



UNIVERSITAT DE BARCELONA



**Variació genètica i evolució d'elements *Alu* recents
en poblacions humanes.
Inferències biodemogràfiques i filogeogràfiques**

**Genetic variation and evolution of recent *Alu* elements in human populations.
Biodemographic and philogeographic inferences**

Memòria presentada per

Emili González Pérez

per optar al grau de
Doctor per la Universitat de Barcelona

Dirigida pel Dr. Pedro Moral Castrillo, Professor Titular d'Antropologia Física
de la Unitat d'Antropologia del Departament de Biologia Animal
de la Universitat de Barcelona.

Programa de Doctorat d'*Antropologia Biològica*, bienni 1999-2001.
Departament de Biologia Animal – Facultat de Biologia

Pedro Moral Castrillo
Director

Clara García Moro
Tutora

Emili González Pérez
Doctorand

Discussió i Conclusions

Creative thinking often means not accepting
views that are generally taken for granted.
There are usually good reasons for persistent ideas,
but history shows that they are often not correct.
Always ask "What if it isn't so?"

Kenneth Weiss

La vida es el arte de sacar conclusiones suficientes a partir de datos insuficientes.

Samuel Butler

13 Discussió global dels resultats

13.1 Sobre les relacions poblacionals a la regió mediterrània

El primer apartat d'aquesta discussió global tractarà els resultats del treball que ha ocupat un major espai de temps i de dedicació en l'elaboració de la tesi. Aquest és el cas dels primers tres manuscrits inclosos com a resultats. Es tracta d'estudis que pretenen abordar una caracterització genètica, de manera parcial o global, de les poblacions mediterrànies i procedir a l'anàlisi de les inferències biodemogràfiques i/o històriques obtingudes.

En el primer article s'ha procedit a una primera aproximació poblacional en base a la variació de 14 elements *Alu* polimòrfics per intentar aclarir fenòmens que haurien modulats les relacions de les poblacions autòctones del Mediterrani Occidental. El segon dels articles publicats ha pretès un enfocament parcial del paper de barrera genètica exercit per l'Estret de Gibraltar a partir de la caracterització de la variació genètica en el *locus* CD4, compost per una inserció *Alu* i un microsatèl·lit de tipus pentanucleòtid. En aquest treball es caracteritzen en detall diferents variants haplotípiques que per edat i distribució geogràfica han estat qualificades de típiques del poblament ancestral de la regió, especialment del nord-oest africà.

El darrer treball d'aquest primer bloc de resultats ha englobat un ampli anàlisi d'una vintena de marcadors *Alu* i de tres sistemes haplotípics *Alu*-STR de naturalesa similar al ja esmentat (CD4, DM i FXIIB), en un total de 17 poblacions no insulars i antropològicament ben caracteritzades de la regió geogràfica estudiada. La discussió d'aquest darrer treball és la més àmplia per l'amplitud dels anàlisis proposats i per la gran cobertura de mostres estudiades. Tot i això, la discussió inclourà certes referències als dos articles parcials i que poden resultar interessants per a una comprensió més adequada dels processos de diferenciació i relació entre les poblacions humanes estudiades.

En conjunt, les relacions humanes a la regió mediterrània van ser avaluades a través de l'anàlisi sistemàtic de 1831 individus autòctons pertanyents a 15 poblacions distribuïdes per quatre països de l'Europa mediterrània i tres països nord-africans amb una particular empremta berber, i a dues poblacions que s'utilitzen com a referències externes: una mostra

centreeuropea d'Alemanya i una altra sud-sahariana originària de Costa d'Ivori. L'amplitud de la mostra utilitzada i la seva acurada caracterització antropològica constitueix probablement una de les contribucions més rellevants de la recerca portada a terme. Així mateix, s'ha de destacar la quantitat de marcadors analitzats: 18 elements *Alu* autosòmics i un lligat al cromosoma Y, que representen un total de 19 unitats hereditàries independents, conjuntament amb tres STR's lligats als *loci* CD4, FXIIIIB i DM. Si els polimorfismes *Alu* han demostrat prèviament ser una eina adequada per inferir relacions humanes ancestrals, la seva combinació amb STR's en forma d'haplotips altament informatius aporta un nou enfocament en forma d'informació qualitativa que pot resultar extremadament útil per avaluar en detall certs aspectes de les històries demogràfiques humanes, com l'existència de particularitats de les poblacions mediterrànies i la quantificació de fenòmens de flux gènic en tota la regió. Complementàriament, aquesta aproximació permet estimar l'antiguitat d'haplotips particulars *Alu*-STR que podrien ser característics del poblament antic del Mediterrani.

Un bon exemple de la utilitat conjunta d'aquests marcadors també ha estat recentment establerta en el quart article que forma part del contingut de resultats d'aquesta tesi i que fa referència a la caracterització genètica de les principals illes del Mediterrani Occidental (González-Pérez et al 2007). Les dades obtingudes per a les poblacions continentals mediterrànies complementen adequadament aquest estudi previ i probablement constitueixen un dels estudis més extensius que s'han dut a terme en aquesta particular regió geogràfica. Estudis previs han contribuït abastament (pel que fa a nombre d'individus i poblacions) a l'aproximació genètica dels grups humans mediterranis (Quintana-Murci et al 2003; Esteban et al 2006; Gerard et al 2006; Olivieri et al 2006; Tomas et al 2008), però només en casos particulars es tracta de dades genètiques autosòmiques, mentre que la majoria opta per aproximacions basades en les línies uniparentals de variació lligada al cromosoma Y o al DNA mitocondrial.

En relació a la reavaluació de les **relacions generals entre poblacions del Mediterrani**, l'anàlisi de les distàncies genètiques obtingudes a partir de les tres bases de dades de marcadors analitzats (*Alu*, STR's i haplotips *Alu*/STR) mostren un patró consistent. De fet, les tres matrius de distàncies estan altament correlacionades, fent indicatiu la consistència genètica dels tres conjunts de dades. La representació en forma d'anàlisi de components principals d'aquestes dades mostren la divergència inicial de la població Sub-Sahariana respecte a la resta. En segon lloc, es diferencia la població centreeuropea (l'altra mostra de referència externa) respecte al conjunt de poblacions

mediterrànies que s'agrupen, al seu torn, en dos conjunts diferenciats: Nord-Africans i Sud-Europeus. Ocupant, el primer d'aquests grups, una posició intermèdia entre la mostra sud-sahariana i el grup europeu. Aquesta visió de conjunt pot ser interpretada com una representació força acurada de la seva posició geogràfica relativa, fet que estaria d'acord amb un procés de diferenciació genètica dependent de la distància física en primer lloc, com constata l'elevada correlació ($p < 0.002$) entre distàncies geogràfiques i genètiques, en base al model d'isolament per distància proposat per Hardy (1999).

Les relacions poblacionals obtingudes no descarten un paper evident de l'Estret de Gibraltar com a obstacle geogràfic durant part de la història d'aquests grups humans. Ara bé, aquesta constatació no permet qualificar-lo d'una barrera contra el flux gènic (Comas et al 2000; Bosch et al 2001) que, en major o menor grau, ha succeït tal com evidencien determinades variants genètiques particulars presents en freqüències polimòrfiques a ambdues ribes del Mediterrani occidental (González-Pérez et al 2009; Esteban et al 2004). És interessant, a més a més, el paper diferencial que podrien haver tingut les poblacions berbers i arabòfones del Nord d'Àfrica en el *pool* gènic del sud d'Ibèria. Així, el flux de poblacions berbers en poblacions de les Alpujarras andaluses podria arribar al 50% calculat en base a la variació per insercions *Alu*, mentre que si considerem un flux estrictament arabòfon aquest component nord-africà es reduiria a la meitat (González-Pérez et al 2003).

A banda d'aquest marc global, l'anàlisi basat en els elements *Alu*, evidencia la peculiaritat genètica de dues mostres respecte a la resta de poblacions mediterrànies: la Vall de Pas al nord de la península Ibèrica i la població berber de l'oasi egipci de Siwa. Aquestes dues poblacions han estat descrites prèviament com a aïllats genètics (Esteban et al 2006; Coudray et al 2007, 2009; Esteban et al 1998) a conseqüència de la seva particular orografia que promou l'aïllament poblacional: valls estrets envoltats de muntanyes abruptes en el primer cas i un petit oasi envoltat de desert entre Líbia i Egipte en el segon cas. Aquestes característiques haurien afavorit l'acció de processos estocàstics promotors de canvis dramàtics en les mides efectives de les poblacions, així com nivells elevats d'endogàmia i, en el cas de l'oasi de Siwa, una considerable contribució de gens d'origen sud-saharià que haurien arribat a través de la vall del Nil (Fakry 1973) i que haurien deixat una marca evident potenciada per l'isolament i la mida reduïda de la població durant bona part de la seva història recent (i que seria consistent amb una contribució sud-sahariana de fins a un 50%, segons es deriva de les dades aportades pels tres sistemes compostos *Alu*/STR).

En el conjunt de les poblacions mediterrànies, la **diversitat genètica** evidenciada pels elements *Alu* polimòrfics és força elevada tenint present les limitacions inherents a un

marcador bial·lèlic. Aquests valors de diversitat són consistents amb el valor global d'heterogeneïtat genètica que mostren les poblacions humanes del Mediterrani (F_{ST} de 3.21%). En general, la diversitat genètica global és més elevada a les poblacions de la riba sud que a les de la riba nord (F_{ST} de 1.6% al Nord i un 2.8% al Sud), fins i tot després de no considerar en l'anàlisi el paper de les dues poblacions més diferenciades (Vall de Pas i Siwa (F_{ST} de 1.4% en ambdós grups). Tot i això, la diferència entre les dues ribes, quan es testa la consistència del Mediterrani com a barrera genètica mitjançant una AMOVA jeràrquica, no es mostra prou consistent, ja que la diferenciació entre els grups del nord i del sud (F_{CT} de 2.02%) té una magnitud equivalent a la que s'observa dins dels mateixos grups (F_{SC} of 2.15%). La diferenciació genètica assoleix valors similars quan s'exclouen les dues poblacions considerades aïllats poblacionals (Vall de Pas i oasi de Siwa), mostrant una manca d'estructuració poblacional directament atribuïble a la barrera marítima (valor F_{SC} dins dels grups de 1.52% i F_{CT} entre els grups de 1.86%). Les dades obtingudes en aquest estudi contrasten amb alguns estudis previs basats en marcadors autosòmics i que pretenien remarcar l'homogeneïtat de tota la regió, especialment de la riba europea (Comas et al 2000; Tomas et al 2008).

La informació que aporten el conjunt de STRs i els sistemes haplotípics *Alu*/STR es mostra d'acord amb l'esmentada anteriorment, indicant diversitats genètiques remarcables en el Mediterrani, sempre amb valors relativament superiors en les poblacions del Nord d'Àfrica. Tanmateix, quan s'utilitzen aquest tipus de marcadors i sistemes compostos les poblacions del nord i del sud del Mediterrani mostren una major homogeneïtat interna. En aquest cas, l'AMOVA jeràrquica evidencia un cert grau de diferenciació genètica entre les dues ribes (amb valors F_{SC} dins de grups tres cops inferior al grau de diferenciació F_{CT} entre grups, 0.7% vs. 2.2%).

Aquesta aparent disparitat entre els resultats d'un i altre conjunt de marcadors probablement es deriva de la naturalesa d'un i altre tipus de polimorfisme i de les diferents habilitats de les insercions *Alu* i dels microsatèl·lits per inferir relacions genètiques particulars. Les primeres són insercions esporàdiques i per tant esdeveniments únics i més estables que probablement reflecteixen millor relacions de tipus ancestral, mentre que els darrers es caracteritzen per una taxa de mutació clarament superior i, per tant, resulten més útils a l'hora de detectar aspectes recents dels contactes poblacionals i sobre els fluxos gènics esdevinguts.

De manera complementària, l'avaluació del grau de **contribució genètica sud-sahariana** pot ajudar ha entendre millor aquests resultats. La presència de gens sud-

saharians en el *pool* gènic nord-africà, tal com evidencien elements *Alu* i haplotips *Alu*/STR, donaria suport a la idea d'una destacada permeabilitat del desert del Sàhara a les migracions humanes, tal com han proposat altres autors utilitzant marcadors clàssics (Coudray et al 2006), uniparentals (Plaza et al 2003, Aredi et al 2004), i autosòmics (Myles et al 2005). Aquest flux gènic sud-saharià –especialment el detectat pels marcadors *Alu* (amb un valor global del 12.9%)– podria ser el resultat d'antics contactes encara detectables en poblacions que han substituït als grups humans que poblaren el Nord d'Àfrica al final del Paleolític i durant el Mesolític.

D'altra banda, la considerable contribució sud-sahariana evidenciada pels haplotips *Alu*/STR (amb un valor global del 39.5% al Nord d'Àfrica), seria el reflex de contactes més o menys continus però especialment afavorits per períodes amb condicions climàtiques favorables (Dubief 1999). Així, les dades geològiques i climatològiques apuntarien a dos períodes especialment actius per afavorir el trànsit de grups humans africans cap al nord aprofitant èpoques d'una menor aridesa del Sàhara (fa entre 12000 i 8000 anys i més recentment entre fa 6000 i 3000 anys). Els valors de barreja genètica mostren valors similars en totes les mostres del Nord d'Àfrica, sempre superiors als del sud d'Europa, i podrien explicar la diferenciació trobada entre les dues ribes del Mediterrani. Concretament, en les mostres del sud d'Europa també es detecta un petit grau de barreja amb gens d'origen sud-saharià que podrien haver arribat a través dels continus contactes trans-mediterranis entre les dues ribes del mar, tal com s'ha descrit prèviament en estudis basats en dades genètiques uniparentals (Cruciani et al 2007, Plaza et al 2003, Scozzari et al 2001).

La constatació de que aquesta empremta sud-sahariana, detectada amb la bateria d'insercions *Alu*, es pot detectar a tota la riba nord del Mediterrani, des de la península Ibèrica (10%) fins a Turquia (9.4%), donaria suport a la hipòtesi de que el flux gènic a la regió hauria tingut lloc antigament i podria relacionar-se amb els primers registres històrics relacionats amb navegants fenicis i romans i que s'hauria mantingut mitjançant un més o menys constant trànsit d'esclaus que va durar fins a finals del segle XVII (Olesa-Muñido 1968) i per la possible marca deixada per les expansions islàmiques entre els segles VIII i XV. Simultàniament, aquest *stock* africà a Europa hagués estat mantingut per residents d'origen africà provinents de la Península Aràbica i les ciutats del sud de l'Àsia Menor durant els primers segles de domini islàmic i otomà, respectivament. De fet, un estudi recent basat en la distribució d'haplotips de STRs del cromosoma Y a la península Ibèrica

(Adams et al 2008) suggereix una destacada contribució de jueus i musulmans a l'actual perfil de moltes poblacions espanyoles (especialment pel que fa al llinatge masculí).

L'existència de trets genètics característics de la regió mediterrània. Tot i que per un costat, un flux gènic sud-saharià diferencial podria explicar almenys en part la diferenciació entre la riba nord i sud del Mediterrani com ja hem explicat, la detecció d'algunes combinacions particulars d'haplotips *Alu*/STR (com ara, CD4 110(-) i DM 107 (-)) que només s'han trobat al Mediterrani), evidenciaria l'existència de certes característiques mediterrànies específiques en el *background* genètic d'aquestes poblacions i presents ja d'antic. L'estimació d'edats per aquests haplotips característics de poblacions mediterrànies, donen uns resultats prou coincidents per ambdós: 36 000 anys per a la variant CD4 110(-) i 48 000 per a la combinació DM 107(-). En conseqüència, es podria hipotetitzar que aquestes variants haurien aparegut en cultures paleolítiques de l'àrea mediterrània, i que encara es detecten en el *pool* gènic actual de la regió com a marques identificatives deixades per les primeres poblacions humanes presents en aquest territori.

Les majors freqüències de les combinacions CD4 110(-) i DM 107(-) s'han trobat en poblacions berbers de la regió de l'Alt Atlas marroquí (7% i 5.5%, respectivament), i es troben en freqüències polimòrfiques a totes les poblacions analitzades al nord d'Àfrica (exceptuant els mozabits per la variant CD4 110(-)). També s'han trobat aquests haplotips particulars dispersos per la riba nord del Mediterrani: a la península Ibèrica, a Grècia, a Turquia i a les principals illes del Mediterrani occidental (Mallorca, Còrsega, Sardenya i Sicília; González-Pérez et al 2007). L'haplotip CD4 110(-) també s'ha detectat en poblacions del Sàhara occidental i de Mauritània (Flores et al 2007) i en cinc de les set illes de l'arxipèlag canari (Flores et al 2001).

La distribució observada d'aquests haplotips, conjuntament amb una sèrie de característiques comunes com ara les edats estimades coincidents amb l'expansió de les cultures paleolítiques a la regió, la seva àmplia distribució per tota l'àrea mediterrània i les Illes Canàries poblades pels guantxes emparentats amb els berbers del Nord d'Àfrica, la seva freqüència elevada en tot el Mediterrani occidental i la seva presència en poblacions aïllades com les regions centrals de Sardenya i Còrsega, semblen suggerir que l'actual *background* genètic de les poblacions mediterrànies encara preserva marques paleolítiques que no han desaparegut, ni per efecte de les expansions neolítiques produïdes a les dues ribes paral·leles ni per processos migratoris recents i de flux gènic homogeneïtzador.

Els resultats d'aquesta tesi basada en marcadors autosòmics estarien d'acord amb estudis previs realitzats sobre marcadors uniparentals (Plaza et al 2003; Achilli et al 2004;

Olivieri et al 2006; Cruciani et al 2007). Aquests treballs suggereixen que alguns haplogrups o clades específicament mediterranis (U6 i M1b al mtDNA; EM78 i EM81 al cromosoma Y) datarien de temps paleolítics i que podrien haver entrat a Europa i al Nord d'Àfrica des del Llevant. Tot i que els resultats d'aquest apartat de la tesi es limiten estrictament a la variació genètica a la regió mediterrània i encara que no pot oferir una conclusió sobre el possible origen oriental de combinacions particulars, sí que són capaços de demostrar la utilitat de l'ús combinat d'elements autosòmics *Alu* amb polimorfismes STR per a detectar relacions poblacionals detallades i rastres del passat en el *pool* genètic de les poblacions. Així, la concordança general entre dades uniparentals i autosòmiques confirma la capacitat dels mètodes inferencials de la genètica de poblacions humanes per reconstruir una mateixa història comuna, la dels pobles mediterranis en el nostre cas.

13.2 Sobre la particularitat genètica de les grans illes mediterrànies

Entre els resultats d'aquesta tesi doctoral sobre la caracterització poblacional de les grans illes del Mediterrani occidental, cal destacar també les dades sobre variació de 18 insercions *Alu* autosòmiques i els tres sistemes haplotípics *Alu*/STR's ja esmentats. La discussió permetrà, així mateix, diferenciar el grau d'heterogeneïtat dins les illes ja que s'han obtingut dades de dues poblacions geogràficament diferenciades a Còrsega, Sardenya i Sicília, i que afegides a la població insular de Mallorca fan un total de set mostres integralment analitzades en aquest projecte. Així, aquest apartat pretén repassar el grau de diversitat i diferenciació interna d'aquestes illes i l'efecte de possibles fenòmens d'aïllament, les afinitats i diferències entre poblacions insulars, el grau de flux gènic extern que haurien rebut i, finalment, l'edat de divergència de les diferents poblacions.

Les set poblacions illenques mediterrànies analitzades mostren, pel que fa als marcadors *Alu*, un considerable **grau de diversitat genètica**, exceptuant la Sicília oriental i la costa occidental de Sardenya on l'heterozigositat observada és menor. La diferenciació genètica entre poblacions i entre illes és altament significativa tal com es dedueix dels resultats obtinguts amb les comparacions poblacionals per parelles (17 de les 21 comparacions possibles són estadísticament significatives) i dels valors globals de F_{ST} (2,2%). A l'illa de Sardenya es presenta una tendència general a una reducció de la diversitat genètica conjuntament amb uns patrons discrepants per a les freqüències al·lèliques d'alguns elements *Alu*, evidenciant una certa diferenciació d'aquesta illa. El valor global de diferenciació entre illes (2.2%, $p < 0.001$) és fins i tot lleugerament superior o de la mateixa

magnitud que el trobat pel conjunt dels europeus (1.9%) o pels nord-africans (2.3%) per a un conjunt equivalent de marcadors i de mostres (Comas et al 2000; González-Pérez et al 2003; González-Pérez et al 2009).

Pel que fa a les **relacions poblacionals globals**, les poblacions sardes i la Sicília occidental es mostren en la representació global de distàncies clarament diferenciades respecte a la resta de poblacions insulars. L'efecte de la deriva genètica en grups poblacionals reduïts pot haver contribuït activament a aquests diferenciació. Tot i que no podem ignorar certes evidències històriques, lingüístiques i també genètiques que suggereixen clares diferències a Sardenya entre el procés de poblament humà de l'interior i de la costa com a conseqüència de l'aïllament a la regió muntanyosa central de la població original de cultura nuràgica, lluny de l'efecte que haurien tingut els invasors romans i cartaginesos a les costes de la illa (Piazza et al 1988). En el cas de l'illa de Sicília, els registres històrics també indiquen un destacat retrocés de l'antiga població sicanià (2500-1200 aC) davant l'arribada a les costes orientals dels sicels provinents de la Itàlia continental (1200 aC). D'aquesta manera, les dades aportades pels elements *Alu* i pels sistemes compostos *Alu*/STR estarien d'acord amb l'evidència que han aportat prèviament el estudis del cromosoma Y (Francalacci et al 2004) i del mtDNA (Morelli et al 2000) que suggereixen que tant Sardenya com Sicília serien les poblacions humanes més diferenciades de la conca occidental mediterrània.

La relativa heterogeneïtat entre la resta de mostres insulars i revelada en les representacions gràfiques d'escalaments multidimensionals basats en el conjunt de 18 marcadors *Alu* es veu relativament reduïda quan les poblacions continentals mediterrànies s'uneixen a l'anàlisi global. En aquest darrer cas, la relativa proximitat de Còrsega, Mallorca i de Sicília oriental a les poblacions continentals mediterrànies (França, Espanya i Itàlia) suggerida per l'anàlisi d'una bateria completa de marcadors *Alu* tornaria a estar d'acord amb els resultats d'estudis basats en el DNA uniparental (Falchi et al 2006; Picornell et al 2005).

Pel que fa a la **informació poblacional revelada per cada un dels grups de marcadors** utilitzats, s'observa que aquest grau de diferenciació genètica a les illes occidentals de la Mediterrània no resulta tan evident quan centrem exclusivament l'anàlisi en els STR's analitzats. Tot i que totes les mostres presenten uns valors destacats de diversitat intrapoblacional, no poden detectar-se diferències significatives entre poblacions ni graus de variació genètica destacables per cap dels tres STR's estudiats. Les discrepàncies aparents entre el conjunt de marcadors *Alu* i els STR's podrien estar relacionats amb la naturalesa i característiques pròpies de cada marcador. Les insercions *Alu* són fenòmens

únics i de variació bial·lèlica, característiques que evitarien l'efecte distorsionador causat per fluctuacions aleatòries lligades a processos de mutació i de relativa reducció poblacional. D'altra banda, l'elevada taxa mutacional dels STR's conjuntament amb l'efecte homogeneïtzador del flux gènic recent farien que aquests marcadors fossin més influenciables per fenòmens recents de la dinàmica poblacional i podrien explicar les aparents discrepàncies observades entre els dos tipus de marcadors genètics estudiats.

El grau de **flux gènic entre les illes** del Mediterrani occidental i més enllà sembla haver estar clarament destacable. En el cas de la variació associada als STR's torna a evidenciar-se una alta heterozigositat en totes les poblacions, en molts casos amb paràmetres de variació que superen els descrits per a poblacions continentals europees (Tishkoff et al 1996, 1998; Esteban et al 2004). D'aquesta manera, la insularitat no sembla haver actuat –en general– com una barrera genètica estricta, al menys entre les illes mediterrànies occidentals i les poblacions de la riba mediterrània del sud d'Europa, tal com indiquen així mateix els nombrosos registres històrics.

Les dades aportades pel nostre treball no poden detectar una afinitat directa entre illes del Mediterrani i poblacions nord-africanes. Ara bé, aquesta dada global, no pot amagar l'evidència de rastres particulars de flux gènic d'origen africà a les illes. Per exemple, el treball ha permès detectar traces de contribució nord-africana en el *pool* genètic d'algunes de les poblacions insulars, especialment evidents en el cas de la diversitat descrita pels haplotips *Alu*/STR del sistema CD4. Així, la població de Sicília mostra freqüències destacables d'haplotips típicament mediterranis suposadament originats al nord-oest africà (al voltant del 7% en el conjunt de l'illa), clarament superiors a les presents en la resta d'illes (amb menys del 2,5%), i que apuntarien a l'estratègica posició geogràfica d'aquesta població i al seu paper històric de pas entre el nord d'Àfrica i l'Europa continental. En contraposició, l'illa de Mallorca, tot i la dominació islàmica de més de tres segles, no mostra restes destacables d'aquests haplotips mediterranis ancestrals amb origen al Nord d'Àfrica. Aquest mateix patró havia estat observat en estudis previs (Picornell et al 2005) que donen suport a l'evidència històrica d'una ben documentada repoblació completa de l'illa per part dels espanyols després de la conquesta catalanoaragonesa.

Resulta interessant, així mateix, aportar algunes dades sobre les **edats de divergència entre poblacions**, tot i que aquestes estimacions representen valors màxims, ja que el mètode utilitzat assumeix que la variació mesurada en els polimorfismes STR's s'ha desenvolupat a escala local. De fet, sabem que l'efecte de les migracions i el flux gènic hauria tingut un paper considerable en la configuració de la variació observada en les

poblacions actuals i per tant aquestes edats de divergència són relativament esbiaixades. Aquestes estimacions de temps configuren un rang màxim de temps de divergència entre dos grups insulars d'al voltant de $24\,259 \pm 6211$ anys. El primer grup estaria configurat per: Sardenya central, Còrsega i Sicília occidental (les poblacions amb un aparent major component de *pool* genètic ancestral). El segon grup el formarien: la costa occidental sarda, Sicília oriental i Mallorca (és a dir, les poblacions insulars amb major contacte amb navegants i migracions recents). Al seu torn, una data aproximada de 5973 ± 2815 anys separaria a les poblacions de la costa occidental corsa i de Sicília occidental. Les dates obtingudes amb aquesta aproximació serien aparentment compatibles amb l'heterogeneïtat poblacional mostrada pel conjunt d'elements *Alu*, i suggeririen que algunes de les diferències entre mostres podrien remuntar-se a l'efecte diferencial dels primers poblaments humans d'aquestes illes, conjuntament amb l'efecte posterior de la deriva genètica i els processos d'isolament.

D'altra banda, i en clar contrast, l'elevada diversitat dins de les poblacions i la remarcable homogeneïtat entre illes obtingudes a partir de les dades provinents d'STR's i haplotips *Alu*/STR indicarien que, com a mínim des de l'època neolítica a la regió, el flux genètic hauria estat un component destacable de la dinàmica poblacional a la conca occidental mediterrània. La deriva genètica a la costa occidental de Sardenya i el flux genètic a la Sicília occidental haurien contribuït de manera destacable a la seva diferenciació genètica.

13.3 Sobre la caracterització genètica de l'Illa de Pasqua i el poblament del Pacífic

L'anàlisi de la variació de 18 polimorfismes d'inserció *Alu* autosòmics també s'ha aplicat a la caracterització genètica d'una població polinèsica particular: l'Illa de Pasqua. Aquest treball pot considerar-se complementari d'altres investigacions sobre la caracterització genètica d'aquesta població tan particular –la població humana insular més aïllada del planeta– (Cruz-Coke et al 1988, Etcheverry 1967, Hagelberg et al 1994, Hurles et al 2003, Lum et al 1998, Serjeantson et al 1989), però amb un parell de remarcables novetats. Es tracta del primer treball publicat basat en la determinació d'un conjunt suficientment ampli d'elements polimòrfics *Alu* en aquesta població del Pacífic. A més, també és el primer cop que dues mostres coetànies però genealògicament diferents i ben

caracteritzades de la mateixa població són analitzades conjuntament per tal de fer inferències sobre el seu passat recent i per avaluar l'efecte del contacte recent amb poblacions colonitzadores. Aquesta diferenciació entre les dues mostres s'ha pogut establir inequívocament en base a la informació personalitzada i confirmada pels registres genealògics disponibles a partir d'un estudi demogràfic portat a terme sobre la mateixa població (Hernández et al. 2000). Les particularitat de cada una de les dues mostres de l'Illa de Pasqua van permetre una aproximació directa tant a l'heterogeneïtat actual de la població com als canvis que hauria pogut patir el *pool* gènic durant el darrer segle d'història de l'illa (segle XX) i que probablement encara es manté.

La diversitat genètica actual a l'Illa de Pasqua. Els marcadors autosòmics estudiats han permès demostrar l'existència d'una clara diferenciació genètica a l'actual població de l'Illa de Pasqua. Aquesta variació resulta evident a partir d'una simple comparació de la distribució dels marcadors i sembla ser estadísticament consistent ja que les diferències s'observen per més del 22% de les comparacions realitzades, un valor àmpliament superior al 5% esperat per l'atzar. Una altra evidència d'aquesta heterogeneïtat poblacional és la distància genètica entre les dues poblacions de l'illa que arriba al 0.051, un valor equivalent a la distància mitjana entre les poblacions del sud-est asiàtic i del Pacífic i tres vegades superior a les distàncies habituals entre poblacions europees, confirmant-se així un grau de diversitat considerable entre els actuals habitants de l'Illa de Pasqua.

D'acord amb les característiques de les mostres estudiades, la diferenciació genètica observada pot ser atribuïda a l'efecte de fenòmens recents de migració i barreja poblacional (concretament d'un segle d'antiguitat). Aquest efecte estaria d'acord tant amb l'increment detectat de diversitat genètica a la mostra amb ancestres no-rapanui (NR) de Pasqua, com amb la reducció de prop del 50% de les distàncies genètiques que mostra aquesta població amb la resta de grups poblacionals, en comparació amb les majors distàncies de la mostra RR (amb tots els ancestres rapanui). Similarment, l'estabilitat mutacional que presenten els marcadors *Alu* i el relatiu curt període de temps (un segle) que cobririen les mostres utilitzades, també donarien suport a l'efecte migratori com l'explicació més probable per a la diferenciació. Aquests resultats concorden amb els estudis demogràfics que s'han dut a terme sobre la població d'Illa de Pasqua i que mostren un increment molt destacable i progressiu del grau d'exogàmia en el decurs del s. XX (Hernández et al 2000).

Tot i que el fet de no poder establir una filogeografia acurada pels marcadors *Alu* ens implica una sèrie de restriccions en la identificació de les principals **fonts d'immigració a l'illa**, podem obtenir una clara evidència indirecta a partir de l'anàlisi de

barreja genètica portat a terme. Aquesta aproximació analítica ha permès estimar la contribució de gens europeus a la mostra NR de l'illa i quantificar-la en valors que pràcticament doblen la de la població RR ($m=0.484$ vs. $m=0.296$). Aquestes xifres suggeririen que el flux genètic europeu hauria estat un important factor de canvi en el *pool* genètic de l'illa observat avui dia. La introducció de variació genètica d'origen europeu és un fenomen ben establert per la gran majoria de poblacions polinèsiques, ben documentat històricament des del segle XVIII i que ha quedat evidenciat en diferents estudis genètics d'illes oceàniques, especialment a partir de l'estudi de la variació del cromosoma Y (Hurles et al. 2002, Underhill et al. 2001).

En moltes ocasions, aquesta entrada de variació europea va ser principalment canalitzada per via masculina i basada en nombres poblacionals baixos dels exploradors i conqueridors, però amb un efecte ràpidament magnificat per les particulars condicions insulars d'aquestes poblacions (mides poblacionals petites, baixa resistència a epidèmies, colls d'ampolla recurrents després del contacte amb els europeus, etc.). En el cas particular de l'Illa de Pasqua, la possible entrada d'emigrants d'origen europeu que suggereix la variació genètica autosòmica estaria directament associada amb l'efecte acumulatiu de les inicialment limitades arribades d'individus provinents de Xile a partir del moment de l'annexió de 1888, i que s'hauria intensificat al llarg del segle XX, especialment a partir dels anys seixanta amb l'obertura de l'aeroport internacional que connectaria l'illa amb la resta de la humanitat. D'acord amb els registres demogràfics disponibles, en aquells moments, la major font d'immigrants seria el Xile continental i, en conseqüència, l'arribada de gens europeus a través del desembarcament de xilens amb un component d'ascendència europea important.

Respecte a les **relacions genètiques i l'origen de la població Rapanui** que vivia a l'Illa de Pasqua fa més d'un segle (representada pel llinatge RR analitzat en l'estudi), els elements *Alu* indicarien una afinitat destacada amb les poblacions humanes del sud-est asiàtic. L'estructuració de la variació genètica en grups poblacionals i geogràfics indica que la bateria d'insercions *Alu* utilitzades en les comparacions, tindrien suficient poder estadístic (mitjançant una anàlisi AMOVA) per resoldre les diferències de constitució entre els següents grups poblacionals: sud-est asiàtic, grup taiwanès (Taiwan+Filipines), Amèrica i Europa. És a dir, per diferenciar entre el conjunt de poblacions relacionats amb els possibles ancestres polinèsics i, particularment, amb l'origen de l'actual panorama genètic detectat a l'Illa de Pasqua. En aquest sentit, tant les distàncies genètiques com l'anàlisi de components principals evidencien la similaritat de Rapanui amb els habitants del sud-est

asiàtic i de localitats de la Wallacea, així com una destacada diferenciació tant amb els europeus com amb els australians i melanesis. L'afinitat de les poblacions polinèsiques i de Rapanui amb el sud-est asiàtic també s'ha observat en altres tipus d'estudis com els basats en variants moleculars de les globines (Hill et al. 1989) i en dades recents provinents del DNA mitocondrial (Kayser et al. 2000; Oppenheimer and Richards, 2001). En contraposició, altres tipus de dades indicarien una major afinitat dels polinesis amb els grups humans de Taiwan i Filipines (com per exemple, en els primers estudis de mtDNA: Lum et al. 1998; i en els basat en el sistema HLA: Hagelberg et al. 1999).

Tot i que no es pot descartar categòricament una possible diferenciació del grup Taiwanès per efecte recent de la deriva i de fenòmens demogràfics particulars esdevinguts durant els darrers 5000 anys (Melton et al. 1998), els resultats mostrats per la bateria de marcadors autosòmics emprats en aquest treball donarien un suport principal al model de poblament polinèsic que ha rebut el nom de *Voyaging Corridor* o, més concretament, la seva variant *Slow Boats to Polynesia* (Oppenheimer and Richards, 2001). Aquesta teoria postula que els ancestres dels actuals polinèsics provindrien de poblacions de la regió Wallacea amb amplis coneixements de navegació.

És probable que l'expansió d'aquests navegants s'hagués produït abans de l'arribada de l'onada neolítica a la regió, tal com es dedueix de la datació de variants genètiques i haplotips específics d'aquestes poblacions. La canoa hauria estat, en aquest cas, l'eina imprescindible per a aquestes expansions poblacionals. Aquest instrument, hauria permès l'avançada de l'onada migratòria d'illa a illa al llarg de la Polinèsia propera en un primer pas, i mitjançant viatges a llarga distància per a l'expansió recent fins a illes remotes com el cas de l'Illa de Pasqua. Aquest model implica una sèrie d'episodis intermedis en que els navegants s'haurien establert durant un cert temps a la costa de la regió melanèsica, fet que implicaria un cert grau de barreja poblacional entre les poblacions locals melanèsiques i els navegants ancestrals polinèsics. Precisament, la bateria de polimorfismes *Alu* emprats en aquest treball donen dades altament consistents amb aquesta hipòtesi: la variació genètica detectada en aquest estudi permet inferir un grau de flux gènic polinèsic a la regió costanera de Nova Guinea (Melanèsia) de fins al 15%, un fet que no es detecta en les poblacions melanèsiques més aïllades de l'interior de la regió.

Una altra evidència genètica a favor de l'origen occidental dels polinesis és la clina detectada en la variació genètica, que es redueix progressivament de poblacions més occidentals a les més orientals del Pacífic, i que ha estat detectada per a diferents marcadors genètics (Martinson et al 1993). Tot i aquesta tendència global, quan centrem l'anàlisi en la

distribució de marcadors *Alu* a la mostra d'Illa de Pasqua amb tots els ancestres rapanui, aquesta població presenta una elevada diversitat genètica tenint en compte el seu remarcable isolament. Aquesta és una característica que també s'ha observat per a d'altres marcadors (Hurles et al 2003), i que aparentment no estaria d'acord amb el seu grau d'aïllament i amb la seva història demogràfica caracteritzada per períodes d'endogàmia i colls d'ampolla.

Tanmateix, aquesta diversitat major de l'esperada a Rapanui al començament del segle XX sembla clarament explicable quan entren en la visió global els esdeveniments històrics pels que va passar l'illa i el paper del trànsit d'esclaus conduït pels peruans al llarg de tota la Polinèsia durant la segona meitat del segle XIX. Aquest tràgic episodi històric hauria suposat inicialment la despoblació de moltes illes (fins al 80% de la població original), inclosa Rapanui, i la desastrosa repatriació que pel camí va emportar-se les vides de milers de nadius polinesis (Maude 1981). Aquesta repatriació va implicar l'arribada a Illa de Pasqua dels esclaus alliberats provinents d'altres poblacions polinèsiques i de les Amèriques, molts d'ells amb un considerable grau de barreja amb gens europeus, i tot just abans de que la població original de l'illa es veiés assolada per les epidèmies portades per aquests immigrants i retornats. Tot i que el nombre de repatriats que realment van arribar a l'illa seria relativament baix, el seu impacte i contribució a les generacions posteriors rapanui seria força considerable com a conseqüència de la seva resistència a les malalties i les epidèmies (ja havien sobreviscut prèviament a les pèssimes condicions de treball, l'esclavatge i al llarg viatge de retorn), tal i com semblen evidenciar diferents registres històrics (Maude 1981; McCall 1996). Aquesta diversificació d'avantpassats que haurien deixat la seva empremta al *pool* gènic de la població rapanui de finals del segle XIX explicaria en bona part la relativament elevada diversitat que ara observem per als elements *Alu* i altres marcadors genètics. El fet de que molts dels esclaus repatriats fossin de diversos orígens polinèsics, fa que les relacions genètiques que s'obtenen a partir de la variació associada als elements *Alu* serien indicatius tant de traces del poblament original de la Polinèsia com dels efectes que han tingut els fenòmens històrics recents en la determinació del patrimoni genètic dels moderns rapanui.

Respecte a les possibles relacions poblacionals de l'Illa de Pasqua i d'altres nadius polinèsics amb els amerindis (Heyerdahl, 1950), que esporàdicament s'han suggerit a partir de dades genètiques (Cann and Lum, 1996) i culturals (Green, 2000), la variació present en els *loci* autosòmics estudiats sembla no estar d'acord amb cap afinitat especial que doni suport a contactes prehistòrics claus entre Rapanui i aquestes poblacions. Les mostres

poblacionals americanes que s'han utilitzat per a l'estudi comparatiu en aquest treball mostren una relació genètica més propera amb poblacions del sud-est asiàtic, deixant oberta la possibilitat de l'existència de diferents orígens i moments per a les migracions humanes que van acabar poblant el continent americà.

En conjunt, aquest treball centrat en l'estudi de la variació genètica associada a elements *Alu* a l'Illa de Pasqua, ha demostrat l'existència d'una considerable heterogeneïtat genètica en la població rapanui actual. Aquesta evidència seria il·lustrativa de l'impacte que els fenòmens migratoris del darrer segle haurien tingut en un procés de canvi del *pool* gènic en un grup poblacional humà petit i inicialment aïllat. Els resultats obtinguts de l'estudi són indicatius també de la necessitat d'integrar adequadament la informació històrica per entendre en tota la seva amplitud la complexitat inherent a la història genètica de molts grups humans, amb independència del grau de variació propi de cada marcador utilitzat. Finalment, el treball també destaca la importància cabdal que representa un mostreig acurat per als estudis de poblacions humanes a l'hora de clarificar la naturalesa i el moment dels canvis que afecten a l'evolució dels grups humans.

14 *Conclusions de l'Estudi*

Les principals conclusions derivades dels treballs que conformen aquesta tesi doctoral s'enumeren a continuació:

1. Les dades genètiques aquí analitzades donen suport, a part d'una clara diferenciació poblacional humana a Àfrica, a una **estructuració de les poblacions mediterrànies** entre ribera nord i sud que no descarta un cert grau de contacte a través de l'Estret de Gibraltar, així com la particularitat i diferenciació de grups humans de l'àmbit mediterrani amb històries particulars com ara la població del Vall de Pas i de l'oasi de Siwa. El patró de diferenciació primari depèn d'una correlació destacada de les distàncies genètiques dels sistemes autosòmics estudiats amb les distàncies geogràfiques entre poblacions de la regió mediterrània. Aquesta evidència donaria suport a la teoria que estipula que en un àmbit amb una història humana tant dinàmica i antiga, el primer factor que explica la variabilitat genètica és la distància geogràfica.
2. En el **Mediterrani occidental**, la variació genètica interpoblacional és considerable i permet una diferenciació global significativa entre poblacions de la regió. Aquesta conclusió il·lustra la importància d'un mostreig adequat i d'una caracterització antropològica suficientment precisa per entendre tota la complexitat biològica inherent a les poblacions humanes mediterrànies.
3. La variació genètica observada per a **diferents sistemes autosòmics** (insercions *Alu*, microsatèl·lits i haplotips *Alu*/STR) mostra una correlació significativa, tot i les diferències de naturalesa mutacional en un i altre tipus de marcador.
4. La **variació genètica associada als haplotips** estudiats en diferents sistemes és concordant amb l'evolució particular de cada *locus*. Així els haplotips associats a elements *Alu* ancestrals (CD4(+), FXIIB(-) i DM(+)) presenten graus de variació superiors als derivats, demostrant que les insercions *Alu*

polimòrfiques poden utilitzar-se com una eina fiable per a l'establiment de la polaritat del canvi evolutiu.

5. El sistema haplotípic **CD4** presenta aquest mateix patró general amb una estructuració geogràfica evident: els cromosomes CD4(+) i CD4(-) no presenten variacions clarament diferenciades en poblacions sud-saharianes, la seva diferenciació és significativa al Nord d'Àfrica i, finalment, altament significativa a la ribera nord del Mediterrani, una tendència que concorda amb l'esperada per un procés d'expansió humana *Out of Africa*. Pel que fa al sistema **FXIIB**, la diferenciació entre cromosomes ancestrals (-) i derivats (+) no és tan evident, presentant diversitats relativament elevades en ambdós grups. Els cromosomes FXIIB(-) presenten al·lels petits del tetranucleòtid en poblacions sud-saharianes (172 i 176), una freqüència major d'al·lels llargs al Nord d'Àfrica (184 i 188) i clarament dominant a la ribera nord del Mediterrani (188). En els cromosomes derivats FXIIB(+), l'estructuració de la variació és menys destacada amb dominància de l'al·lel 180 al Mediterrani i del 184 a la regió sud-sahariana. En el sistema haplotípic **DM**, la variació observada concordaria amb l'aparició de la deleció (*Alu*(-)) lligada a la variant 98 bp del trinucleòtid, mentre que en haplotips ancestrals *Alu*(+) dominen els al·lels de baix nombre de repeticions (77 bp).
6. Així mateix, cal destacar que la variació en els sistemes haplotípics *Alu*/STR estudiats atribuïble a la diferenciació entre **Nord i Sud del Mediterrani** és globalment significativa (variació entre grups del 2,2% i global del 2,9%).
7. S'han pogut detectar dos **haplotips particulars**, CD4 110(-) i DM 107(-), relativament freqüents en poblacions berbers i presents al Nord del Mediterrani però absents en poblacions sud-saharianes i centreeuropees, que podrien estar relacionats amb l'origen dels primers pobladors de la regió mediterrània i que s'haurien originat en el Paleolític, fa entre 30 000 i 40 000 anys.
8. S'han establert valors de **flux gènic** sud-saharià al Nord d'Àfrica d'al voltant del 13% per al conjunt de marcadors *Alu* (valors coincidents amb estudis que utilitzen SNP's i, en general, també amb els marcadors uniparentals). La quantificació del flux calculat a partir d'haplotips *Alu*/STR pot arribar al 35%.
9. El flux gènic a través de l'**Estret de Gibraltar** és quantificable quan es testa per a la bateria d'elements *Alu* la influència nord-africana al sud de la

península Ibèrica: aquesta influència és el doble de marcada quan ens centrem en gens berbers en comparació amb poblacions nord-africanes arabòfones. Al conjunt de la riba nord mediterrània, el flux gènic arribat del Nord d'Àfrica es pot quantificar també a través del rastreig del *pool* gènic sud-saharià que hauria hagut de travessar la regió: en el cas de les dades obtingudes pels haplotips *Alu*/STR, aquest flux torna a ser coincident amb estudis previs (6-15%).

10. La població de l'illa de **Sardenya** presenta una variació genètica molt destacada amb valors mitjans de F_{ST} (3,7%) per a la bateria de polimorfismes *Alu* que tripliquen els del conjunt de grans illes de la Mediterrània occidental. Aquest valor és coincident amb les distàncies genètiques mitjanes (quatre vegades superiors entre illes mediterrànies que pel conjunt de les poblacions continentals mediterrànies).
11. La diversitat i nombre d'haplotips particulars per al sistema **CD4** són especialment destacats a les illes de Sardenya i Sicília, gràcies a la presència de variants típicament africanes (85(-) i 100(-)) i d'altres característiques de poblacions berbers (110(-)). Els patrons de distribució dels sistemes **DM** i **FXIIIB** mostren diversitats intrapoblacionals força elevades i una certa homogeneïtat global per a les illes mediterrànies.
12. La representació gràfica de les distàncies genètiques entre **illes** i en **relació amb altres poblacions** mediterrànies evidencia la diferenciació de la població sarda, especialment el centre de l'illa, i de la Sicília occidental (aquestes serien les poblacions que haurien mantingut un *pool* gènic ancestral més antic per isolament en el cas de Sardenya i un contacte amb el nord d'Àfrica més acusat en el cas de l'occident sicilià). Mallorca, Còrsega i la Sicília oriental han estat molt més relacionades històricament amb les poblacions continentals europees.
13. Pel que fa a la població polinèsica de l'**Illa de Pasqua**, la bateria d'elements *Alu* autosòmics estudiats permet la caracterització d'una diversitat genètica relativament elevada a la població, superior a l'esperada pel seu aïllament geogràfic extrem. El test global de diferenciació poblacional indica l'heterogeneïtat entre els dos llinatges caracteritzats a Rapa Nui: el llinatge amb tots els ancestres descrits rapanui i el llinatge amb grau de mestissatge recent amb no polinèsics.

14. L'anàlisi de **barreja genètica a l'Illa de Pasqua** resulta en un domini de la contribució polinèsica (50%) al llinatge rapanui, però amb presència destacada de contribució europea (29%) i americana (21%). En el llinatge mestís s'observa una contribució europea molt elevada (49%), i una proporció similar de gens polinèsics (27%) i amerindis (25%). Aquesta tendència general confirmaria el flux gènic del darrer segle provinent d'individus europeus.
15. La comparació global amb altres poblacions mundials mitjançant una anàlisi AMOVA permet diferenciar significativament els principals **grups humans** (de la **regió asiàtico-pacífica**): grup taiwanès, illes del sud-est asiàtic i polinèsia, aborígens australians i melanesis de Nova Guinea. Aquesta estructura explica un 10,3% de variació entre grups i un 6,6% dins dels mateixos. El llinatge aparentment no mestís de Rapanui es relaciona primàriament amb poblacions del sud-est asiàtic i en segon lloc amb poblacions americanes, mentre que les distàncies genètiques es doblen quan es compara amb altres poblacions mundials. Les diferents aproximacions gràfiques a la distribució de distàncies genètiques per la bateria d'elements *Alu* relacionen l'Illa de Pasqua i altres poblacions polinèsiques amb illes del sud-est asiàtic i de la Wallacea, deixant a part el grup format per Taiwan i Filipines que sembla estar més proper a l'expansió humana que hauria travessat l'Estret de Bering per colonitzar Amèrica. Aquest patró dona suport al model de colonització del Pacífic anomenat *Slow Boats to Polynesia*.
16. Rapanui presenta una **diversitat genètica superior a l'esperada** per la seva situació geogràfica a l'extrem d'una clina oest-est des del sud-est asiàtic. Aquesta diversitat detectada al llinatge representatiu de l'illa a principis de segle XX, evidenciaria la influència que haurien tingut el tràfic d'esclaus exercit pel Perú durant el segle XIX i l'alliberament atzarós dels mateixos en illes polinèsiques durant la segona meitat d'aquell segle.