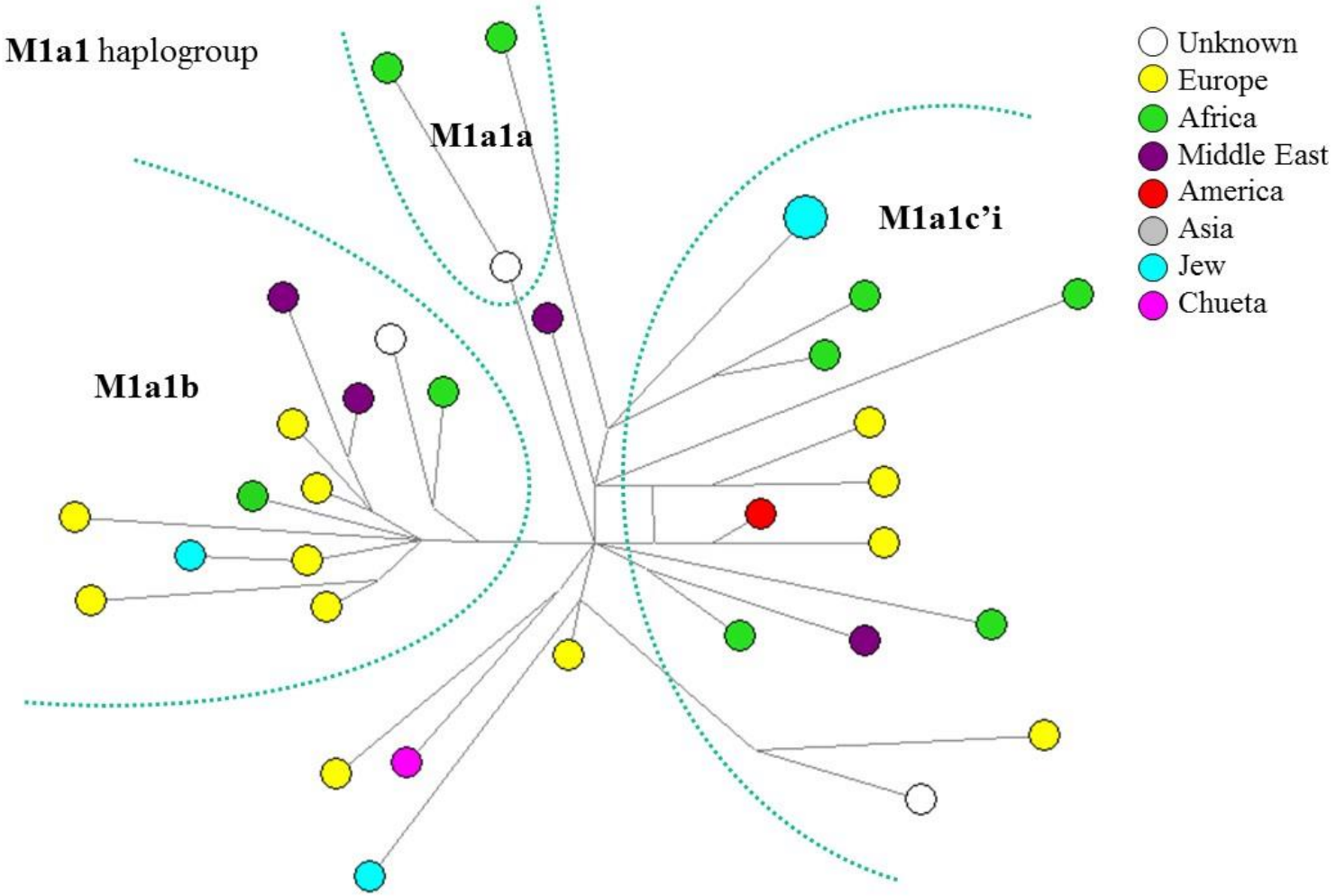
Population	Sample	Mx.1	Mx.2	Mx.3	Haplogroup	Coding region mutations					
Chueta	X03	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X10	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X12	x	x	x	H2a2a	No mutations detected					
Chueta	X17	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X19	x	x		H3	750G	1438G	4769G	6776C		
Chueta	X22	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X27	x	x	x	H2a2a	No mutations detected					
Chueta	X33	x	x		H3	750G	1438G	4769G	6776C		
Chueta	X36	x	x	x	H2a2a	No mutations detected					
Chueta	X38	x	x		H3	750G	1438G	4769G	6776C		
Chueta	X39	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X43	x	x		H11a	750G	961G	1438G	4769G		
Chueta	X60	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X68	x	x	x	H2a2a	No mutations detected					
Chueta	X126	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X136	x	x		H1	750G	1438G	3010A	4769G		
Chueta	X146	x	x		H6a1	750G	1438G	3915A	4727G	4769G	
Chueta	X149	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M02	x	x	x	H*	750G	1438G	4769G			
Majorca	M05	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M07	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M10	x	x	x	H*	750G	1438G	4769G			
Majorca	M14	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M15	x	x		H11a	750G	961G	1438G	4769G		
Majorca	M18	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M21	x	x		H3	750G	1438G	4769G	6776C		
Majorca	M22	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M24	x	x		H3	750G	1438G	4769G	6776C		
Majorca	M25	x	x		H3	750G	1438G	4769G	6776C		
Majorca	M26	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M28	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M30	x	x		No H	7028T					
Majorca	M40	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M42	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M46	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M50	x	x	x	H2a2a	No mutations detected					
Majorca	M58	x	x	x	H*	750G	1438G	4769G			
Majorca	M59	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M61	x	x		H11a	750G	961G	1438G	4769G		
Majorca	M65	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M66	x	x		H4a1a1a	750G	3992T	4769G	8269A	10044G	14365T
Majorca	M68	x	x	x	H2a2a	No mutations detected					
Majorca	M70	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M71	x	x		H6a1	750G	1438G	3915A	4727G	4769G	
Majorca	M74	x	x	x	H*	750G	1438G	4769G			
Majorca	M75	x	x		H1	750G	1438G	3010A	4769G		
Majorca	M78	x	x	x	H*	750G	1438G	4769G			
Majorca	M79	x	x		H6a1	750G	1438G	3915A	4727G	4769G	
Majorca	M81	x	x		H6a1	750G	1438G	3915A	4727G	4769G	
Majorca	M104	x	x		H4a1a1a	750G	3992T	4769G	8269A	10044G	14365T

Supplementary Table 5. Complete genome haplotypes and haplogroups of Balearic origin samples. Haplogroups are named following the mtDNA tree Build 17 (PhyloTree).

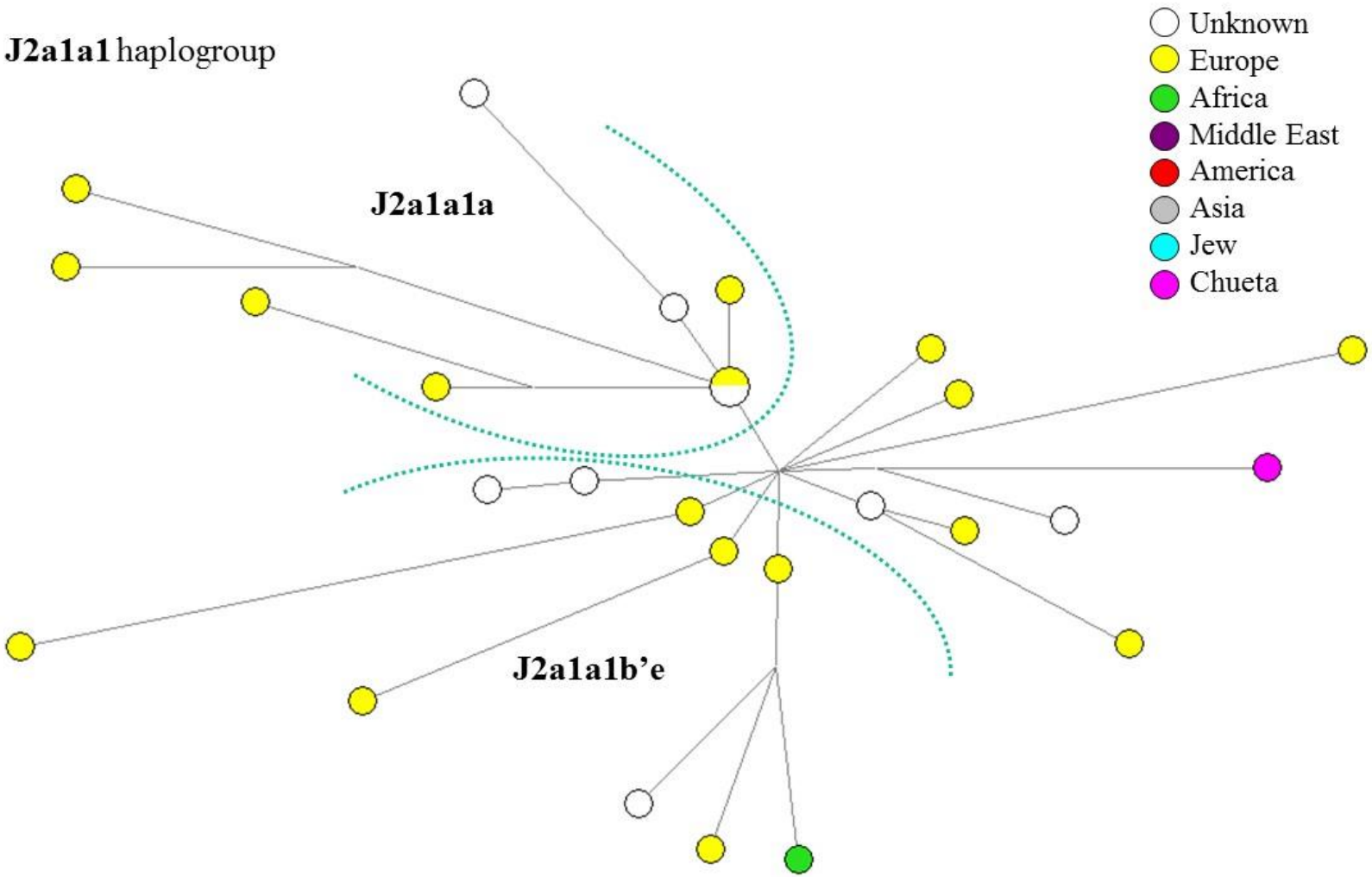
Sample	Haplogroup	Haplotypes																	
X02	U1a1a1	73G	263G	285T	385G	522del	523del	709A	750G	930A	1438G	2218T	2706G	3591A	3736A	4769G	4991A	6026A	7028T
		7581C	8860G	9377G	9575A	9716C	11467G	11719A	12308G	12372A	12879C	13104G	13422G	14070G	14071G	14364A	14766T	15148A	15326G
		15954C	16182C	16183C	16189C	16249C													
X24	R0a2m	58C	60.1T	64T	263G	309.1C	315.1C	750G	1438G	2355G	2442C	2706G	3847C	4767G	4769G	7028T	8860G	13188T	13858G
		14766T	15326G	15674C	15734A	16126C	16362C	16519C											
X36	H	263G	315.1C	750G	1438G	3621C	4769G	8860G	12712G	15326G	16223T	16519C							
X49	M5a1b1	73G	263G	315.1C	489C	524.1A	524.2C	709A	750G	1303A	1438G	1888A	2706G	3921T	3954T	4769G	4916G	6461G	7028T
		8701G	8860G	9540C	9833C	10398G	10400T	10873C	11719A	12477C	12705T	14323A	14766T	14783C	15043A	15287C	15301A	15326G	16129A
		16223T	16291T	16298C	16519C														
X52	M1a1	73G	195C	263G	489C	750G	813G	1438G	2706G	3705A	4769G	6446A	6671C	6680C	7028T	8701G	8860G	9540C	10398G
		10400T	10506G	10873C	11719A	12346T	12403T	12705T	12810G	12950C	14110C	14766T	14783C	15043A	15301A	15326G	16129A	16183C	16189C
		16223T	16249C	16311C	16359C	16519C													
X53	J2a1a1	73G	150T	152C	195C	203A	215G	263G	295T	319C	489C	513A	750G	1438G	2706G	4216C	4769G	7028T	7476T
		7789A	8860G	8958T	10398G	10499G	11251G	11377A	11719A	12612G	13395G	13708A	13722G	14133G	14766T	15257A	15326G	15452A	16069T
		16126C	16145A	16231C	16261T														
X70	R0a2m	58C	60.1T	64T	263G	309.1C	315.1C	750G	1438G	2355G	2442C	2706G	3847C	4767G	4769G	7028T	8860G	13188T	13858G
		14766T	15326G	15674C	16126C	16362C	16519C												
X118	K1a4a1	73G	263G	497T	750G	1189C	1438G	1811G	2706G	3480G	4769G	6260A	7028T	8860G	9055A	9698C	10398G	10550G	11299C
		11467G	11485C	11719A	11840T	12308G	12372A	13740C	14167T	14766T	14798C	15326G	16129A	16224C	16256T	16311C	16519C		
X127	R0a2m	58C	60.1T	64T	263G	309.1C	315.1C	750G	1438G	2355G	2442C	2706G	3847C	4767G	4769G	7028T	8860G	13188T	13858G
		14766T	15326G	15674C	16126C	16362C	16519C												
X132	R0a2m	58C	60.1T	64T	263G	309.1C	315.1C	750G	1438G	2355G	2442C	2706G	3847C	4767G	4769G	7028T	8860G	13188T	13858G
		14766T	15326G	15674C	16126C	16362C	16519C												
X133	K1a1b1a	73G	114T	263G	315.1C	497T	750G	1189C	1438G	1811G	2706G	3480G	4769G	7028T	8029T	8860G	9055A	9698C	10398G
		10550G	10978G	11299C	11467G	11470G	11719A	11914A	12308G	12372A	12954C	14167T	14766T	14798C	15326G	15924G	16223T	16224C	16234T
		16311C	16519C																
X135	T1a	73G	263G	709A	750G	1438G	1888A	2706G	4216C	4769G	4917G	6656T	7028T	8697A	8860G	10116G	10463C	11251G	11719A
		12633A	13368A	14766T	14905A	15326G	15452A	15607G	15928A	16126C	16163G	16186T	16189C	16294T	16298C	16319A	16519C		
X139	T2c1d1	73G	146C	263G	279C	709A	750G	1438G	1888A	2706G	4216C	4769G	4917G	5187T	6261A	7028T	7873T	8475T	8697A
		8860G	10463C	10822T	11251G	11719A	11812G	13056T	13368A	14233G	14544A	14766T	14905A	15326G	15452A	15607G	15928A	16126C	16292T
		16294T	16519C																
X142	R0a2m	58C	60.1T	64T	263G	309.1C	315.1C	750G	1438G	2355G	2442C	2706G	3847C	4767G	4769G	7028T	8860G	13188T	13858G
		14766T	15326G	15674C	16126C	16362C	16519C												
X155	K1a1b1a	73G	114T	263G	315.1C	497T	750G	1189C	1438G	1811G	2706G	3480G	4769G	7028T	8029T	8860G	9055A	9698C	10398G
		10550G	10978G	11299C	11467G	11470G	11719A	11914A	12308G	12372A	12954C	14167T	14766T	14798C	15326G	15924G	16223T	16224C	16234T
		16311C	16519C																
M33	T2c1d1a	73G	146C	263G	279C	709A	750G	1438G	1888A	2706G	3027C	4216C	4769G	4917G	5187T	6261A	7028T	7873T	8697A
		8860G	8911C	8980A	10463C	10822T	11251G	11719A	11812G	12363T	13368A	14233G	14766T	14905A	15326G	15452A	15569T	15607G	15928A
		16126C	16292T	16294T	16519C														
Caldés	R0a2m	58C	60.1T	64T	263G	309.1C	309.2C	315.1C	750G	1438G	2355G	2442C	2706G	3847C	4767G	4769G	7028T	8860G	13188T
		13858G	14766T	15326G	15674C	16126C	16362C	16519C											



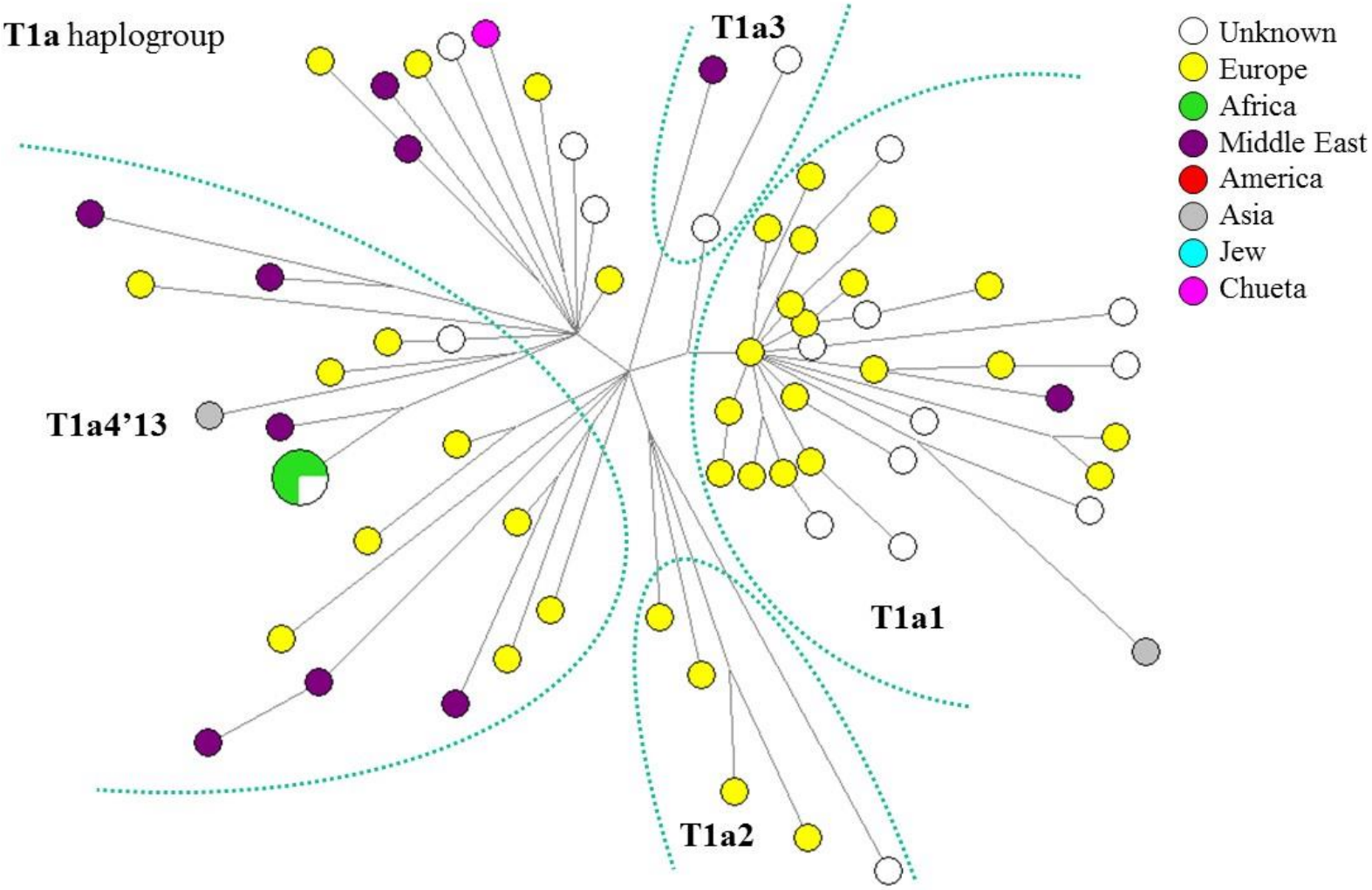
Supplementary Figure 1. Phylogenetic network of the M5a1b1 haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



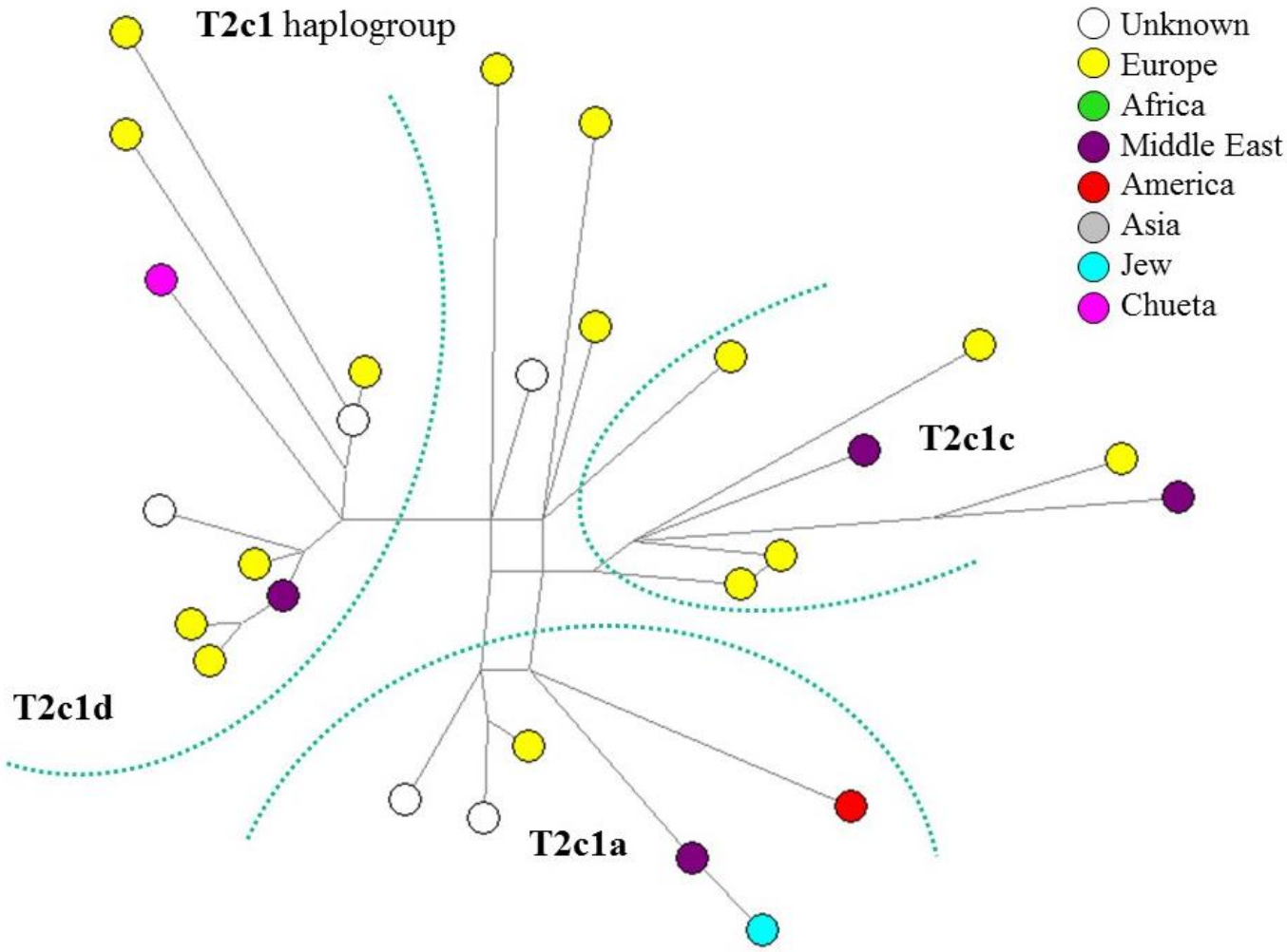
Supplementary Figure 2. Phylogenetic network of the M1a1 haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



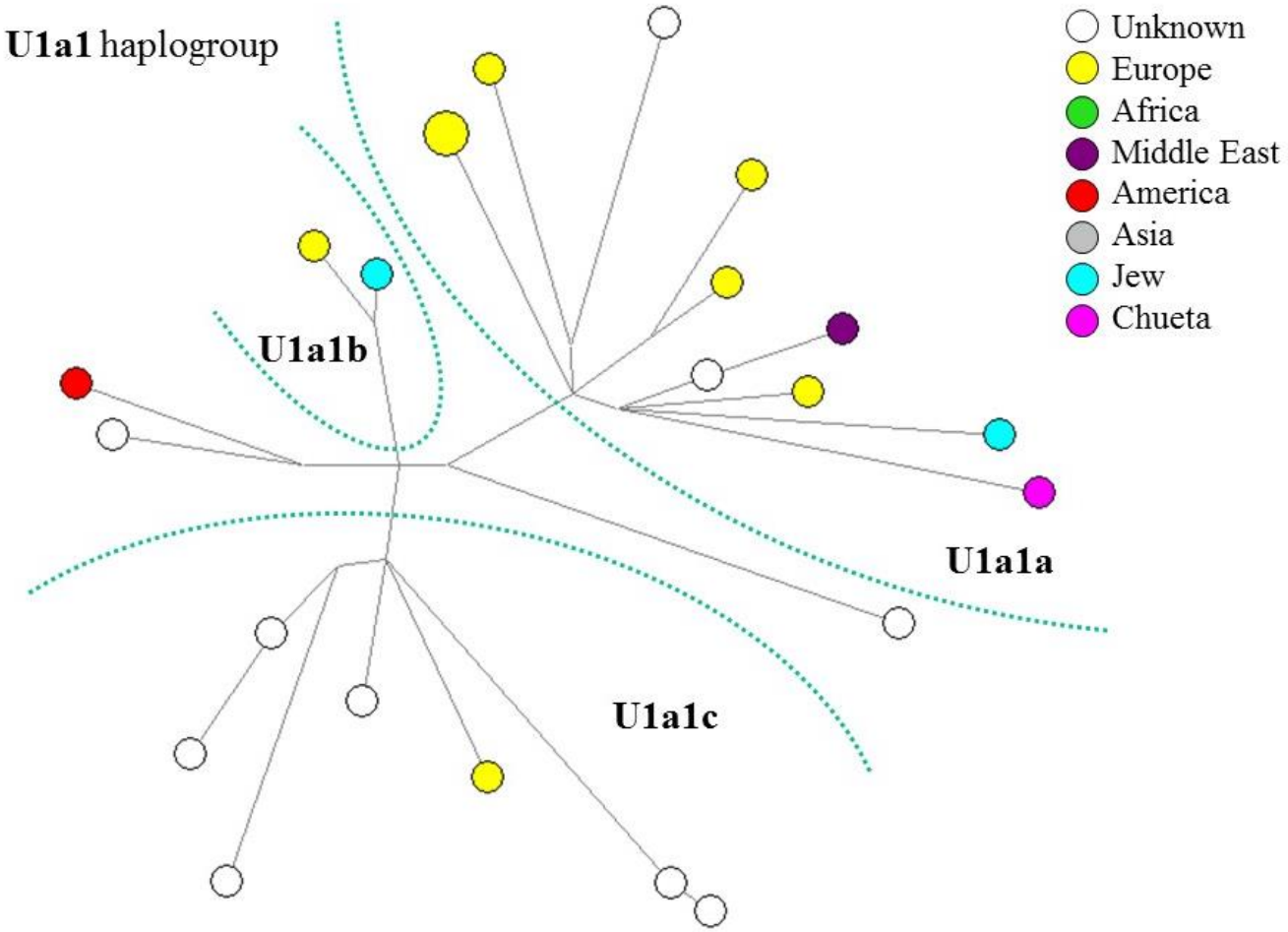
Supplementary Figure 3. Phylogenetic network of the J2a1a1 haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



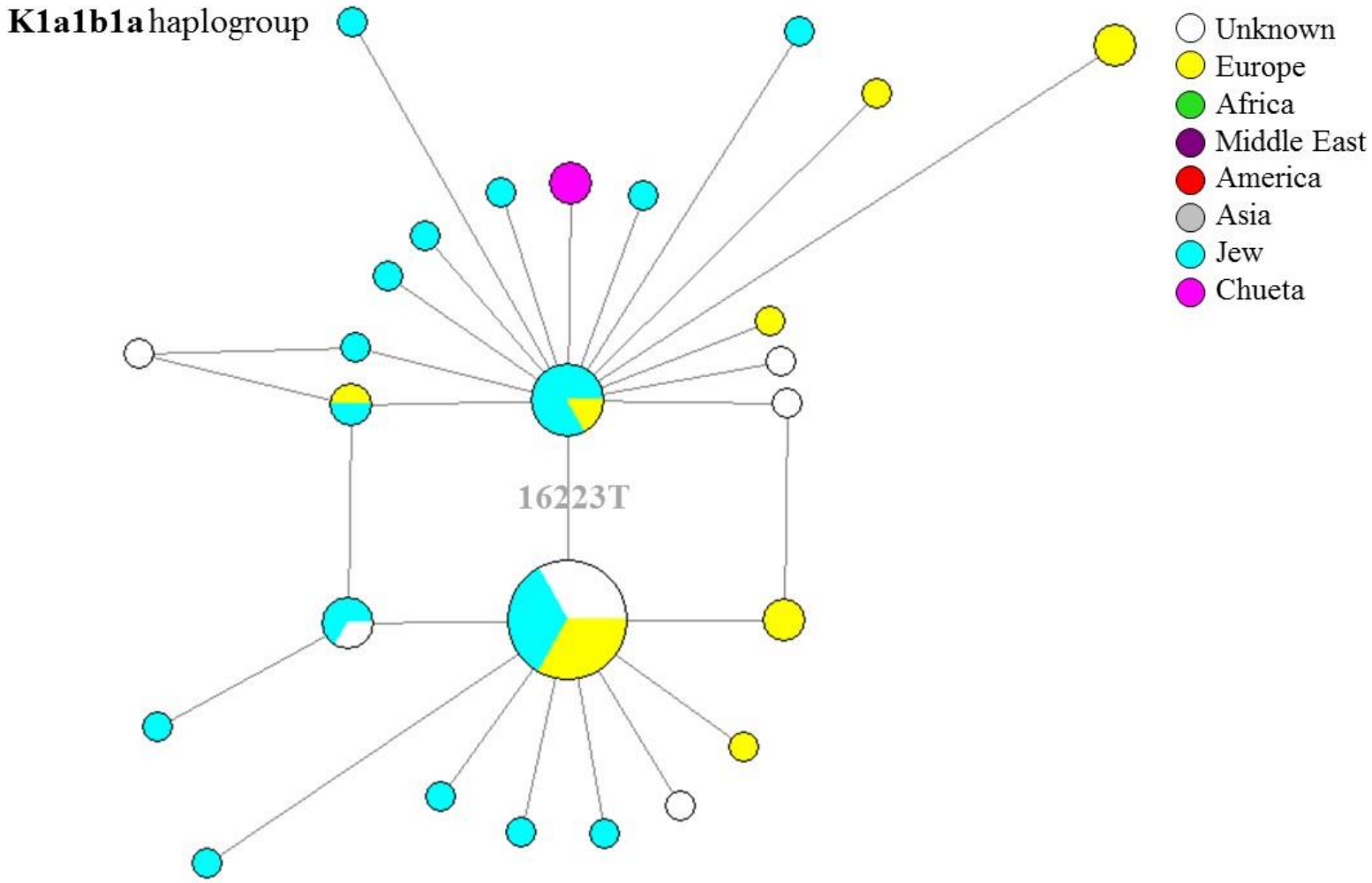
Supplementary Figure 4. Phylogenetic network of the T1a haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



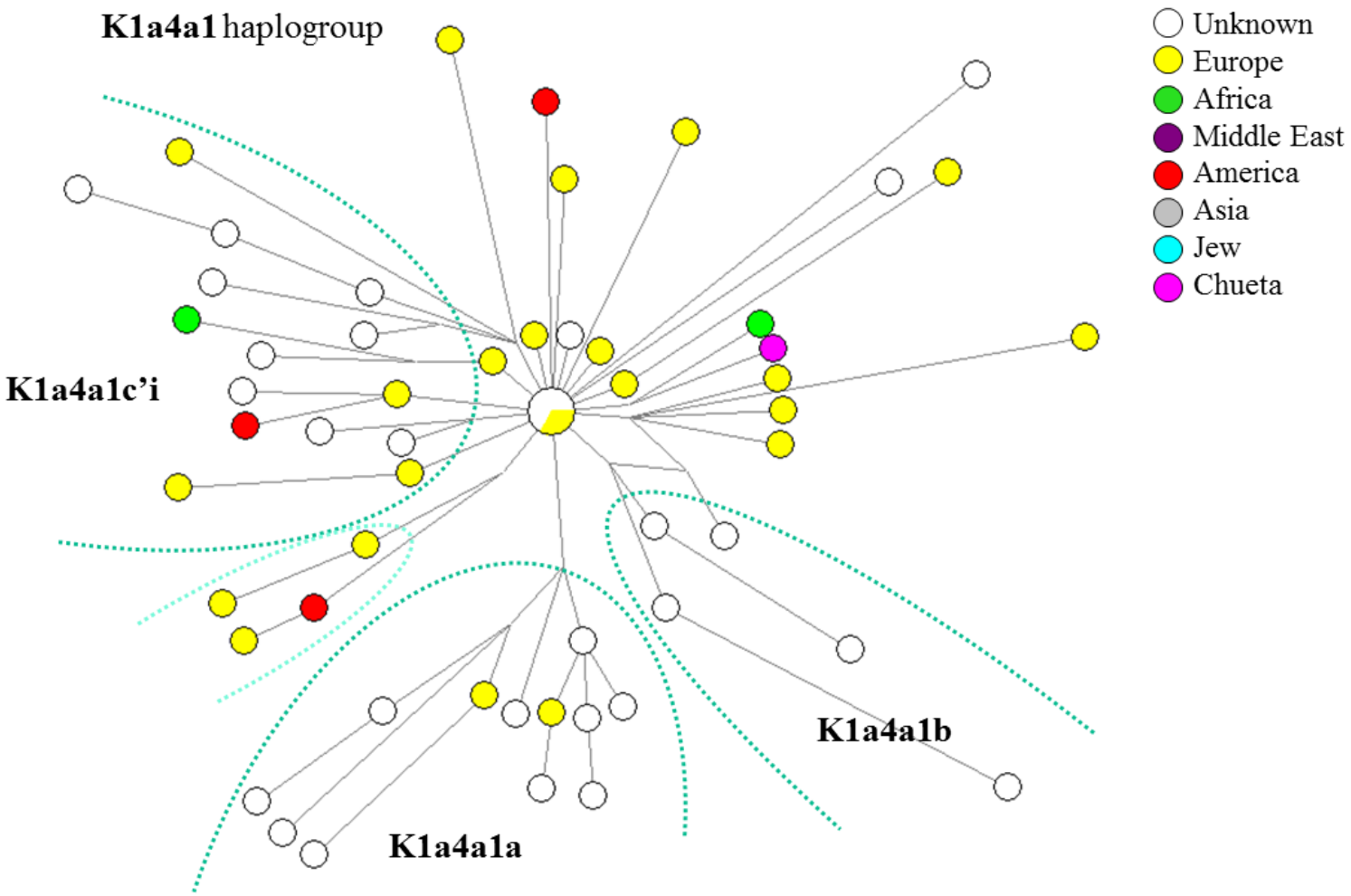
Supplementary Figure 5. Phylogenetic network of the T2c1d haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



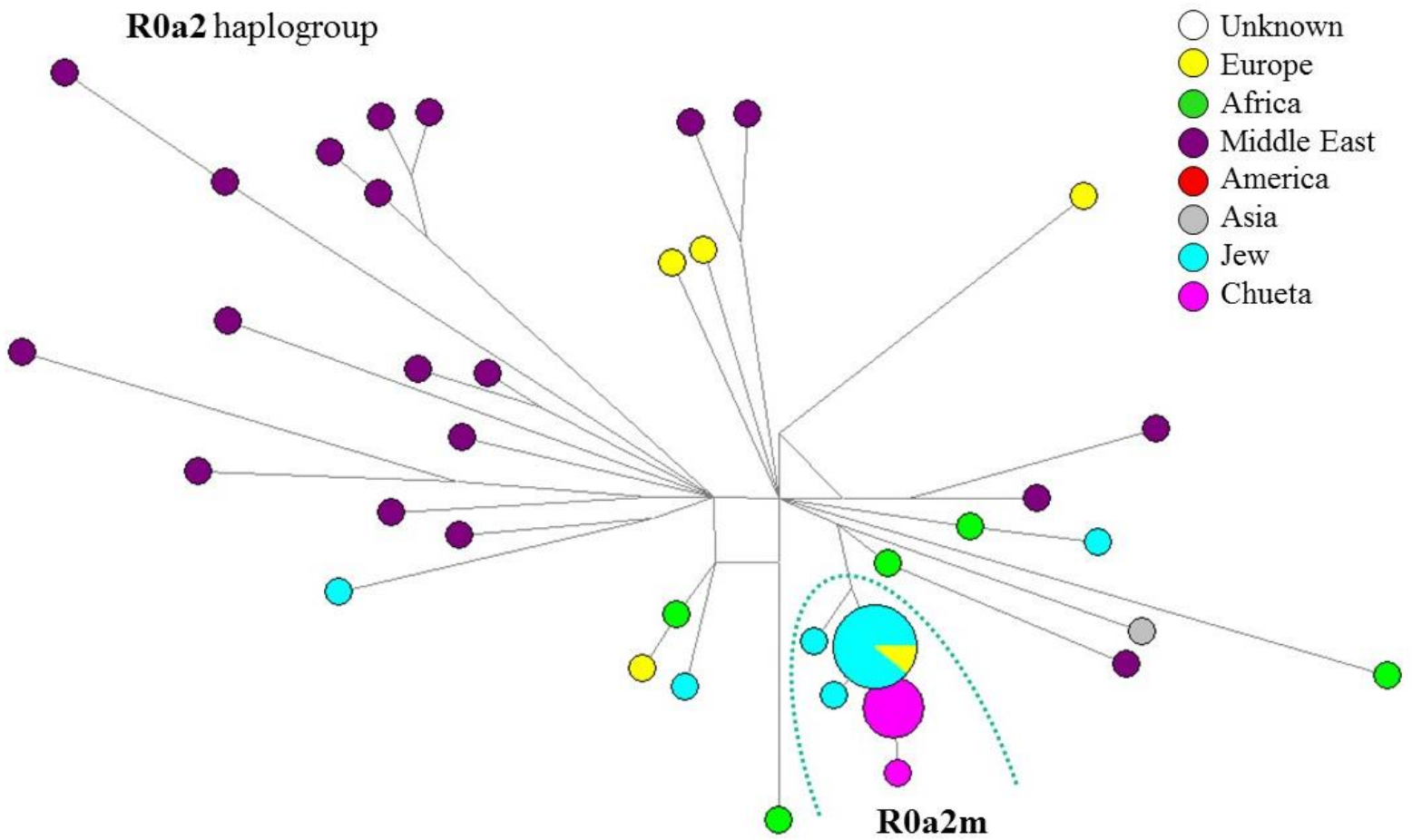
Supplementary Figure 6. Phylogenetic network of the U1a1 haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



Supplementary Figure 7. Phylogenetic network of the K1a1b1a haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.



Supplementary Figure 8. Phylogenetic network of the K1a4a1j haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend. Lighter blue separation indicates the proposal of a new cluster.



Supplementary Figure 9. Phylogenetic network of the M5a1b haplogroup. The smallest circles are singletons and the remaining are proportional of the number of samples included. Colours are assigned as indicated in the figure legend.

Chapter 4: Y-Chromosome

1. Y-chromosome lineages in Chuetas reveal the maintenance of their Middle Eastern genetic ancestry

In preparation

Introduction

The fourth and last chapter is devoted to the Y-chromosome, which enables us to characterize the paternal contribution to the Chueta population gene pool. In this work, we have studied the Chueta population and its host population (Majorca). In the literature, there are many studies on the Y-chromosome of different Jewish communities (Hammer et al., 2000; 2009; Thomas et al., 2000; Nebel et al., 2001; 2005; Behar et al. 2003; 2004a; 2010; Picornell et al., 2004; Shen et al., 2004; Adams et al., 2008; Nogueiro et al., 2010; Rootsi et al., 2013) which will enable a comparison with the results in Chuetas.

Taking into account that the surname is one the most important conditions to be considered Chueta and that surnames in Spain are "inherited" patrilineally, our hypothesis is that the Chuetas will be closer to other Jews (especially to Sephardic Jews) in Y-chromosome lineages than in mitochondrial ones. Since mtDNA (and X-chromosome) have also been studied in Chuetas, in its host population and in Jewish populations with whom they share origin, a comparison between results in these different uniparental markers will enable us to analyse whether or not there has been asymmetrical sex-biased admixture with the surrounding population.

The study was carried out by genotyping a set of 17 Y-STR markers, included in the Y-filer kit (Applied Biosystems). The haplotypes defined by these STRs were classified in haplogroups by analysing 38 SNPs in order to define the paternal lineages in Chuetas.

Although there is a previous study based on Y-STRs in the Chueta's host population (Majorca) (Rodríguez et al., 2009), a different set of markers was used. Therefore, in order to simplify comparisons, a representative subset (46 out of 91) of the previously studied Majorcan individuals have now been typed with the Y-filer kit and the set of 38 SNPs.

To delve deeper into the knowledge of the Y-chromosome lineages that define the paternal gene pool of the Chuetas, it was considered very interesting to sequence entire Y-chromosomes of the most characteristic lineages found in the present study. For this reason, there is an ongoing collaboration with Dr. DM Behar who has done it for other Jewish populations (Behar et al., 2010; Roosti et al., 2013), which will complete the results presented in the current PhD thesis.

Y-chromosome lineages in Chuetas reveal the maintenance of their Middle Eastern genetic ancestry

(In preparation)

Key words: Y-chromosome, Y-filer, STR haplotype, SNP haplogroup, Chuetas, Sephardic Jews, Middle East

1. Introduction

The presence of Jewish settlements in the Balearic Islands seems to date from the first century AD, although the oldest remains found (some tombs with Hebrew inscriptions and a letter of the Minorcan bishop) date back to the 5th century (Cortés, 1985). During the Muslim domination of the island (10th to 13th centuries) it is well-known that an important Jewish community lived in Majorca. The Catalan conquest, in the 13th century, entailed the protection of the Jews by the King (due to their importance in sea trade) and an increase in Majorcan Jewry with the arrival of Catalan Jews (Assis 1996-1997). The medieval Jewish community in Majorca was very prosperous with an important economic and cultural activity, as reflected by the well-known cartographic work of the Cresques family (Harwood, 2006; Llompart, 2011; Chacón 2013).

Like in other Iberian cities, in 1391, a pogrom took place against the Majorcan Jewry. Over 300 Jews were murdered but 800 managed to save themselves in a fortress. A number of Jews escaped to the nearest North African coast, and others submitted to baptism (Pérez, 2006). The conversions and exodus continued from 1391 to 1435, due to the pressure from the Roman Catholic Church and especially to the preaching of Friar Vincent Ferrer. The last massive conversion occurred in Majorca in 1435, so officially there were no Jews in the island fifty-seven years before the Catholic Kings' expulsion decree (1492).

After the conversions, the so-called new Christians appeared. Most of them mixed with the surrounding population and, as in many other cities in Spain, their track has been lost. However, in Majorca, a group of converted Jews continued to live in the same neighbourhood (*Call*), maintaining the relationships and occupations they had had before the conversion. They kept as a closed community and practised, to a greater or lesser extent, their Jewish beliefs and traditions. This situation of hidden religious practises (known as Crypto-Judaism) was maintained until the end of the 17th century when the Inquisition managed to put an end to it (Braunstein, 1936).

After these events, this community, rather than being assimilated into Majorcan society, began its greatest era of isolation, namely the Chueta period. Thereby, the collective consciousness of its origin was preserved, in spite of submitting to the Catholic faith.

Historically this people was stigmatized and segregated, and practised strict endogamy until the first half of the twentieth century. However there was a definite point of inflection when the island opened to tourism in the 1950s, as the arrival of newcomers (Spaniards or foreigners) who had no knowledge of the status of Chuetas led to a decrease in anti-Chueta prejudice.

The word "Chueta" was first documented in the inquisitorial processes in the 17th century, as an expression used by the prosecuted for themselves (Porqueres, 2001). Although its etymology is disputed and has several hypotheses, the most accepted is that it derives from a diminutive of Jew (de Muntaner, 2002). One of the most important defining elements of this group is that they are carriers of any of the 15 surnames of converso lineage (Aguiló, Bonnín, Cortès, Fortesa, Fuster, Martí, Miró, Picó, Pinya, Pomar, Segura, Tarongí, Valentí, Valleriola and Valls), affected by the inquisitorial sentences for Crypto-Judaism in the last quarter of the seventeenth century (Forteza, 1972). Some of these surnames are common in other Spanish regions, but are not related to Judaism. In Majorca, however, they have been fixed in the collective memory by their identification of Chueta families.

Genetic studies have been performed in Chuetas in autosomal markers (eg. Picornell et al., 1997; Tomàs et al., 2000; Ferragut et al., 2016) and mtDNA (Picornell et al., 2005), showing that this population presents evidence for a Middle Eastern ancestry, therefore there is a significant persistence of its Jewish heritage as well as signs of introgression from its host non-Jewish population. It would be especially interesting for the knowledge of the Chueta population to study the Y-chromosome, which is associated with the first surname of inheritance, because of the importance of the surname as a defining element of the Chueta social group.

Therefore, we undertook a study of the Y-chromosome in Chuetas and Majorcans (its host population) in order to study the Y phylogeny at different resolutions. Single-nucleotide polymorphisms (SNPs) – which are slowly evolving – and short tandem repeat (STR) loci – which are more quickly evolving – were used to distinguish different time scales. With this work, we aimed: to analyse the paternal gene pool of the Chuetas in the context of the available Y-chromosome data from Jewish and non-Jewish populations from Europe, North Africa, and the Middle East in the literature; and to estimate the likely ancestry of the present Y-chromosome lineages of this population.

2. Material and Methods

2.1. Samples

For this study, samples from 146 unrelated males were obtained after informed consent: 100 from the Chueta population, and 46 from Majorca (Balearic Islands, Spain), included in the study as the host population of Chuetas. The Majorcan samples constitute a subset of a larger sample previously genotyped for 12 Y-STRs (Rodríguez et al., 2009). The

Majorcan and 20 Chueta samples were anonymised DNA extracts belonging to the collection of the Genetics Laboratory, University of the Balearic Islands. DNA extraction of the rest of the samples was carried out from buccal swabs by standard phenol-chloroform method. Extracted DNA was quantified using a NanoVue™ Plus Spectrophotometer (GE Healthcare, Buckinghamshire) and normalised to a concentration of 2 ng/µl for Y-STR typing, and to 20 ng/µl for Y-SNP typing.

2.2. PCR analysis of Y-STRs and Y-SNPs

Seventeen Y-chromosome STR markers (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a, DY385b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA_H4, DYS437, DYS438 and DYS448x) were amplified using the Y-filer kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's recommendations.

Thirty-eight SNPs were typed to define the major male lineages. Thirty-two of them were genotyped using SNaPshot™ kit (Applied Biosystems) in five multiplexes as previously described by Brion et al. (2005), Gomes et al. (2010), Roewer et al. (2013) and Marques et al. (2016). M1 and M269 were genotyped with conventional PCR followed by agarose gel electrophoresis; S116, M17 and M18 were genotyped by sequencing (primers described in section 3.3.6); and DYS458.2 was used to determine the J1 chromosomes (Myres et al., 2007).

SNaPshot reaction was performed by adding 1 µl SNapShot mix, 1 µl SBE (single-base extension) primers in 5 µl final reaction volume to the purification tubes. The minisequencing conditions were: 25 cycles at 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s. Then a final purification was implemented by adding 1 µl SAP in each tube and incubating at 37 °C for 60 min and 85 °C for 15 min.

2.3. Genotyping of Y-STRs and Y-SNPs

The Y-STR amplification products and Y-SNP purified minisequencing products were separated in an ABI PRISM 3130 Genetic Analyser, and electropherograms were analysed using GeneMapper ID Sofware v3.2 and Peak Scanner software (Applied Biosystems). Y-STR alleles were designated according to ISFG recommendations (Gusmão et al., 2006), and Y-SNP haplogroups according to the latest ISOGG update (2017).

2.4. Statistical analysis

Allele and haplotype frequencies were estimated by gene counting. Molecular diversity indices were estimated for 17 Y-STRs, as well as for the minimum haplotype using

Arlequin v.3.5.1.2 software (Excoffier and Lischer, 2010). Discrimination capacity was calculated as the percentage of different haplotypes and haplotype match probability as 1-haplotype diversity.

In order to examine the relationship of the populations under study with other published population data, F_{ST} genetic distances, Analysis of Molecular Variance (AMOVA), as well as the corresponding non-differentiation p-values were assessed using Arlequin v3.5.1.2. For an easier visualization of the observed genetic distances, a multidimensional scaling (MDS) plot of the pairwise F_{ST} matrix was represented using SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA).

Median joining networks of Y-STR haplotypes were constructed using Network 4.5.1.0 (www.fluxus-technology.com; Bandelt et al., 1999).

Admixture proportions were estimated with mY statistics implemented in Admix 2.0 (Dupanloup and Bertorelle, 2001). This coalescent-based estimator takes into account allele frequencies, as well as molecular information. Two parental populations were used: Majorcan (n= 46) and Sephardic Jews (n= 174) (Adams et al., 2008) to create molecular divergence matrices. Admixture analysis was carried out with 7 microsatellites following software guidelines. The bootstrap procedure was set to 1 000 000 repetitions.

3. Results and Discussion

3.1. Y-STR diversity

Allele frequencies and gene diversities of each Y-STR of the populations under study are shown in Supplementary Table 1. A total of 81 different alleles were detected among the single allelic loci typed with Y-filer kit. Of them, 70.37% were observed in both populations. The median number of alleles per locus was 5.4 (ranging from 9 for DYS458 to 3 for DYS437). In DYS385, 25 allelic combinations were found.

Gene diversity was below 50% at two loci in Majorcans and at three loci in Chuetas, with especially low values in DYS389I (0.224) and DYS392 (0.288). As expected due to its diallelic structure, DYS385 loci showed the highest GD values (>0.83 in both populations).

These results are consistent with other studies showing DYS392 and DYS458 as the least and more polymorphic single allelic loci, respectively (e.g. Elmrghni et al., 2012; Olofsson et al., 2015)

Amongst the 146 males analysed, 97 Y-filer haplotypes were observed (Table 1 and Supplementary Table 2), 81 of which were only observed once (singletons). The most

frequent haplotypes were h43 and h38 observed in 12 and 9 Chueta individuals, respectively. No haplotype was shared between Chuetas and Majorcans.

Table 1 shows diversity and forensic parameters for the Y-filer haplotypes compared with minimum haplotypes. The overall haplotype diversity increased by 0.5% in Majorca and 2.5% in Chuetas by using the 17 loci in Y-filer instead of the 9 loci in minimal haplotype.

Table 1. Molecular diversity indices for haplotypes and haplogroups in Chuetas and Majorcans.

	N	K	UH	HD	h	MPD	DC(%)	MP(%)
Yfiler								
Chuetas	100	53	39	0.965± 0.008	0.611±0.311	10.387±4.774	53.00	3.49
Majorca	46	44	44	0.998± 0.005	0.606±0.312	10.308±4.783	95.65	0.20
Minimum haplotype								
Chuetas	100	40	26	0.940± 0.012	0.567±0.307	5.107±2.495	40.00	6.04
Majorca*	91	74	53	0.993±0.004	0.600±0.328	4.802±2.365	81.32	0.73
Y-SNPs								
Chuetas	100	13	3	0.827±0.023	0.106±0.059	4.033±2.029	--	--
Majorca	47	12	5	0.771±0.058	0.121±0.067	4.614±2.302	--	--

N: number of individuals; K: number of haplotypes/haplogroups; UH: unique haplotypes/haplogroups; HD: Haplotype diversity ± SD; h: Gene diversity over loci; MPD: Mean pairwise differences; DC: Discrimination capacity; MP: Match probability. *Rodríguez et al. (2009).

Haplotype diversity (and consequently power of discrimination) in Chuetas is significantly lower than in Majorcans or other populations in the literature (eg. Aboukhalid et al., 2010; Ambrosio et al., 2010; Tokdemir and Tunçet, 2017). These results are in accordance with the historical and demographic data of this population and with the reduced genetic diversity in some markers found in previous genetic studies (Picornell et al., 1992; 2005). Therefore, the high haplotype match probability (3.49%) must be taken into account in forensic practice.

3.2. Y-SNP haplogroups diversity

Chueta samples were classified into 13 different haplogroups according to the 38 genotyped SNPs (Table 1 and Figure 1). The six most frequent haplogroups (≥4%), E1b1b1a1-M78, G-M201, J1-DYS458.2 (used in this study as synonymous of J1-M267), J2-M172, Q1-P36.2, and R1a1a-M17 accounted together for up to 87%, with J2-M172 as the most frequent (33%). Three out of the seven remaining haplogroups occured only in one individual.

High diversity values, based on haplogroup frequencies, were found in Chuetas (HD: 0.827±0.023) in comparison with other Spanish populations, but in the same range as the Sephardic Jew population (Adams et al., 2008; Álvarez et al., 2014).

AMOVA analyses at both haplogroup and haplotypes (Y-STRs) level showed significance when taking Chuetas and Majorcans as a whole (F_{ST}=0.1959; $p<10^{-5}$), indicating the differentiation of the Chuetas with respect to their host population. In order to identify putative Chueta founding lineages, the Bayesian 0.95 credible regions (95% CR) for the haplogroup distribution were calculated using a Bayesian approach for binomial sampling, by means of SAMPLING software (Macaulay, personal communication) with frequency values obtained for Chuetas and Majorcans. This criterion indicated three differential haplogroups: R1b1a1a2-M269, J1-DYS458.2, and J2-M172, pointing towards the lack of R1b1a1a2-M269 and the presence of J1-DYS458.2 and J2-M172 as Chuetas' putative founding lineages. The frequency in Chuetas of haplogroups rarely found in neighbouring populations, E1b1b1a1-M78, Q1-P36.2, G-M201, and R1a1a-M17 (14, 10, 8, and 4%, respectively), could also mean that they could have been present in the original Jewish Majorcan gene pool.

3.3. Phylogeographic analysis

- Haplogroup R1b1a1a2

The most frequent subclade of haplogroup R in Europe, with the mutation M-269 which defines the haplogroup R1b1a1a2, has a controversial origin in terms of place and date. First studies (Semino et al., 2000; Soares et al., 2010) pointed to its origin in the Franco-Cantabrian Refugia in the Palaeolithic era, in part, due to its high frequency levels in this area. A parallel expansion from an Eastern Europe refugia was proposed by Cinnioğlu et al. (2004). In 2010, Balaresque et al. suggested a completely different origin. They proposed an Eastern European origin in the Neolithic period, due to the high diversity of haplotypes they found in the area. In this study, germinal mutation rates were used instead of evolutionary ones, leading to younger dating. Trying to solve the question, studies analysing SNPs inside the M-269 clade aimed to understand the structure of the haplogroup (Cruciani et al., 2011a; Myres et al., 2011). The idea of the origin in the East was maintained but the dating seems to be early in the Mesolithic. In 2015, Batini et al. dated this cluster as much more recent, about 5Kya in the late Neolithic (Bronze Age) and situated its origin in the steppe region north of the Caspian Sea. The latest studies again situated the origin in the East of Europe, in the Palaeolithic era, suggesting possible mixing with Western origin R haplogroups later on in the Neolithic expansions (Valverde et al., 2016).

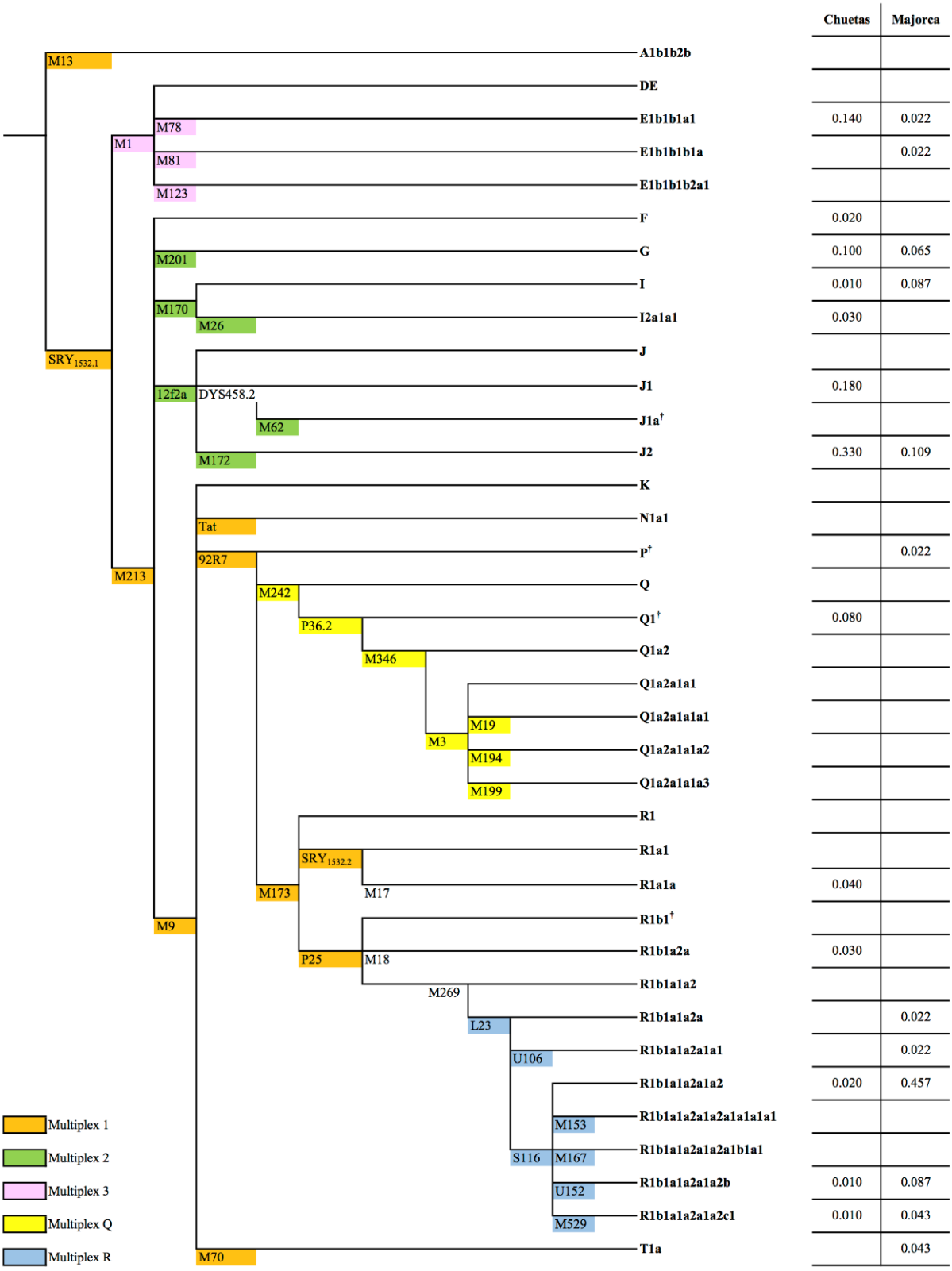


Figure 1. Phylogenetic tree of the 38 Y-SNPs typed, and haplogroup frequencies for Chuetas and Majorca populations. Haplogroups were named in accordance with the latest Update of ISOGG 2017. Haplogroups labelled with † are named as in their original description: J1a-M62, P-92R7, and R1b1-P25 (Brion et al., 2005), and Q1-P36.2 (Roewer et al., 2013).

For the time being, due to differences in opinion in calculating TMRCAs, it is impossible to accurately date these evolutionary episodes, without soundly dated and genotyped archaeological remains and/or more complete Y-chromosome data (Valverde et al., 2016).

Taking into account the frequencies found in our populations, Majorca shows similar values (63%) to the rest of Spanish populations, whereas the Chueta population (4%), has similar values to Middle Eastern and North African populations (means of 3% and 2%, respectively). Libyan Jews and Israelites show moderate frequencies (5% and 8%, respectively), while higher values are found in Sephardic Jews (11.5%), and especially in Bragança Jews (28%) (Supplementary Table 3). These results support some degree of Iberian admixture in Sephardic Jews (Adams et al., 2008) and an important introgression from the host population in Bragança Jews, as suggested by Nogueiro et al. (2010).

- Haplogroup J

The most likely place of origin of haplogroup J-12f2a has been proposed to be the Fertile Crescent (Cinnioğlu et al., 2004) and it shows an East to West gradient in Europe. Its two branches, J1-M267 and J2-M172, have a different distribution, showing different peopling pathways (Francalacci and Sanna, 2008). J2-M172 has its high frequencies in the Levant decreasing to North Africa and Arabia and it is the most frequent branch in Europe, mainly throughout the Mediterranean basin, which could be associated to demic diffusion of the Neolithic agriculturalists (Semino et al., 2004; Francalacci and Sanna, 2008) or, more recently, the Phoenician and other historical, mainly maritime, migrations (Hammer et al., 2000; Di Giacomo et al., 2004; Zalloua et al., 2008). The J1-M267 haplogroup has its maximum frequencies in Arabia and decreases beyond the Middle East and North Africa (El-Sibai et al., 2009). The frequencies found in North Africa have been associated with its Arabization (Francalacci and Sanna, 2008).

Frequencies among Jewish populations are suggested to be higher in J2-M172 than J1-M267 (Capelli et al., 2006). Taking into account mean values, it is also confirmed in Supplementary Table 3 (24% vs 19%), even though there are a number of populations (such as the Sephardim) who do not follow this statement. Mean values for frequencies for J2-M172 reveal higher values in Jews (24%), followed by other populations in the Middle East (18%). Frequencies in Europe are about 10%, in Spain, specifically, the frequency of J2-M172 is 8%. In the Chueta population this haplogroup had a frequency of 33%, and was the modal haplogroup. This value is above the mean for Jewish populations (24%), and higher than the Sephardic percentage (21%). This high frequency of J2-M172 in Chuetas, together with the comparison with their host population values (10%), clearly indicates that this haplogroup is evidence that Chuetas maintain, in male lineages, relics of their Jewish ancestry.

For haplogroup J1-M267 (considered in Supplementary Table 3, J(xJ2-M172)), Chuetas have similar values (18%) to those in Jews (mean value 19%). In the Middle East, mean values are 30%, with the highest frequencies found in Arabian countries (72% in Yemen, for example). Taking into account that no sample was found in the host population, belonging to this haplogroup, we could also suggest that J1 in Chuetas is a signature of their Jewish/Middle Eastern origin. Within the Jews, the Cohanim have the highest frequencies (46%) of J1-M267. The commonly-known Cohen Modal Haplotype (CMH), described extensively in Hammer et al. (2009), belongs to this haplogroup. No haplotype in Chuetas matches with the CMH, although two of them share 10 out of the 12 alleles with those described in the Cohanim sample.

- Other lineages

Haplogroup E probably originated from the East African populations that generated the Out of Africa expansion. Most of these settlements involve the later mutation of E-M35 varieties such as M78, M81, and M123, which extended to Arabia and the Northern Mediterranean Coast (Chiaroni et al., 2009). The branch E1b1b1a1-M78 originated in Eastern Africa around 13.1-14.8 Kya (Semino et al., 2004; Trombetta et al., 2015), and from there, started its dispersion to other African regions, as well as to outside Africa (Cruciani et al., 2007). The arrival to Europe of the M78 mutation is attributed to the spread of the Neolithic farmers from the Fertile Crescent (Francalacci and Sanna, 2008). Chuetas showed a differential frequency of this haplogroup (14%) with respect to their host population (2%). Other studies analysing populations with Jewish origin have found a variety of percentages: ranging between 3% and 12.5% in Sephardic Jews (Semino et al., 2004; Adams et al., 2008; Hammer et al., 2009), between 5% and 10% in Ashkenazim (Semino et al., 2004; Shen et al., 2004), between 10 and 15% in North African Jews (Shen et al., 2004), and 3.5% in Crypto-Jews of Bragança (Nogueiro et al., 2010). Frequencies in Middle East populations (Jewish and non-Jewish) ranged between 5% and 16.7% (Shen et al., 2004; Semino et al., 2004). So, it seems likely that the differential presence of this haplogroup in the Chueta population came from their well-known historical Jewish origin and/or by admixture with North African Jews, due to commercial contact between both communities (Chacón, 2009).

Q-M242 haplogroup is relatively frequent in Central Asia and Central Siberia, and is the main haplogroup in Native American populations. A recent study analysing in depth the phylogeny of Q-M242 haplogroup supports its Central Asian origin (Balanovsky et al., 2017), previously suggested by other authors (Malyarchuk et al., 2011). Native American branches are defined by Q1a2-M346 (also found in Siberia in low frequencies) and especially Q1a2a1a1-M3 (formerly known as Q1a3a) (Roewer et al., 2013). Since the haplogroup Q1-P36.2(xM346) is practically absent in Europe and Africa (Supplementary Table 3), its frequency in Chuetas (8%) was unexpected. Taking into account that Q1-P36.2(xM346) has also been found in Jewish populations with percentages ranging from 1% to 5% (Behar et al., 2004a; Adams et al., 2008; Hammer et al., 2009), its presence in

Chuetas might be associated with their Jewish ancestry. Balanovsky et al., (2017) have described an Ashkenazi branch, so it could be interesting to delve further into the study of this haplogroup, to infer its origin in Chuetas.

The most common haplotype in Eastern Europe is R1a-M420 (Kayser et al., 2005), reaching frequencies greater than 55% in Belarus, Poland and South Russia (Supplementary Table 3). Its main subclade R1a1a-M17 (Underhill et al., 2010) is found in high percentages in Ashkenazim, especially Levites, where it was believed to be an introgression from the host populations. This haplogroup was also present in other non-Ashkenazi Jewish populations, at a much lower frequency (1.8% to 8%) (Behar et al., 2003; 2004a; Nebel et al. 2005; Supplementary Table 3). In 2013, Rootsi et al. studied in depth complete R1a Y-chromosomes, describing new sub-branches. The Ashkenazi branch was defined by the SNP M582, which was also shared by other Middle Eastern populations, but was absent in Eastern Europeans. Hence, the hypothesis of an introgression from the Ashkenazi host populations was refuted and they suggested an original Middle Eastern origin. In Chuetas, R1a1a-M17 was found in 4% of the individuals analysed, but not in the Majorcan samples. It could be interesting to assess whether the Chueta R1a1a chromosomes belong to the Near Eastern subclades.

Haplogroup G, defined by SNP M201, has its highest frequencies in present day Georgia and Turkey, with frequencies of 31% and 29%, respectively (Supplementary Table 3). For this reason, its origin was suggested to be in the Caucasus region (Cinnioğlu et al., 2004). Chuetas and Majorcans display relatively similar values (10% and 6.5%, respectively). In Jewish populations, Moroccan and Sephardic Jews have the highest values, 16% and 19% (Supplementary Table 3). Under these conditions, Adams et al. (2008) suggested that the frequency found in the Iberian Peninsula (5%) could be the result of an introgression of the Sephardic population in Iberia. At any rate, Nogueiro et al. (2010) analysing the STR variation inside this haplogroup confirm the existence of gene flow between Sephardim and Iberians, but the direction cannot be determined. Further analysis of this haplogroup in Jewish and non-Jewish populations from Europe and the Middle East would be necessary to solve this question.

3.4. Relationship between surnames and Y-chromosome

The relationship between the Y-chromosome and surnames has been explored in different populations (e.g. McEvoy and Bradley, 2006; King and Jobling, 2009; Solé-Morata et al., 2015; Calderón et al., 2015; Martinez-Cadenas et al., 2016). Taking into account the fact that, in most societies, surnames are passed down from fathers to sons, just like the Y chromosome, men sharing the same surnames would also be expected to share related Y chromosomes.

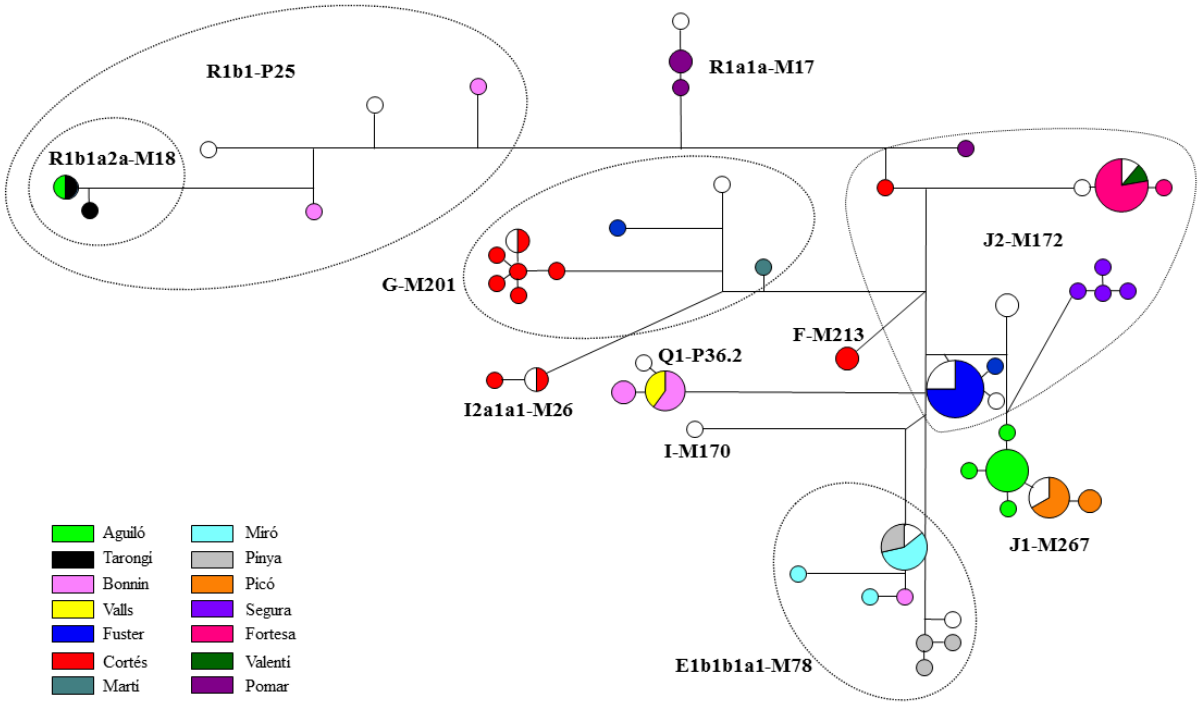


Figure 2. Median-Joining network of Chueta haplotypes. Colour code indicates the Chueta surname of the sample, for anonymised individuals the colour white has been used. Smaller circles are singletons and size is proportional to haplotype frequency. Haplogroup assignment is indicated in each group of haplotypes.

Although studying this relationship was not an initial aim of our work on the Y-chromosome in Chuetas and, therefore, the sampling was not designed for it, with a median joining network we assigned and compared haplotypes and haplogroups found with the surnames of the individuals when available (Figure 2 and 3). Fourteen of the 15 Chueta surnames are represented, but not the surname Valleriola, which has become recently extinguished in male lines.

The majority of this surname set contains a highly reduced haplogroup diversity (h=0.000-0.222). The surname Pomar revealed two different haplogroups with intermediate h value (0.500) and Cortés was the most diverse with 4 haplogroups and an h value of 0.691; although the sample size should be enlarged to confirm these results.

Haplogroups with high frequency in Sephardic populations, J1-M267, J2-M172, and G-201, were found to be associated: J1-M267 to Picó and Aguiló; J2-M172 to Segura, Fortesa, and Fuster; and G-M201 to Cortès and Fuster. R1a1a-M17 and Q1-P36.2, frequently found in Ashkenazi Jews, were detected in individuals with the surname Pomar, and Valls and Bonnín, respectively. Pinya and Miró had E1b1b1a1-M78, which is typically from the Horn of Africa and has been found in North African Jews. Finally, the surname Tarongí (and one Aguiló individual) showed the rare haplogroup R1b1a2a-M18.

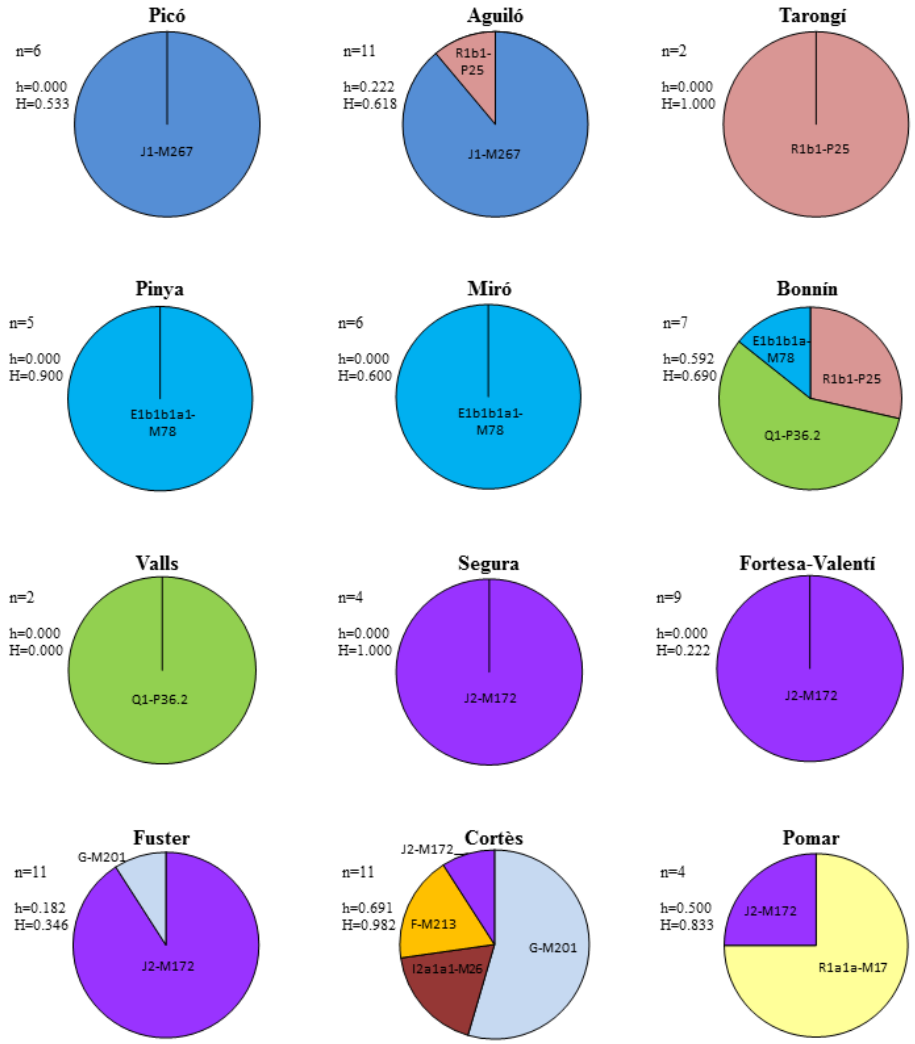


Figure 3. Y-chromosome haplogroup composition in Chueta surnames with more than one occurrence. For each surname, gene diversity based on haplogroups (h), haplotypes (H), and number of individuals are provided.

In Figure 2, the haplotypes within each haplogroup can be observed. In most of the surnames haplotypes show a star-like distribution with only one or two mutational steps between them, except Pinya (E1b1b1a1-M78) and Bonnín (R1b1-P25).

Valentí shared the almost unique haplotype found in Fortesa individuals, in accordance with the origin of this surname. Valentí was a nickname of a family with the surname Fortesa, which started to be used in the 18th century as a surname in a branch of the family in an attempt not to be considered Chuetas any more (Porqueres 2001; Planas, 2003; Aguiló, unpublished).

The fact that different surnames shared the same haplotype, such as Valls and Bonnin, could be due to different reasons, such as the fact that there is evidence that different male members of the same family were baptized with different Christian names (Contreras,

1997; Pérez, 2005; de Muntaner, 2012). The case of males from different Jewish families being baptized with the same surname has also been documented, which could explain very distant Y-lineages within a surname, as is the case of Bonnín or Cortès. However, the scarcity of historical documents with the Christian names that converted Jews adopted does not allow us to assess to what extent these cases were common or not.

3.5. Comparison with other populations

- Comparison with Iberian populations

The results were compared with those of Sephardic Jews (Adams et al., 2008), Crypto-Jews from Bragança (Nogueiro et al., 2010), and the non-Jewish populations from Majorca (this work), and from the Iberian Peninsula (Adams et al., 2008).

In Figure 4 it can be seen that the haplogroup distribution in Chuetas is quite similar to that found in Sephardic Jews, with J1-M267 and J2-M172 as the modal haplogroups, in contrast with non-Jewish populations where R1b1-P25 is the main haplogroup. Bragança's Crypto-Jews seem to have a higher introgression from neighbouring non-Jewish populations, leading to similar frequencies in Middle Eastern J1-M267 and J2-M172 haplogroups, and the Western European haplogroup R1b1.

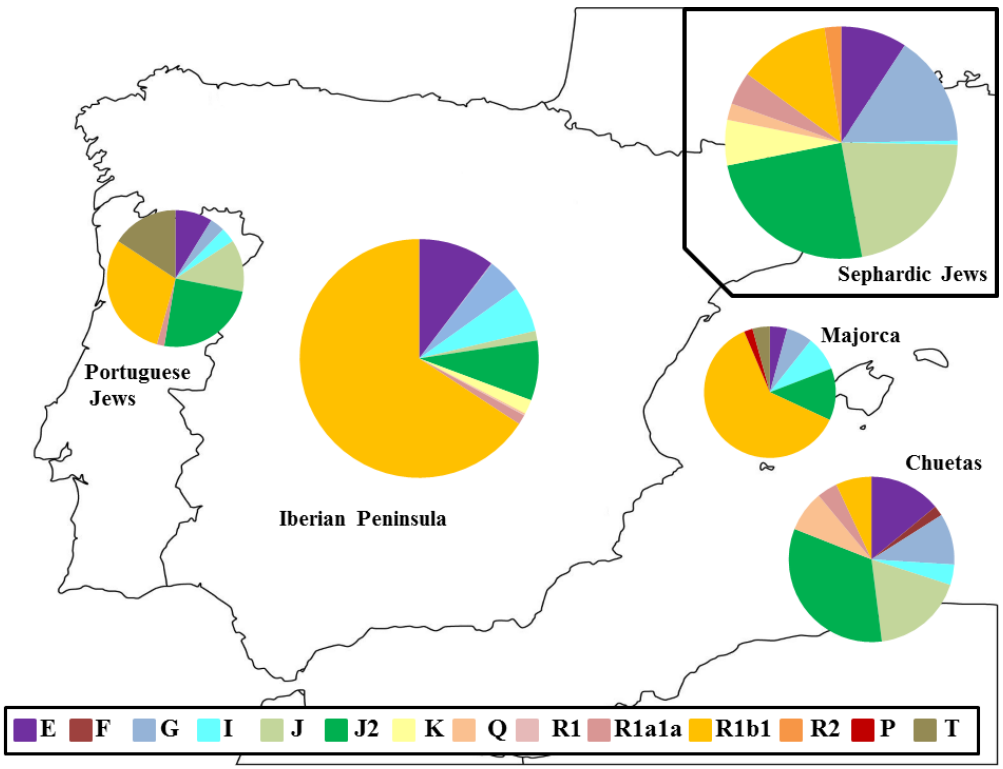


Figure 4. Map showing haplogroup diversity in 5 different populations. Pie chart sizes are proportional to sample size. Haplogroup nomenclature is defined by the following SNPs: E-M35, F-M213, G-M201, I-M170, J-12f2 (xM172), J2-M172, K-M9, Q-M242, R1-M173 (xM17, xP25), R1a1a-M17, R1b1-P25, R2-M124, P-92R7 and T-M70.

Pairwise F_{ST} analyses confirmed these observations. With the exception of Majorcan and Iberian Peninsula populations, all the pairwise comparisons showed statistically significant differences. F_{ST} values were highly significant ($p<10^{-5}$), except in the pair Chuetas-Sephardim, where a value was at the limit of significance found (F_{ST}=0.0115; p=0.0090. Significant level after Bonferroni correction $p\geq0.010$).

Solé-Morata et al. (2015) studied Y-chromosome diversity within different Catalan surnames. Four of them are related to a Jewish origin: Salom and Maymó (with a Hebrew linguistic root), Vidal (often mentioned as a surname taken by Jews who converted to Christianity), all three of which are common in Majorca, and Estruch (a common byname among medieval Jews in Catalonia). Using Sephardic Jews as a parental population, they found a signature of a possible Jewish origin in the surname Estruch, due to the fact that about two-thirds of Estruch men carried J2-M172 haplogroup, but no such signature was found in the other Jewish related surnames. Chuetas did not share any haplotype with the individuals of these four surnames and, in haplogroup distribution (Figure 5), were more similar to Estruch men than to the other Hebrew surnames, due to the fact that the modal surname in Chuetas is also J2-M172.

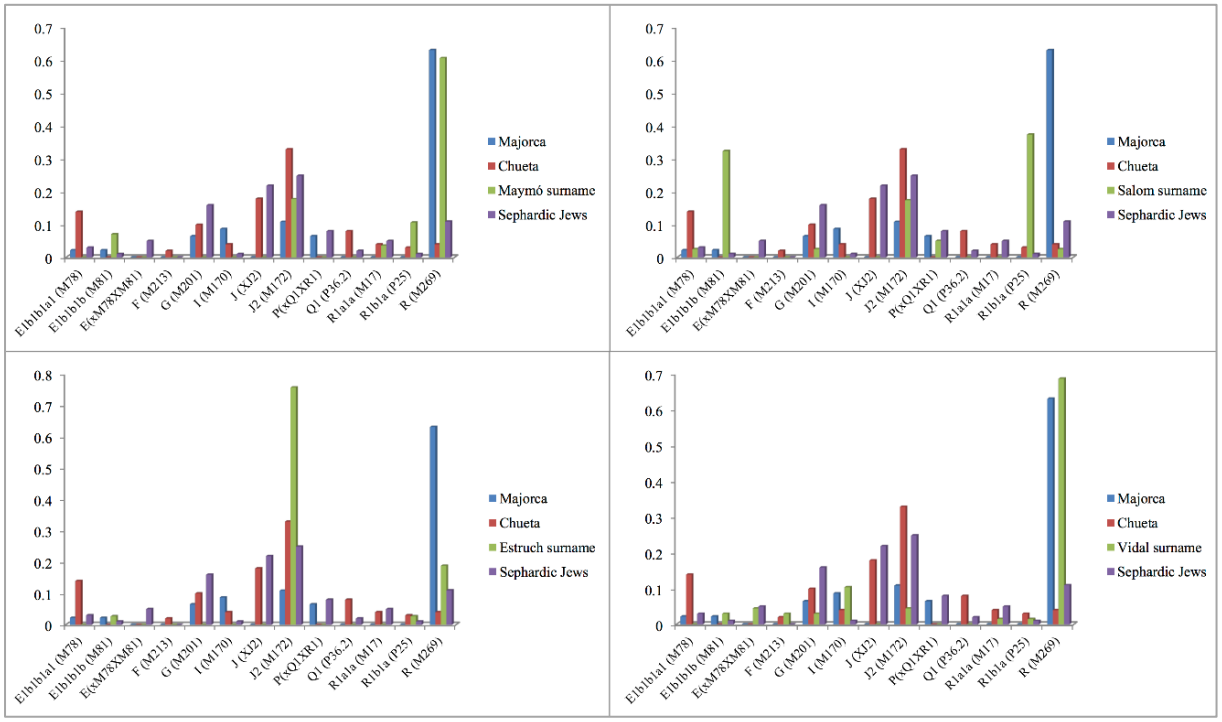


Figure 5. Comparison between haplogroup diversity of Chueta, Majorcan, and Sephardic populations with four Catalan surnames of Jewish origin (Solé-Morata et al., 2015).

- Comparison with worldwide populations

To evaluate the genetic relationship between Chuetas and previously reported data on other Jewish and non-Jewish populations (Supplementary Table 3), pairwise F_{ST} genetic distances were computed and represented in MDS plots. Since no populations in the literature have been studied with the same set of SNPs and STRs, a different comparison has been done with 14 and 11 SNPs (Figure 6 and 7), and with 7 and 17 STRs (Figure 8 and 9). Stress values in all the MDS analyses showed a greater structure than would be obtained from a random dataset (Sturrock and Rocha, 2000).

In Figure 6 the distribution of the European populations can be observed along the X-axis, ranging from the Western to the Eastern region. Middle Eastern populations are located on the upper right side, close to some South Eastern European ones, including populations with a Jewish origin. In spite of the tradition among Pathans (from Pakistan) of being descended from the exiled lost tribes of Israel, the position in this MDS analysis does not support their Jewish origin, in accordance with Lee et al. (2014).

In order to be able to compare with more Jewish populations and their respective host populations, 11 SNPs were used (Figure 7). Along the X-axis three different groups can be observed, ranging from the one including European populations to the group comprised of African populations. Most Jewish groups, together with Middle Eastern populations, show an intermediate position between European and African clusters, indicating that most of the Jewish groups are more similar to each other and to the populations in the geographical region where they originated, than to their host populations. Anyhow, some Jewish populations show a more significant influence from their host population, such as Libyan and Ethiopian Jews, on the one hand, and Bragança and Georgian Jews on the other. Chuetas showed a very distant position with their host population, being situated in the Middle Eastern group, close to Sephardim and Ashkenazim.

Although there are fewer data on Jewish groups based on STRs, MDS (Figure 8 and 9) confirms that Chuetas are closer to Middle Eastern populations than to their neighbouring non-Jewish populations in Y-chromosome lineages. Also, they seem to have more influence from North African populations than other Jewish populations, in accordance with historical data. There was an important commercial relationship between Majorcan and North African Jews (Chacón, 2009) and there is even evidence of North African Jewish merchants settling in Majorca (Aguiló, unpublished).

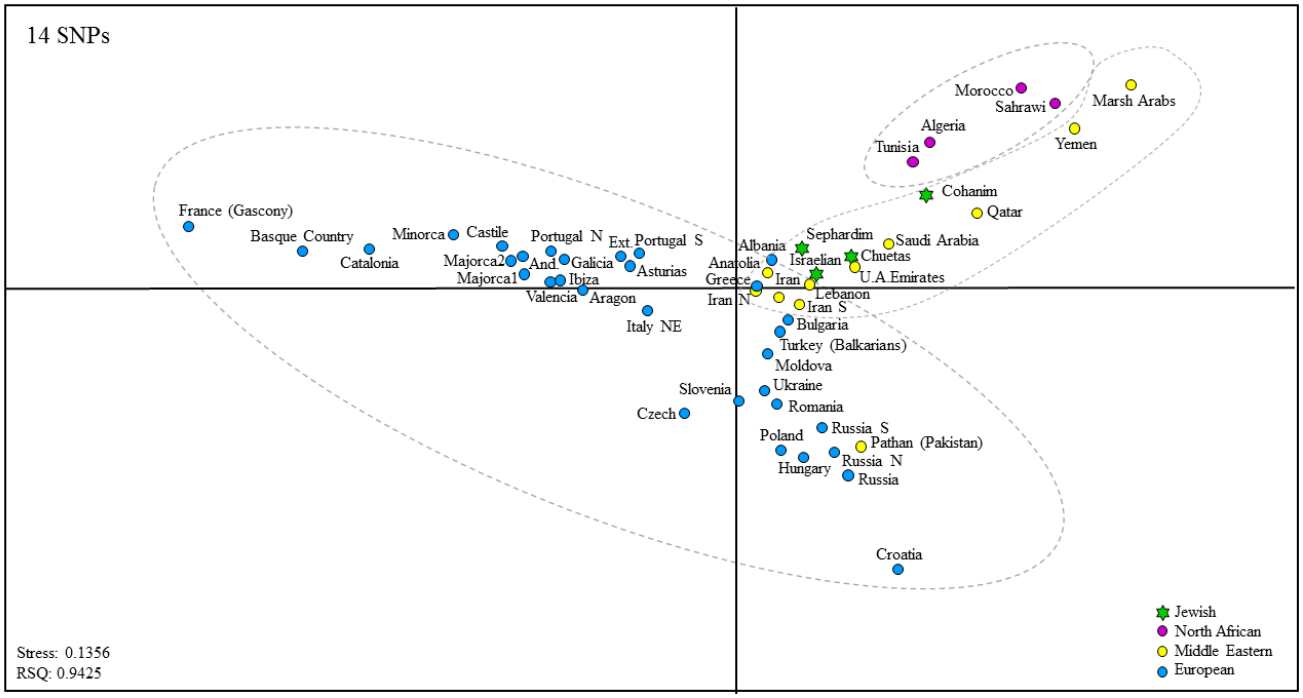


Figure 6. Multidimensional Scaling plot performed with populations, and SNP distribution from supplementary Table 3. Names shortened for better viewing of the figure are: And. (Andalucia); Ext. (Extremadura); and U.A. Emirates (United Arab Emirates). Circles defining each population are coloured following the legend code. Jewish populations are labelled with a Star of David.

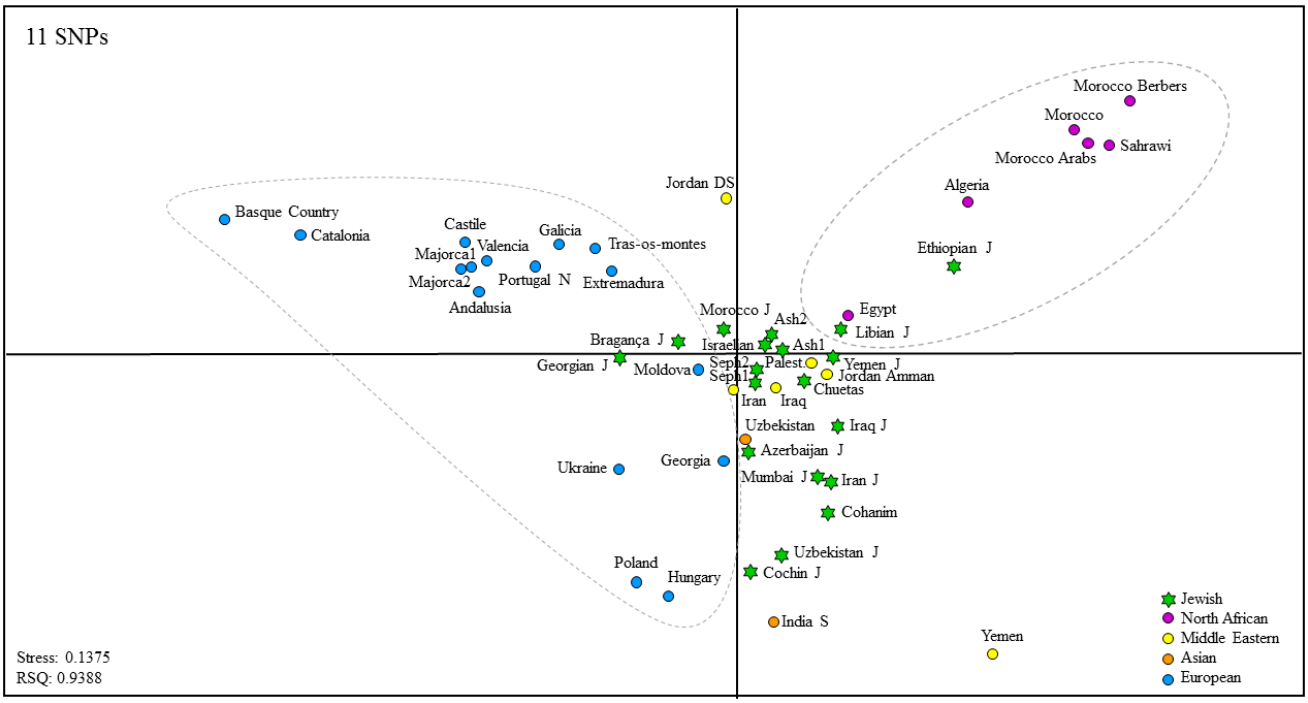


Figure 7. Multidimensional Scaling plot performed with populations and SNP distribution from Supplementary Table 3. To have a larger population available, the SNP classification was modified from Supplementary Table 3 as follows: Q-M242 was included in K-M9, and R1b1-P25 and R1b1a1a2-M269 were included in R1-M173. Names shortened for better viewing in the figure are: Jordan DS (Jordan Dead Sea) and Ash (Ashkenazi Jews). Circles defining each population are coloured following the legend code. Jewish populations are labelled with a Star of David.

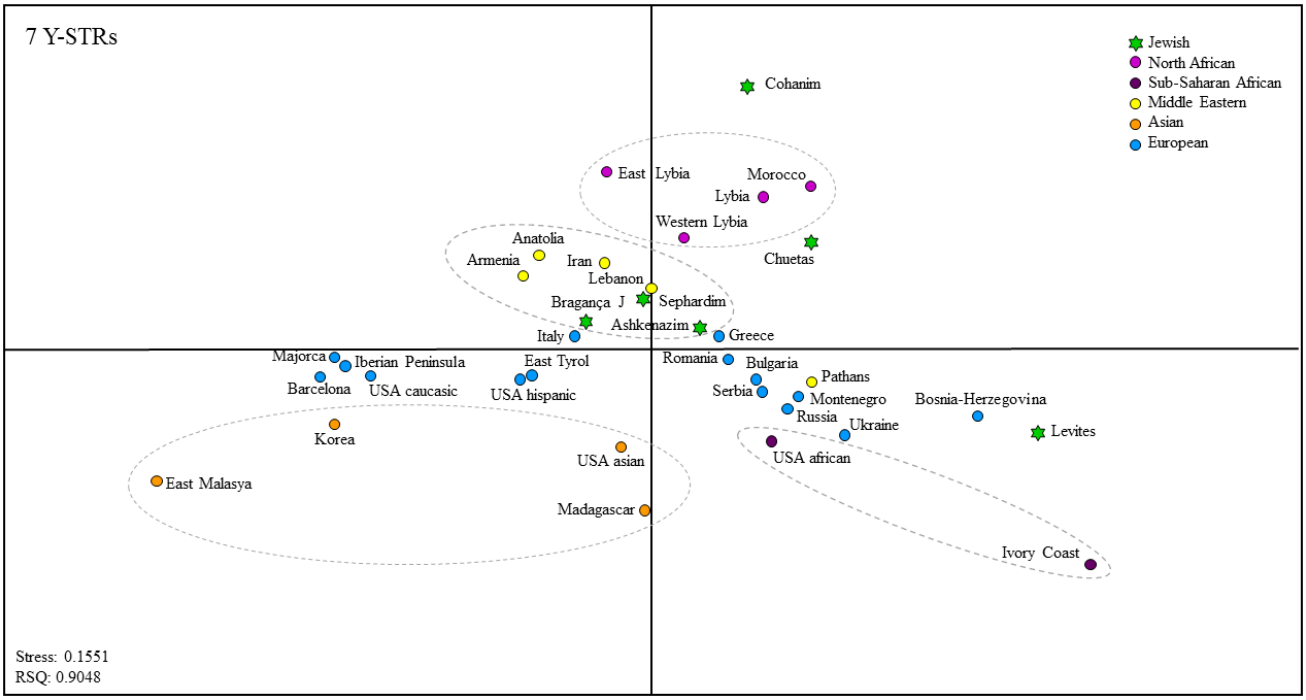


Figure 8. Multidimensional Scaling plot performed with populations from supplementary Table 3, with 7 available STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393). Circles defining each population are coloured following the legend code. Jewish populations are labelled with a Star of David.

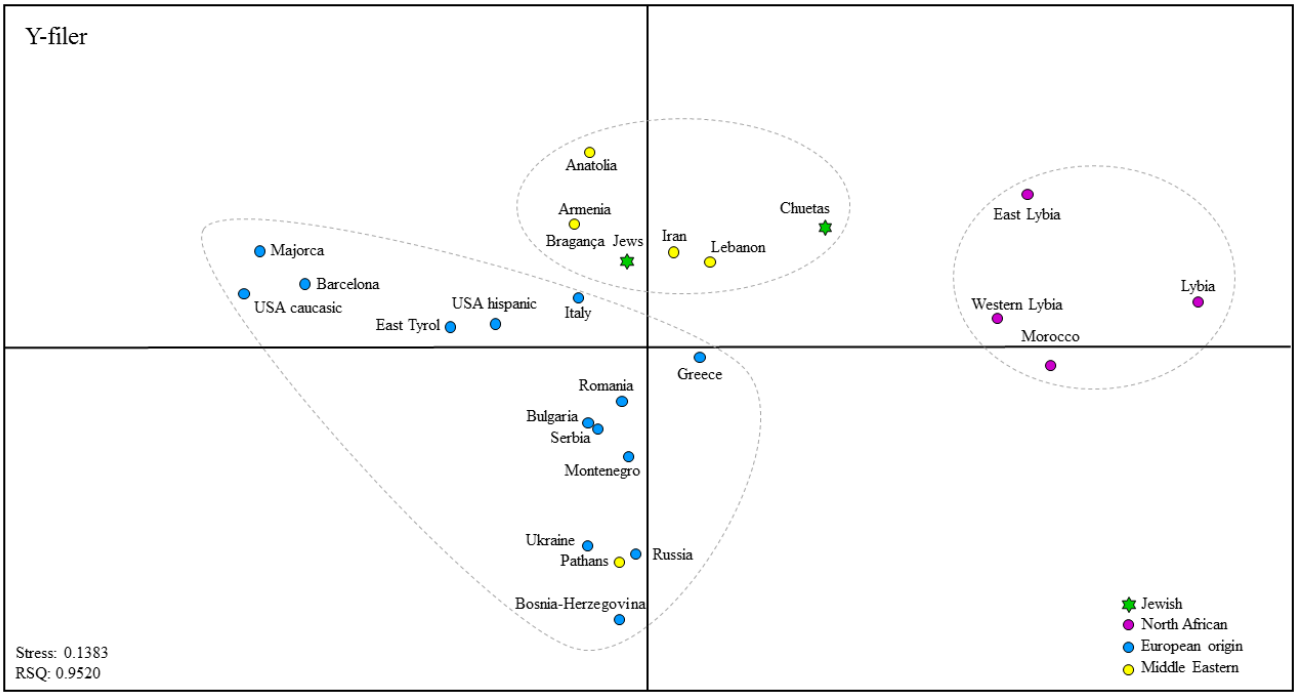


Figure 9. Multidimensional Scaling plot performed with populations from supplementary Table 3, with Y-filer STRs. Circles defining each population are coloured following the legend code. Jewish populations are labelled with a Star of David.

- Admixture

To assess the impact of the contributions of the initial Jewish population and the Majorcan host population to the current Chueta population, we carried out an admixture analysis, employing the mY estimators and treating the Chueta population as a hybrid of two parental populations: Majorcan and Sephardic Jews.

Admixture analysis based on Y-STR haplotypes indicated an extremely high proportion of ancestry from Sephardic Jewish (99.4%) sources.

It is important to consider factors that might influence the apparent proportions of the parental populations' ancestry that are estimated with admixture analysis. Choice of parental populations can have a major effect on the results. In this sense, we have to take into account: a) Majorcan individuals were sampled from a sizeable population that has maintained its existence *in situ*, but with a probable level of admixture with the other parental population. Adams et al. (2008) estimated a proportion of Sephardic Jewish admixture of 21.5%; b) Sephardic Jewish sample was taken from a comparatively small group of self-defined individuals whose ancestors lived in various parts of the Iberian Peninsula, and after the expulsion in the 15th century, settled in countries where pre-existing Jewish communities possibly lived. Therefore, they were probably subject to some degree of admixture with Iberians and/or with other Jewish communities; c) Jewish population settlement in different Iberian regions may not have been genetically homogeneous in origin or might have had different admixture levels with other Jews, for instance the Jewish community from Majorca (parental population of current Chuetas) probably had a greater North African Jewish admixture than other Sephardic groups. Other factors that could have influenced the results are the small sample size of the Majorcan parental population (n=46), and the available number of Y-STRs in the three populations.

4. Conclusions

In summary, Y-chromosome results show that most of Jewish communities are more similar to each other and to Middle Eastern populations, than to their host populations. The Chueta population has the same behaviour, which can be observed by the high prevalence of J2-M172 and J1-M267 haplogroups, and the lack of R1b1a1a2-M269. Haplogroup distribution and diversity in Chuetas is very similar to the Sephardic community and, indeed, the p-values of their genetic distances are on the limit of significance. Admixture analyses also manifest a minimal contribution of its host population. These differences with the Majorcan populations are also highlighted in the genetic distance values, and in the fact that the two populations do not share any haplotype. Even though the similarities with Sephardic Jews seem clear, there might be signatures of other Jewish communities' contribution, such as North African and Ashenazim, in the Chueta gene pool, which can be presumed, with the presence of haplogroups like E1b1b1a1-M78, Q1-P36.2, and R1a1a-M17. More extensive studies should be performed in order to better characterise the origin of all the Chueta Y-chromosome lineages.

Supplementary Table 1. Allele frequencies for the 17 STR loci of the Y-filer kit in Chueta (CHU) and Majorcan (MAJ) populations.

Allele	DYS456		DYS389I		DYS390		DYS389II		DYS458		DYS19		DYS393		DYS391		DYS439		DYS635		DYS392		Y GATA H4		DYS437		DYS438		DYS448		Alleles	DYS385	
N	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 99	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	CHU 100	MAJ 46	N	CHU 100	MAJ 46
8																								0.022							13-18	0.230	
9															0.180	0.087											0.320	0.130			13-16	0.110	0.065
10			0.040										0.010		0.720	0.543	0.050	0.022					0.090	0.022			0.490	0.196			14-16	0.240	0.028
11													0.010		0.090	0.370	0.400	0.283			0.840	0.261	0.510	0.413			0.150	0.043			11-14	0.050	0.391
12	0.010		0.030	0.239									0.420	0.196	0.010		0.430	0.543			0.010	0.065	0.320	0.500			0.030	0.609			16-16	0.120	
13	0.090	0.021	0.880	0.521							0.172	0.043	0.460	0.696			0.120	0.130			0.070	0.630	0.070	0.043			0.010	0.022			13-13	0.050	
14	0.140	0.043	0.040	0.239					0.110		0.465	0.674	0.090	0.109				0.022				0.043	0.010		0.630	0.370					12-12	0.050	0.022
15	0.370	0.587	0.010						0.090	0.043	0.253	0.261													0.260	0.500					11-15	0.030	0.087
16	0.290	0.326							0.130	0.283	0.071	0.022									0.060				0.110	0.130					13-15	0.010	
17	0.100	0.021							0.420	0.326	0.040										0.020										11-12	0.010	
17.2									0.100																						14-14	0.020	0.022
18									0.050	0.326																				0.130	19-19	0.010	
18.2									0.060																						13-14	0.010	0.043
19									0.020	0.022																			0.280	0.587	15-15	0.010	0.022
19.2									0.020																						15-16	0.010	0.022
20																			0.080	0.043									0.420	0.130	18-19	0.030	
21																			0.510	0.196									0.300	0.065	12-13	0.010	
22					0.220	0.109													0.270	0.109										0.087	12-14		0.043
23					0.540	0.217													0.140	0.500											14-20		0.022
24					0.190	0.565														0.109											11-13		0.087
25					0.050	0.087														0.043											13-17		0.065
26						0.022																									18-21		0.022
27							0.040																								10-14		0.022
28							0.150	0.174																							11-16		0.022
29							0.190	0.478																							14-18		0.022
30							0.320	0.326																									
31							0.220																										
32							0.080	0.022																									
NA	6	5	5	3	4	5	6	4	9	5	5	4	5	3	4	3	4	5	4	6	5	4	5	5	3	3	5	5	3	5	NA	17	17
D	0.749	0.558	0.224	0.627	0.628	0.627	0.791	0.648	0.777	0.721	0.691	0.486	0.600	0.476	0.445	0.573	0.645	0.620	0.647	0.700	0.288	0.540	0.631	0.589	0.529	0.610	0.640	0.585	0.662	0.623	D	0.861	0.833

N: Number of individuals, NA: number of different alleles, D: Gene Diversity

Supplementary Table 2. Haplotype (Ht) frequencies found in the Chueta (CHU) and Majorcan (MAJ) populations.

Ht	Haplogroup	CHU	MAJ	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a	DYS385b	DYS393	DYS391	DYS439	DYS635	DYS392	Y GATA H4	DYS437	DYS438	DYS448
01	E1b1b1a1-M78	1	-	15	13	23	29	17	14	18	19	13	10	13	20	11	12	14	10	21
02	E1b1b1a1-M78	1	-	15	13	23	30	17	14	18	19	13	10	13	20	11	12	14	10	21
03	E1b1b1a1-M78	1	-	15	13	23	30	17	14	18	19	13	11	13	20	11	12	14	10	21
04	E1b1b1a1-M78	1	-	16	13	24	30	16	13	16	16	13	10	11	22	11	12	14	10	20
05	E1b1b1a1-M78	1	-	17	13	24	30	16	13	15	16	13	10	11	22	11	12	14	10	20
06	E1b1b1a1-M78	1	-	17	15	24	32	16	13	16	16	13	10	11	23	11	12	14	10	20
07	E1b1b1a1-M78	1	-	15	13	23	30	17	14	19	19	13	10	12	20	11	12	14	10	21
08	E1b1b1a1-M78	7	-	17	13	24	30	16	13	16	16	13	10	11	22	11	12	14	10	20
09	E1b1b1a1-M78	-	1	15	14	24	32	18	14	18	21	12	10	12	20	13	13	14	10	20
10	E1b1b1b-M81	-	1	15	14	24	30	18	13	13	16	13	9	10	21	11	11	14	10	20
11	F-M213	2	-	15	14	22	30	18	15	16	16	12	10	11	20	11	12	14	9	20
12	G-M201	1	-	13	13	22	32	17	15	13	13	14	9	11	22	11	11	16	10	21
13	G-M201	1	-	12	13	22	32	17	15	13	13	14	9	11	22	11	11	16	10	21
14	G-M201	1	-	13	13	22	30	17	15	13	13	14	9	11	22	11	11	16	10	21
15	G-M201	1	-	13	13	22	32	17	15	13	13	13	9	11	22	11	11	16	10	21
16	G-M201	2	-	13	13	22	32	17	15	13	13	14	9	11	23	11	11	16	10	21
17	G-M201	1	-	13	13	22	32	18	15	13	13	14	9	11	22	11	11	16	10	21
18	G-M201	1	-	14	12	22	29	17	15	14	14	14	10	10	21	11	12	15	9	20
19	G-M201	1	-	15	12	22	29	16	15	14	14	14	10	12	20	11	12	16	10	21
20	G-M201	1	-	15	12	22	29	18	15	14	14	10	10	11	20	11	12	16	10	21
21	G-M201	-	1	14	12	22	29	17	15	15	15	14	10	12	20	11	11	16	11	21
22	G-M201	-	1	15	12	22	28	16	15	15	16	12	10	12	21	11	11	16	10	22
23	G-M201	-	1	15	12	23	28	17	15	13	14	14	10	12	21	12	11	16	10	22
24	I-M170	1	-	15	14	23	32	16	17	15	15	14	10	12	21	12	11	14	10	20
25	I-M170	-	1	13	12	26	28	16	15	13	17	13	10	11	21	11	10	15	10	19
26	I-M170	-	1	14	12	22	28	15	14	13	14	13	10	11	21	11	11	15	10	20
27	I-M170	-	2	15	14	23	30	18	15	11	13	13	11	11	21	11	11	16	10	20
28	I2a1a1-M26	1	-	14	13	23	28	17	17	11	12	13	10	11	21	11	13	15	10	21
29	I2a1a1-M26	2	-	14	13	23	29	17	17	12	12	13	10	11	21	11	13	15	10	21
30	J1-DYS458.2	1	-	16	13	23	30	17.2	14	13	18	12	10	12	21	11	11	14	10	20
31	J1-DYS458.2	1	-	16	13	23	31	17.2	14	13	18	11	10	12	21	11	11	14	10	20
32	J1-DYS458.2	7	-	16	13	23	31	17.2	14	13	18	12	10	12	21	11	11	14	10	20
33	J1-DYS458.2	1	-	16	13	23	31	17.2	14	13	18	12	10	12	21	11	11	14	10	21
34	J1-DYS458.2	6	-	16	13	23	31	18.2	14	13	18	12	10	13	21	11	11	14	10	20

35	J1-DYS458.2	2	-	16	13	23	31	19.2	14	13	18	12	10	13	21	11	11	14	10	20
36	J2-M172	1	-	15	13	22	31	15	16	13	14	12	10	11	23	11	10	15	9	21
37	J2-M172	2	-	15	13	23	30	17	14	13	18	12	10	11	22	11	11	16	9	20
38	J2-M172	9	-	14	13	23	28	14	15	13	16	13	9	12	21	11	12	14	9	21
39	J2-M172	1	-	14	13	23	28	14	16	13	16	13	9	12	21	11	12	14	9	21
40	J2-M172	1	-	15	13	22	31	15	16	13	15	12	10	11	21	11	12	14	9	21
41	J2-M172	1	-	15	13	23	28	14	15	13	16	13	9	12	21	11	12	14	9	21
42	J2-M172	1	-	15	13	23	29	17	14	14	16	12	10	11	21	11	11	15	9	19
43	J2-M172	12	-	15	13	23	30	17	14	14	16	12	10	11	21	11	11	15	9	19
44	J2-M172	1	-	15	13	23	30	17	14	14	16	12	10	12	21	11	11	15	9	19
45	J2-M172	1	-	16	10	24	27	17	14	13	18	12	10	12	21	11	11	15	10	20
46	J2-M172	1	-	16	10	24	27	17	14	13	18	12	10	12	22	11	11	15	10	20
47	J2-M172	1	-	16	10	24	27	18	14	13	18	12	10	12	22	11	11	15	10	20
48	J2-M172	1	-	16	10	25	27	17	14	13	18	12	10	12	22	11	11	15	10	20
49	J2-M172	-	1	15	14	24	30	16	14	14	18	12	10	11	22	11	11	16	10	20
50	J2-M172	-	1	16	12	22	30	18	14	13	17	12	9	11	22	11	12	15	9	22
51	J2-M172	-	1	16	12	22	30	18	14	13	17	12	10	11	22	11	12	15	9	22
52	J2-M172	-	1	16	13	23	29	16	15	13	16	12	9	12	22	11	12	14	9	21
53	J2-M172	-	1	16	13	23	29	17	15	13	16	12	9	12	22	11	12	14	9	21
54	P-92R7	-	1	15	14	23	30	16	14	14	20	13	10	11	25	12	11	15	11	19
55	Q1-P36.2	1	-	15	13	22	28	17	13	14	16	13	10	12	22	16	10	14	11	19
56	Q1-P36.2	4	-	15	13	22	29	17	13	14	16	13	10	12	22	16	10	14	11	19
57	Q1-P36.2	2	-	15	13	22	29	17	13	14	16	13	10	12	22	17	10	14	11	19
58	Q1-P36.2	1	-	15	13	22	29	17	?	14	16	13	10	12	22	16	10	14	11	19
59	R1a1a-M17	1	-	16	13	25	30	15	16	11	14	13	11	10	23	11	14	14	11	20
60	R1a1a-M17	2	-	16	13	25	31	15	16	11	14	13	11	10	23	11	13	14	11	20
61	R1a1a-M17	1	-	16	13	25	31	15	16	11	15	13	11	10	23	11	13	14	11	20
62	R1b1a2a-M18	2	-	13	13	24	28	15	15	12	12	13	11	12	23	13	11	14	11	19
63	R1b1a2a-M18	1	-	13	14	24	29	15	15	12	12	13	11	12	23	13	11	14	11	19
64	R1b1a1a2a-L23	-	1	15	13	24	29	16	15	11	15	13	11	11	23	13	13	15	12	19
65	R1b1a1a2a1a1-U106	-	1	15	13	23	29	18	14	11	14	13	10	12	23	14	12	15	12	19
66	R1b1a1a2a1a2-S116	1	-	16	13	23	29	19	14	11	14	13	12	12	21	13	12	15	12	19
67	R1b1a1a2a1a2-S116	1	-	16	13	24	29	17	14	11	14	12	10	12	23	13	11	14	12	19
68	R1b1a1a2a1a2-S116	-	1	15	12	24	28	16	14	11	15	13	10	12	23	13	12	15	12	19
69	R1b1a1a2a1a2-S116	-	1	15	12	25	28	17	14	11	13	13	10	14	24	13	11	15	12	19
70	R1b1a1a2a1a2-S116	-	1	15	12	25	28	17	14	11	13	13	11	13	24	13	11	15	12	19
71	R1b1a1a2a1a2-S116	-	1	15	13	24	29	16	14	11	14	13	11	12	23	13	11	14	12	18
72	R1b1a1a2a1a2-S116	-	1	15	13	24	29	16	14	11	15	13	11	12	23	13	12	15	12	19

73	R1b1a1a2a1a2-S116	-	1	15	13	24	29	17	14	11	14	14	10	12	25	13	11	14	12	18
74	R1b1a1a2a1a2-S116	-	1	15	13	24	29	18	14	11	14	13	10	12	23	13	12	14	12	19
75	R1b1a1a2a1a2-S116	-	1	15	13	24	30	16	14	11	14	13	11	13	23	13	12	15	12	19
76	R1b1a1a2a1a2-S116	-	1	15	13	24	30	16	16	11	14	13	11	12	24	13	12	15	12	19
77	R1b1a1a2a1a2-S116	-	1	15	13	24	30	18	14	11	14	12	10	11	23	13	11	14	12	18
78	R1b1a1a2a1a2-S116	-	1	15	14	24	29	17	14	11	14	13	10	12	23	13	11	14	12	18
79	R1b1a1a2a1a2-S116	-	1	15	14	24	30	17	14	11	15	13	11	12	23	13	12	15	12	19
80	R1b1a1a2a1a2-S116	-	1	16	12	25	28	17	14	11	14	13	10	12	23	13	12	15	13	19
81	R1b1a1a2a1a2-S116	-	1	16	13	23	29	17	15	10	14	13	10	12	23	13	12	14	12	19
82	R1b1a1a2a1a2-S116	-	1	16	13	24	29	16	14	11	14	12	11	12	23	13	11	14	12	19
83	R1b1a1a2a1a2-S116	-	1	16	13	24	29	17	14	12	14	13	11	13	23	13	11	14	12	18
84	R1b1a1a2a1a2-S116	-	1	16	13	24	29	18	14	11	14	13	11	12	23	13	12	15	12	19
85	R1b1a1a2a1a2-S116	-	1	16	13	24	29	19	14	11	14	13	11	13	23	13	12	15	12	19
86	R1b1a1a2a1a2-S116	-	1	16	13	24	30	18	14	11	14	14	10	12	23	12	12	15	12	19
87	R1b1a1a2a1a2-S116	-	1	16	14	24	30	17	14	11	16	13	11	13	23	13	12	14	12	18
88	R1b1a1a2a1a2-S116	-	1	17	13	24	29	18	14	11	14	13	11	11	23	13	12	14	12	19
89	R1b1a1a2a1a2b-U152	1		15	13	24	29	16	14	11	15	13	10	13	23	13	13	15	13	20
90	R1b1a1a2a1a2b-U152	-	1	15	13	25	29	18	14	12	14	13	10	12	23	13	12	14	12	19
91	R1b1a1a2a1a2b-U152	-	1	15	14	24	30	18	14	12	12	14	11	12	24	13	12	15	12	19
92	R1b1a1a2a1a2b-U152	-	2	16	13	24	29	17	14	11	14	13	10	12	23	13	12	15	12	19
93	R1b1a1a2a1a2c1-M529	1	-	17	13	24	29	19	14	12	13	13	11	12	23	13	11	15	12	19
94	R1b1a1a2a1a2c1-M529	-	1	15	13	24	29	17	15	11	14	13	10	13	23	13	12	15	12	19
95	R1b1a1a2a1a2c1-M529	-	1	16	13	24	29	16	14	11	14	13	11	11	24	13	8	15	12	19
96	T-M70	-	1	15	13	23	29	18	15	14	16	13	10	11	21	13	11	14	9	19
97	T-M70	-	1	15	14	23	30	15	13	14	14	13	10	12	21	14	11	14	9	19

Supplementary Table 3. Haplogroup frequencies of Jewish, European, Near-Middle Eastern, and North African populations (na= not available).

			E(xE1b1b1)	**E1b1b1**	**F**	**G**	**I**	**J(xJ2)**	**J2**	**K(xQxR1)**	**Q**	**R1(xR1a1axR1b1 xR1b1a1a2)**	**R1a1**	**R1b1 (xR1b1a1a2)**	**R1b1a1a2**	**Other**	**Reference**
		N	YAP/M1	M35	M213	M201	M170	12f2a	M172	M9	M242	M173	SRY10831.2/*M17	P25	M269		
Jews	Chueta	100	--	0.140	0.020	0.100	0.040	0.180	0.330	--	0.080	--	0.040	0.030	0.040	--	1
	Sephardic	174	0.006	0.098	--	0.167	0.011	0.230	0.190	0.109	na	0.138	0.052	na	na	--	2
	Sephardic	174	0.006	0.086	--	0.155	0.006	0.218	0.247	0.086	0.023	--	0.046	0.011	0.115	--	3
	Bragança	57	--	0.087	--	0.035	0.035	0.123	0.245	0.158	na	--	0.018	0.018	0.281	--	4
	Ashkenazi	856	0.004	0.183	0.007	0.071	0.039	0.217	0.190	0.067	na	0.092	0.130	na	na	--	2
	Ashkenazi	442	0.007	0.197	0.009	0.097	0.041	0.190	0.190	0.029	0.052	0.014	0.075*	0.100	na	--	5
	Azerbaijani	57	0.105	--	--	0.158	0.018	0.140	0.298	0.193	na	0.053	0.035	na	na	--	2
	Cochin	162	--	--	0.170	0.050	0.020	--	0.070	0.300	na	0.010	0.360*	na	na	0.020	6
	Cohanim	215	--	0.070	0.019	0.033	0.005	0.460	0.293	0.028	0.009	--	0.023	0.005	0.056	--	7
	Djerba	19	--	0.053	0.105	na	--	0.316	0.526	--	--	--	--	--	--	--	8
	Ethiopian	27	0.185	0.259	--	--	--	--	0.037	0.037	na	--	--	na	na	0.481	9; 2
	Georgian	62	0.226	--	--	0.048	--	0.016	0.129	0.161	na	0.339	0.081	na	na	--	2
	Iranian	49	--	0.122	0.082	--	--	0.163	0.122	0.449	na	0.020	0.041	na	na	--	2
	Iraqi	79	--	0.165	--	0.101	--	0.165	0.114	0.418	na	0.038	--	na	na	--	2
	Israelites	737	0.003	0.185	0.008	0.098	0.030	0.151	0.210	0.103	0.047	0.005	0.042	0.035	0.080	0.004	7
	Kurdish	95	0.121	na	na	na	na	0.222	0.152	0.404	na	na	0.040*	na	na	0.061	10
	Libyan	20	--	0.300	--	0.100	--	0.100	0.400	--	--	--	0.050	na	0.050	--	11
	Moroccan	83	--	0.133	--	0.193	--	0.205	0.133	0.120	na	0.217	--	na	na	--	2
	Mumbai (Bene Israel)	31	--	0.065	0.161	0.065	--	0.194	0.419	--	--	--	0.065	na	na	0.032	2
	Urbekistani	15	--	0.067	0.067	--	--	0.067	0.600	0.067	na	0.067	0.067	na	na	--	2
	Yemenite	74	--	0.216	--	0.068	--	0.338	0.162	0.122	na	0.054	0.027	na	na	0.014	2
Europe	Spain (Majorca)	46	--	0.044	--	0.065	0.087	--	0.109	0.065	--	--	--	--	0.631	--	1
	Spain (Majorca)	62	0.020	0.070	--	0.060	0.080	0.020	0.080	0.020	--	--	--	--	0.660	--	3
	Spain (Minorca)	37	--	0.190	--	--	0.030	--	0.030	--	--	--	0.030	--	0.730	--	3
	Spain (Ibiza)	54	--	0.080	--	0.130	0.020	--	0.040	0.170	--	--	--	--	0.570	--	3
	Spain (Aragon)	34	--	0.060	--	--	0.180	--	0.120	0.060	--	--	0.030	--	0.560	--	3
	Spain (Andalusia)	167	--	0.090	--	0.036	0.060	0.024	0.108	--	0.006	0.006	0.024	--	0.647	--	3
	Spain (Asturias)	20	--	0.150	--	0.050	0.100	--	0.150	0.050	--	--	--	--	0.500	--	3
	Spain (Basque country)	116	--	0.010	--	--	0.080	0.010	0.030	--	0.010	--	--	--	0.870	--	3
	Spain (Castile)	194	--	0.124	--	0.062	0.026	0.010	0.067	0.026	--	--	0.016	--	0.670	--	3
	Spain (Catalonia)	80	--	0.030	--	0.060	0.030	--	0.060	--	--	--	--	--	0.810	--	3
	Spain (Extremadura)	52	--	0.180	0.020	0.040	0.100	--	0.120	0.060	--	--	--	--	0.500	--	3
	Spain (Galicia)	88	--	0.170	--	0.060	0.100	0.010	0.070	0.010	--	--	--	--	0.570	--	3
	Spain (Valencia)	73	0.010	0.100	--	0.010	0.100	0.030	0.050	0.010	--	--	0.030	0.030	0.610	--	3
	Portugal (North)	60	0.030	0.120	--	0.120	0.020	0.020	0.070	0.020	--	--	0.030	--	0.590	--	3
	Portugal (South)	78	0.010	0.160	--	0.090	0.040	0.030	0.150	0.050	--	--	0.010	--	0.470	--	3
	Portugal (Tras-os-Montes)	30	--	0.200	--	0.033	0.134	0.033	--	0.033	na	--	0.033	--	0.534	--	4
	Albania	55	--	0.254	--	0.018	0.217	0.036	0.199	--	--	--	0.091	--	0.182	--	12
	Albania	93	--	0.430	0.011	0.011	0.247	--	0.226	0.022	--	--	0.011*	na	0.043	--	13
	Belarus	196	--	0.026	--	0.005	0.189	0.005	0.010	0.082	na	0.071	0.612	na	na	--	2
	Bosnia	85	--	0.129	0.035	0.035	0.482	0.024	0.095	0.012	na	--	0.153	--	0.035	--	14

	Bulgaria	808	0.004	0.217	0.006	0.048	0.266	0.034	0.105	0.024	0.004	--	0.175*	0.002	0.105	0.005	15
	Croatia	90	--	0.089	0.011	0.011	0.733	--	0.011	--	--	--	0.122	--	0.022	--	14
	Cyprus	42	--	0.381	0.214	0.024	--	na	0.286	0.071	na	--	--	0.024	na	--	16
	Cyprus	65	--	0.200	0.092	na	0.077	0.062	0.369	0.046	na	0.092	0.031*	na	na	0.031	8
	Czech	75	--	--	--	0.040	0.280	--	0.053	0.040	--	--	0.413	0.013	0.333	--	12
	France	64	--	0.078	0.016	0.031	0.141	0.016	0.047	0.031	na	0.641	--	na	na	--	2; 17
	France (Gascony)	24	--	--	--	--	--	--	0.040	--	--	--	--	--	0.970	--	3
	Georgia	66	--	0.030	--	0.318	0.015	0.045	0.318	0.075	--	--	0.106	--	0.091	--	12
	Germany	1215	0.002	0.062	0.043	na	0.236	na	0.040	0.045	na	0.389	0.179*	na	na	0.003	18
	Greece	92	--	0.218	--	0.033	0.098	0.022	0.207	0.054	--	--	0.163	0.022	0.174	0.011	12
	Greece (Corinthia)	104	--	0.298	0.010	0.029	0.192	--	0.163	0.029	--	0.010	0.173*	na	0.096	--	13
	Hungary	53	--	0.094	--	0.019	0.179	--	0.019	0.019	--	--	0.566	--	0.151	--	12
	Italy (Central)	88	--	0.102	--	0.091	0.102	0.011	0.307	0.080	--	--	0.023*	na	0.284	--	13
	Italy (Ionian)	125	--	0.240	--	0.144	0.096	0.040	0.176	0.056	--	--	0.040*	na	0.208	--	13
	Italy (North)	88	--	0.114	--	0.091	0.159	--	0.091	0.045	--	0.068	--	na	0.432	--	13
	Italy (North-East)	67	0.015	0.030	--	0.119	0.199	--	0.135	0.078	--	--	0.104	--	0.418	--	12
	Italy (Sicily)	314	--	0.131	0.016	0.099	0.115	--	0.213	0.045	--	0.019	0.029*	na	0.334	--	13
	Italy (South)	117	--	0.171	0.017	0.137	0.094	--	0.214	0.060	--	0.017	0.026*	na	0.265	--	13
	Macedonia	52	--	0.231	--	0.038	0.288	--	0.115	0.038	na	0.019	0.135*	0.135	na	--	19
	Moldova	125	--	0.128	--	0.008	0.288	0.040	0.040	0.024	0.008	--	0.304*	--	0.160	--	20
	Poland	99	--	0.040	--	--	0.040	0.010	0.010	0.010	--	--	0.566	--	0.182	--	12
	Romania	54	--	0.075	0.019	0.056	0.463	--	0.057	--	--	--	0.204*	--	0.130	--	20
	Russia (Central)	364	--	0.050	0.008	--	0.175	0.010	0.024	0.189	0.003	--	0.465	0.004	0.071	0.002	21
	Russia (Northern)	380	--	0.002	0.008	0.012	0.131	0.002	0.016	0.430	0.007	--	0.342	--	0.054	--	21
	Russia (Southern)	484	--	0.018	0.003	0.010	0.210	0.005	0.030	0.117	0.003	--	0.554	--	0.048	0.002	21
	Serbia	81	--	0.223	0.049	0.012	0.358	--	0.087	0.074	na	--	0.136	--	0.062	--	14
	Slovakia	594	0.002	0.066	0.045	0.045	0.253	na	0.029	0.040	na	--	0.372*	0.140	na	0.008	16
	Slovenia	75	--	0.027	--	0.026	0.307	0.013	0.026	--	--	--	0.387	--	0.213	--	12
	Turkey (Balkarian)	38	--	0.026	0.026	0.289	0.026	--	0.237	0.132	--	--	0.132	0.053	0.079	--	12
	Tyrol (East)	270	0.011	0.044	--	0.078	0.196	0.089	na	0.015	0.004	--	0.141*	0.004	0.418	--	22
	Ukraine	53	--	--	--	--	0.246	0.019	0.076	0.057	--	--	0.397*	0.057	0.151	--	20
Near-Middle East	Anatolia	523	0.006	0.107	0.006	0.109	0.054	0.092	0.243	0.117	0.019	0.004	0.069*	0.013	0.145	0.017	23
	Armenia	57	--	0.088	0.018	0.211	0.053	0.105	0.246	0.105	na	0.035	0.140	na	na	--	2
	Bedouin	34	--	0.118	--	0.029	--	0.676	--	--	na	--	0.147	na	na	0.029	2
	Druze	329	0.003	0.185	0.000	0.125	0.006	0.167	0.167	0.182	na	0.146	0.015	na	na	0.003	2
	Iran	938	0.021	0.070	0.014	0.117	0.005	0.089	0.225	0.127	0.056	0.012	0.143	0.008	0.093	0.017	24
	Iran (North)	33	--	--	--	0.152	--	0.091	0.242	0.152	0.091	--	0.091	--	0.152	0.030	25
	Iran (South)	117	0.017	0.051	0.043	0.128	--	0.120	0.232	0.111	0.026	--	0.188	--	0.060	0.026	25
	Iraq	154	0.013	0.123	--	0.019	0.006	0.311	0.240	0.071	0.019	--	0.084	na	0.110	--	26
	Jordan (Amman)	101	--	0.179	0.010	0.059	0.050	0.406	0.158	0.030	na	0.020	0.020*	na	0.059	0.010	27
	Jordan (Dead Sea)	45	--	0.444	--	--	--	0.089	0.067	--	na	0.400	--	na	--	--	27
	Lebanon	914	0.014	0.162	--	0.066	0.048	0.201	0.259	0.125	0.020	0.002	0.025*	0.005	0.073	0.001	28
	Marsh Arabs	143	--	0.063	--	0.014	--	0.811	0.035	0.021	0.028	--	--	--	0.028	--	26
	Muslim Kurds	95	0.074	na	na	na	na	0.116	0.284	0.242	na	na	0.116*	na	na	0.168	10
	Oman (Arabs)	121	0.090	0.140	0.020	0.020	--	0.380	0.100	0.110	na	0.010	0.090*	--	0.010	0.040	29
	Palestinian	292	0.058	0.164	0.017	0.086	0.007	0.360	0.134	0.065	na	0.079	0.007	na	na	0.024	2; 27
	Pakistan (Pathan)	270	--	--	0.052	0.141	--	--	0.096	0.159	0.052	--	0.493	--	--	0.007	30

	Qatar	72	0.084	0.056	0.014	0.028	--	0.583	0.084	0.042	--	--	0.069	--	0.014	0.028	31
	Samaritan	12	--	0.167	--	--	--	0.333	0.500	--	--	--	--	na	--	--	11
	Saudi Arabia	157	0.089	0.077	0.019	0.032	--	0.420	0.159	0.077	0.026	--	0.051*	--	0.019	0.032	32
	Syria	111	0.045	0.072	0.036	0.027	0.009	0.324	0.225	0.072	na	0.099	0.090	na	na	0.000	27
	United Arab Emirates	164	0.055	0.116	0.048	0.042	--	0.348	0.103	0.128	0.018	0.006	0.073	0.006	0.037	0.012	31
	Uzbekistan	140	--	0.043	0.071	0.050	0.007	0.021	0.179	0.257	na	0.064	0.229	na	na	0.079	2
	Yemen	62	0.048	0.113	--	0.016	--	0.726	0.096	--	--	--	--	--	--	--	31
North Africa	Algeria	46	0.020	0.630	0.090	--	--	0.150	0.040	--	--	--	--	--	0.070	--	33
	Algeria	201	0.104	0.622	0.035	na	na	0.164	0.005	--	0.015	0.045	na	na	na	0.010	34
	Egypt (Arabs)	147	0.030	0.360	--	0.090	0.010	0.200	0.120	0.100	na	0.020	0.03*	--	0.020	0.030	29
	Libya	43	0.050	0.550	0.225	0.050	--	na	0.050		na	--	--	0.025	na	--	16
	Morocco	147	0.080	0.750	0.010	0.010	0.010	0.080	0.030	0.010	--	--	--	--	0.030	--	35
	Morocco (Arabs)	49	--	0.760	--	0.020	--	0.100	0.100	--	na	--	--	--	0.020	--	9
	Morocco (Berbers)	64	0.070	0.800	--	0.050	--	0.060	--	--	na	--	--	--	--	0.030	9
	Saharawi	29	0.060	0.760	--	--	--	0.170	--	--	--	--	--	--	--	--	35
	Tunisia	139	0.020	0.540	0.040	--	--	0.280	0.030	0.010	--	--	--	--	0.080	--	33
Other	Ethiopia	98	0.337	0.378	0.051	--	--	na	0.020	0.061	na	--	--	--	--	0.153	16
	India (South)	102	--	--	0.461	--	--	--	0.059	0.245	na	--	0.157	na	na	0.078	2

1. Present study; 2. Behar et al. (2010) Nature 466: 238-242; 3. Adams et al. (2008) Am. J. Hum. Genet. 83: 725-736; 4. Nogueiro et al. (2010) Am. J. Phys. Anthropol. 141: 373-381; 5. Behar et al. (2004) Hum. Genet. 114: 354-365; 6. Chaubey et al. (2016) Sci. Rep. 6: 19166; 7. Hammer et al. (2009) Hum. Genet. 126: 707-717; 8. Capelli et al. (2006) Ann. Hum. Genet. 70(2): 207-225; 9. Cruciani et al. (2002) Am. J. Hum. Genet. 70(5): 1197-1214; 10. Nebel et al. (2001) Am. J. Hum. Genet. 69: 1095-1112; 11. Shen et al. (2004) Hum. Mutat. 24:248-260; 12. Bataglia et al. (2009) Eur. J. hum. Genet. 17: 820-830; 13. Tofanelli et al. (2016) Eur. J. Hum. Genet. 24(3): 429-436; 14. Marjanovic et al. (2005) Ann. Hum. Genet. 69(6): 757-763; 15. Karachanak et al. (2013) PLoS One 8(3): e56779; 16. Badro et al. (2013) PloS One 8(1): e54616.; 17. Semino et al. (2000) Science 290(5494): 1155-1159; 18. Kayser et al. (2005) Hum. Genet. 117(5): 428-443; 19. Bosch et al. (2006) Ann. Hum. Genet. 70(4): 459-487; 20. Varzaki et al. (2013) PloS One 8(1): e53731; 21. Balanovsky et al. (2008) Am. J. Hum. Genet. 82: 236-250; 22. Niedersttater et al. (2012) PloS One 7(7): e41885.; 23. Cinnioğlu et al. (2004) Hum. Genet. 114(2):127-148; 24. Crugni et al. (2012) PloS One 7(7): e41252; 25. Regueiro et al. (2006) Hum. Hered. 61:132–143; 26. Al-Zahery et al. (2011) BMC Evol. Biol. 11(1): 288; 27. Flores et al. (2005) J. Hum. Genet. 50(9): 435-441; 28. Zalloua et al. (2008) Am. J. Hum. Genet. 83(5): 633-642; 29. Luis et al. (2004) Am. J. Hum. Genet. 74: 532–544; 30. Lee et al. (2014) Forensic Sci. Int. Genet. 11: 111-116; 31. Cadenas et al. (2008) Eur. J. Hum. Genet. 16: 374–386; 32. Abu-Amero et al. (2009) BMC Genet. 10(1): 59; 33. Arredi et al. (2004) Am. J. Hum. Genet. 75: 338–345; 34. Bekada et al. (2015) PloS One 10(9): e0138453; 35. Bosch et al. (2001) Am. J. Hum. Genet. 68(4): 1019-1029.

Population used for comparison based on STRs that are not included in Supplementary Table 3, are the following:

Anatolia (Alakoc (2010) Forensic Sci. Int. Genet. 4: e135–e137); Armenia (Lowery et al. (2013) Legal Med. 15: 85-90); Barcelona (Sánchez et al. (2007) Forensic Sci. Int. 172: 211–217); Bosnia-Herzegovina (Kovacekic et al. (2013) Croat. Med. J. 54: 286-290); East Libya (Elmrghni et al. (2012) Forensic Sci. Int. Genet. 6(2): 224-227); East Malaysia (Meng Chang et al. (2009) Forensic Sci. Int. Genet. 3: e77–e80); Greece (Kovatsi et al. (2013) Forensic Sci. Int. Genet. 4: e21–e22); Iran (Roewer et al. (2009) Forensic Sci. Int. Genet. 4: e53–e55); Italy (Onofri et al. (2007) Int. J. Legal Med. 121: 234–237); Ivory Coast (Brucato et al. (2010) BMC Evol. Biol. 10: 314); Korea (Park et al. (2012) Int. J. Legal Med. 126: 589–599); Lebanon (Haber et al. (2011) J. Hum. Genet. 56(1): 29–33); Libya and Morocco (Fadhlaoui-Zid et al. (2013) PLoS One 8(11): e80293); Madagascar (Capredón et al. (2013) PLoS One 8(11): e80932); Montenegro (Mirabal et al. 2010) Am. J. Phys. Anthropol. 142: 380–390); Romania (Stanciu et al. (2010) Leg. Med. 12: 259-264); Russia (Roewer et al. (2008) Int. J. Legal Med. 122: 219–223); Serbia (Vaselinovik et al. (2008) Forensic Sci. Int. 176: e23–e28); USA populations (Coble et al. (2013) Forensic Sci. Int. Genet. 7: e66–e68); Ukraine (Mielnik-Sikorska et al. (2013) Forensic Sci. Int. Genet. 7: 200–203) and West Libya (Triki-Fendry et al. (2013) Forensic Sci. Int. Genet. 7: e59–e61).

Discussion

5. Discussion

Genetic structural patterns of human populations are usually a combination of long-term evolutionary forces and short-term social, cultural, and demographic processes. Historical events in the past, such as migrations, population expansions and colonisations in the last few thousand years, mould genetic diversity and leave their signature in the genome (Jobling, 2012). Therefore, on the one hand, historical information can help to understand the current genetic make-up of populations; and, on the other hand, genetic diversity can shed light on relatively recent past historical and demographic events.

An important factor shaping the genetic structure of populations involves barriers to mating and/or dispersal, which impact gene flow rates and might lead to genetic divergence of neighbouring populations (Jobling et al. 2014). In addition to geography; cultural, linguistic, ethnic, and religious features represent the most obvious barriers between human groups, as individuals do not tend to easily cross language or cultural boundaries when choosing a partner (Barbujani, 1997). Hence, populations separated by such barriers can result in genetic isolation. Consequently, one widely investigated topic in Human Population Genetics entails populations that have been considered to be geographically and/or culturally marginal (either currently or in the recent past). The term coined for these characteristics, usually connected to a small population size, is "isolate" (Boattini et al., 2011). Genetic characterization of these isolates is relevant to accomplish several scientific goals related to disease mapping or human diversity. Indeed, these studies are also interesting from an "archaeogenetics" point of view, as traces of ancient background or migration events are more likely to be found in these populations than in non-isolated communities (Shlush et al., 2008; Boattini et al., 2010). Their isolation may have helped these communities to preserve a more direct genetic link with their original background (Larruga et al., 2001).

Studies on the impact of cultural, religious, or political barriers have been carried out in Indian, Middle Eastern, North African, South American, and European populations, among others (e.g. Buhler et al., 2006; Khan et al., 2008; Ennafaa et al., 2009; Román-Busto et al., 2010; Haber et al., 2011; Lashgary et al., 2011; Tamang et al., 2012; Rojas et al., 2013; Pardiñas et al., 2014, Sarno et al., 2016). In Spain, there are some interesting human cultural isolates, such as Pasiegos, Maragatos, Vaqueiros, Caprarian, and Chuetas. Their endogamy, either voluntary or due to social discrimination because of their supposed different origin and/or different way of life, has created a barrier against external marriages, presumably causing genetic differentiation between these groups and their neighbours (Picornell et al., 1997; Sánchez-Velasco et al., 1999; 2003; Larruga et al., 2001; Maca-Meyer et al., 2003; Boattini et al., 2007; Gómez et al., 2014).

Chuetas are a social/cultural isolate inside a geographical island. Their ancestors had religious barriers with their Majorcan neighbours since they were Jews or converted Jews

with Crypto-Judaic practices; at that time they practised voluntary endogamy. After the end of the 17th century (Chueta period), social discrimination forced them into endogamy (Porqueres, 2001). The fact that this community lives in an island has helped to preserve the collective consciousness of their origin and to maintain the historical segregation they have suffered. Therefore, this isolation and the fact that Chuetas are, together with some Crypto-Jewish communities in Portugal, the only current Iberian populations whose ancestors can be traced back to the original Sephardic Jewish populations, make the Chuetas an interesting object of study in the field of Population Genetics.

In this study, we aimed to go further into the characterization of the surviving Chueta population in Majorca until nowadays. The criteria that were applied to the individuals to be included in the study were: a) self-identification as belonging to the community; b) having two Chueta surnames; and c) having at least three generations of maternal and paternal Chueta lineages. As the isolation of this community started, fortunately, to vanish in the mid twentieth century, finding people who accomplished the criteria was not easy, and most of them were born in the early 20th century. Therefore, it is important to highlight the relevance of this study, since in a few decades, Chuetas as a socially identifiable group will have disappeared.

Results of the present work demonstrate that, in spite of the opening of the community from the second half of the 20th century, current Chuetas still display a differentiation from their host population (Majorca), as has previously been observed (Picornell et al., 1994; 2005; Nevo et al., 1996). Genetic differences between Chuetas and their host population, and also between them and other Jewish populations, are confirmed in all chapters. For this reason, we have established appropriate forensic databases on new markers starting to be used in Forensic Genetics (autosomal Indels and X-chromosome markers) for populations with Jewish origin and Chuetas, since it is important to have specific databases for differentiated populations, in order to correctly weight the value of evidence in forensic casework in these populations (Zhivotovsky et al., 2001; Zarrabeitia et al. 2009).

- **Genetic diversity**

Considering that Chuetas are an isolated population with Jewish origin inside an island, their genetic diversity was presumed to be reduced, since Majorca is already an isolated population, with slightly reduced genetic diversity compared with other Spanish populations (Picornell et al., 2005; Rodríguez et al., 2009; Ferragut et al., 2015a). Autosomal Indels results did not show reduced diversity either in Majorca, Chuetas, or other Jewish populations compared with other worldwide populations (Pereira et al., 2009; Manta et al., 2012; Saiz et al., 2014). With respect to X-chromosome results, the Balearic Islands showed lower genetic diversity compared with a mainland reference population, Valencia, which has similar values to other European populations studied (Tillmar, 2012; Tomàs et al., 2012; Cainé et al., 2013). Chuetas did not present lower values than Majorcans

or other Jewish populations. Regarding uniparental markers, in maternal lineages diversity has not suffered a substantial reduction compared with Majorcans or in Y-chromosome haplogroups either, although haplotype diversity in Y-chromosome has indeed been considerably reduced. The entrance of new haplogroups in the Chuetas was probably done by one or a very few individuals, so the current people carrying a particular haplogroup would be descendants of the same individual. This fact could explain that all of them carry the same haplotype (or haplotypes with one or two mutational steps in STRs). In agreement with this hypothesis, most of the 15 Chueta surnames were found to be associated with a very reduced number of haplotypes. Nevertheless, other factors such as genetic drift due to bottlenecks cannot be ruled out as a cause of the reduced haplotype diversity observed.

Mitochondrial diversity levels in Jewish groups have been demonstrated to be different depending on the history of each community; that is, Ashkenazim show clear signs of a bottleneck in the beginning of their history, with just 4 major haplogroups as founders (Behar et al., 2004b; 2006; Brandstätter et al., 2008; Costa et al., 2013; Tian et al., 2015); in contrast, diversity in non-Ashkenazi populations is, in general, higher (Behar et al., 2008b). The Belmonte community in Portugal would be an exception, having extremely low diversity values. Genetic drift probably led to an impoverishment of the genetic diversity in this isolated Crypto-Jewish population. However, Bragança, another Crypto-Jewish Portuguese community, displayed high diversity values (Nogueiro et al. 2015a), supporting the theory that each Jewish community has undergone different historical processes that have led to distinct genetic structures. With regards to the Y-chromosome, studies on Jewish priest castes have also shown differentiation in terms of diversity: while Cohanim show a high prevalence of haplotype CMH, suggesting low diversity levels caused by founder effects (Skorecki et al., 1997; Hammer et al., 2009), Levites show greater diversity levels (Behar et al., 2003). Bottlenecks causing loss of diversity have also been identified in clinical studies, showing a high prevalence of specific illnesses in Jewish communities, mainly in Ashkenazim (Ostrer, 2001; Behar et al., 2004a; Rosner et al., 2009). A differential prevalence of illnesses such as Familial Mediterranean Fever and Hereditary Hemochromatosis has also been seen in the Chueta community (Buades et al., 1995; Domingo et al., 2000; Matas et al. 2006), suggesting signs of founder effect and genetic drift also in this community.

On the whole, Jewish populations show less genetic diversity compared with the host populations in mitochondrial lineages (Thomas et al., 2002; Jobling et al., 2014), but this is not always accomplished in Y-chromosome lineages. Thomas et al. (2002) suggested that an explanation could be that the different communities were founded by a small number of women (also confirmed in other studies: Behar et al., 2008b; Brandstätter et al., 2008; Costa et al., 2013); and, also, as Jewishness is traditionally inherited from mothers to their children, more admixture processes could have taken place in paternal lineages. Since one of the main conditions to be considered Chueta are the surnames and these are inherited from fathers (the mother's surname is lost in two generations), we could expect the Chuetas' diversity to act, in part, in the opposite way. In accordance with this

hypothesis, diversity is especially lower in Y-chromosome haplotypes, even though there is no substantial loss of diversity either in mtDNA or Y-chromosome in terms of haplogroups.

- **Comparison with Sephardic populations**

Regarding the research of the genetic legacy of Sephardic Jews in the paternal lineages of the Chueta population, results showed the differentiation of this population, not only from their host population, but also from other Jewish populations, although in the Y-chromosome Chuetas appear at the limit of significance with Sephardim. Taking into account the Y-chromosome haplogroup composition, Chuetas have three haplogroups with significant differences from their host population. On the one hand, the low frequency of haplogroup R (4%, similar to Middle Eastern and North African populations: Cruciani et al., 2002; Luis et al., 2004; Regueiro et al., 2006; Cadenas et al., 2008; Abu-Amero et al., 2009; Behar et al., 2010; Al-Zahery et al., 2011) compared with Majorcans (63%) who have typical values for Western European populations (Semino et al., 2000; Beleza et al., 2006; Adams et al., 2008; Nogueiro et al., 2010); and on the other hand, the presence of haplogroups J1 (18% vs 0% in Majorcans) and J2 (33% vs 10% in Majorcans). The pattern with these two main haplogroups followed by minor contributors, is the same as is found in other Jewish populations, mainly Sephardim (Adams et al., 2008). In addition to this clear Sephardic signature, Chuetas show evidence of contact with other Jewish communities such as Ashkenazim (haplogroups R1a1a and Q1a) (Behar et al., 2004a; Hammer et al., 2009; Underhill et al., 2010; Balanovsky et al. 2017) and North African Jews (haplogroup E1b1b1a1) (Shen et al., 2004). When the results are compared with the only two communities with Jewish origin still living in the Iberian Peninsula, in Belmonte and Bragança, major differences can be seen in the frequency of haplogroup R, which is the modal in Bragança Jews and represents 25% in Belmonte (Nogueiro et al., 2010; 2015b). Although a Middle Eastern signature can also be found in the paternal lineages of these Portuguese Crypto-Jewish communities, the high frequencies of haplogroup R suggest significant introgression from their neighbouring non-Jewish population, contrary to Chuetas.

Concerning maternal lineages, mtDNA results do not reveal the same levels of similarity with Sephardic Jews, but neither do they with the Majorcan population. In fact, Chuetas have a very unique mtDNA gene pool, as their modal lineage R0a2m (20.19%) is a very rare haplogroup, only described in three individuals in the literature to date (Gandini et al., 2016). Moreover, all the Chuetas' samples belonging to the R0a2m lineage display a private mutation, 13858G, defining a new sub-branch, R0a2m1, not previously described. R0a2m1 is, therefore, a Chueta exclusive branch that has been dated, obtaining a time of origin of ~0.43 Kya. This value points towards a very recent origin of the branch, in the late 16th century. This dating would indicate that this new branch appeared in the converted Majorcan Jewish community. Expansion following the bottleneck produced by the

inquisitorial actions on the Crypto-Jewish group, in the last quarter of the 17th century, could explain why this specific branch of this infrequent haplogroup has become the modal maternal lineage in the descendants of that group, the current Chuetas.

- **Inferring Sephardic original gene pool**

In this thesis, we also aimed to find out whether the original gene pool of the Sephardic population that lived on the Iberian Peninsula until their expulsion could be inferred from Chueta data, combined with data of other Iberian Jewish descendant populations (Bragança and Belmonte, in Portugal). Even though several studies pointed towards distinct communities being founded by different women after the Diaspora (Thomas et al., 2002; Picornell et al., 2006; Behar et al., 2004b; 2006; 2008b), the original Sephardic community is not in fact known, because the current Sephardic groups settled in other Mediterranean countries, probably admixed with previous Jewish populations living in those countries. The founder lineages in Chuetas were found to be haplogroups R0a2m and T1a (20.19% and 5.77% respectively), differentially from Portuguese Crypto-Jewish communities, where haplogroup HV0b is the main founder of the Belmonte population and also a founder in the Bragança community, together with N1, T2 and U2 (Behar et al., 2008b; Nogueiro et al., 2015a; 2015b). Of these four founder haplogroups in Portuguese Jews, three of them (HV0b, N1 and U2) are absent in Chuetas. All in all, the original maternal gene pool of Iberian Jews cannot be inferred from the current populations dating back to them. In fact, two different scenarios could explain this lack of genetic homogeneity: on the one hand, a heterogeneous original Sephardic community that settled the Iberian Peninsula differentially; or, on the other hand, the differential loss of original lineages in the different communities studied, due to genetic drift.

- **Evidence of gene flow**

One of the topics in Population Genetics in Jewish communities has been to try to unravel the impact of the complex system of interrelations between Jewish groups and, especially, between Jews and their host people. Different studies have inferred a variable degree of admixture and introgression from the corresponding Diaspora host populations (Ostrer and Skorecki, 2013). The highest values were found in Ethiopian, Lemba, and Indian Jews, in accordance with their history (Hammer et al., 2000; Thomas et al., 2002; Behar et al., 2010; Non et al., 2011; Chaubey et al., 2016; Waldman et al., 2016), although a Middle Eastern ancestry component is still found in their gene pool. Previous studies in Chuetas estimated the extent of the admixture with the local population at approximately 50% (Tomàs, 2002), a greater rate than the one found in other European Jews, but smaller than in other Jewish communities, where even conversions to Judaism from local people are known (Picornell et al., 1997). The social discrimination to which they were subjected must have restricted the influx of genes from the host population, but mixed marriages with the local population could have taken place more easily than mixed marriages

between other Jewish European populations and their host people where there were strict barriers of religion. At any rate, these values were calculated with some limitations to be taken into account: firstly, they are based on few markers; and, secondly, the current Sephardic population (from Bulgaria and Turkey) were used as a parental, without indeed knowing what the original Majorcan Jewish population gene pool was like.

In the present work, we also aimed to evaluate the admixture rate in Y-chromosome and mtDNA lineages. In Y-chromosome STRs, an extremely high proportion of Sephardic ancestry was found (99.4%) and, therefore, an insignificant contribution from the host population. However, as has been explained before, these results may be biased, since on the one hand, we do not know what the parental Sephardic population was exactly, and the population used might not be representative of the real parental population; and, on the other hand, signs of introgression (in X-chromosome and mtDNA results) from other Jewish and non-Jewish populations, other than Sephardim and Majorcans, have been detected. In mtDNA, admixture rate could not be estimated, as the unique Chuetas pattern cannot be explained as a mix between Majorcans and Sephardim, leaving the question of admixture levels unsolved.

Another question that we endeavoured to answer was whether there has been asymmetrical sex-biased contribution from the Sephardic and Majorcan populations to the Chuetas' gene pool, since in human populations it is not uncommon to find asymmetric patterns of mating (e.g. Carvajal-Carmona et al., 2000; Wood et al., 2005; Tomàs et al., 2006). Considering the fact that the mtDNA results did not turn out to be adequate to calculate admixture, X-chromosome results were used, since indeed two thirds of the X-chromosomes descend from maternal origin, therefore, X-chromosome polymorphisms mostly behave as matrilineal markers. Autosomal and X-chromosome Indels were the markers used in this analysis, since they are the same type of marker and we had genotyped a similar number of loci. Genetic distances between Jewish populations were expected to be 3/4 in autosomal with respect to the X-chromosome, taking into account that effective population size of the X-chromosome is 3/4 that of the autosomes, if only random genetic drift was responsible for this differentiation. However, the observed ratio suggested a gender-biased effective population size generated by the admixture process over time, probably resulting in greater male mobility, related to the well-known commercial relations between different Jewish populations (Mea, 2007; Chacón, 2009) which would have favoured genetic flow mediated preferentially by males. Otherwise, the comparison between populations with Jewish origin and their respective host population suggested larger introgression from female of the respective host non-Jewish populations than from males. All in all, the data suggest a generally sex-biased demographic history, not only in Chuetas but in all Jewish populations studied, with asymmetry between female and male effective population sizes both in the admixture processes among Jewish communities, and between them and their respective non-Jewish host populations.

Apart from the Sephardic heritage, a signature of other Jewish communities' influence was found in the Chuetas' gene pool. Genetic contact with North Africa was especially clear. This is witnessed in Y-chromosome by the presence of some haplogroups (E1b1b1a1) and in the Chuetas' position in MDS plots, where even though they cluster with the Middle Eastern populations, they tend towards North African populations; and also by the presence of haplogroup L (4.8%) in mtDNA results. This is probably because being a community in the middle of the Mediterranean and because of their good communications through trade, Chuetas had contact with North African Jews and even some Jews from those communities established in the island, as historical data confirm (Pons, 1984; Pérez, 2005; Font, 2007). Neither can certain introgression of Ashkenazim be ruled out with regard to our results: the presence of the already mentioned Y-chromosome haplogroups (R1a1a and Q1a), the K1a1b1a haplogroup in mtDNA, and the modal mitochondrial haplogroup in Chuetas, R0a2m, because although R0a originated in the Horn of Africa and spread to the Middle East and North Africa, the R02am branch has been found only in Ashkenazi Jews until now, apart from in Chuetas.

- **State of the art**

In the last few years some researchers have argued against using haplotype motifs as reliable Jewish ancestry DNA predictors (Tofanelli et al., 2014; Falk, 2015). We completely agree that it can be controversial to look for Jewishness in DNA, because in fact complex demographic histories are behind the genetic make-up of each particular population with Jewish origin. In any case though, in Chuetas we have detected a substantial Middle Eastern ancestry in their gene pool, clearly higher than the Middle East signals that are present in their host population, since both populations have lived in the Mediterranean area, and both have received gene influence from Neolithic, Phoenician and other events that spread from Levant to the entire Mediterranean region (Zalloua et al., 2008). The differential Middle Eastern footprint in Chuetas could not readily be explained by these expansions, since a similar signature would be expected to have been left in both populations. Thus, considering the unquestionable historical evidence that relate the Chuetas with the Jewish populations who settled in Majorca long ago in the past, the most reasonable explanation for the differences found between Majorcan non-Chueta population and Chuetas is their Jewish origin.

On-going research on full Y-chromosome sequences, genome-wide analyses and the increase of the number of mitogenomes sequenced in Chuetas and other Jewish populations, will enable much more precise comparisons and inferences of the original source of the haplogroups found in the Chueta population; and autosomal information to be complemented, thereby leading to the possibility of studying more recent demographic events, answering unsolved questions as regards the admixture rate, and also the possibility of studying predisposition to diseases in this group.

Conclusions

6. Conclusions

All in all, the results achieved during this doctoral thesis make it possible for the following foremost conclusions to be drawn:

1. The 38-plex panel of autosomal Indels analysed demonstrates that the Chueta population is statistically different from all the other populations studied (host population, and populations with Jewish origin). This panel of markers proved to be very useful for general forensic purposes and for molecular anthropology studies of populations from different continents, but unsuitable to distinguish populations at the subcontinental level, namely in the Mediterranean region.

2. In X-chromosome, Jewish populations tend to be closer to their respective host populations than to other Jewish groups. Chuetas are the most differentiated population, with statistical differences with respect to other Jewish populations and also to Majorca, their host population. The combined use of the 53 X-chromosome markers demonstrated their usefulness in complex kinship cases, since the greater the number of markers genotyped, the greater the improvement observed in paternity indexes.

3. Specific databases for Forensic Genetics purposes were established for 38 autosomal Indels and 53 X-chromosomal markers (9 Alu insertions, 12 STRs and 32 Indels) in Chuetas, Jewish populations, and Balearic populations, including the description of new alleles in X-STRs.

4. In paternal lineages, Chuetas differ significantly from other Jewish populations and Majorcans. However, they are at the limit of significance with Sephardic Jews, with whom they share a very similar haplogroup composition, with high proportions of the Middle Eastern haplogroups J2 (33%) and J1 (18%), and low frequencies (4%) of haplogroup R (the most frequent in Majorca, and in Western European populations). Therefore Chuetas' paternal lineages reflect a strong ancestral Sephardic component.

5. In mitochondrial DNA founder lineages, Chuetas present a distinctive gene pool, with no similarities to other Sephardic communities. The hallmark is the modal haplogroup in Chuetas, the rare R0a2m, found in 21% of the genotyped individuals. A private mutation found in all Chuetas' R0a samples enabled us to describe a new sub-branch, R0a2m1, which is very recent and originated in the island probably at the end of the 17th century. Present results support the hypothesis that each Jewish community was founded by different Middle Eastern maternal lineages.

6. Besides the Sephardic ancestry evidenced by the results, especially of the Y-chromosome, other contributions were detected. Mitochondrial haplogroups K1a1b1a and R0a2m, and Y-chromosome haplogroups Q1a and R1a1a could suggest Ashkenazim introgression; and L lineages in mtDNA, together with E1b1b1a1 in Y-chromosome, contacts with North African Jews. Non-Jewish contributions from outside the island cannot be ruled out either.

7. Isolation of the Chueta community has not led to the impoverishment of genetic diversity, which displayed high values in all the markers studied. A loss of haplotype diversity was only detected in the Y-chromosome. The low heterogeneity within each of the Chueta surnames would point towards a low number of founder individuals for each of the haplogroups.

8. Admixture estimate in Y-chromosome indicate an extremely high Sephardic contribution to the Chueta gene pool, leaving Majorcan contribution as residual. Even though in X-chromosome and mtDNA there are signs of non-Jewish contribution, the admixture level in maternal lineages could not be estimated, as Chuetas' mtDNA profile cannot be explained as a mix between Sephardic and Majorcan ones.

9. X-chromosomal markers suggest a sex-biased contribution to the Chueta gene pool, reflecting a genetic flow between Jewish communities mediated preferentially by males, and greater introgression from non-Jewish females of the host population.

10. Overall, combining the historical information available with the results achieved in this thesis will enable going further into the knowledge of the genetic legacy of Sephardic Jews in the Chueta population. This legacy is unquestionable in paternal lineages, while the maternal counterpart is more difficult to address due to the heterogeneity of mtDNA profiles in non-Ashkenazi Jewish populations. At any rate, the differential Middle Eastern mtDNA signature in Chuetas compared to their host population reflects their known historical Jewish origin. On-going research on full Y-chromosome sequences and genome-wide analysis would surely enable to complement the findings herein disclosed.

References

7. References

1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467(7319): 1061–1073.

1000 Genomes Project Consortium; Auton A; Brooks LD; Durbin RM; Garrison EP; Kang HM et al. (2015) A global reference for human genetic variation. Nature 526(7571): 68–74.

Aboukhalid R; Bouabdellah M; Abbassi M; Bentayebi K; Elmzibri M; Squalli D et al. (2010) Haplotype frequencies for 17 Y-STR loci (AmpFlSTR® Y-filer™) in a Moroccan population sample. Forensic Sci. Int. Genet. 4(3): e73–e74.

Abu-Amero KK; González AM; Larruga JM; Bosley TM and Cabrera VM (2007) Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol. Biol. 7(1): 32.

Abu-Amero KK; Hellani A; González AM; Larruga JM; Cabrera VM and Underhill PA (2009) Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC Genet. 10(1): 59.

Adams SM; Bosch E; Balaresque PL; Ballereau SJ; Lee AC; Arroyo E et al. (2008) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. Am. J. Hum. Genet. 83(6): 725–736.

Aguiló B (unpublished) Grup de genealogía. Memòria del carrer. www.memoriadelcarrer.com.

Aken BL; Ayling S; Barrell D; Clarker L; Curwen V; Fairley S et al. (2016) The Ensembl gene annotation system. Database 2016: 1–19.

Alonso A; Albarrán C; Martín P; Garcıa P; Garcıa O; de la Rúa C et al. (2003) Multiplex–PCR of short amplicons for mtDNA sequencing from ancient DNA. In: Brinkmann B and Carracedo A (Eds.) International Congress Series 1239: 585–588.

Álvarez L; Ciria E; Marques SL; Santos C and Aluja MP (2014) Y-chromosome analysis in a Northwest Iberian population: Unraveling the impact of Northern African lineages. Am. J. Hum. Biol. 26(6): 740–746.

Álvarez-Iglesias V; Mosquera-Miguel A; Cerezo M; Quintans B; Zarrabeitia MT; Cusco I et al. (2009) New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. PLoS One 4(4): e5112.

Al-Zahery N; Pala M; Battaglia V; Grugni V; Hamod MA; Kashani BH et al. (2011) In search of the genetic footprints of Sumerians: a survey of Y-chromosome and mtDNA variation in the Marsh Arabs of Iraq. BMC Evol. Biol. 11(1): 288.

Ambrosio B; Hernández C; Novelletto A; Dugoujon JM; Rodríguez JN; Cuesta P et al. (2010). Searching the peopling of the Iberian Peninsula from the perspective of two Andalusian subpopulations: a study based on Y-chromosome haplogroups J and E. Collegium antropologicum 34(4): 1215–1228.

Anderson S; Bankier AT; Barrell BG; de Bruijn MH; Coulson AR; Drouin J et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290(5806): 457–465.

Andrews RM; Kubacka I; Chinnery PF; Lightowlers RN; Turnbull DM and Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23(2): 147.

Antonarakis SE (2010) Human genome sequence and variation. In: Speicher et al. (Eds.) Vogel and Motulsky's Human Genetics. Berlin, Springer.

Assis TY (1992) Los judíos de la Corona de Aragón y sus dominios. In: Beinart H (Ed.) Moreset Sefarad. El legado de Sefarad. Jerusalem, Magnes-Universidad Hebrea.

Assis TY (1996-1997) Jaime II y los judíos en la corona de Aragón. Anales de la Universidad de Alicante. Historia Medieval 11: 331–342.

Athanasiadis G; Esteban E; Via M; Dugoujon JM; Moschonas N; Chaabani H et al. (2007) The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups. Eur. J. Hum. Genet. 15(5): 578–583.

Athey TW (2013) Whit Athey's Haplogroup Predictor. www.hprg.com/hapest5.

Atzmon G; Hao L; Pe'er I; Velez C; Pearlman A; Palamara PF et al. (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with Shared Middle Eastern Ancestry. Am. J. Hum. Genet. 86(6): 850–859.

Balanovsky O; Gurianov V; Zaporozhchenko V; Balaganskaya O; Urasin V; Zhabagin M et al. (2017) Phylogeography of human Y-chromosome haplogroup Q3-L275 from an academic/citizen science collaboration. BMC Evol. Biol. 17(1): 18.

Balaresque PL; Ballereau SJ and Jobling MA (2007) Challenges in human genetic diversity: demographic history and adaptation. Hum. Mol. Genet. 16(R2): R134–R139.

Balaresque P; Bowden GR; Adams SM; Leung HY; King TE; Rosser ZH et al. (2010) A predominantly neolithic origin for European paternal lineages. PLoS Biol. 8(1): e1000285.

Bandelt HJ; Forster P and Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16(1): 37–48.

Bandelt HJ; Kong QP; Richards M and Macaulay V (2006) Estimation of mutation rates and coalescence times: some caveats. In: Bandelt et al. (Eds.) Human Mitochondrial DNA and the Evolution of *Homo sapiens*. Berlin Heidelberg, Springer.

Barbaro A; Cormaci P; Falcone G; Votano S and La Marca A (2012) Distribution of 8 X-chromosomal STR loci in an Italian population sample (Calabria). Forensic Sci. Int. Genet. 6(6): e174–e175.

Barbieri C; Vicente M; Rocha J; Mpoloka SW; Stoneking M and Pakendorf B (2013) Ancient substructure in early mtDNA lineages of southern Africa. Am. J. Hum. Genet. 92(2): 285–292.

Barbujani G (1997) DNA Variation and Language Affinities. Am. J. Hum. Genet. 61(5):1011–1014.

Baron SW (1937) Social Religious History of the Jews. New York, Columbia University Press.

Bastos-Rodrigues L; Pimenta JR and Pena SDJ (2006) The Genetic Structure of Human Populations Studied through Short Insertion-Deletion Polymorphisms. Annals Hum. Genet. 70(5): 658–665.

Batini C; Hallast P; Zadik D; Delser PM; Benazzo A; Ghirotto S et al. (2015) Large-scale recent expansion of European patrilineages shown by population resequencing. Nat. Commun. 6: 7152.

Batzer MA and Deininger PL (2002) Alu Repeats and Human Genomic Diversity. Nature Rev. Genet. 3(5): 370–379.

Bedford FL; Yacobi D; Felix G and Garza FM (2013) Clarifying mitochondrial DNA subclades of T2e from Mideast to Mexico. J. Phylogen. Evol. Biol. 1: 121.

Behar DM; Thomas MG; Skorecki K; Hammer MF; Bulygina E; Rosengarten D et al. (2003) Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am. J. Hum. Genet. 73(4): 768–779.

Behar DM; Garrigan D; Kaplan ME; Mobasher Z; Rosengarten D; Karafet TM et al. (2004a) Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. Hum. Genet. 114(4): 354–365.

Behar DM; Hammer MF; Garrigan D; Villems R; Bonné-Tamir B; Richards M et al. (2004b) MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. Eur. J. Hum. Genet. 12(5): 355–364.

Behar DM; Metspalu E; Kivisild T; Achilli A; Hadid Y; Tzur S et al. (2006) The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am. J. Hum. Genet. 78(3): 487–497.

Behar DM; Villems R; Soodyall H; Blue-Smith J; Pereira L; Metspalu E et al. (2008a) The dawn of human matrilineal diversity. Am. J. Hum. Genet. 82(5): 1130–1140.

Behar DM; Metspalu E; Kivisild T; Rosset S; Tzur S; Hadid Y et al. (2008b) Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3(4): e2062.

Behar DM; Yunusbayev B; Metspalu M; Metspalu E; Rosset S; Parik J et al. (2010) The genome-wide structure of the Jewish people. Nature 466(7303): 238–242.

Behar DM; van Oven M; Rosset S; Metspalu M; Loogväli EL; Silva NM et al. (2012) A "copernican" reassessment of the human mitochondrial DNA tree from its root. Am. J. Hum. Genet. 90(5): 675–684.

Behar DM; Metspalu M; Baran Y; Kopelman NM; Yunusbayev B; Gladstein A et al. (2013) No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. Hum. Biol. 85(6): 859–900.

Bekada A; Benhamamouch S; Boudjema A; Fodil M; Menegon S; Torre C et al. (2010) Analysis of 21 X-chromosomal STRs in an Algerian population sample. Int. J. Legal Med. 124(4): 287–294.

Beleza S; Gusmão L; Lopes A; Alves C; Gomes I; Giouzeli M et al. (2006) Micro-Phylogeographic and Demographic History of Portuguese Male Lineages. Ann. Hum. Genet. 70(2): 181–194.

Ben-Sasson HH; Malamat A; Tadmor H; Stern M; Safrai S and Ettinger S (1976) A history of the Jewish people. Cambridge, MA: Harvard University Press.

Benson DA; Cavanaugh M; Clark K; Karsch-Mizrachi I; Lipman DJ; Ostell J et al. (2013) GenBank. Nucleic Acids Res. 41(D1): D36–D42.

Bentayebi K; Ramon MM; Castro JA; Barbaro A; Bouabdeallah M; Amzazi S et al. (2011) Inferring ethnicity from the X-chromosome ALU insertions: Data from Western Mediterranean human groups. Forensic Sci. Int. Genet. Sup. Ser. 3(1): e27–e28.

Bentayebi K; Picornell A; Bouabdeallah M; Castro JA; Aboukhalid R; Squalli D et al. (2012) Genetic diversity of 12 X-chromosomal short tandem repeats in the Moroccan population. Forensic Sci. Int. Genet. 6(1): e48–e49.

Bertorelle G and Excoffier L (1998) Inferring admixture proportions from molecular data. Mol. Biol. Evol. 15(10): 1298–1311.

Bestard J (1985) Los chuetas de Mallorca. In: Izard M (Ed.) Marginados, fronterizos, rebeldes y oprimidos. Barcelona, Serbal.

Bibiloni A (1992) Mercaders i navegants a Mallorca durant el segle XVII. Palma de Mallorca, El Tall.

Boattini A; Villegas MJB and Pettener D (2007) Genetic structure of La Cabrera, Spain, from surnames and migration matrices. Hum. Biol. 79(6): 649–666.

Boattini A; Pedrosi ME; Luiselli and Pettener D (2010) Dissecting a human isolate: Novel sampling criteria for analysis of the genetic structure of the Val di Scalve (Italian Pre-Alps). Ann. Hum. Biol. 37(4): 604–609.

Boattini A; Griso C and Pettener D (2011) Are ethnic minorities synonymous for genetic isolates? Comparing Walser and Romance populations in the Upper Lys Valley (Western Alps). J. Anthropol. Sci. 89: 161–173.

Bonné-Tamir B; Ashbel S and Bar-Shani S (1978) Ethnic communities in Israel: the genetic blood markers of the Moroccan Jews. Am. J. Phys. Anthropol. 49(4): 465–474.

Bonné-Tamir B; Johnson MJ; Natali A; Wallace DC and Cavalli-Sforza LL (1986) Human mitochondrial DNA types in two Israeli populations—a comparative study at the DNA level. Am. J. Hum. Genet. 38(3): 341–351.

Bonné-Tamir B and Adam A (1992) Genetic Diversity Among Jews: Diseases and Markers at the DNA Level. USA, Oxford University Press.

Botstein D; White RL; Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32(3): 314–331.

Brandstätter A; Niederstätter H; Pavlic M; Grubwieser P and Parson W (2007) Generating population data for the EMPOP database—an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. Forensic Sci. Int. 166(2): 164–175.

Brandstätter A; Egyed B; Zimmermann B; Tordai A; Padar Z and Parson W (2008) Mitochondrial DNA control region variation in Ashkenazi Jews from Hungary. Forensic Sci. Int. Genet. 2(1): e4–e6.

Braunstein B (1936) The Chuetas of Majorca. Conversos and the Inquisition of Majorca. New York, Ktav Pub Inc.

Brenner C and Morris J (1989) Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. Proceedings for the international symposium on human identification, Promega Corporation.

Brion M; Sobrino B; Blanco-Verea A; Lareu MV and Carracedo A (2005) Hierarchical analysis of 30 Y-chromosome SNPs in European populations. Int. J. Legal Med. 119(1): 10–15.

Buades J; Ben-Chetrit E and Levy M (1995) Familial Mediterranean fever in the" Chuetas" of Mallorca—origin in inquisition? Isr. J. Med. Sci. 31(8): 497–499.

Buhler S; Megarbane A; Lefranc G; Tiercy JM and Sánchez-Mazas A (2006) HLA-C molecular characterization of a Lebanese population and genetic structure of 39 populations from Europe to India–Pakistan. Tissue antígens 68(1): 44–57.

Bustamante CD and Ramachandran S (2009) Evaluating signatures of sex-specific processes in the human genome. Nat. Genet. 41(1): 8–10.

Butler JM (2005) Forensic DNA Typing: biology, technology, and genetics of STR markers. Burlington, Academic Press.

Butler JM (2009) Fundamentals of Forensic DNA Typing. Burlington, Academic Press.

Butler JM (2011) Advanced Topics in Forensic DNA Typing: Methodology. Burlington, Academic Press.

Butler JM (2014) Advanced Topics in Forensic DNA Typing: Interpretation. Burlington, Academic Press.

Cadenas AM; Zhivotovsky LA; Cavalli-Sforza LL; Underhill PA and Herrera RJ (2008) Y-chromosome diversity characterizes the Gulf of Oman. Eur. J. Hum. Genet. 16(3): 374–386.

Cainé L; Costa S and Pinheiro MF (2013) Population data of 12 X-STR loci in a North of Portugal sample. Int. J. Legal Med. 127(1): 63–64.

Calafell F and Larmuseau MH (2016) The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Hum. Genet. 1: 15.

Calderón R; Hernández CL; Cuesta P and Dugoujon JM (2015) Surnames and Y-chromosomal markers reveal low relationships in Southern Spain. PloS One 10(4): e0123098.

Callinan PA; Hedges DJ; Salem AH; Xing J; Walker JA; Garber RK et al. (2003) Comprehensive analysis of Alu-associated diversity on the human sex chromosomes. Gene 317(1-2): 103–110.

Calo CM; Garofano L; Mameli A; Pizzamiglio M and Vona G (2003) Genetic analysis of a Sicilian population using 15 short tandem repeats. Hum. Biol. 75(2): 163–178.

Cambra A; Muñoz-Saá I; Crespí C; Serra A; Etxagibel A; Matamoros N et al. (2009) MICA-HLA-B haplotype diversity and linkage disequilibrium in a population of Jewish descent from Majorca (the Balearic Islands). Hum. Immun. 70(7): 513–517.

Campbell CL; Palamara PF; Dubrovsky M; Botigué LR; Fellous M; Atzmon G et al. (2012) North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. Proc. Natl. Acad. Sci. USA 109(34): 13865–13870.

Cann RL; Stoneking M and Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325: 31–36.

Canturk KM; Emre R; Kınoglu K; Başpınar B; Sahin F and Ozen M (2014) Current status of the use of single-nucleotide polymorphisms in Forensic practices. Genet. Test. Mol. Biomarkers 18(7): 455–460.

Capelli C; Redhead N; Romano V; Calì F; Lefranc G; Delague V et al. (2006) Population structure in the Mediterranean basin: A Y chromosome perspective. Ann. Hum. Genet. 70(2): 207–225.

Cardoso S; Sevillano R; Gamarra D; Santurtún A; Martínez-Jarreta and de Pancorbo MM (2017) Population genetic data of 38 insertion-deletion markers in six populations of the northern fringe of the Iberian Peninsula. Forensic Sci. Int. Genet. 27: 175–179.

Carmelli D and Cavalli-Sforza LL (1979) The genetic origin of the Jews: a multivariate approach. Hum. Biol. 51(1): 41–61.

Carracedo A; Butler JM; Gusmao L; Linacre A; Parson W; Roewer L et al. (2013) New guidelines for the publication of genetic population data. Forensic Sci. Int. Genet. 7(2): 217–220.

Carrel L and Willard HF (2005) X inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434(7031): 400–404.

Carvajal-Carmona LG; Soto ID; Pineda N; Ortíz-Barrientos D; Duque C; Ospina-Duque J et al. (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. Am. J. Hum. Genet. 67(5): 1287–1295.

Casanova M; Leroy P; Boucekkine C; Weissenbach J; Bishop C; Fellous M et al. (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230(4732): 1403–1406.

Cavalli-Sforza LL and Bodmer WF (1971). The Genetics of Human Populations. San Francisco, WH Freeman.

Cerezo M; Gusmão L; Černý V; Uddin N; Syndercombe-Court D; Gómez-Carballa A et al. (2016) Comprehensive Analysis of Pan-African Mitochondrial DNA Variation Provides New Insights into Continental Variation and Demography. J. Genet. Genomics 43(3): 133–143.

Černý V; Mulligan CJ; Fernandes V; Silva NM; Alshamali F; Non A et al. (2011) Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South Arabia. Mol. Biol. Evol. 28(1): 71–78.

Chacón JM (2009) Los judíos mallorquines en el comercio y en las redes de intercambio valencianas y mediterráneas del medievo. Anales de la Universidad de Alicante. Revista de Historia Medieval 15: 75–85.

Chacón JM (2013) Viure al marge. La vida quotidiana dels jueus de Mallorca (segles XIII- XIV). Palma, Lleonard Muntaner ed.

Chakravarti A (1999) Population genetics—making sense out of sequence. Nat. Genet. 21(S1): 56–60.

Chaubey G; Singh M; Rai N; Kariappa M; Singh K; Singh A et al. (2016) Genetic affinities of the Jewish populations of India. Sci. Rep. 6: 19166.

Chiaroni J; Underhill PA and Cavalli-Sforza LL (2009) Y chromosome diversity, human expansion, drift, and cultural evolution. Proc. Natl. Acad. Sci. USA 106(48): 20174–20179.

Chinnery PF (2006) The transmission and segregation of mitochondrial DNA in Homo sapiens. In: Bandelt et al. (Eds.) Human Mitochondrial DNA and the Evolution of Homo sapiens. Berlin Heidelberg, Springer.

Chinnery PF and Hudson G (2013) Mitochondrial genetics. Br. Med. Bull. 106: 135–159.

Chinniah R; Vijayan M; Thirunavukkarasu M; Mani D; Raju K; Ravi PM et al. (2016) Polymorphic Alu Insertion/Deletion in Different Caste and Tribal Populations from South India. PLoS One 11(6): e0157468.

Chouraqui A (1968) Between East and West: A History of the Jews of North Africa. Philadelphia, Jewish Publication Society of America.

Cinnioğlu C; King R; Kivisild T; Kalfoğlu E; Atasoy S; Cavalleri GL et al. (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum. Genet. 114(2): 127–148.

Comas D; Calafell F; Benchemsi N; Helal A; Lefranc G; Stoneking M et al. (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. Hum. Genet. 107(4): 312–319.

Contreras A (1997) Los médicos judíos en la Mallorca bajomedieval. Siglos XIV-XV. Palma de Mallorca, M. Font ed.

Cordaux R and Batzer MA (2009) The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. 10(10): 691–703.

Cortés G (1985) Historia de los judíos mallorquines y sus descendientes cristianos. Palma de Mallorca, M. Font ed.

Costa MD; Pereira JB; Pala M; Fernandes V; Olivieri A; Achilli A et al. (2013) A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nat. Commun. 4: 2543.

Crespí C; Milà J; Martínez-Pomar N; Etxagibel A; Muñoz-Saa I; Priego D et al. (2002) HLA polymorphism in a Majorcan population of Jewish descent: comparison with Majorca, Minorca, Ibiza (Balearic Islands) and other Jewish communities. Tissue Antigens 60(4): 282–291.

Cruciani F; Santolamazza P; Shen P; Macaulay V; Moral P; Olckers A et al. (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am. J. Hum. Genet. 70(5): 1197–1214.

Cruciani F; La Fratta R; Trombetta B; Santolamazza P; Sellitto D; Colomb EB et al. (2007) Tracing past human male movements in northern/eastern Africa and western Eurasia: new

clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol. Biol. Evol. 24(6): 1300–1311.

Cruciani F; Trombetta B; Antonelli C; Pascone R; Valesini G; Scalzi V et al. (2011a) Strong intra-and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. Forensic Sci. Int. Genet. 5(3): e49–e52.

Cruciani F; Trombetta B; Massaia A; Destro-Bisol G; Sellitto D and Scozzari R (2011b) A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. Am. J. Hum. Genet. 88(6): 814–818.

Dale AW (1882) The Synod of Elvira and Christian Life in the Fourth Century. London, Macmillan.

Das R; Wexler P; Pirooznia M and Elhaik E (2016) Localizing Ashkenazic Jews to primeval villages in the ancient Iranian lands of Ashkenaz. Genome Biol. Evol. 8(4): 1132–1149.

Davidovic S; Malyarchuk B; Aleksic J; Derenko M; Topalovic V; Litvinov A et al. (2017) Mitochondrial super-haplogroup U diversity in Serbians. Ann. Hum. Biol. 19:1–11.

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am. J. Hum. Genet. 67(5): 1055–1061.

de Muntaner P (2002) Martí: Una familia del brazo noble mallorquín durante el siglo XVII en Homenaje a Guillem Rosselló Bordoy. (Vol. 2). Palma, Govern de les Illes Balears.

de Muntaner P (2012) La descendencia del juez Mossé Xabi, judío de Mallorca del siglo XIV. Memòries de la Reial Acadèmia Mallorquina d'Estudis Genealògics, Heràldics i Històrics (22): 23–61.

Deininger P (2011) Alu elements: know the SINEs. Genome Biol. 12(12): 236.

Desmarais D; Zhong Y; Chakraborty R; Perreault C and Busque L (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J. Forensic Sci. 43(5): 1046–1049.

Di Giacomo F; Luca F; Popa LO; Akar N; Anagnou N; Banyko J et al. (2004) Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum. Genet. 115(5): 357–371.

Diegoli TM (2015) Forensic typing of short tandem repeat markers on the X and Y chromosomes. Forensic Sci. Int. Genet. 18: 140–151.

Diegoli TM and Coble MD (2015) Characterization of X Chromosomal Short Tandem Repeat Markers for Forensic Use. Washington DC, National Institute of Justice (NIJ).

Domingo C; Touitou I; Bayou A; Ozen S; Notarnicola C; Dewalle M et al. (2000) Familial Mediterranean fever in the 'Chuetas' of Mallorca: a question of Jewish origin or genetic heterogeneity. Eur. J. Hum. Genet. 8(4): 242–246.

Dupanloup I and Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol. Biol. Evol. 18(4): 672–675.

Edelmann J; Hering S; Augustin C and Szibor R (2008) Characterisation of the STR markers DXS10146, DXS10134 and DXS10147 located within a 79.1 kb region at Xq28. Forensic Sci. Int. Genet. 2(1): 41–46.

Edelmann J; Lutz-Bonengel S; Naue J and Hering S. (2012) X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit. Forensic Sci. Int. Genet. 6(1): e24–e34.

Edelmann J; Kohl M; Dressler J and Hoffmann A (2016) X-chromosomal 21-indel marker panel in German and Baltic populations. Int. J. Legal Med. 130(2): 357–360.

Edholm OG and Samueloff S (1973) Biological studies of Yemenite and Kurdish Jews in Israel and other groups in southwest Asia. I. Introduction, background and methods. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 266(876): 85–95.

Egeland T and Sheehan N (2008) On identification problems requiring linked autosomal markers. Forensic Sci. Int. Genet. 2(3): 219–225.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nature Rev. Genet. 5(6): 435–445.

Elmrghni S; Coulson-Thomas YM; Kaddura M; Dixon RA and Williams DR (2012) Population genetic data for 17 Y STR markers from Benghazi (East Libya). Forensic Sci. Int. Genet. 6(2): 224–227.

El-Sibai M; Platt DE; Haber M; Xue Y; Youhanna SC; Wells RS et al. (2009) Geographical Structure of the Y-chromosomal Genetic Landscape of the Levant: A coastal-inland contrast. Ann. Hum. Genet. 73(6): 568–581.

Ennafaa H; Cabrera VM; Abu-Amero KK; González AM; Amor MB; Bouhaha R et al. (2009) Mitochondrial DNA haplogroup H structure in North Africa. BMC Genet. 10(1): 8.

Excoffier L; Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2): 479–491.

Excoffier L and Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10(3): 564–567.

Falk R (2015) Genetic markers cannot determine Jewish descent. Front. Genet. 5: 462.

Falush D; Stephens M and Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4): 1567–1587.

Fernandes V; Triska P; Pereira JB; Alshamali F; Rito T; Machado A et al. (2015) Genetic stratigraphy of key demographic events in Arabia. PloS One 10(3): e0118625.

Ferragut JF; Bentayebi K; Castro JA; Ramon C and Picornell A (2015a) Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations. Int. J. Legal Med. 129(2): 253–255.

Ferragut JF; Castro JA; Ramon C and Picornell A (2015b) Genetic diversity of 12 X-chromosomal short tandem repeats in Jewish populations. Forensic Sci. Int. Genet. Sup. Ser. 5: e327–e329.

Ferragut JF; Marques SL; Ramon C; Castro JA; Amorim A; Álvarez, L et al. (2015c) Founding mothers of Chueta population. Forensic Sci. Int. Genet. Sup. Ser. 5: e492–e494.

Ferragut JF; Pereira R; Castro JA; Ramon C; Nogueiro I; Amorim A et al. (2016) Genetic diversity of 38 insertion–deletion polymorphisms in Jewish populations. Forensic Sci. Int. Genet. 21: 1–4.

Flegontov P; Kassian A; Thomas MG; Fedchenko V; Changmai P and Starostin G (2016) Pitfalls of the geographic population structure (GPS) approach applied to human genetic history: A case study of Ashkenazi Jews. Genome Biol. Evol. 8(7): 2259–2265.

Font M (1993) Anonymous. Muntaner, Lleonard (introducció). Relación de los Sanbenitos: 1755. Palma, M. Font ed.

Font MS (2007) La fe vençuda. Jueus, conversos i xuetes a Mallorca. Palma, M. Font ed.

Forster P; Harding R; Torroni A and Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. 59(4): 935–945.

Forteza M (1972) Els descendens dels jueus conversos de Mallorca. Palma de Mallorca, Moll.

Francalacci P; Morelli L; Underhill PA; Lillie AS; Passarino G; Useli A et al. (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. Am. J. Phys. Anthropol. 121(3): 270–279.

Francalacci P and Sanna D (2008) History and geography of human Y-chromosome in Europe: a SNP perspective. J. Anthropol. Sci. 86: 59–89.

Fregel R; Cabrera V; Larruga JM; Abu-Amero KK and González AM (2015) Carriers of mitochondrial DNA macrohaplogroup N lineages reached Australia around 50,000 years ago following a northern Asian route. PLoS One 10(6): e0129839.

Freitas NS; Resque RL; Ribeiro-Rodrigues EM; Guerreiro JF; Santos NP; Ribeiro-dos-Santos Â et al. (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. Int. J. Legal Med. 124(6): 589–593.

Fu Q; Mittnik A; Johnson PL; Bos K; Lari M; Bollongino R et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. 23(7): 553–559.

Gandini F; Achilli A; Pala M; Bodner M; Brandini S; Huber G et al. (2016) Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. Sci. Rep. 6: 25472.

Garau RF (1931) La fe triunfante en quatro autos: Celebrados en Mallorca por el Santo Oficio de la Inquisición en que han salido ochenta y ocho reos, y de treinta y siete relajados sólo hubo tres pertinaces. Palma, Library of Alexandria.

Garrigan D and Hammer MF (2006) Reconstructing human origins in the genomic era. Nature Rev. Genet. 7(9): 669–680.

Gayà-Vidal M; Dugoujon JM; Esteban E; Athanasiadis G; Rodríguez A; Villena M et al. (2010) Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. Am. J. Hum. Biol. 22(2): 154–162.

Gerber S; Rozet JM; Takezawa SI; dos Santos LC; Lopes L; Gribouval O et al. (2000) The photoreceptor cell-specific nuclear receptor gene (PNR) accounts for retinitis pigmentosa in the Crypto-Jews from Portugal (Marranos), survivors from the Spanish Inquisition. Hum. Genet. 107(3): 276–284.

Ghiani ME; Stefano Piras I; Mitchell RJ and Vona G (2004) Y-chromosome 10 locus short tandem repeat haplotypes in a population sample from Sicily Italy. Leg. Med. (Tokyo) 6(2): 89–96.

Giguère G (2006) Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS. Tutorials in Quantitative Methods for Psychology 2(1): 27–38.

Gomes V; Sanchez-Diz P; Amorim A; Carracedo A and Gusmão L (2010) Digging deeper into East African human Y chromosome lineages. Hum. Genet. 127(5): 603–613.

Gómez P; Gómez J; Corao AI; De Canga J and Coto E (2014) Effect of mitochondrial, APOE. ACE and NOS3 gene polymorphisms on cardiovascular risk factors among the Vaqueiros de Alzada, a Northern Spain human isolate. Ann. Hum. Biol. 41(1): 94–97.

Gómez-Carballa A; Pardo-Seco J; Fachal L; Vega A; Cebey M; Martinón-Torres N et al. (2013) Indian signatures in the westernmost edge of the European romani diaspora: new insight from mitogenomes. PloS One 8(10): e75397.

Gómez-Pérez L; Alfonso-Sánchez MA; Pérez-Miranda AM; de Pancorbo MM and Peña JA (2007) Utilidad de las inserciones *Alu* en los estudios de mestizaje. Antropo. 14: 29–36.

Gonçalves R; Freitas A; Branco M; Rosa A; Fernandes AT; Zhivotovsky LA et al. (2005) Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry. Ann. Hum. Genet. 69(4): 443–454.

González AM; Larruga JM; Abu-Amero KK; Shi Y; Pestano J and Cabrera VM (2007) Mitochondrial lineage M1 traces an early human backflow to Africa. BMC Genomics 8(1): 223.

González-Pérez E; Via M; Esteban E; López-Alomar A; Mazieres S; Harich N et al. (2003) Alu Insertions in the Iberian Peninsula and North West Africa–Genetic Boundaries or Melting Pot? Coll. Antropol. 27(2): 491–500.

González-Pérez E; Moral P; Via M; Vona G; Varesi L; Santamaria J et al. (2007) The ins and outs of population relationships in west-Mediterranean islands: data from autosomal Alu polymorphisms and Alu/STR compound systems. J. Hum. Genet. 52(12): 999–1010.

Goodman RM (1979) Genetic disorders among the Jewish people. Baltimore, The Johns Hopkins University Press.

Goodwin W; Linacre A and Hadi S (2007) An Introduction to Forensic Genetics. Chichester, John Wiley & Sons.

Graen D (2012) Oldest Jewish Archaeological Evidence on the Iberian Peninsula. https://www.uni-jena.de/en/News/PM120525_Schrifttafel.html.

Griffiths AJ; Wessler SR; Carroll SB and Doebley J (2012) Introduction to genetic analysis. New York, W. H. Freeman and Company.

Grover A and Sharma PC (2016) Development and use of molecular markers: past and present. Crit. Rev. Biotechnol. 36(2): 290–302.

Guix P; Picornell A; Parera M; Galmes A; Obrador A; Ramon MM et al. (2002) Distribution of HFE C282Y and H63D mutations in the Balearic Islands (NE Spain). Clinical Genet. 61(1): 43–48.

Gusmão L; Butler JM; Carracedo A; Gill P; Kayser M; Mayr WR et al. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int. J. Leg. Med. 120(4): 191–200.

Haasl RJ and Payseur BA (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106(1): 158–171.

Haber M; Platt DE; Badro DA; Xue Y; El-Sibai M; Bonab MA et al. (2011) Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. Eur. J. Hum. Genet. 19(3): 334–340.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series 41(41): 95-98.

Hallast P; Batini C; Zadik D; Delser PM; Wetton JH; Arroyo-Pardo E et al. (2015) The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. Mol. Biol. Evol. 32(3): 661–673.

Hammer MF; Redd AJ; Wood ET; Bonner MR; Jarjanazi H; Karafet T et al. (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc. Natl. Acad. Sci. USA 97(12): 6769–6774.

Hammer MF; Karafet TM; Redd AJ; Jarjanazi H; Santachiara-Benerecetti S; Soodyall H et al. (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol. Biol. Evol. 18(7): 1189–1203.

Hammer MF; Behar DM; Karafet TM; Mendez FL; Hallmark B; Erez T et al. (2009) Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. Hum. Genet. 126(5): 707–717.

Harwood J (2006) To the Ends of the Earth: 100 Maps that Changed the World. London, Marshall editions.

Hedrick PW (1985) Genetics of Populations. Boston, Jones and Barlett Publishers, Inc.

Henn BM; Gravel S; Moreno-Estrada A; Acevedo-Acevedo S and Bustamante CD (2010) Fine-scale population structure and the era of next-generation sequencing. Hum. Mol. Genet. 19(R2): R221–R226.

Hering S; Augustin C; Edelmann J; Heidel M; Dressler J; Rodig H et al. (2006) DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases. Int. J. Legal Med. 120(6): 337–345.

Hirschberg HZ (1974) A History of the Jews in North Africa. Leiden, Brill.

Holland MM and Parsons TJ (1999) Mitochondrial DNA Sequence Analysis—Validation and Use for Forensic Casework. Forensic Sci. Rev. 11(1): 21–50.

Horváth G; Zalán A; Kis Z and Pamjav H (2012) A genetic study of 12 X-STR loci in the Hungarian population. Forensic Sci. Int. Genet. 6(1): e46–e47.

Houck M (2015) Forensic Biology. London, Academic Press.

Hughes JF and Rozen S (2012) Genomics and genetics of human and primate Y chromosomes. Annual Rev. Genom. Hum. Genet. 13: 83–108.

Hughes JF and Page DC (2015) The Biology and Evolution of Mammalian Y Chromosomes. Ann. Rev. Genet. 49: 507–527.

Huynh KD and Lee JT (2005) X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. Nature Rev. Gene. 6(5): 410–418.

Ibarra A; Restrepo T; Rojas W; Castillo A; Amorim A; Martínez B et al. (2014) Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms. PLoS One 9(1): e87202.

Ingman M; Kaessmann H; Pääbo S and Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408(6813): 708–713.

Iniesta A; Martínez A; Ponce J (Eds.) (2009) Lorca. Luces de Sepharad. Museo arqueológico de Murcia. Murcia, Tres Fronteras.

Inturri S; Menegon S; Amoroso A; Torre C and Robino C (2011) Linkage and linkage disequilibrium analysis of X-STRs in Italian families. Forensic Sci. Int. Genet. 5(2): 152–154.

ISOGG (International Society of Genetic Genealogy) (2017) Y-DNA Haplogroup Tree 2017, Version: 12.69, Date: 16 March 2017. http://www.isogg.org/tree/ Date of access: 17, March, 2017.

Jobling MA (2012) The impact of recent events on human genetic diversity. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367(1590): 793–799.

Jobling MA and Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4(8): 598–612.

Jobling M; Hollox E; Hurles M; Kisivild T and Tyler-Smith C (2014) Human evolutionary genetics. New York, Garland Science, Taylor and Francis Group.

Johnson NA and Lachance J (2012) The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. Ann. NY Acad. Sci. 1256: E1–E22.

Jones DA (1972) Blood samples: probability of discrimination. J. Forensic Sci. Soc. 12(2): 355–359.

Jorde LB (1992) Population diseases in the Ashkenazi population: evolutionary considerations. In: Bonné-Tamir B and Adam A (Eds.) Genetic Diversity Among Jews: diseases and markers at the DNA level. New York, Oxford University Press.

Jurka J (2004) Evolutionary impact of human Alu repetitive elements. Curr. Opin. Genet. Dev. 14(6): 603–608.

Karachanak S; Carossa V; Nesheva D; Olivieri A; Pala M; Kashani BH et al. (2012) Bulgarians vs the other European populations: a mitochondrial DNA perspective. Int. J. Legal Med. 126(4): 497–503.

Karafet TM; Mendez FL; Meilerman MB; Underhill PA; Zegura SL and Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. 18(5): 830–838.

Karlin S; Kenett R and Bonné-Tamir B (1979) Analysis of biochemical genetic data on Jewish populations II. Results and interpretations of heterogeneity indices and distance measures with respect to standards. Am. J. Hum. Genet. 31(3): 341–365.

Kayser M; Lao O; Anslinger K; Augustin C; Bargel G; Edelmann J et al. (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum. Genet. 117(5): 428-443.

Kayser M and de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat. Rev. Genet. 12(3): 179–192.

Kayser M and Ballantyne KN (2014) Y chromosome in forensic science. In: Primorac D and Schanfield M (Eds.) Forensic DNA applications: an interdisciplinary perspective. New York, CRC Press, Taylor and Francis Group.

Kearse M; Moir R; Wilson A; Stones-Havas S; Cheung M; Sturrock S et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647–1649.

Khan F; Pandey AK; Borkar M; Tripathi M; Talwar S; Bisen PS et al. (2008) Effect of sociocultural cleavage on genetic differentiation: a study from North India. Hum. Biol. 80(3): 271–286.

Kidd KK; Pakstis AJ; Speed C and Kidd JR (2004) Understanding human DNA sequence variation. J. Heredity 95(5): 406–420.

King TE and Jobling MA (2009) What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends Genet. 25(8): 351–360.

Kishida T; Wang W; Fukuda M and Tamaki Y (1997) Duplex PCR of the Y-27H39 and HPRT loci with reference to Japanese population data on the HPRT locus. Nihon Hoigaku Zasshi 51(2): 67–69.

Kivisild T (2015) Maternal ancestry and population history from whole mitochondrial genomes. Investig. Genet. 6: 3.

Kloss-Brandstätter A; Pacher D; Schönherr S; Weissensteiner H; Binna R; Specht G et al. (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. 32(1): 25–32.

Kobyliansky E; Micle S; Goldschmidt-Nathan M; Arensburg B and Nathan H (1982) Jewish populations of the world: genetic likeness and differences. Ann. Hum. Biol. 9(1): 1–34.

Kopelman NM; Stone L; Wang C; Gefel D; Feldman MW; Hillel J et al. (2009) Genomic microsatellites identify Shared Jewish ancestry intermediate between Middle Eastern and European populations. BMC Genet. 10: 80.

Krüger J; Fuhrmann W; Lichte KH and Steffens C (1968) Zur Verwendung des Polymorphismus der sauren Erythrocytenphosphatase bei der Vaterschaftsbegutachtung. Deutsche Zeitschrift für die gesamte gerichtliche Medizin 64(2): 127–146.

Kundu S and Ghosh SK (2015) Trend of different molecular markers in the last decades for studying human migrations. Gene 556(2): 81–90.

Labuda D; Lefebvre JF; Nadeau P and Roy-Gagnon MH (2010) Female-to-male breeding ratio in modern humans—an analysis based on historical recombinations. Am. J. Hum. Genet. 86(3): 353–363.

Lahn BT and Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286(5441): 964–967.

Lahn BT; Pearson NM and Jegalian K (2001) The human Y chromosome, in the light of evolution. Nature Rev. Genet. 2: 207–216.

Larruga JM; Díez F; Pinto FM; Flores C and González AM (2001) Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain. Eur. J. Hum. Genet. 9(9): 708–716.

Lashgary Z; Khodadadi A; Singh Y; Houshmand SM; Mahjoubi F; Sharma P et al. (2011) Y chromosome diversity among the Iranian religious groups: A reservoir of genetic variation. Ann. Hum. Biol. 38(3): 364–371.

Laub E and Laub JF (1987) El mito triunfante: Estudio Antropologico-social de los Chuetas mallorquines. Palma, M. Font ed.

Lee EY; Shin KJ; Rakha A; Sim JE; Park MJ; Kim NY et al. (2014) Analysis of 22 Y chromosomal STR haplotypes and Y haplogroup distribution in Pathans of Pakistan. Forensic Sci. Int. Genet. 11: 111–116.

Levy H (1999) Comprehensive History of the Jews of Iran. Costa Mesa, Mazda Publishers.

Levy-Coffman E (2005) A mosaic of people: the Jewish story and a reassessment of the DNA evidence. J. Genet. Geneal. 1: 12–33.

Li C; Zhao S; Zhang S; Li L; Liu Y; Chen J et al. (2011) Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. Forensic Sci. Int. Genet. 5(1): e27–e30.

Librado P and Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11): 1451–1452.

Lippold S; Xu H; Ko A; Li M; Renaud G; Butthof A et al. (2014) Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Invest. Genet. 5: 13.

Listman JB; Hasin D; Kranzler HR; Malison RT; Mutirangura A; Sughondhabirom A et al. (2010) Identification of population substructure among Jews using STR markers and dependence on reference populations included. BMC Genet. 11: 48.

Liu N; Chen L; Wang S; Oh C and Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genet. 6(S1): S26.

Liu J; Wang J; Zhang X; Li Z; Yun K; Liu Z et al. (2017) A mixture detection method based on separate amplification using primer specific alleles of INDELs-a study based on two person's DNA mixture. J. Forensic Legal Med. 46: 30–36.

Livshits G; Sokal RR and Kobyliansky E (1991) Genetic affinities of Jewish populations. Am. J. Hum. Genet. 49(1): 131–146.

Llompart G (2011) Los Judíos Mallorquines y la Cartografía Medieval. In: Durán R (Ed.) Mallorca Judaica, Cuadernos de Historia, No. 5 (pp. 113-131). Palma, Asociación Amigos del Castillo de San Carlos.

López-Escribano H; Parera M; Guix P; Serra J; Gutierrez A; Balsells D et al. (2013) Balearic archipelago: three islands, three beta-thalassemia population patterns. Clin. Genet. 83(2): 175–180.

Lovell A; Moreau C; Yotova V; Xiao F; Bourgeois S; Gehl D et al. (2005) Ethiopia: between Sub-Saharan Africa and Western Eurasia. Ann. Hum. Genet. 69(3): 275–287.

Lucotte G and David F (1992) Y-chromosome-specific haplotypes of Jews detected by probes 49f and 49a. Hum. Biol. 64(5): 757–761.

Lucotte G; Smets P and Ruffié J (1993) Y-chromosome-specific haplotype diversity in Ashkenazic and Sephardic Jews. Hum. Biol. 65(5): 835–840.

Lucotte G and Smets P (1999) Origins of Falasha Jews studied by haplotypes of the Y chromosome. Hum. Biol. 71(6): 989–993.

Luis JR; Rowold DJ; Regueiro M; Caeiro B; Cinnioğlu C; Roseman C et al. (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am. J. Hum. Genet. 74(3): 532–544.

Maca-Meyer N; González AM; Larruga JM; Flores C and Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. 2: 13.

Maca-Meyer N; Sánchez-Velasco P; Flores C; Larruga JM; González AM; Oterino A et al. (2003) Y-chromosome and mitocondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). Ann. Hum. Genet. 67(4): 329–339.

Macaulay V; Hill C; Achilli A; Rengo C; Clarke D; Meehan W et al. (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308(5724): 1034–1036.

Mairal Q; Santos C; Silva M; Marques SL; Ramos A; Aluja MP et al. (2013) Linguistic isolates in Portugal: insights from the mitochondrial DNA pattern. Forensic Sci. Int. Genet. 7(6): 618–623.

Malyarchuk BA; Rogozin IB; Berikov VB and Derenko MV (2002) Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. Hum. Genet. 111(1): 46–53.

Malyarchuk B; Derenko M; Denisova G; Maksimov A; Wozniak M; Grzybowski T et al. (2011) Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. J. Hum. Genet. 56(8): 583–588.

Manta F; Caiafa A; Pereira R; Silva D; Amorim A; Carvalho EF et al. (2012) Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. Forensic Sci. Int. Genet. 6(5): 658–661.

Margulis L (1981) Symbiosis in Cell evolution. New York, Freeman.

Marques SL; Goios A; Rocha AM; Prata MJ; Amorim A; Gusmão L et al. (2015) Portuguese mitochondrial DNA genetic diversity—An update and a phylogenetic revision. Forensic Sci. Int. Genet. 15: 27–32.

Marques SL; Gusmão L; Amorim A; Prata MJ and Álvarez L (2016) Y chromosome diversity in a linguistic isolate (Mirandese, NE Portugal). Am. J. Hum. Biol. 28(5): 671–680.

Marrero P; Abu-Amero KK; Larruga JM and Cabrera VM (2016) Carriers of human mitochondrial DNA macrohaplogroup M colonized India from southeastern Asia. BMC Evol. Biol. 16(1): 246.

Martínez-Cadenas C; Blanco-Verea A; Hernando B; Busby GB; Brion M; Carracedo A et al. (2016) The relationship between surname frequency and Y chromosome variation in Spain. Eur. J. Hum. Genet. 24(1): 120–128.

Martins JC (2006) Portugal e os judeus: Judaísmo e anti-semitismo no século XX. Lisbon, Vega.

Matas M; Guix P; Castro JA; Parera M; Ramon MM; Obrador A et al. (2006) Prevalence of HFE C282Y and H63D in Jewish populations and clinical implications of H63D homozygosity. Clinical Genet. 69(2): 155–162.

McEvoy B and Bradley DG (2006) Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. Hum. Genet. 119(1-2): 212–219.

Mea E (2007) Problemática do Judaísmo. In: Séculos XVI-XVII: Congresso Internacional Inquisição Portuguesa. Tempo, Razão e Circunstância. Lisboa, S Paulo.

Meier RJ (2010) The nature of human biological and genetic variability. Physical (Biological) Anthropology. Bloomington, EOLSS.

Meier RJ and Raff JA (2010) Genetics in Human Biology. In: Muehlenbein MP (Ed.) Human evolutionary biology. Cambridge, Cambridge University Press.

Mendez FL; Krahn T; Schrack B; Krahn AM; Veeramah KR; Woerner AE et al. (2013) An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. Am. J. Hum. Genet. 92(3): 454–459.

Mendizabal I; Valente C; Gusmão A; Alves C; Gomes V; Goios A et al. (2011) Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective. PloS One 6(1): e15988.

Messina F; Finocchio A; Rolfo MF; De Angelis F; Rapone C; Coletta M et al. (2015) Traces of forgotten historical events in mountain communities in Central Italy: A genetic insight. Am. J. Hum. Biol. 27(4): 508–519.

Mills RE; Luttig CT; Larkins CE; Beauchamp A; Tsui C; Pittard WS et al. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 16(9): 1182–1190.

Mills RE; Pittard WS; Mullaney JM; Farooq U; Creasy TH; Mahurkar AA et al. (2011) Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res. 21(6): 830–839.

Moore K (1976) Los de la calle. Un estudio sobre los Chuetas. Madrid, Siglo Veintiuno.

Mullaney JM; Mills RE; Pittard WS and Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. Hum. Mol. Genet. 19(R2): R131–R136

Muntaner L (1986) La inquisició espanyola a Mallorca, un model d'activitat peculiar? In: Muntaner L and Colom M (Eds.) La inquisició a les Illes Balears. Segles XV al XIX. Palma, Conselleria d'Educació i Cultura de les Illes Balears.

Myres NM; Ekins JE; Lin AA; Cavalli-Sforza LL; Woodward SR and Underhill PA (2007) Y-chromosome short tandem repeat DYS458.2 non-consensus alleles occur independently in both binary haplogroups J1-M267 and R1b3-M405. Croat. Med. J. 48(4): 450–459.

Myres NM; Rootsi S; Lin AA; Järve M; King RJ; Kutuev I et al. (2011) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Hum. Genet. 19(1): 95–101.

Nachman MW and Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156(1): 297–304.

Nebel A; Filon D; Brinkmann B; Majumder PP; Faerman M and Oppenheim A (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. Am. J. Hum. Genet. 69(5): 1095–1112.

Nebel A; Filon D; Faerman M; Soodyall H and Oppenheim A (2005) Y chromosome evidence for a founder effect in Ashkenazi Jews. Eur. J. Hum. Genet. 13(3): 388–391.

Nei M and Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. Genetics 97(1): 145–163.

Nei M (1987) Molecular Evolutionary Genetics. New York, Columbia University Press.

Nevo S; Picornell A; Miguel A; Castro JA; Joel A; Heno N et al. (1996) Orosomucoid (ORM1) polymorphism in Arabs and Jews of Israel: more evidence for a Middle Eastern origin of the Jews. Hum. Biol. 68(2): 217–229.

Nogueiro I; Manco L; Gomes V; Amorim A and Gusmão L (2010) Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. Am. J. Phys. Anthropol. 141(3): 373–381.

Nogueiro I; Teixeira J; Amorim A; Gusmão L and Álvarez L (2015a) Echoes from Sepharad: signatures on the maternal gene pool of crypto-Jewish descendants. Eur. J. Hum. Genet. 23(5): 693–699.

Nogueiro I; Teixeira JC; Amorim A; Gusmão L and Álvarez L (2015b) Portuguese crypto-Jews: the genetic heritage of a complex history. Front. Genet. 6: 12.

Nogueiro MIP (2015) Tracing Sephardic Jewry Through Genetics: Crypto-Jews and the Second Diaspora. PhD Thesis. University of Porto, Portugal.

Non AL; Al-Meeri A; Raaum RL; Sanchez LF and Mulligan CJ (2011) Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia. Am. J. Phys. Anthropol. 144(1): 1–10.

Nothnagel M; Schmidtke J and Krawczak M (2010) Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. Int. J. Legal Med. 124(3): 205–215.

Oefner PJ; Holzi G; Shen P; Shpirer I; Gefel D; Lavi T et al. (2013) Genetics and the history of the Samaritans: Y-chromosomal microsatellites and genetic affinity between Samaritans and Cohanim. Hum. Biol. 85(6): 825–858.

Olaisen B; Bär W; Brinkmann B; Budowle B; Carracedo A; Gill P et al. (1998) DNA recommendations 1997 of the International Society for Forensic Genetics. Vox Sang. 74(1): 61–63.

Oliveira EJ; Pádua JG; Zucchi MI; Vencovsky R and Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. Genet. Mol. Biol. 29(2): 294–307.

Olivieri A; Achilli A; Pala M; Battaglia V; Fornarino S; Al-Zahery N et al. (2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314(5806): 1767–1770.

Olofsson JK; Mogensen HS; Buchard A; Børsting C and Morling N (2015) Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit. Forensic Sci. Int. Genet. 16: 232–236.

Ostrer H (2001) A genetic profile of contemporary Jewish populations. Nat. Rev. Genet. 2(11): 891–898.

Ostrer H (2011) Changing the game with whole exome sequencing. Clinical Genet. 80(2): 101–103.

Ostrer H and Skorecki K (2013) The population genetics of the Jewish people. Hum. Genet. 132(2): 119–127.

Pakendorf B and Stoneking M (2005) Mitochondrial DNA and human evolution. Annu. Rev. Genom. Hum. Genet. 6: 165–83.

Pala M; Olivieri A; Achilli A; Accetturo M; Metspalu E; Reidla M et al. (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am. J. Hum. Genet. 90(5): 915–924.

Pardiñas AF; Roca A; García-Vázquez E and López B (2014) Evaluation of large-scale genetic structure in complex demographic and historical scenarios: The mitochondrial DNA and Y-chromosome pools of the Iberian Atlantic façade. Am. J. Phys. Anthropol. 153(4): 617–626.

Parson W and Bandelt H (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Sci. Int. Genet. 1(1): 13–19.

Parson W and Dür A (2007) EMPOP—a forensic mtDNA database. Forensic Sci. Int. Genet. 1(2): 88–92.

Patai R and Patai J (1975) The Myth of the Jewish Race. New York, C. Scribner's Sons.

Perdigó (1946) Anonymous. La Inquisición de Mallorca. Reconciliados y Relajados, 1488–1691. Barcelona, Perdigó.

Pereira R; Phillips C; Alves C; Amorim A; Carracedo A and Gusmão L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30(21): 3682–3690.

Pereira V; Moncada E; Diez IE; Tomas C; Amorim A; Morling N et al. (2011) Genetic characterization of Somali and Iraqi populations using a set of 33 X-chromosome Indels. Forensic Sci. Int. Genet. Sup. Ser. 3(1): e137–e138.

Pereira R and Gusmão L (2012) Capillary electrophoresis of 38 noncoding biallelic mini-indels for degraded samples and as complementary tool in paternity testing. DNA Methods Mol. Biol. 830: 141–157.

Pereira R; Pereira V; Gomes I; Tomas C; Morling N; Amorim A et al. (2012a) A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. Int. J. Legal Med. 126(1): 97–105.

Pereira R; Phillips C; Pinto N; Santos C; dos Santos SE; Amorim A et al. (2012b) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One 7(1): e29684.

Pereira V; Gusmão L; Valente C; Pereira R; Carneiro J; Gomes I et al. (2012c) Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. Am. J. Phys. Anthropol. 148(3): 389–394.

Pereira V; Tomàs C; Sánchez JJ; Syndercombe-Court D; Amorim A; Gusmão L et al. (2015) The peopling of Greenland: further insights from the analysis of genetic diversity using autosomal and X-chromosomal markers. Eur. J. Hum. Genet. 23(2): 245–251.

Pérez E (Ed.) (2005) Fonts per a l'estudi de la comunitat jueva de Mallorca. Catalonia Hebraica VI. Barcelona, Editorial PPU.

Pérez J (2006) The Spanish Inquisition: A history. Encyclopedia Judaica, Jewish Encyclopedia. Yale, University Press.

Phillips C (2015) Forensic genetic analysis of bio-geographical ancestry. Forensic Sci. Int. Genet. 18: 49–65.

Piazza A; Matullo G; Romano V; Ayala GF; Bonanno CT; Calì F et al. (2000) Towards a genetic history of Sicily. Journal of Cultura Heritage 1(S2): S39–S42.

Picornell A; Miguel A; Castro JA and Ramon MM (1990) Enzymatic polymorphisms in the Jewish community (Chuetas) from the Majorca Island. Gene Geogr. 4(3): 165–171.

Picornell A; Castro JA and Ramon MM (1991) Blood groups in the Chueta community (Majorcan Jews). Hum. Hered. 41(1): 35–42.

Picornell A (1992) Caracterització genètica de la població xueta mitjançant polimorfismes hemàtics. PhD thesis. Palma, Universitat de les Illes Balears.

Picornell A; Castro JA and Ramon MM (1992) PI and TF subtypes in Chuetas (Majorcan Jews). Hum. Hered. 42(5): 321–323.

Picornell A; Castro JA and Ramon MM (1994) Serum protein polymorphism in Chuetas (Majorcan Jews) —GC, A2HS, ORM, ITI and HP. Gene Geogr. 8(2): 137–145.

Picornell A; Castro JA and Ramon MM (1997) Genetics of the Chuetas (Majorcan Jews): a comparative study. Hum. Biol. 69(3): 313–328.

Picornell A; Tomàs C; Jiménez G; Castro JA and Ramon MM (2002) Jewish population genetic data in 20 polymorphic loci. Forensic Sci. Int. 125(1): 52–58.

Picornell A; Jiménez G; Castro JA; Ramon MM (2004) Minimal Y-chromosome haplotypes plus DYS287 in Jewish populations. J. Forensic Sci. 49(2): 410–412.

Picornell A; Gómez-Barbeito L; Tomàs C; Castro JA and Ramon MM (2005) Mitochondrial DNA HVRI variation in Balearic populations. Am. J. Phys. Anthropol. 128(1): 119–130.

Picornell A; Gimenez P; Castro JA and Ramon MM (2006) Mitochondrial DNA sequence variation in Jewish populations. Int. J. Legal Med. 120(5): 271–281.

Pimenta JR and Pena SDJ (2010) Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. Genet. Mol. Res. 9(1): 601–607.

Pinto N; Magalhaes M; Conde-Sousa E; Gomes C; Pereira R; Alves C et al. (2013) Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. Forensic Sci. Int. Genet. 7(1): 16–21.

Pinto JC; Pereira V; Marques SL; Amorim A; Alvarez L and Prata MJ (2015) Mirandese language and genetic differentiation in Iberia: a study using X chromosome markers. Ann. Hum. Biol. 42(1): 20–25.

Planas R (2003) Els malnoms dels xuetes de Mallorca. Palma, Lleonard Muntaner ed.

Pons A (1984) Los Judíos del Reino de Mallorca durante los siglos XIII y XIV. Palma de Mallorca, M. Font ed.

Pons J (1988) Assegurances i canvis marítims a Mallorca: les companyies (1660-1680). Estudis d'Història Econòmica 2: 43–68.

Porqueres E (2001) L'endogàmia dels xuetes de Malloca. Identitat i matrimoni en una comunitat de conversos (1435-1750). Palma, Lleonard Muntaner ed.

Porqueres E (2011) Redefinición y secretismo en la comunidad xueta de Mallorca. In: Durán R (Ed.) Mallorca Judaica, Cuadernos de Historia, No. 5 (pp. 15–64). Palma, Asociación Amigos del Castillo de San Carlos.

Poulsen L; Farzad MS; Borsting C; Tomas C; Pereira V and Morling N (2015) Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris. Forensic Sci. Int. Genet. 17: 43–46.

Poznik GD; Xue Y; Mendez FL; Willems TF; Massaia A; Sayres MAW et al. (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nature Genet. 48(6): 593–599.

Pritchard JK; Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2): 945–959.

Pugach I and Stoneking M (2015) Genome-wide insights into the genetic history of human populations. Invest. Genet. 6: 6.

Purps J; Geppert M; Nagy M and Roewer L (2015) Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases. Forensic Sci. Int. Genet. 19: 238–242.

Qiagen (2010) Investigator Argus 12 X-STR Kit handbook. Spain, Madrid SA.

Qian Y; Kehr B and Halldórsson BV (2015) PopAlu: population-scale detection of Alu polymorphisms. PeerJ 3: e1269.

Qualitype GMbH (2014) Genoproof3Theory Manual. Dresden, Edition Qualitype GMbH.

Ramos A; Santos C; Álvarez L; Álvarez L; Nogués R and Aluja MP (2009) Human mitochondrial DNA complete amplification and sequencing: a new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. Electrophoresis 30(9): 1587–1593.

Ramos A; Santos C; Barbena E; Mateiu L; Álvarez L; Nogués R et al. (2011) Validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification: a revision based on the New Human Genome Reference Sequence (GRCh37). Electrophoresis 32(6-7): 782–783.

Regueiro M; Cadenas AM; Gayden T; Underhill PA and Herrera RJ (2006) Iran: tricontinental nexus for Y-chromosome driven migration. Hum. Hered. 61(3): 132–143.

Rejwan N (1985) The Jews of Iraq: 3000 Years of History and Culture. Boulder, Westview Press.

Relethford JH (2012) Human Population Genetics. Hoboken, NJ, USA, John Wiley & Sons, Inc.

Riera F and Melià J (1973) Les lluites antixuetes del segle XVIII. Palma, Moll.

Riera F and Porqueres E (1996) La causa xueta a la cort de Carles III. Palma, Lleonard Muntaner ed.

Rito T; Richards MB; Fernandes V; Alshamali F; Cerny V; Pereira L et al. (2013) The first modern human dispersals across Africa. PLoS One 8(11): e80031.

Ritte U; Neufeld E; Broit M; Shavit D and Motro U (1993) The differences among Jewish communities—maternal and paternal contributions. J. Mol. Evol. 37(4): 435–440.

Robino C; Inturri S; Gino S; Torre C; Di Gaetano C; Crobu F et al. (2006) Y-chromosomal STR haplotypes in Sicily. Forensic Sci. Int. 159(2–3): 235–240.

Rodríguez V; Tomàs C; Sánchez JJ; Castro JA; Ramon MM; Barbaro A et al. (2009) Genetic sub-structure in western Mediterranean populations revealed by 12 Y-chromosome STR loci. Int. J. Legal Med. 123(2): 137–141.

Roewer L (2009) Y chromosome STR typing in crime casework. Forensic Sci. Med. Pathol. 5(2): 77–84.

Roewer L; Nothnagel M; Gusmão L; Gomes V; González M; Corach D et al. (2013) Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. PLoS Genet. 9(4): e1003460.

Rojas MY; Alonso LA; Sarmiento VA; Eljach LY and Usaquén W (2013) Structure analysis of the La Guajira-Colombia population: A genetic, demographic and genealogical overview. Ann. Hum. Biol. 40(2): 119–131.

Román-Busto J; Fuster V; Colantonio S; Zuluaga P; Blanco MJ and Guardado-Moreira MJ (2010) Mate choice in Olivenza: influence of border change on Spanish–Portuguese lineages. J. Biosoc. Sci. 42(01): 129–140.

Romanini C; Catelli ML; Borosky A; Pereira R; Romero M; Puerto MS et al. (2012) Typing short amplicon binary polymorphisms: Supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. Forensic Sci. Int. Genet. 6(4): 469–476.

Roostalu U; Kutuev I; Loogväli EL; Metspalu E; Tambets K; Reidla M et al. (2007) Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. Mol. Biol. Evol. 24(2): 436–448.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M et al. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am. J. Hum. Genet. 75(1): 128–137.

Rootsi S; Behar DM; Järve M; Lin AA; Myres NM; Passarelli B et al. (2013) Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. Nat. Commun. 4: 2928.

Rosenberg NA; Woolf E; Pritchard JK; Schaap T; Gefel D; Shpirer I et al. (2001) Distinctive genetic signatures in the Libyan Jews. Proc. Natl. Acad. Sci. USA 98(3): 858–863.

Rosner G; Rosner S and Orr-Urtreger A (2009) Genetic testing in Israel: an overview. Ann. Rev. Genomics Hum. Genet. 10: 175–192.

Ross MT; Grafham DV; Coffey AJ; Scherer S; McLay K; Muzny D et al. (2005) The DNA sequence of the human X chromosome. Nature 434(7031): 325–337.

Roth C and Novella J (1979) Los judíos secretos: historia de los marranos. Madrid, Altalena.

Rozenberg D (2010) La España contemporánea y la cuestión judía: retejiendo los hilos de la memoria y de la historia. Madrid, Marcial Pons Historia.

Rubicz R and Crawford M (2007) Molecular Markers. In: Crawford MH (Ed.) Anthropological Genetic Studies in Anthropological genetics: theory, methods and applications. Cambridge, Cambridge University Press.

Saillard J; Forster P; Lynnerup N; Bandelt HJ and Nørby S (2000) MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. J. Hum. Genet. 67(3): 718–726.

Saiz M; Álvarez-Cubero MJ; Martínez-González LJ; Álvarez JC and Lorente JA (2014) Population genetic data of 38 insertion-deletion markers in South East Spanish population. Forensic Sci. Int. Genet. 13: 236–238.

Salas A; Richards M; De la Fe T; Lareu MV; Sobrino B; Sánchez-Diz P et al. (2002) The making of the African mtDNA landscape. Am. J. Hum. Genet. 71(5): 1082–1111.

Salem AH; Bahri R; Jarjanazi H and Chaabani H (2014) Geographical and social influences on genetic diversity within the Egyptian population: analyses of Alu insertion polymorphisms. Annals Hum. Biol. 41(1): 61–66.

Sánchez-Velasco P; Escribano de Diego J; Paz-Miguel JE; Ocejo-Vinyals G and Leyva-Cobián F (1999) HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. Tissue Antigens 53(1): 65–73.

Sanchez-Velasco P; Gomez-Casado E; Martínez-Laso J; Moscoso J; Zamora J; Lowy E et al. (2003) HLA alleles in isolated populations from North Spain: origin of the Basques and the ancient Iberians. Tissue Antigens 61(5): 384–392.

Santaquiara Benerecetti AS; Semino O; Passarino G; Torroni A; Brdicka R; Fellous M et al. (1993). The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. Ann. Hum. Genet. 57(1): 55–64.

Santos C; Fregel R; Cabrera VM; Gonzalez AM; Larruga JM and Lima M (2010) Mitochondrial DNA patterns in the Macaronesia islands: variation within and among archipelagos. Am. J. Phys. Anthropol. 141(4): 610–619.

Santos C; Phillips C; Oldoni F; Amigo J; Fondevila M; Pereira R et al. (2015) Completion of a worldwide reference panel of samples for an ancestry informative Indel assay. Forensic Sci. Int. Genet. 17: 75–80.

Sarno S; Tofanelli S; De Fanti S; Quagliariello A; Bortolini E; Ferri G et al. (2016). Shared language, diverging genetic histories: high-resolution analysis of Y-chromosome variability in Calabrian and Sicilian Arbereshe. Eur. J. Hum. Genet. 24(4): 600–606.

Saunier JL; Irwin JA; Strouss KM; Ragab H; Sturk KA and Parsons TJ (2009) Mitochondrial control region sequences from an Egyptian population sample. Forensic Sci. Int. Genet. 3(3): e97–e103.

Schaffner SF (2004) The X chromosome in population genetics. Nat. Rev. Genet. 5(1): 43–51.

Schiffels S and Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. Nature Genet. 46(8): 919–925.

Schwartz M and Vissing J (2002) Paternal inheritance of mitochondrial DNA. N. Engl. J. Med. 347(8): 576–580.

Scozzari R; Cruciani F; Pangrazio A; Santolamazza P; Vona G; Moral P et al. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. Hum. Immunol. 62(9): 871–884.

Semino O; Passarino G; Oefner PJ; Lin AA; Arbuzova S; Beckman LE et al. (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. Science 290(5494): 1155–1159.

Semino O; Magri C; Benuzzi G; Lin AA; Al-Zahery N; Battaglia V et al. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am. J. Hum. Genet. 74(5): 1023–1034.

Shanks H (1988) Ancient Israel: A Short History from Abraham to the Roman Destruction of the Temple. Engelwood Cliffs, New Jersey, Prentice-Hall Biblical Archaeological Society.

Shapiro S (1984) Jews in Old China, Studies by Chinese Scholars. Nueva York, Hippocrene Books.

Shen P; Lavi T; Kivisild T; Chou V; Sengun D; Gefel D et al. (2004) Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum. Mutat. 24(3): 248–260.

Sherry ST; Ward MH; Kholodov M; Baker J; Phan L; Smigielski EM et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29(1): 308–311.

Shlush LI; Behar DM; Yudkovsky G; Templeton A; Hadid Y; Basis F et al. (2008) The Druze: a population genetic refugium of the Near East. PLoS One 3: e2105.

Skare Ø; Sheehan N and Egeland T (2009) Identification of distant family relationships. Bioinformatics 25(18): 2376–2382.

Skorecki K; Sellg S; Blazer S; Bradman R; Bradman N; Waburton PJ et al. (1997) Y-Chromosome of Jewish Priests. Nature 385(6611): 32.

Soares P; Ermini L; Thomson N; Mormina M; Rito T; Röhl A et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet. 84(6): 740–759.

Soares P; Achilli A; Semino O; Davies W; Macaulay V; Bandelt HJ et al. (2010) The archaeogenetics of Europe. Current Biol. 20(4): R174-R180.

Soares P; Alshamali F; Pereira JB; Fernandes V; Silva NM; Afonso C et al. (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. Mol. Biol. Evol. 29(3): 915–927.

Solé-Morata N; Bertranpetit J; Comas D and Calafell F (2015) Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. Eur. J. Hum. Genet. 23(11): 1549–1557.

Song WH; Yi YJ; Sutovsky M; Meyers S and Sutovsky P (2016) Autophagy and ubiquitin-proteasome system contribute to sperm mitophagy after mammalian fertilization. Proc. Natl. Acad. Sci. USA. 113(36): E5261–E5270.

Soodyall H (2013) Lemba origins revisited: tracing the ancestry of Y chromosomes in South African and Zimbabwean Lemba. SAMJ: South African Medical Journal 103(12): 1009–1013.

Soundararajan U; Yun L; Shi M and Kidd KK (2016) Minimal SNP overlap among multiple panels of ancestry informative markers argues for more international collaboration. Forensic Sci. Int. Genet. 23: 25–32.

Spurdle AB and Jenkins T (1996) The origins of the Lemba" Black Jews" of southern Africa: evidence from p12F2 and other Y-chromosome markers. Am. J. Hum. Genet. 59(5): 1126–1133.

Steiper ME (2010) DNA Markers of Human Variation. In: Muehlenbein MP (Ed.) Human evolutionary biology. Cambridge, Cambridge University Press.

Stewart JB and Chinnery PF (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat. Rev. Genet. 16(9): 530–542.

Stillman NA (1991) The Jews of Arab Lands in modern times. Philadelphia, Jewish Publication Society.

Strachan T and Read A (2006) Genética Humana. México, McGraw-Hill Interamericana.

Sturrock K and Rocha J (2000) A multidimensional scaling stress evaluation table. Field methods 12(1): 49–60.

Sumita DR and Whittle MR (2009) Updated allelic structures of the DXS10135 and DXS10078 STR loci. Forensic Sci. Int. Genet. Sup. Ser. 2(1): 51–52.

Sun JX; Mullikin JC; Patterson N and Reich DE (2009) Microsatellites are molecular clocks that support accurate inferences about history. Mol. Biol. Evol. 26(5): 1017–1027.

Sun S and Heitman J (2012) Should Y stay or should Y go? The evolution of non-recombining sex chromosomes. BioEssays 34(11): 938–942.

Sutovsky P; Moreno RD; Ramalho-Santos J; Dominko T; Simerly C and Schatten G (1999) Ubiquitin tag for sperm mitochondria. Nature 402(6760): 371–372.

Sutton WK; Knight A; Underhill PA; Neulander JS; Disotell TR and Mountain JL (2006) Toward resolution of the debate regarding purported crypto-Jews in a Spanish-American population: evidence from the Y chromosome. Ann. Hum. Biol. 33(1): 100–111.

Szibor R; Krawczak M; Hering S; Edelmann J; Kuhlisch E and Krause M (2003) Use of X-linked markers for forensic purposes. Int. J. Legal Med. 117(2): 67–74.

Szibor R (2007) X-chromosomal markers: past, present and future. Forensic Sci. Int. Genet. 1(2): 93–99.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3): 585–595.

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135(2): 599–607.

Takezaki N; Nei M and Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol. Biol. Evol. 27(4): 747–752.

Tamang R; Singh L and Thangaraj K (2012) Complex genetic origin of Indian populations and its implications. J. Biosci. 37(5): 911–919.

Tamura K and Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10(3): 512–526.

Teixeira JC; Nogueiro I; Goios A; Gusmão L; Amorim A and Álvarez L (2011) Mitochondrial DNA-control region sequence variation in the NE Portuguese Jewish community. Forensic Sci. Int. Genet. Sup. Series 3(1): e51–e52.

Tereba A (1999) Tools for analysis of population statistics. Profiles in DNA 2(3): 14–16.

Thomas MG; Parfitt T; Weiss DA; Skorecki K; Wilson JF; Le Roux M et al. (2000) Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba—the "Black Jews of Southern Africa". Am. J. Hum. Genet. 66(2): 674–686.

Thomas MG; Weale ME; Jones AL; Richards M; Smith A; Redhead N et al. (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. Am. J. Hum. Genet. 70(6): 1411–1420.

Tian JY; Wang HW; Li YC; Zhang W; Yao YG; van Straten J et al. (2015) A genetic contribution from the Far East into Ashkenazi Jews via the ancient Silk Road. Sci. Rep. 5: 8377.

Tillmar AO (2012). Population genetic analysis of 12 X-STRs in Swedish population. Forensic Sci. Int. Genet. 6(2): e80–e81.

Tofanelli S; Taglioli L; Bertoncini S; Francalacci P; Klyosov A and Pagani L (2014) Mitochondrial and Y chromosome haplotype motifs as diagnostic markers of Jewish ancestry: a reconsideration. Front. Genet. 5: 384.

Tokdemir M and Tunçez FT (2017) Genetic polymorphisms of 17 Y-STR loci in Eastern Turkey population. Gene Rep. 6: 15–18.

Tomàs C; Picornell A; Castro JA; Ramon MM; Gusmão L; Lareu MV et al. (2000) Genetic variability at nine STR loci in the Chueta (Majorcan Jews) and the Balearic populations investigated by a single multiplex reaction. Int. J. Legal Med. 113(5): 263–267.

Tomàs C (2002) Aplicació dels STR a l'estudi de la població humana de les Illes Balears, i a l'anàlisis de lligament de la miopia i el transtron bipolar. PhD thesis. Palma, Universitat de les Illes Balears.

Tomàs C; Jimenez G; Picornell A; Castro JA and Ramon MM (2006) Differential maternal and paternal contributions to the genetic pool of Ibiza Island, Balearic Archipelago. Am. J. Phys. Anthropol. 129(2): 268–278.

Tomàs C; Sánchez JJ; Barbaro A; Brandt-Casadevall C; Hernández A; Ben Dhiab M et al. (2008) X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans). BMC Evol. Biol. 8: 75.

Tomàs C; Pereira V and Morling N (2012) Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. Int. J. Legal Med. 126(1): 121–128.

Trombetta B; D'Atanasio E; Massaia A; Ippoliti M; Coppa A; Candilio F et al. (2015) Phylogeographic refinement and large scale genotyping of human Y chromosome haplogroup E provide new insights into the dispersal of early pastoralists in the African continent. Genome Biol. Evol. 7(7): 1940–1950.

Underhill PA; Shen P; Lin AA; Jin L; Passarino G; Yang W et al. (2000) Y chromosome sequence variation and the history of human populations. Nat. Genet. 26(3): 358–361.

Underhill PA and Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Ann. Rev. Genet. 41: 539–564.

Underhill PA; Myres NM; Rootsi S; Metspalu M; Zhivotovsky LA; King RJ et al. (2010) Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. Eur. J. Hum. Genet. 18(4): 479–484.

Urquhart A; Kimpton CP; Downes TJ and Gill P (1994) Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. Int. J. Legal Med. 107(1): 13–20.

Valverde L; Illescas MJ; Villaescusa P; Gotor AM; García A; Cardoso S et al. (2016) New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. Eur. J. Hum. Genet. 24(3): 437–441.

van Oven M; Geystelen A; Kayser M; Decorte R and Larmuseau MH (2014) Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum. Mutat. 35(2): 187–191.

van Oven M (2015) PhyloTree Build 17: growing the human mitochondrial DNA tree. Forensic Sci. Int. Genet. Sup. Ser. 5: e392–e394.

Velez C; Palamara PF; Guevara-Aguirre J; Hao L; Karafet T; Guevara-Aguirre M et al. (2012) The impact of Converso Jews on the genomes of modern Latin Americans. Hum. Genet. 131(2): 251–263.

Waldman YY; Biddanda A; Dubrovsky M; Campbell CL; Oddoux C; Friedman E et al. (2016) The genetic history of Cochin Jews from India. Hum. Genet. 135(10): 1127–1143.

Weber JL; David D; Heil J; Fan Y; Zhao C and Marth G (2002) Human Diallelic Insertion/Deletion Polymorphisms. Am. J. Hum. Biol. 71(4): 854–862.

Weissensteiner H; Pacher D; Kloss-Brandstätter A; Forer L; Specht G; Bandelt HJ et al. (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 44 (W1): W58–W63.

Willems T; Gymrek M; Poznik GD; Tyler-Smith C; 1000 Genomes Project Chromosome Y Group and Erlich Y (2016) Population-Scale Sequencing Data Enable Precise Estimates of Y-STR Mutation Rates. Am. J. Hum. Genet. 98(5): 919–933.

Wilson Sayres MA and Makova KD (2011) Genome analysis substantiate male mutation bias in many species. Bioessays 33(12): 938–945.

Wood ET; Stover DA; Ehret C; Destro-Bisol G; Spedini G; McLeod H et al. (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur. J. Hum. Genet. 13(7): 867–876.

Wright S (1920) The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. Proc. Natl. Acad. Sci. USA 6(6): 320–332.

Wright S (1969) The theory of gene frequencies. In: Evolution and the Genetics of Populations. (Vol. 2). Chicago, University of Chicago Press.

Xiao FX; Yotova V; Zietkiewicz E; Lovell A; Gehl D; Bourgeois S et al. (2004) Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. Eur. J. Hum. Genet. 12(4): 301–311.

YCC (Y Chromosome Consortium) (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. 12(2): 339–348.

Yang N; Li H; Criswell LA; Gregersen PK; Alarcon-Riquelme ME; Kittles R et al. (2005) Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. Hum. Genet. 118(3–4): 382–392.

Zalloua PA; Platt DE; El-Sibai M; Khalife J; Makhoul N; Haber M et al. (2008) Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. Am. J. Hum. Genet. 83(5): 633–642.

Zarrabeitia MT; Pinheiro F; de Pancorbo MM; Cainé L; Cardoso S; Gusmão L et al. (2009) Analysis of 10 X-linked tetranucleotide markers in mixed and isolated populations. Forensic Sci. Int. Genet. 3(2): 63–66.

Zhivotovsky LA; Ahmed S; Wang W and Bittles AH (2001) The forensic DNA implications of genetic differentiation between endogamous communities. Forensic Sci. Int. 119(3): 269–272.

Zidkova A; Horinek A; Kebrdlova V and Korabecna M (2013) Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. Int. J. Legal Med. 127(1): 7–10.

Zietkiewicz E; Yotova V; Gehl G; Wambach T; Arrieta I; Batzer M et al. (2003) Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am. J. Hum. Genet. 73(5): 994–1015.

Zoossmann-Diskin A; Ticher A; Hakim I; Goldwitch Z; Rubinstein A and Bonné-Tamir B (1991) Genetic affinities of Ethiopian Jews. Isr. J. Med. Sci. 27(5): 245–251.