



# Anàlisi multilocus del polimorfisme nucleotídic al llarg del cromosoma J de *Drosophila subobscura*

Roser Pratdesaba i Moreno

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

# Universitat de Barcelona

Departament de Genètica

Memòria presentada per optar al títol de Doctora per la  
Universitat de Barcelona

Programa de doctorat en Genètica

## **Anàlisi multilocus del polimorfisme nucleotídic al llarg del cromosoma J de *Drosophila subobscura***

Tesi doctoral presentada per:

**Roser Pratdesaba i Moreno**

Directores de Tesi:

**Dra. Carmen Segarra i Robert**

**Dra. Montserrat Agudé i Porres**

Barcelona, Febrer 2014



**A la memòria del meu avi**



## Agraïments

Hi ha moltes persones que m'han ajudat i acompanyat durant aquests anys i d'aquesta experiència m'enduc, entre altres coses, un grapat de bons amics. En aquest sentit, voldria agrair a totes aquestes persones el fet d'haver-me ensenyat i d'haver estat al meu costat durant aquest llarg viatge, i en especial a:

La Montserrat Agudé i la Carme Segarra, per haver-me donat l'oportunitat de fer la tesi doctoral, per la seva paciència, confiança, recolzament i esforç durant tots aquests anys.

El Julio Rozas, per la seva amistat, bons consells i tantes "comilones" passades.

La Montserrat Papaceit, per la seva disponibilitat, paciència i ajuda en el microscopi.

L'Elvira Juan i la Dorcas, per la seva companyia al laboratori.

El Pablo, per haver-me ajudat tantes vegades, sempre tenint un moment per a mi, i per la seva amistat incondicional.

La Eva i el Michael, per ser tan bons companys de laboratori i per totes les experiències passades junts.

El Juanma i la Cris, els meus companys d'aventures, per ser tan bones persones i haver fet aquest camí molt més planer.

L'Àlex, per tantes preguntes sobre la coalescència respostes i tots els bons moments.

La Sara i el Filipe, per la seva amistat.

La Gemma i el David, per haver-me ensenyat els camins experimentals.

El Sebas, el Carlos, l'Àngel i la resta dels membres que han passat pel grup de Genètica Molecular Evolutiva durant aquests anys.

La Maria Pilar Aluja i el Paolo Francalacci, per haver-me iniciat en aquest món.

Els meus pares i la meva família, per tota l'ajuda i el recolzament incondicional.

Finalment la Vero, pel passat, el present i tot el futur que ens espera.



El present treball s'ha realitzat gràcies a una beca de Formació de Personal Investigador del Ministerio de Educación y Ciencia (BES-2008-009526). A més a més, el treball ha contat amb les subvencions de la Secretaría de Estado de Universidades e Investigación del Ministerio de Educación y Ciencia (BFU2007-63228) i de la Comissió Interdepartamental de Recerca i Innovació Tecnològica de la Generalitat de Catalunya (2009SGR-1287).





## ÍNDEX

---

INTRODUCCIÓ	1
I1. <i>DROSOPHILA</i> COM A ORGANISME MODEL	3
I2. EVOLUCIÓ MOLECULAR	4
I2.1 Variabilitat genètica	4
I2.2 Detecció de la variabilitat i significat evolutiu	4
I2.3 Història evolutiva de les poblacions: factors selectius i demogràfics	6
I3. <i>DROSOPHILA SUBOBSCURA</i>	12
I3.1 Taxonomia de l'espècie	12
I3.2 Distribució geogràfica	13
I3.3 Biologia de l'espècie	13
I3.4 Els cromosomes de <i>D. subobscura</i>	13
I3.5 Polimorfisme cromosòmic	15
I3.6 Variabilitat nucleotídica associada al polimorfisme cromosòmic	17
I3.7 El sistema cromosòmic $J_{st}/J_1$ a <i>D. subobscura</i>	20
I3.8 Influència de factors demogràfics en la variabilitat nucleotídica de <i>D. subobscura</i>	21
OBJECTIUS	25
MATERIALS I MÈTODES	29
M1. OBTENCIÓ DE DADES	31
M1.1 Mostra poblacional	31
M1.2 Caracterització dels inserits dels fags recombinants	31

<b>M1.3 Purificació del DNA genòmic i seqüenciació</b>	<b>34</b>
<b>M1.4 Assemblatge i alineament de seqüències</b>	<b>35</b>
<b>M2. ANÀLISI DE LES DADES</b>	<b>36</b>
<b>M2.1 Estimes dels nivells i patrons de variació nucleotídica</b>	<b>36</b>
- Nivells de variació nucleotídica	36
- Patrons de variació nucleotídica	37
- Desequilibri de lligament entre posicions	41
- Test de Hudson-Kreitman-Aguadé (HKA)	42
<b>M2.2 Anàlisi de la variació nucleotídica a les regions afectades per la inversió <math>J_1</math></b>	<b>42</b>
- Distància al punt de trencament	42
- Diferenciació genètica entre les ordenacions $J_1$ i $J_{st}$	43
- Conversió gènica	44
- Genealogia i xarxa de la mostra de seqüències	45
- Desequilibri de lligament amb la inversió	45
<b>M3. INFERÈNCIES EVOLUTIVES</b>	<b>46</b>
<b>M3.1 Estimació del temps de divergència entre diversos llinatges</b>	<b>46</b>
<b>M3.2 Mutacions recurrents i l'espectre de freqüències desdoblada</b>	<b>47</b>
<b>M3.3 Inferències demogràfiques (<i>Approximate Bayesian Computation</i>)</b>	<b>48</b>
- Introducció al procés d'ABC	49
- Dades observades	50
- Disseny de models	50
- Simulacions de Monte Carlo	56
- Tria d'estadístics	57
- Distribucions predictives posteriors	57

- Tria de models	58
- Control de qualitat en la tria de models	59
- Inferència dels paràmetres del model més probable	60
- Control de qualitat en la inferència dels paràmetres: espectre de freqüències desdoblament simulat <i>a posteriori</i>	60
<b>M3.4 Estima de l'edat de la inversió <math>J_1</math> mitjançant una anàlisi d'ABC</b>	60
<b>M3.5 Detecció d'una regió candidata de ser diana de la selecció</b>	64
<b>MS4. MATERIAL SUPLEMENTARI</b>	65
<b>MS4.1 Encebadors d'amplificació i seqüenciació</b>	65
<b>MS4.2 Condicions de PCR</b>	69
<b>RESULTATS</b>	73
<b>R1. OBTENCIÓ DE LÍNIES HOMOCARIOTÍPIQUES PER AL CROMOSOMA J DE <i>D. SUBOBSCURA</i> A PARTIR D'UNA POBLACIÓ NATURAL DE BARCELONA</b>	75
<b>R2. CARACTERITZACIÓ DE LES REGIONS LOCALITZADES AL LLARG DEL CROMOSOMA J DE <i>D. SUBOBSCURA</i></b>	77
<b>R3. ESTUDI DELS NIVELLS I PATRONS DE VARIACIÓ A <i>D. SUBOBSCURA</i> EN REGIONS NO CODIFICADORES LOCALITZADES AL LLARG DEL CROMOSOMA J I NO AFECTADES PER INVERSIONS CROMOSÒMIQUES</b>	81
<b>R3.1 Nivells de variació nucleotídica</b>	81
<b>R3.2 Patrons de variació nucleotídica</b>	83
- Anàlisi dels patrons de polimorfisme mitjançant els estadístics de classe I	84
- Anàlisi dels patrons de polimorfisme mitjançant els estadístics de classe II	86

<b>R3.3 Polimorfisme i divergència</b>	<b>88</b>
- $\pi/K$	<b>88</b>
- Test d'HKA	<b>89</b>
<b>R4. ANÀLISI DE LA VARIACIÓ EN REGIONS NO CODIFICADORES AFECTADES PER LA INVERSIÓ <math>J_1</math> DE <i>D. SUBOBSCURA</i></b>	<b>91</b>
<b>R4.1 Nivells de variació nucleotídica</b>	<b>92</b>
- $\pi/K$	<b>94</b>
<b>R4.2 Patrons de variació nucleotídica</b>	<b>94</b>
<b>R4.3 Comparació dels nivells i patrons de polimorfisme en regions localitzades dins i fora de la inversió <math>J_1</math></b>	<b>96</b>
<b>R4.4 Diferenciació genètica entre les ordenacions <math>J_1</math> i <math>J_{st}</math></b>	<b>98</b>
<b>R4.5 Conversió gènica</b>	<b>100</b>
<b>R4.6 Desequilibri de lligament</b>	<b>103</b>
<b>R4.7 Genealogia de la mostra de seqüències</b>	<b>104</b>
<b>R5. INFERÈNCIES DEMOGRÀFIQUES EN L'ESPÈCIE <i>D. SUBOBSCURA</i></b>	<b>106</b>
<b>R5.1 Estimació del temps de divergència entre diversos llinatges</b>	<b>106</b>
- Model de substitució	<b>106</b>
- Temps de divergència	<b>107</b>
<b>R5.2 Història demogràfica recent de <i>D. subobscura</i> estimada a través de metodologies d'<i>Approximate Bayesian Computation</i></b>	<b>108</b>
- Disseny de la distribució <i>a priori</i> pel paràmetre poblacional de mutació ( $\theta$ )	<b>108</b>
- Efecte de les mutacions recurrents en l'espectre de freqüències desdoblats	<b>109</b>
- Estadístics simulats	<b>111</b>

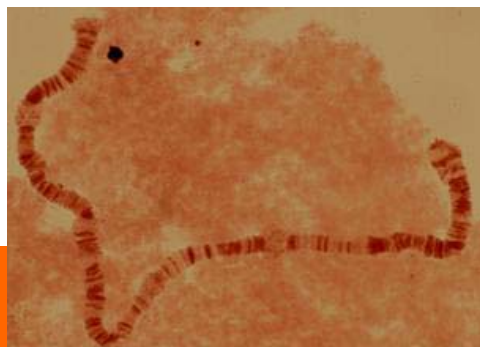
- Tria d'estadístics	114
- Distribucions predictives posteriors	115
- Tria de models	117
- Control de qualitat en la tria de models	119
- Inferència dels paràmetres del model més probable	120
- Controls de qualitat en les inferències dels paràmetres del model més probable	122
<b>R5.3 Estima de l'edat de la inversió <math>J_1</math> mitjançant una anàlisi d'ABC</b>	<b>124</b>
- Control de qualitat en la inferència de l'edat de la inversió $J_1$ a través de validacions predictives posteriors	125
<b>R5.4 Detecció d'una regió candidata de ser diana de la selecció</b>	<b>126</b>
<b>RS6. MATERIAL SUPLEMENTARI</b>	<b>128</b>
- <b>RS6.1 Taules de polimorfisme de les regions localitzades al llarg del cromosoma J i no afectades per la inversió <math>J_1</math></b>	<b>128</b>
- <b>RS6.2 Taules de polimorfisme de les regions localitzades dintre de la inversió <math>J_1</math></b>	<b>148</b>
- <b>RS6.3 Distribucions posteriors dels dos models menys probables</b>	<b>158</b>
- Model de coll d'ampolla (M3)	158
- Model de barreja (M4)	160
<b>DISCUSSIÓ</b>	<b>163</b>
<b>D1. OBTENCIÓ DE LÍNIES HOMOCARIOTÍPIQUES PER AL CROMOSOMA J DE <i>D. SUBOBSCURA</i></b>	<b>165</b>
<b>D2. ESTUDI DELS NIVELLS I PATRONS DE VARIACIÓ A <i>D. SUBOBSCURA</i> EN REGIONS NO CODIFICADORES LOCALITZADES AL LLARG DEL CROMOSOMA J I NO AFECTADES PER INVERSIONS CROMOSÒMIQUES</b>	<b>167</b>

D2.1 Nivells i patrons de variació	167
<b>D3. INFERÈNCIES DEMOGRÀFIQUES EN L'ESPÈCIE <i>D. SUBOBSCURA</i></b>	<b>170</b>
D3.1 Temps de divergència	170
D3.2 Metodologies d' <i>Approximate Bayesian Computation</i> : pros i contres	171
- Informació prèvia disponible en relació al paràmetre poblacional de recombinació	171
- Tria d'estadístics resum en la tria de models	172
- Validació de la tria de models	174
- Avaluació de diferents metodologies en les estimes de les distribucions posteriors dels paràmetres del model demogràfic més probable	174
D3.3 Detecció d'una regió possiblement afectada per un esdeveniment selectiu recent	178
D3.4 Història demogràfica de <i>D. subobscura</i>	179
- Influència dels canvis climàtics ocorreguts durant els últims 200.000 anys	179
- Temps de generació i grandària efectiva	185
- Termotolerància	185
<b>D4. ANÀLISI DE LA VARIACIÓ EN REGIONS NO CODIFICADORES AFECTADES PER LA INVERSIÓ <math>J_1</math> DE <i>D. SUBOBSCURA</i></b>	<b>188</b>
D4.1 Nivells i patrons de variació	188
D4.2 Diferenciació genètica entre ordenacions	195
D4.3 Estimació de l'edat de la inversió $J_1$	200
<b>CONCLUSIONS</b>	<b>207</b>
<b>BIBLIOGRAFIA</b>	<b>213</b>









# INTRODUCCIÓ



## **I1. DROSOPHILA COM A ORGANISME MODEL**

Les espècies del gènere *Drosophila* han prevalgut com a espècies model clau en els grans avenços que han caracteritzat a la genètica al llarg dels darrers 100 anys - des dels primers estudis amb al·lels mutants (Bridges i Morgan 1910), als posteriors mapes d'ordenació dels gens al llarg dels cromosomes (Sturtevant 1917), els estudis de regulació gènica (Bender et al. 1983), la seqüenciació del primer genoma complet (Adams et al. 2000) i el desenvolupament de la proteòmica (Giot et al. 2003) i de la genòmica comparada (Richards et al. 2005; Clark et al. 2007). De fet, les espècies de *Drosophila* posseeixen en general una sèrie de característiques, tals com un temps de generació curt, un elevat nombre de descendents per parella i una fàcil manipulació en el laboratori, que les fan ideals per a la realització d'aquests estudis. S'han escrit monografies i volums sencers descrivint aquestes espècies (Lindsey i Zimm 1992; Krimbas, 1993; Powell, 1997; Sullivan et al. 2000; Ashburner et al. 2005, entre d'altres). El nombre de publicacions científiques on es cita el gènere *Drosophila* ha crescut exponencialment i s'estima que hi ha més de 1000 grups de recerca arreu del món que continuen utilitzant les espècies del gènere *Drosophila* com a organisme model (Clark et al. 2003).

## **I2. EVOLUCIÓ MOLECULAR**

### **I2.1 Variabilitat genètica**

Es podria definir l'evolució, en un sentit ampli, com el conjunt de canvis que es donen en les característiques dels organismes en períodes llargs de temps. La presència de variabilitat genètica és un requisit perquè els organismes puguin evolucionar, ja que únicament les modificacions de les característiques amb una base genètica són les que es transmeten d'una generació a la següent. De fet, els organismes en les poblacions naturals presenten diferències genètiques tant a nivell intraespecífic (polimorfisme) com interespecífic (divergència). Un dels objectius principals de la genètica de poblacions ha estat i segueix sent l'estudi de la variabilitat genètica, per intentar entendre tant el seu significat evolutiu com les forces que la influeixen i la modelen. Els mecanismes principals que generen variabilitat genètica a nivell de les seqüències de nucleòtids de DNA són: la mutació (errors en el procés de replicació o de reparació del DNA i d'altres processos que donarien lloc a substitucions nucleotídiques, delecions i insercions) i la recombinació (generació de noves combinacions de variants preexistents). Entre les forces evolutives que modelen aquesta variabilitat genètica trobem la selecció natural que dóna lloc a canvis adaptatius i la deriva genètica que és un procés estocàstic que genera canvis aleatoris no adaptatius, tenint un major efecte quant menor és la grandària efectiva poblacional. Cal remarcar que els canvis demogràfics (migració, colonització, creixement poblacional, extinció, estructuració poblacional, etc.) també actuen en el manteniment o la pèrdua de la variabilitat genètica i, per tant, s'han de tenir en compte alhora d'inferir els mecanismes que modelen la variabilitat genètica de les poblacions.

### **I2.2 Detecció de la variabilitat i significat evolutiu**

Prèviament al desenvolupament de les tècniques moleculars, les estimes de variabilitat genètica només es podien obtenir a partir de les diferències morfològiques, cromosòmiques o en l'eficàcia biològica dels individus de les poblacions. Els estudis de Theodosius Dobzhansky

amb *D. pseudoobscura*, on descriu que el 10% dels cromosomes són letals en homozigosi, són pioners en els estudis de la variació genètica (Dobzhansky 1933). Al 1937, amb la publicació del llibre "Genetics and the Origin of Species", Dobzhansky dóna un gran impuls a la genètica de poblacions experimental basant-se en la teoria sintètica de l'evolució que unifica la selecció natural proposada per Darwin amb la genètica mendeliana. La invenció als anys 60 de la tècnica d'electroforesi de proteïnes permet el desenvolupament de l'estudi de la variació proteica i es poden obtenir per primera vegada bones estimes de la variabilitat genètica de les poblacions (Lewontin et al. 1966; Harris 1966). Kimura al 1968, publica les bases de la teoria de l'evolució molecular coneguda com "neutralisme", la qual proposa que una gran proporció de la variació molecular existent dintre de les poblacions és neutra en front de la selecció natural i que la major part de les substitucions nucleotídiques que es produeixen durant l'evolució són degudes a la deriva genètica i no a la selecció. La seqüenciació directa de DNA, juntament amb el desenvolupament de noves tècniques tals com els mapes de restricció a través de la utilització d'enzims de restricció i la visualització en gels d'agarosa o de poliacrilamida (Langley et al. 1982; Kreitman i Aguadé 1986), donen un impuls important a l'estudi de la variabilitat nucleotídica de les poblacions. Als anys vuitanta, Kreitman publica un treball pioner en l'estudi de la variabilitat genètica directament a nivell de seqüències de DNA, concretament al locus *Adh* de *D. melanogaster* (Kreitman 1983). En l'actualitat, la genètica de poblacions teòrica juntament amb els grans avenços experimentals de les tècniques moleculars proporcionen el marc de referència ideal per el desenvolupament de mètodes robusts per l'anàlisi i la interpretació del polimorfisme de DNA en les poblacions i les espècies. Aquests mètodes es basen tant en estadístics descriptius, l'objectiu dels quals és quantificar els nivells i patrons de polimorfisme, com en els tests de neutralisme els quals utilitzen la teoria neutralista com a hipòtesi nul·la per detectar desviacions d'aquest model produïdes per l'acció de la selecció

natural o per canvis demogràfics (Hudson et al. 1987; Tajima 1989; McDonald i Kreitman 1991; Fay i Wu 2000).

### **12.3 Història evolutiva de les poblacions: factors selectius i demogràfics**

Es podria definir la història evolutiva de les poblacions, en un sentit general, com tots aquells factors que han influenciat i modelat la variabilitat genètica de les poblacions fins l'actualitat. Inferir la història evolutiva de les poblacions a través de la variabilitat de les seqüències de DNA és un dels objectius principals de la genètica de poblacions molecular, englobant tant estudis que pretenen detectar la petjada deixada per la selecció natural com estudis que pretenen desentrellar els principals factors demogràfics que han patit les poblacions al llarg de la seva història. Dues teories han estat clau en el desenvolupament de models teòrics per a la realització d'inferències evolutives. Aquestes són la teoria neutralista de l'evolució molecular (Kimura 1980) abans esmentada i el desenvolupament posterior de la teoria de la coalescència (Kingman 1982; Hudson 1991).

El model neutre estàndard de l'evolució molecular considera que la major part de les mutacions que contribueixen a la divergència i al polimorfisme nucleotídic són neutres o pràcticament neutres i que la fracció de mutacions beneficioses que hi contribueixen és tan petita que es pot menysprear (Kimura 1968, 1983). Aquest model també accepta l'existència de mutacions deletèries i, per tant, que el tipus de selecció que preval d'acord al neutralisme és la selecció purificadora, tot i que aquesta no comporti canvis evolutius. No obstant això, la selecció natural pot actuar de diverses formes no només contribuint a la pèrdua de variabilitat eliminant mutacions deletèries (selecció purificadora). Per una banda, la selecció natural pot actuar augmentant la freqüència d'una mutació beneficiosa en la població fins que arriba a fixar-la, sent el cas de la selecció direccional positiva. L'acció de la selecció positiva es veu associada al fenomen d'escombrada selectiva, el qual consisteix en que les mutacions neutres adjacents a la mutació beneficiosa també augmenten de freqüència en la població acabant-se

per fixar (Maynard i Haigh 1974). Després d'aquesta escombrada selectiva, sorgeixen noves mutacions que donaran lloc a un patró de variació nucleotídica marcat per un excés de polimorfismes amb variants a baixa freqüència. La regió afectada per l'escombrada selectiva dependrà de la magnitud de la selecció i de la freqüència de recombinació en aquella regió. Per una altra banda, la selecció natural també pot actuar mantenint més d'una variant en la població (selecció equilibradora) fet que es reflecteix en uns patrons de variació amb un excés de variants a freqüències intermèdies. A nivell experimental, una de les estratègies que es segueix per intentar detectar la petjada de la selecció natural és la d'utilitzar gens candidats, és a dir, gens que per les seves funcions són rellevants per l'adaptació dels organismes (per exemple, Aguadé et al. 1992; Aguadé 1999; Guirao-Rico et al. 2009). Aquesta estratègia va ser modificada posteriorment en els estudis multilocus en els que s'analitza la variació nucleotídica a múltiples regions genòmiques independentment de la seva funció. En aquests estudis primer es detecta la petjada d'un possible esdeveniment selectiu en la variabilitat nucleotídica d'una regió i després s'intenta identificar quin és el gen candidat responsable (Glinka et al. 2003; Orengo i Aguadé 2004, 2010; Ometto et al. 2005; Pavlidis et al. 2012). En l'actualitat, els avenços associats als projectes genòmics han fet possible l'anàlisi de la variabilitat nucleotídica a nivell de tot el genoma (Begun et al. 2007; Turner et al. 2008; Kulathinal et al. 2009; Kolaczowski et al. 2010; Langley et al. 2012). Aquests estudis de rastreig genòmic continuen amb la línia iniciada amb els estudis multilocus.

Sota el model neutre de l'evolució molecular, i considerant que les poblacions són panmíctiques i es troben en equilibri, els paràmetres poblacionals de recombinació ( $R=4 N_e r$ ) i de mutació ( $\theta=4 N_e \mu$ ) (sent  $N_e$  la grandària efectiva de l'espècie,  $r$  la taxa de recombinació per parell de bases per generació i  $\mu$  la taxa de mutació per nucleòtid per generació) poden ser estimats a través de les dades del polimorfisme de les seqüències de DNA (Hudson 1991; Eanes et al. 1996; Andolfatto et al. 2001). No obstant, s'ha de tenir en compte que les

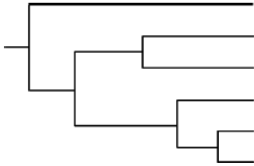
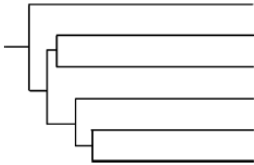
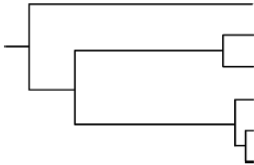


poblacions naturals no són sistemes estàtics sinó que es veuen influenciades pel medi natural que les envolta, de manera que al llarg del temps poden patir fluctuacions en la seva grandària efectiva (expansions, reduccions poblacionals, fenòmens migratoris, etc.) (Li et al. 2006; Thornton et al. 2006). Els períodes glacials i interglacials ocorreguts durant el Pleistocè i l'Holocè són un clar exemple de que variacions climàtiques poden produir canvis en la grandària de les poblacions (Hewitt 2000, 2004; Schmitt 2007). Aquests canvis demogràfics afectaran a la variabilitat genètica de les poblacions determinant que aquestes no estiguin en equilibri. La teoria de la coalescència (Kingman 1982) proporciona mètodes que permeten no tan sols estimar paràmetres poblacionals tals com els de mutació, recombinació i migració, sinó també detectar canvis en la grandària efectiva en funció de la genealogia dels gens. D'aquesta manera, el neutralisme juntament amb la teoria de la coalescència ha permès, gràcies a mètodes de màxima versemblança i mètodes bayesians, el desenvolupament d'estadístics que detecten desviacions de la neutralitat, la realització de simulacions a través de mètodes computacionals, la inferència de models evolutius i l'estimació de paràmetres poblacionals. S'ha d'afegir, que els mètodes basats en aproximacions bayesianes són dels més poderosos i flexibles alhora d'estimar paràmetres demogràfics (Tavaré et al. 1997; Pritchard et al. 1999; Beaumont et al. 2002). La particularitat dels mètodes d'ABC (*Approximate Bayesian Computation*) és que permeten obtenir la probabilitat posterior dels models i la distribució posterior dels paràmetres que componen aquests models sense la necessitat d'avaluar analíticament les funcions de probabilitat, que esdevenen pràcticament impossibles de calcular quan els models demogràfics són complexes.

En referència als estadístics que tenen com a objectiu detectar desviacions del model neutre estàndard, d'ara en endavant referit com a SNM o condicions de neutralitat, es poden classificar àmpliament en tres grups (Achaz 2009): els basats en l'espectre de freqüències de les mutacions (Tajima 1989; Fu i Li 1993; Fay i Wu 2000), els que ho fan en la distribució dels

haplotips (Fu 1997; Depaulis i Veuille 1998) i els que intenten capturar deformacions específiques dels arbres genealògics (*tree shape tests*) (Ramos-Onsins i Rozas 2002). Diferents estadístics són sensibles a diferents violacions del model neutre estàndard. Per exemple, l'estadístic *D* de Tajima (Tajima 1989) és sensible per detectar polimorfismes amb variants a baixa freqüència mentre que l'estadístic *H* normalitzat (*H*<sub>norm</sub>) de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) ho és per les variants derivades a elevada freqüència. Per això, alhora de realitzar inferències evolutives, és important recollir a partir de les dades un conjunt d'estadístics suficientment ampli perquè pugui explicar la variabilitat nucleotídica de l'espècie o les espècies en estudi. D'aquesta manera, la topologia dels arbres genealògics pot ser indicativa dels factors que han conformat la història evolutiva de les espècies (Taula I1).

**Taula I1.** Diferents escenaris demogràfics en funció de la topologia dels arbres genealògics obtinguts per simulacions de coalescència. Modificat de Ramirez-Soriano et al. (2008).

Escenari	Característiques	Topologia de l'arbre
<b>Neutralitat</b>	Els arbres tenen topologies neutrals regulars.	
<b>Expansió</b>	Els arbres tenen una topologia amb branques externes llargues en relació a les internes donant lloc a l'acumulació d'un nombre de mutacions a baixa freqüència més elevat de l'esperat. Els estadístics basats en l'espectre de freqüències acostumen a ser negatius.	
<b>Contracció</b>	Els arbres tenen branques internes llargues en relació a les externes donant lloc a un nombre de mutacions a freqüències intermèdies més elevat de l'esperat. Els estadístics basats en l'espectre de freqüències acostumen a ser positius.	

Hi ha certes situacions, però, en què els factors demogràfics i els selectius poden donar una topologia de l'arbre genealògic similar. Per exemple, un procés d'escombrada selectiva o d'expansió poblacional estarà representat en la genealogia per un arbre estrella amb branques externes llargues en relació a les internes donant lloc a un excés de mutacions recents. Al

contrari, un procés de subdivisió poblacional o selecció equilibradora provocarà que les branques externes de la genealogia siguin curtes en relació a les internes donant lloc a un excés de mutacions antigues. La realització d'estudis multilocus és un bon plantejament per diferenciar els possibles efectes selectius dels demogràfics ja que s'espera que els canvis demogràfics afectin de manera similar a totes les regions del genoma, mentre que els canvis selectius estiguin localitzats en regions concretes (Orenge i Agudé 2004; Schmid et al. 2005; Hadrill et al. 2005; Li i Stephan 2006).

Una bona metodologia per a l'estudi dels canvis demogràfics que han conformat la història evolutiva de les espècies és la d'escollir regions del genoma que en principi no hagin estat sotmeses a pressions selectives, sent el DNA no codificador un bon candidat per a aquest tipus d'estudis. Les regions no codificadores del genoma engloben totes aquelles seqüències de DNA que no codifiquen per proteïnes, és a dir formen part d'aquesta categoria els introns, UTRs i les regions intergèniques. A finals dels anys 60 es descobreix que la major part del genoma no codifica per proteïnes (Britten et al. 1968), considerant-se aquest DNA no codificador com a DNA porqueria (Ohno 1972) i veient-se el seu estudi com una pèrdua de temps i diners (Lewin 1986; Robertson 1986). Clàssicament, el DNA no codificador s'havia considerat mancat de funció biològica i per tant, al no tenir limitacions funcionals, la seva evolució es considerava neutra, és a dir, que no estava afectada per processos de selecció natural. Recentment, però, han sorgit un nombre important d'estudis multilocus i de genòmica comparada que examinen els nivells i patrons de variació del DNA no codificador en varies espècies de *Drosophila* i en altres organismes (Khon et al. 2004; Andolfatto 2005; Gaffney et al. 2006; Hadrill et al. 2008; Sella et al. 2009; Parsch et al. 2010). Aquests estudis, en contra de la visió clàssica, suggereixen que el DNA no codificador pot estar sotmès a limitacions funcionals i, per tant, que aquest DNA pot estar afectat per pressions selectives. De fet, si les regions no codificadores tenen un paper important en la regulació dels gens i si les mutacions que

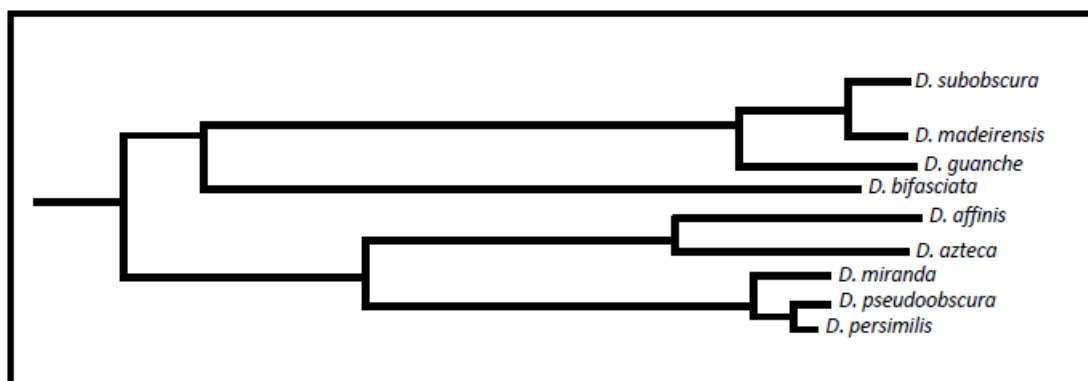
sorgeixen *de novo* produeixen un efecte perjudicial per la funció d'aquests gens, aquestes regions tendiran a evolucionar més lentament en estar sotmeses a selecció purificadora (Andolfatto 2005). S'ha d'afegir que el projecte *ENCODE* (*Encyclopedia of DNA Elements*) (*ENCODE Project Consortium* 2004, Dunham et al. 2012) el qual desenvolupa un catàleg de tots els elements que es poden trobar en el genoma humà ha canviat de forma definitiva la visió que es tenia sobre les regions no codificadores del genoma i sobre la seva possible influència en processos de regulació. No obstant, aquest canvi de visió ha estat la base d'un ampli debat (Graur et al. 2013). Al projecte *ENCODE* s'han d'afegir, entre d'altres, els projectes *modENCODE* en relació als organismes *D. melanogaster* (*modENCODE Consortium* 2010) i *Caenorhabditis elegans* (*C. elegans Sequencing Consortium* 2012) els quals recullen dades relacionades amb l'estructura gènica del genoma, l'expressió dels gens i la regulació de la cromatina entre d'altres al llarg de totes les etapes del desenvolupament.

Tot i així, les regions no codificadores continuen sent bones candidates per a la realització d'estudis demogràfics. Si s'estudia la variabilitat nucleotídica a nivell de múltiples loci i aquesta variabilitat és similar en totes les regions d'estudi, molt possiblement indicarà que els nivells i patrons de polimorfisme observats no són deguts a fenòmens de selecció natural. A més a més, en el cas que aquesta variabilitat es desvii significativament del model neutre estàndard, es podrà considerar, que la història demogràfica de l'espècie estarà influenciant la diversitat genètica de la població en estudi.

### 13. DROSOPHILA SUBOBSCURA

#### 13.1 Taxonomia de l'espècie

*Drosophila subobscura* va ser descrita el 1936 per J. Collin i està àmpliament distribuïda per la regió paleàrtica. Aquesta espècie es podria considerar l'equivalent europeu de *D. pseudoobscura* en els estudis de genètica evolutiva donada la seva gran abundància en molts països d'Europa i el seu ric polimorfisme cromosòmic. *D. subobscura* és una espècie de dípter que pertany a la família *Drosophilidae*, al gènere *Drosophila*, al subgènere *Sophophora*, al grup *obscura* i al subgrup *obscura* del vell món. A. H. Sturtevant va ser el primer en definir la taxonomia de la família *Drosophilidae* l'any 1921, una família composta per més de 3200 espècies, 2000 de les quals pertanyen al gènere *Drosophila*. El gènere *Drosophila* es divideix en tres subgèneres: *Sophophora*, *Drosophila* i *Idiomyia*. Dintre del subgènere *Sophophora*, el grup *obscura* està format per més de 35 espècies. Clàssicament Sturtevant (1942) divideix el grup *obscura* en dos subgrups *l'affinis* i *l'obscura*, tot i que recentment s'hagi canviat la nomenclatura del grup *obscura*, dividint-se en dos subgrups: subgrup *obscura* del vell món i subgrup *pseudoobscura* del nou món (Figura I1).



**Figura I1.** Arbre filogenètic basat en la divergència nucleotídica de la regió *rp49* en diferents espècies del grup *obscura*. Modificat de Ramos-Onsins et al. (1998).

### 13.2 Distribució geogràfica

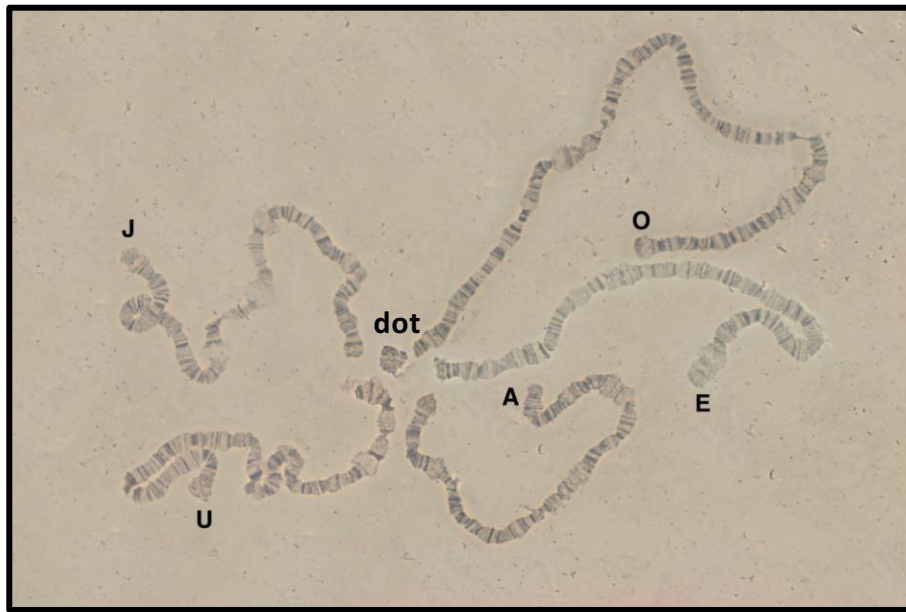
*D. subobscura* es distribueix per tota Europa excepte Islàndia i la part central i septentrional d'Escandinàvia. També es troba a les illes Canàries, Açores i Madeira, i s'ha observat al nord d'Àfrica fins al Sàhara i al pròxim orient fins a Iran. La seva distribució per l'antiga Unió Soviètica no està ben establerta. Recentment ha colonitzat el continent Americà. Hi ha registres de la seva presència a Xile a l'any 1978 (Brncic et al. 1981) i, posteriorment al 1982 es detectà a la costa Oest Nord-americana (Beckenbach i Prevosti 1986).

### 13.3 Biologia de l'espècie

*D. subobscura* és un insecte holometàbol amb un cicle biològic caracteritzat pels estadis d'embrió, 3 estadis larvals, pupa i adult. Tot i que no es coneix clarament la seva ecologia, es considera una espècie polífaga i generalista, sent la seva font principal d'alimentació els llevats i els bacteris que fermenten matèria orgànica vegetal. A la natura es troba a la vora dels marges dels boscos de *Quercus*, *Abies*, *Pinus*, *Castanea*, *Ulmus*, etc. El límit nord de la seva àrea de distribució a Europa coincideix amb el de *Quercus robur*. És capaç de viure en altituds que varien des del nivell del mar fins a més de 2000 metres (Krimbas 1993). La densitat poblacional és superior en les àrees centrals de la distribució i varia en funció de les estacions, depenent de les condicions climàtiques, especialment de la temperatura. En condicions òptimes de laboratori i a 18<sup>o</sup> C el seu cicle biològic és aproximadament de 20 dies (Orengo i Prevosti 1994).

### 13.4 Els cromosomes de *D. subobscura*

El cariotip de *D. subobscura* consta de 6 parells de cromosomes que es divideixen en 5 parells d'acrocèntrics anomenats A, J, U, E i O i un parell de puntiformes (Mainx et al. 1953). Els cromosomes J, U, E i O són els autosomes i l'A és el cromosoma sexual (Figura I2).



**Figura I2.** Fotografia dels 6 cromosomes (5 acrocèntrics i el puntiforme) en una preparació de cromosomes politènics d'un individu de *D. subobscura*. Cedida per M. Papaceit.

Muller (1940) i Sturtevant i Novitski (1941) proposen que el cariotip bàsic ancestral del gènere *Drosophila* ha d'estar format per cinc cromosomes acrocèntrics llargs i un cromosoma puntiforme. A més a més, basant-se en estudis comparatius observen que les diferents espècies del gènere *Drosophila* tendeixen a tenir mutants fenotípics similars en els mateixos grups de lligament. Això els va permetre establir un sistema de nomenclatura comú per tot el gènere en el qual els diferents cromosomes o braços cromosòmics es classifiquen en sis elements que van anomenar elements de Muller A-F (Taula I2). Anàlisis comparatives posteriors de l'organització dels cromosomes, a nivell de sintènia i d'ordre dels gens, corroboren que el contingut gènic dintre dels elements de Muller està altament conservat en les diferents espècies del gènere *Drosophila* (Segarra et al. 1992, 1995, 1996; Ranz et al. 1997, 2001, 2003; Papaceit et al. 2006; Schaeffer et al. 2008; Powell et al. 2010). A més a més, aquests estudis demostren que els intercanvis genètics entre els diferents elements produïts per inversions pericèntriques, translocacions o transposicions són molt rars (no obstant, veure Conceição i Aguadé 2010) i que les inversions paracèntriques són un dels mecanismes principals pel qual l'ordre dels gens pot ser alterat dintre dels diferents elements de Muller.

**Taula 12.** Homologies entre els elements de Muller i els cromosomes o braços cromosòmics de *D. subobscura*, *D. melanogaster* i *D. pseudoobscura*.

Element de Muller	<i>D. subobscura</i>	<i>D. melanogaster</i>	<i>D. pseudoobscura</i>
<b>A</b>	A	X	XL
<b>B</b>	U	2L	4
<b>C</b>	E	2R	3
<b>D</b>	J	3L	XR
<b>E</b>	O	3R	2
<b>F</b>	D	4	5

### 13.5 Polimorfisme cromosòmic

Sturtevant (1917) és el primer que descriu l'existència de les inversions cromosòmiques en observar la supressió de la recombinació en heterozigots en l'espècie *D. melanogaster*. En l'estudi de les inversions han jugat un paper clau el descobriment dels cromosomes politènics (Painter 1933) i les posteriors anàlisis de la variació cromosòmica en poblacions naturals de *D. pseudoobscura* i *D. persimilis* realitzades per Dobzhansky (Sturtevant i Dobzhansky 1936a, 1936b; Dobzhansky 1937). Una inversió es defineix com un tipus d'alteració estructural dels cromosomes en la qual un segment cromosòmic canvia d'orientació a la mateixa posició dintre del propi cromosoma. Les inversions cromosòmiques poder ser pericèntriques o paracèntriques en funció de si el segment invertit engloba o no el centròmer. Tres quartes parts de les espècies del gènere *Drosophila* són polimòrfiques per inversions paracèntriques detectables a nivell citològic (Krimbas 1992). En el subgènere *Sophophora*, les taxes de fixació de les inversions paracèntriques no són iguals per tots els elements de Muller. La major taxa de fixació es troba a l'element de Muller A (el cromosoma X), mentre que s'observen taxes intermèdies pels elements D i E i la més baixa és per l'element B (Papaceit et al. 2006).

El polimorfisme cromosòmic en l'espècie *D. subobscura* ha estat àmpliament analitzat, sent els estudis realitzats per Stumm-Zollinger (1953), Mainx et al. (1953) i Kunze-Mühl i Müller (1958) els pioners en aquest camp. *D. subobscura* es caracteritza per presentar un elevat



polimorfisme per inversions cromosòmiques en poblacions naturals. S'han caracteritzat més de 60 inversions diferents que formen més de 85 ordenacions (Krimbas i Powell 1992; Powell 1997). No es coneixen els mecanismes responsables de l'establiment del polimorfisme cromosòmic per inversions en la naturalesa, tot i que s'hagin suggerit tres models selectius diferents i no excloents entre ells: coadaptació (Dobzhansky 1950), selecció de complexos gènics adaptats (Wasserman 1968) i adaptació local (Kirkpatrick i Barton, 2006) (per revisió, Santos 2009). El caràcter adaptatiu del polimorfisme per inversions cromosòmiques a *D. subobscura* està ben establert, sent un clar exemple l'associació entre el polimorfisme cromosòmic i la longitud de l'ala (Orengo i Prevosti 1999; Huey et al. 2000) o també les clines latitudinals presents en l'àrea de distribució de l'espècie (Prevosti et al. 1985, 1988; Krimbas i Powell 1992; Mestres et al. 2001). De fet, hi ha estudis que demostren que la presència de clines latitudinals podria estar fortament relacionada amb factors ambientals i ecològics, especialment els climàtics (Prevosti et al. 1988, 1990; Krimbas i Powell 1992), observant-se també variacions en la freqüència de diferents ordenacions en funció de les temperatures de les diferents estacions anuals (Fontdevila et al. 1983; Frutos i Prevosti, 1984; Orengo i Prevosti 1996; Santos et al. 2005). La colonització d'Amèrica per *D. subobscura* ha estat un experiment natural únic que ha permès corroborar el caràcter adaptatiu de les inversions cromosòmiques, ja que les clines latitudinals presents a Europa per determinades ordenacions s'han reproduït en el mateix sentit a Amèrica (Prevosti et al. 1985, 1988; Ayala et al. 1989). A més a més, hi ha algunes evidències que indiquen que el polimorfisme cromosòmic per inversions a *D. subobscura* s'està adaptant al canvi climàtic actual, en concret a l'increment de les temperatures, ja que s'ha observat un augment de les ordenacions típiques de regions càlides (Balanyà et al. 2006; Rodríguez-Trelles et al. 2007). Les evidències de selecció pel polimorfisme de les inversions s'han detectat no només a *D. subobscura* i a d'altres espècies de *Drosophila*

(Powell 1997) sinó també a altres dípters (Ayala et al. 2011; Rottschaefer et al. 2011) i a organismes com el ratolí (Lyon 2003) i els humans (Stefansson et al. 2005, Steiberg et al. 2012).

Hi ha diferents estudis que mostren que les inversions són monofilètiques (Rozas i Aguadé 1990, Andolfatto et al. 1999). De fet, si es considera que els punts de trencament dels cromosomes es produeixen a l'atzar, la probabilitat de que dues inversions diferents abarquin exactament el mateix fragment de cromosoma és molt baixa i, tot i que fos així, encara seria més baixa la probabilitat de que les dues inversions s'arribessin a establir en la població. S'han descrit varis models per explicar l'origen de les inversions. En el model que podríem definir com a clàssic (Krimbas i Powell 1992), la formació de noves inversions és deguda a dues ruptures aleatòries i simultànies en un mateix cromosoma i una posterior reparació errònia del fragment. Ranz et al. (2007) hipotetitzen que en realitat les inversions es poden originar per ruptures desiguals de les dues cadenes de DNA generant-se extrems protuberants de cadena senzilla. Les reparacions errònies posteriors estan per tant associades a la duplicació amb orientació invertida de regions genòmiques localitzades a la vora dels punts de trencament (Guillén i Ruiz 2012; Papaceit et al. 2013). Per últim s'ha descrit com a mecanisme generador de noves inversions la recombinació ectòpica que es produiria entre seqüències homòlogues invertides localitzades en diferents posicions del cromosoma (Lim i Simons 1994), havent evidències sobre la implicació dels elements transponibles en aquest mecanisme (Cáceres et al. 1999, 2001; Casals et al. 2005; Marzo et al. 2008).

### **13.6 Variabilitat nucleotídica associada al polimorfisme cromosòmic**

Si una nova inversió confereix un avantatge selectiu als individus portadors, l'acció de la selecció positiva podria portar ràpidament a que la nova ordenació arribi a una freqüència d'equilibri. Com a conseqüència d'aquest increment tan ràpid de freqüència, totes les regions incloses dintre de la inversió podrien tenir un nivell de variació nucleotídica pràcticament nul just després del seu establemient. És a dir, l'establiment d'una nova inversió es podria

contemplar com un procés d'escombrada selectiva parcial (Maynard-Smith i Haigh 1974) amb una important pèrdua de variació en els cromosomes invertits al llarg de la regió afectada per la inversió.

Posteriorment, la variabilitat nucleotídica de les inversions s'anirà recuperant lentament per la introducció de noves variants per mutació. La mutació serà l'únic mecanisme generador de variabilitat si el bloqueig de l'intercanvi genètic a la regió invertida és total en els heterozigots. Com més antiga sigui la inversió, la probabilitat de que hagin sorgit noves mutacions serà més gran, fet que farà que, en absència de recombinació, es vagin incrementant les diferències entre les ordenacions cromosòmiques i així la seva diferenciació genètica (Navarro et al. 2000). Per tant, no només és la generació de noves mutacions el que jugarà un paper destacat en la variabilitat nucleotídica de les inversions, sinó que també és molt important el paper del bloqueig de la recombinació.

La forta reducció de la recombinació en la regió invertida en individus heterocariotípics pot ser deguda a diferents causes. Si en un heterozigot es produeix un entrecreuament senzill en la regió invertida es generen gàmetes recombinants amb cromosomes desequilibrats que no són funcionals. Per altra banda, la recombinació està reduïda en els heterocariotips per la dificultat d'aparellament o sinapsi entre els cromosomes homòlegs principalment en regions pròximes als punts de trencament de les inversions (per una revisió, Roberts 1976). Aquesta inhibició de la recombinació en les regions afectades per inversions cromosòmiques, fa que a nivell poblacional es puguin diferenciar els cromosomes en dues classes: els portadors de les ordenacions estàndard i els portadors de les ordenacions amb la inversió. Les inversions cromosòmiques, per tant, protegeixen llargues regions del genoma del flux genètic degut a una forta reducció de la recombinació en les regions invertides, fet que podria facilitar el procés d'especiació (Noor et al. 2001; Navarro et al. 2003; Rieseberg et al. 2001, Stevison et al. 2011, Faria et al. 2010).

Malgrat el fort bloqueig de la recombinació, l'aïllament genètic entre els dos tipus de cromosomes en la regió invertida no és total ja que hi ha mecanismes, com els dobles entrecreuaments que afectin a les mateixes cromàtides i la conversió gènica, que permeten un cert nivell de flux gènic entre cromosomes homòlegs dintre d'una inversió (Rozas i Aguadé 1994; Schaeffer et al. 2005; Machado et al. 2007; McGaugh et al. 2012; Cheng et al. 2012). Estudis teòrics prediuen com aquest intercanvi genètic no té per què ser homogeni al llarg de tota la inversió. Així, el flux gènic produït per conversió gènica estarà distribuït uniformement al llarg de tota la regió invertida menys en les regions molt properes al punt de trencament on podria ser inexistent, mentre que la freqüència dels dobles entrecreuaments serà més elevada en les regions centrals de la inversió. Sota aquest escenari, s'espera que la diferenciació genètica entre ordenacions cromosòmiques sigui superior en les regions properes als punts de trencament (Navarro et al. 1997; Laayouni et al. 2003). S'ha de tenir en compte també que l'intercanvi genètic variarà en funció de la grandària de la inversió, ja que es prediu que els dobles entrecreuaments només es produeixen en inversions més grans de 20 cM (Navarro et al. 2000).

L'elevat polimorfisme cromosòmic que presenta *D. subobscura* per inversions paracèntriques la fa ser una espècie ideal per l'estudi de com les inversions afecten la variació en les poblacions naturals. Els primers estudis empírics realitzats en *D. subobscura* que relacionen els nivells de polimorfisme per inversions i la variabilitat nucleotídica són a nivell de gens individuals, com per exemple el del gen *rp49* que es troba a la vora del punt de trencament de la inversió O<sub>3</sub> (Rozas i Aguadé 1991, 1994, 1999) i on per primera vegada es quantifica la conversió gènica entre ordenacions gràcies a l'algoritme desenvolupat per Bertrán et al. (1997). Altres exemples corresponen al gen *Acph1* (Navarro-Sabaté et al. 2003, 2003), *yellow* (Munté et al. 2000) i *Rpl215* (Llopart i Aguadé 2000). Posteriorment, es realitzen estudis multilocus en diferents elements de Muller els quals pretenen abastar tota la regió

afectada per la inversió en poblacions naturals amb ordenacions cromosòmiques diferents (Munté et al. 2005; Nóbrega et al. 2008). Els estudis de les inversions cromosòmiques a *D. subobscura* han permès corroborar que l'origen de les inversions és monofilètic i que en aquesta espècie el contingut gènic de les ordenacions està fortament diferenciat, tot i que s'observin fenòmens de flux gènic deguts a dobles entrecreuaments o conversió gènica. En general, s'observen nivells de variabilitat nucleotídica menors en l'ordenació derivada, detectant-se en alguns casos una major diferenciació genètica a la vora del punt de trencament (Nóbrega et al. 2008). Tot i així, i en contra del que prediuen els estudis teòrics, Munté et al. (2005) han mostrat, que els nivells de diversitat nucleotídica i de diferenciació genètica presenten una distribució bastant homogènia al llarg de la inversió  $O_3$  en comparar cromosomes  $O_{st}$  i  $O_{3+4}$ . Així doncs, les inversions cromosòmiques produeixen reestructuracions del genoma, afectant a les taxes d'especiació, a l'adaptació i a l'evolució dels cromosomes en poblacions naturals, fet que les ha convertit en un dels objectius principals en els estudis de genètica de poblacions i genètica evolutiva durant molts anys.

### 13.7 El sistema cromosòmic $J_{st}/J_1$ a *D. subobscura*

L'element D de Muller es correspon amb el cromosoma J de *D. subobscura*. Aquest element és autosòmic i acrocèntric en les espècies *D. subobscura*, *D. guanche* i *D. madeirensis*, mentre que en el cas de *D. pseudoobscura*, *D. persimilis* i *D. miranda* correspon al braç dret (XR) del cromosoma sexual que és metacèntric, i en el cas de *D. melanogaster*, l'element D de Muller està fusionat amb l'element E sent, respectivament, els braços 3L i 3R del cromosoma metacèntric 3 (Buzzati i Traverso, 1955; Taula 12). El cromosoma J de *D. subobscura* és el que presenta un menor polimorfisme cromosòmic a aquesta espècie, si exceptuem el cromosoma puntiforme. Les ordenacions majoritàries de la regió Paleàrtica i pràcticament les úniques presents tant al Nord d'Àfrica com a Europa són les ordenacions  $J_{st}$  i  $J_1$ . Les altres dues ordenacions del cromosoma J descrites a l'espècie són:  $J_{3+4}$  i  $J_{1+5}$ . L'ordenació  $J_{3+4}$  només es

troba a freqüències apreciables en àrees de l'est del Mediterrani, Àsia menor i el sud dels Balcans i l'ordenació  $J_{1+5}$  és exclusiva d'Israel (Krimbas 1992).

La inversió cromosòmica  $J_1$  es troba delimitada per les seccions 22B/C-25A2/A3 (Krimbas i Powell 1992) i és una inversió relativament petita. Les espècies *D. guanche* i *D. madeirensis* estretament relacionades amb *D. subobscura* presenten pel cromosoma J l'ordenació estàndard ( $J_{st}$ ) o una ordenació derivada de la  $J_{st}$ , fet que fa pensar que l'ordenació  $J_{st}$  de *D. subobscura* és l'ordenació ancestral. La freqüència de l'ordenació cromosòmica  $J_1$  presenta una clina de Sud Oest cap a Nord Est, trobant-se les freqüències més elevades en àrees del Nord Est d'Àfrica. Per contra, l'ordenació  $J_{st}$  presenta una clina de Nord cap a Sud i predomina en àrees del Nord d'Europa (Krimbas 1992).

### **13.8 Influència de factors demogràfics en la variabilitat nucleotídica de *D. subobscura***

Mentre s'han realitzat alguns estudis l'objectiu del quals és desentrellar quins factors demogràfics han conformat la variabilitat nucleotídica de *D. subobscura* després del procés de colonització iniciat al final dels anys setanta al continent Americà (Rozas i Aguadé 1991; Pascual et al. 2007), poc es sap del paper que ha jugat la demografia en la història evolutiva de l'espècie en la regió Paleàrtica. La disponibilitat de dades sobre el polimorfisme per ordenacions cromosòmiques en diferents poblacions naturals de *D. subobscura* ha fet possible una anàlisi a nivell de components principals per intentar validar la hipòtesi de que la distribució geogràfica de les ordenacions cromosòmiques pot haver estat influenciada per factors històrics (Krimbas i Loukas 1980). Així és com Menozzi i Krimbas (1992) realitzen una anàlisi de components principals (variables sintètiques que engloben informació de les freqüències de totes les ordenacions cromosòmiques), estimant els tres components principals primers, i situant els seus valors en mapes geogràfics. El primer component principal explica un 18% del total de la informació, el segon un 12% i el tercer un 8%. Mentre el primer component proporciona informació sobre la relació entre el polimorfisme cromosòmic i els factors

climàtics, el segon i tercer component aporten informació sobre possibles factors històrics. Així és com el segon component presenta valors màxims a Àsia Menor, el Caucas i els Balcans, observant també un valor màxim inferior a Espanya. Aquest patró va fer hipotetitzar que després de la última glaciació hi va haver un procés de recolonització d'Europa a partir de diferents refugis glacials, sent la major font de recolonització els refugis localitzats en els Balcans, Àsia Menor i el Caucas i com a segona font majoritària la península Ibèrica. Així mateix, els Alps podrien haver actuat com a barrera natural no permetent una colonització extensa de les poblacions del nord d'Europa a partir del refugi Italià. A més a més, els registres fòssils d'espècies de *Quercus*, un dels boscos on viu *D. subobscura*, reflecteixen patrons similars de l'avançament de l'espècie, fet que permet hipotetitzar que la història demogràfica de *D. subobscura* no està tant lligada a les migracions humanes com seria el cas de *D. melanogaster* (Li i Stephan 2006; Thornton i Andolfatto 2006). Tot i així, Menozzi i Krimbas (1992) també postulen que una elevada freqüència d'una determinada ordenació en una regió concreta no té per què predir que aquesta ordenació sigui l'ancestral en aquella regió, ja que els diferents períodes glacials ocorreguts en la regió Paleàrtica durant el període del Quaternari associats amb possibles reduccions poblacionals, poden haver fet que per fenòmens de deriva aquestes ordenacions hagin arribat a freqüències elevades.

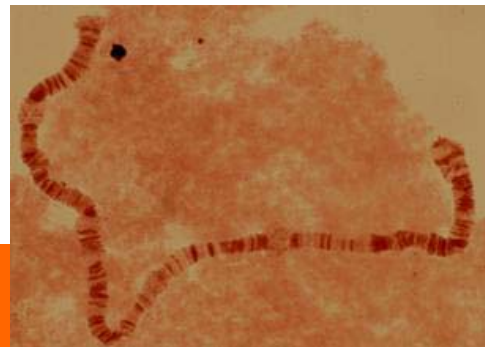
Malgrat s'hagi hipotetitzat sobre els possibles esdeveniments que han perfilat la variabilitat cromosòmica actual de *D. subobscura* en la regió Paleàrtica, no s'han analitzat models demogràfics concrets que puguin explicar la història evolutiva de l'espècie. No es disposa, per tant, d'un model demogràfic que sigui el més probable ni dels paràmetres demogràfics associats a aquest model. Castro et al. (2010) analitzen la variació d'un fragment de 942 parells de bases d'un gen mitocondrial en una mostra de 48 individus de *D. subobscura* i arriben a la conclusió que l'explicació més plausible dels nivells i patrons de polimorfisme observats en la regió Paleàrtica és la d'una possible expansió poblacional ocorreguda després

de l'últim període glacial i precedida per un fenomen de reducció poblacional. Tot i així, en aquest estudi només s'utilitza un gen mitocondrial i per tant no poden diferenciar si els nivells i patrons de polimorfisme observats són deguts a factors demogràfics o selectius.

Alhora de realitzar inferències sobre la història demogràfica de l'espècie *D. subobscura* s'ha d'anar amb cura amb la utilització de regions afectades per inversions cromosòmiques ja que aquestes podrien presentar nivells i patrons de variació nucleotídica determinats per les pròpies inversions i no per la història demogràfica de l'espècie. Per tant, les regions genòmiques afectades per inversions no serien bones candidates alhora d'estudiar la demografia de l'espècie, fet que dificulta aquests estudis ja que *D. subobscura* és una de les espècies que presenta un major polimorfisme per inversions cromosòmiques. L'anàlisi de regions genòmiques del cromosoma J serà un bon punt de partida per a la realització d'inferències evolutives, ja que aquest és el menys polimòrfic de tot el seu complement cromosòmic. Per tant, l'estudi de múltiples regions genòmiques no codificadores i no incloses dintre de la inversió  $J_1$  (present a la població d'estudi), farà possible en primer lloc diferenciar els efectes demogràfics dels selectius. S'espera que els efectes selectius estiguin localitzats en regions concretes i els demogràfics afectin la variabilitat a totes les regions d'estudi. En segon lloc, estudis de polimorfisme a nivell multilocus permetran comparar models senzills i la tria del més probable per explicar la història demogràfica recent de *D. subobscura* en la regió Paleàrtica. Un cop obtingut aquest model demogràfic, es podrà utilitzar per obtenir prediccions sobre els nivells i patrons de variació en absència de selecció. Així mateix, permetrà detectar regions, la variabilitat de les quals presenti una desviació respecte a aquest model i per tant que no es pugui explicar en absència de selecció positiva. Es pot destacar que els estudis multilocus de les regions no codificadores del cromosoma J per intentar desentrellar la història evolutiva de l'espècie, permetran disposar d'una informació valuosa



ahora d'estudiar la variabilitat nucleotídica d'aquesta espècie i detectar la possible empremta deixada per la selecció natural.



## OBJECTIUS

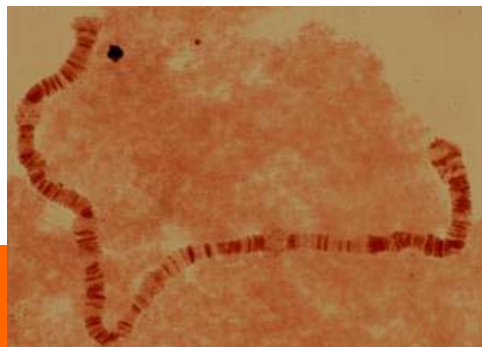


La història evolutiva dels organismes queda impresa d'alguna manera en el seu genoma a nivell de la seqüència nucleotídica, per tant, l'estudi dels nivells i patrons de polimorfisme nucleotídic en una espècie com *D. subobscura* constitueix una bona estratègia per inferir la importància relativa dels diferents factors evolutius que modelen aquesta variació. A més a més, l'elevat polimorfisme per inversions cromosòmiques que presenta *D. subobscura*, fa que sigui una espècie idònia per estudiar l'efecte d'aquests canvis estructurals del genoma a nivell de la variació nucleotídica. L'anàlisi de la variabilitat genètica en múltiples regions no codificadores localitzades al llarg del cromosoma J en una població natural de *D. subobscura* de Barcelona, ha permès abordar un doble objectiu global. Per una banda, inferir la història demogràfica recent de l'espècie en la regió Paleàrtica ja que el cromosoma J és el que presenta un menor polimorfisme per inversions cromosòmiques. Per altra banda, estudiar l'efecte de pràcticament l'única inversió que segrega en el cromosoma J a la població analitzada ( $J_1$ ) sobre la variabilitat nucleotídica. Per assolir aquests objectius, s'han establert els següents objectius específics:

1. Identificar diferents regions no codificadores a partir de fags recombinants aïllats aleatòriament d'una genoteca de l'espècie *D. subobscura* i que havien estat localitzats per hibridació *in situ* en el cromosoma J d'aquesta espècie (Papaceit et al. 2006).
2. Obtenir diferents línies altament consanguínies i homocariotípiques en relació al cromosoma J (element D de Muller) de *D. subobscura* per poder seqüenciar les regions identificades en individus homozigòtics.
3. Analitzar els nivells i patrons de variació nucleotídica en múltiples regions no codificadores i no afectades per inversions cromosòmiques distribuïdes al llarg del cromosoma J en una població natural de *D. subobscura* de Barcelona.
4. Estudiar a nivell multilocus el segment cromosòmic afectat per la inversió cromosòmica  $J_1$  en diferents línies homocariotípiques que es diferencien per la

presència o absència d'aquesta inversió en una població natural de *D. subobscura* de Barcelona.

5. Detectar la possible petjada deixada per la selecció natural en la variabilitat nucleotídica de les diferents regions en estudi.
6. Inferir els possibles canvis demogràfics en la història evolutiva recent de *D. subobscura*.
7. Analitzar l'efecte d'una inversió de grandària mitjana-petita, com la inversió  $J_1$ , sobre la variabilitat nucleotídica.
8. Estimar l'edat de la inversió  $J_1$ .



## MATERIALS I MÈTODES



## M1. OBTENCIÓ DE DADES

### M1.1 Mostra poblacional

La població natural de *D. subobscura* en estudi es va capturar a l'Observatori Fabra de Barcelona al 2007. A partir de les femelles del camp es van establir 83 línies isomaternes amb la finalitat d'obtenir línies altament consanguínies. A tal propòsit es van realitzar almenys 13 encreuaments consecutius germà-germana de cada una de les línies isomaternes. Després dels 13 encreuaments, es va comprovar primer que les línies fossin homocariotípiques pel cromosoma J mitjançant l'anàlisi dels cromosomes politènics de les glàndules salivals de com a mínim 8 larves de tercer estadi descendents d'un encreuament germà-germana. La metodologia emprada per la visualització al microscopi dels cromosomes politènics va ser la tinció estàndard amb orceïna acètico-làctica. Subseqüentment es va determinar l'ordenació cromosòmica del cromosoma J encreuant mascles de la descendència de l'encreuament anterior amb femelles verges de la soca marcadora *chcu* amb l'ordenació estàndard pel cromosoma J. Es van obtenir 27 línies altament consanguínies que es diferencien per la presència o absència de la inversió  $J_1$ : 9 d'elles presenten l'ordenació  $J_{st}$  i 18 l'ordenació  $J_1$ . A més, es va utilitzar com a *outgroup* una línia altament consanguínea de l'espècie *D. guanche* obtinguda prèviament al laboratori mitjançant la mateixa metodologia. Les diferents línies es mantenen en medi estàndard de farina de blat de moro-agar i a 18°C.

### M1.2 Caracterització dels inserits dels fags recombinants

En iniciar el treball es disposava de diversos fags recombinants aïllats aleatòriament d'una genoteca de l'espècie *D. subobscura* i que havien estat localitzats per hibridació *in situ* en el cromosoma J d'aquesta espècie (Papaceit et al. 2006). Alguns d'aquests fags recombinants es localitzen dintre de la inversió  $J_1$ . A partir del DNA purificat dels fags recombinants es va realitzar una digestió amb l'enzim de restricció *EcoRI*, una posterior purificació del DNA digerit mitjançant columnes *Milipore* "Microcon-PCR" i un gel



d'electroforesi de comprovació al 0,8% de concentració d'agarosa. Els diferents fragments que es van generar, es van clonar per *shotgun cloning* que consisteix en la lligació aleatòria dels fragments de DNA de l'inserit del fag en un plasmidi i la posterior transformació de cèl·lules competents. Les reaccions de digestió, lligació i transformació es van realitzar seguint protocols estàndards (Maniatis et al. 1983). El plasmidi que es va utilitzar com a vector és el *pBluescriptIIISK* (Stratagene) que té un gen de resistència a ampil·lina i un lloc de clonació múltiple (*polylinker*) en l'interior del gen *LacZ*. La soca d'*E. coli* que es va utilitzar per la transformació és l'*XL1-Blue* (Stratagene) resistent a la tetraciclina. El medi que es va utilitzar per sembrar les cèl·lules transformades és LB+ampicil·lina+tetraciclina+X-gal+IPTG. Els antibiòtics només permeten el creixement de colònies *XL1-Blue* transformades. A més el sistema IPTG+X-gal permet diferenciar les colònies transformades que han incorporat el plasmidi amb l'inserit (colònies blanques) de les que han incorporat el plasmidi sense inserit (colònies blaves). Per a cada fag recombinant inicial es va amplificar l'inserit del plasmidi de 12 colònies blanques mitjançant la tècnica de PCR amb els encebadors universals T3 i T7 presents en els extrems del lloc de clonació múltiple i seguint les condicions següents:

Desnaturalització	94°C	3 min	
<i>Annealing</i>	96°C	10 seg	30 cicles
Extensió	46°C	10 seg	
Finalització	65°C	10 min	
	4°C	∞	

Posteriorment, es va realitzar un gel d'electroforesi de comprovació a l'1% d'agarosa. Per a cada fag inicial es van seleccionar el major nombre possible d'inserits amb longituds diferents. En considerar tots els fags, aquestes longituds variaven de 500 pb a 8 kb. Una vegada feta la selecció, es va purificar el DNA plasmídic mitjançant columnes *Miniprep* de la casa comercial QIAGEN per obtenir quantitat suficient de DNA plasmídic en estoc i es va fer una prova de

confirmació digerint el producte purificat amb l'enzim de restricció *EcoRI*. Per una altra banda, el DNA dels inserits amplificats seleccionats es va purificar mitjançant una precipitació d'acetat amònic i EtOH absolut. Posteriorment es van seqüenciar els extrems dels inserits purificats amb els encebadors universals T3 i T7, utilitzant el *kit* de seqüenciació comercial *BigDye Terminator 3.1* (Applied Biosystems) amb les següents condicions:

Desnaturalització	94°C	3 min	
<i>Annealing</i>	96°C	10 seg	30 cicles
Extensió	50°C	5 seg	
Finalització	60°C	4 min	
	4°C	∞	

Els productes de seqüenciació van ser analitzats amb el seqüenciador automàtic *ABI PRISM 3700* (Applied Biosystems) disponible en el Servei de Genòmica dels Serveis Científico-Tècnics de la Universitat de Barcelona (SCT-UB). Una vegada es van obtenir les seqüències nucleotídiques, es van anotar els inserits dels fags recombinants a *D. subobscura* a través de la cerca BLAST en els genomes complets de *D. melanogaster* i *D. pseudoobscura* disponibles a [www.flybase.com](http://www.flybase.com). Es van seleccionar les regions genòmiques amb homologia entre les espècies (corresponents a l'element D de Muller) d'una longitud que variava entre 500 i 3500 parells de bases i localitzades en regions no codificadores al llarg de tot el cromosoma J. Per aquells fags recombinants seleccionats els inserits dels quals no es localitzaven en regions no codificadores, es van dissenyar encebadors d'amplificació a partir del genoma de *D. pseudoobscura* en zones properes a les regions genòmiques identificades. A més a més, es va incrementar el nombre de regions en estudi, seguint dues metodologies. Per una banda, es van seleccionar 3 regions localitzades al cromosoma J d'acord a Laayouni et al. (2007) per les que es van dissenyar encebadors sobre la seqüència genòmica de *D. pseudoobscura* disponible a [www.flybase.com](http://www.flybase.com). Totes les sondes van ser posteriorment comprovades i localitzades

citològicament a través de la tècnica d'hibridació *in situ*, d'acord al protocol descrit a Segarra i Aguadé (1992). Per una altra banda i donat que es volia incrementar el nombre de regions dins de la inversió  $J_1$  es va utilitzar el genoma de *D. pseudoobscura* com a guia i les regions ja identificades com anclatge. Es van seleccionar noves regions allunyades de les d'anclatge i es va comprovar la seva localització mitjançant hibridació *in situ*. En aquest cas, els encebadors per amplificar aquestes regions es van dissenyar sobre els *contigs* preliminars de la seqüència del genoma de *D. subobscura* (esborrany 0) obtinguda per la BSI (*Barcelona Subobscura Initiative*) i que s'havien fet accessibles poc temps abans.

Seguint les diferents aproximacions es van obtenir un total de 24 regions genòmiques, sis de les quals es troben dintre de la inversió  $J_1$  i les restants en regions no codificadores repartides al llarg del cromosoma J. Mitjançant la tècnica de *primer walking* es va seqüenciar la totalitat de les regions seleccionades, dissenyant encebadors d'una longitud de 18 nucleòtids amb una distància entre ells d'aproximadament 600 parells de bases. Tots els encebadors, tant els d'amplificació com els de seqüenciació es van dissenyar mitjançant el programa OLIGO 6.0 (Rychlik 2007).

### **M1.3 Purificació del DNA genòmic i seqüenciació**

Es va extreure per a cada una de les línies homocariotípiques obtingudes a partir de la població natural de *D. subobscura* de Barcelona i per una línia de *D. guanche*, el DNA genòmic d'un sol individu mitjançant el *kit Puregen®Core Kit B* (QIAGEN). Es van dissenyar amb el programa OLIGO 6 (Rychlik 2007) encebadors d'amplificació de 20 nucleòtids de les 24 regions genòmiques en estudi. Seguidament, es va procedir a l'amplificació mitjançant la tècnica de PCR, la posterior purificació a través de plaques *Multiscreen* de Millipore® i la seqüenciació de doble filament en les diferents línies de *D. subobscura* i una línia de *D. guanche*. En les Taules MS4.1.a-MS4.1d del material suplementari del final de l'apartat dels Materials i Mètodes hi ha disponibles les seqüències nucleotídiques dels encebadors dissenyats per les diferents regions

en estudi en les espècies *D. subobscura* i *D. guanche*. Les condicions d'amplificació per PCR de les diferents regions en estudi es mostren en les Taules MS4.2.a- MS4.2.c.

#### **M1.4 Assemblatge i alineament de seqüències**

Les seqüències parcials de les diferents regions genòmiques en estudi es van assemblar independentment per a cada individu homocariotípic de l'espècie *D. subobscura* i per l'individu de l'espècie *D. guanche* mitjançant el programa *SeqMan* (del paquet informàtic DNASTar, *Lasergene*). Els alineaments múltiples, en els que s'inclou l'*outgroup*, es van realitzar amb el programa *MegAlign* (DNASTar, *Lasergene*) i/o el programa MAFFT (Kato i Standley 2013) disponible *online* a <http://mafft.cbrc.jp/alignment/software>. Aquests alineaments van ser visualitzats i editats mitjançant el programa *MacClade* versió 4.05 (Maddison i Maddison 2000). En obtenir la seqüència consens, totes les posicions polimòrfiques es van comprovar manualment (*SeqMan*).

## M2. ANÀLISI DE LES DADES

### M2.1 Estimes dels nivells i patrons de variació nucleotídica

Per les anàlisis dels nivells i patrons de variació nucleotídica de les 24 regions genòmiques en estudi tant a nivell individual com multilocus es va utilitzar el programa DnaSP v5 (Librado i Rozas 2009) i el programa MANVA (Ramos-Onsins, disponible en <http://bioinformatics.cragenomica.es/numgenomics//people/sebas/software/software.html>). Les posicions afectades per insercions i delecions es van excloure de les anàlisis.

#### Nivells de variació nucleotídica

Els nivells de variació nucleotídica es van analitzar mitjançant els següents estimadors.

#### Diversitat nucleotídica ( $\pi$ ) (Nei 1987, equació 10.5)

Nombre promig de diferències nucleotídiques entre parells de seqüències per posició

$$\pi = \frac{k}{L}$$

sent  $k$  el nombre promig de diferències nucleotídiques entre les seqüències (Tajima 1983, equació A3) i  $L$  el nombre de posicions analitzades. El paràmetre  $k$  és sensible al nombre de posicions polimòrfiques, a la freqüència que presenten les variants nucleotídiques i a la longitud de la regió analitzada, però no depèn de la grandària de la mostra. En canvi, el paràmetre  $\pi$  no és sensible ni a la grandària de la mostra ni a la longitud de la regió analitzada.

#### L'estimador de Watterson ( $\theta$ ) (Watterson 1975)

S'obté a partir del nombre de posicions polimòrfiques que segreguen ( $S$ ) en una mostra de  $n$  seqüències nucleotídiques. Aquest estimador ( $\theta$  o estimador de Watterson per seqüència) correspon a  $k$  quan les mutacions són selectivament neutres, la població és panmíctica i es troba en equilibri mutació-deriva, sent  $\theta = 4N_e\mu$  ( $N_e$  grandària efectiva de la població i  $\mu$  la taxa de mutació neutra per generació).

$$\theta_{\text{(per seqüència)}} = \frac{S}{\sum_{i=1}^{n-1} 1/i}$$

Si es considera el nombre de posicions analitzades ( $L$ ), s'obté una estima de l'heterozigositat per nucleòtid [ $\theta$  o estimador de Watterson per posició (Watterson 1975 equació 1.4a)].

$$\theta_{\text{(per posició)}} = \frac{S/L}{\sum_{i=1}^{n-1} 1/i}$$

#### Diversitat haplotípica ( $H_w$ ) (Nei 1987, equació 8.4)

Considera el nombre i la freqüència dels diferents haplotips presents en una mostra de  $n$  seqüències de DNA. De fet, correspon a la probabilitat de que en agafar a l'atzar dues seqüències de la mostra, aquestes siguin diferents.

$$H_w = \frac{n}{n-1} (1 - \sum X_i^2)$$

on  $n$  és el nombre de seqüències i  $X_i$  és la freqüència de la seqüència (haplotip)  $i$ .

#### **Patrons de variació nucleotídica**

Els patrons de variació nucleotídica es van analitzar mitjançant els següents estimadors.

#### Estadístic $D$ de Tajima (Tajima 1989)

L'estadístic  $D$  de Tajima està basat en la comparació de dos estimadors del nivell de variabilitat:  $\theta_{\text{(per seqüència)}}$  estimat a partir del nombre de llocs segregants o polimòrfics ( $S$ ) i del nombre de seqüències (Watterson 1975), i el nombre promig de diferències nucleotídiques en comparar les seqüències dos a dos ( $k$ , Tajima 1983; Nei 1987)

$$D = \frac{k - S/a_n}{\sqrt{\text{Var}(k - S/a_n)}}$$

on

$$a_n = \sum_{i=1}^{n-1} 1/i$$

Donat que l'estimador  $k$  té en compte la freqüència a la qual segreguen les variants nucleotídiques mentre que  $\theta_{(\text{per seqüència})}$  no, l'estadístic  $D$  de Tajima pot detectar desviacions de l'espectre de freqüències observat respecte del que s'esperaria sota el model neutre estàndard. L'estadístic  $D$  de Tajima seria igual a zero si es complís el model neutre estàndard ja que els estimadors que compara són equivalents sota aquest model. Així mateix, valors positius de l'estadístic  $D$  de Tajima indicarien un excés de polimorfismes amb variants a freqüències intermèdies, com a conseqüència de l'existència d'estructuració poblacional o de l'acció de selecció equilibradora. Valors negatius de l'estadístic  $D$  de Tajima denotarien un excés de polimorfismes amb variants a baixa freqüència com a conseqüència d'una expansió poblacional o després de l'acció de selecció direccional.

#### Tests de Fu i Li (1993)

Els tests de Fu i Li (1993) es fonamenten en la distribució de les mutacions esperades en les branques internes i externes d'una genealogia neutra considerant equilibri mutació-deriva. En una genealogia, les mutacions antigues tendiran a acumular-se en les branques internes, mentre que les mutacions més recents tendiran a acumular-se en les branques externes. Fu i Li proposen dos tests estadístics ( $D$  i  $F$ ) els quals es poden aplicar amb *outgroup* (estadístics  $D$  i  $F$ ) o sense (estadístics  $D^*$  i  $F^*$ ). Els tests que no tenen en compte l'*outgroup* estan basats en el nombre de mutacions úniques en la mostra (*singletons*), mentre que els tests amb *outgroup* estan basats en el nombre de mutacions en les branques externes de la genealogia. L'estadístic  $D$  es basa en el nombre total de mutacions i el nombre de mutacions en les branques externes de la genealogia (nombre de *singletons*), mentre que l'estadístic  $F$  està basat en el nombre promig de diferències entre seqüències i el nombre de mutacions en les branques externes de la genealogia (nombre de *singletons*).

#### Estadístic $D$ de Fu i Li amb *outgroup*

$$D = \frac{\eta - a_n \eta_e}{\sqrt{u_D \eta - v_D \eta^2}}$$

on  $\eta$  és el nombre total de mutacions,  $\eta_e$  és el nombre de mutacions en les branques externes i  $a_n$ ,  $u_D$  i  $v_D$  són funcions de la grandària mostral.

Estadístic  $D^*$  de Fu i Li sense *outgroup*

$$D^* = \frac{(n/n-1) \eta - a_n \eta_s}{\sqrt{u_D^* \eta - v_D^* \eta^2}}$$

on  $\eta$  és el nombre total de mutacions,  $\eta_s$  és el nombre de mutacions úniques en la mostra i  $a_n$ ,  $u_D^*$  i  $v_D^*$  són funcions de la grandària mostral.

Estadístic  $F$  de Fu i Li amb *outgroup*

$$F = \frac{k - \eta_e}{\sqrt{u_F \eta - v_F \eta^2}}$$

on  $k$  és el promig de diferències nucleotídiques entre seqüències,  $\eta_e$  el nombre de mutacions úniques en la mostra i  $u_F$  i  $v_F$  són funcions de la grandària mostral.

Estadístic  $F^*$  de Fu i Li sense *outgroup*

$$F^* = \frac{k - (n/n-1) \eta_s}{\sqrt{u_F^* \eta - v_F^* \eta^2}}$$

on  $k$  és el nombre promig de diferències nucleotídiques entre seqüències,  $\eta_s$  és el nombre de mutacions úniques en la mostra i  $u_F^*$  i  $v_F^*$  són funcions de la grandària mostral.

Estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006)

Estadístic basat en la comparació de dues estimes del nivell de variabilitat:  $\theta_\pi = k$  (Tajima 1983) i  $\theta_L$  (el nombre promig de canvis acumulats des de l'ancestre comú més recent en una mostra de  $n$  seqüències (Zeng et al. 2006)).

$$H_{\text{norm}} = \frac{\theta_\pi - \theta_L}{\sqrt{(\theta_\pi - \theta_L)}}$$

L'estadístic infereix les variants ancestrals a través de la seqüència de l'*outgroup*, permetent diferenciar les variants ancestrals de les derivades. Aquest estadístic és molt útil ja que permet detectar desviacions de l'espectre de freqüències desdoblada observat respecte l'espectre



esperat sota condicions de neutralitat. Igual que l'estadístic  $D$  de Tajima, s'espera que sota aquestes condicions, l' $H$  normalitzada de Fay i Wu ( $H_{norm}$ ) tingui valors propers a zero.

#### Estadístic $R_2$ de Ramos-Onsins i Rozas (2002)

Estadístic basat en el nombre de polimorfismes amb variants a freqüència única i el promig de diferències nucleotídiques

$$R_2 = \frac{(\sum_{i=1}^n (U_i - k/2)^2 / n)^{1/2}}{S}$$

on  $n$  és la grandària de la mostra,  $S$  el nombre de llocs segregants,  $k$  el promig de diferències nucleotídiques entre dues seqüències i  $U_i$  el nombre de polimorfismes amb variants a freqüència única en la seqüència  $i$ . La idea de l'estadístic és que el nombre esperat de polimorfismes amb variants a freqüència única en una branca de la genealogia després d'un esdeveniment recent d'expansió poblacional severa és  $k/2$ ; conseqüentment, s'esperarien valors baixos d'aquest estadístic sota aquest escenari demogràfic.

#### Estadístic $F_s$ de Fu (Fu 1997)

Estadístic basat en la probabilitat de tenir un nombre determinat d'al·lels en una mostra de  $n$  seqüències, donat un valor de  $\theta$ . S'espera que tendeixi a tenir valors propers a zero en condicions de neutralitat.

#### Divergència nucleotídica ( $K$ ) (Nei 1987)

La divergència nucleotídica entre l'espècie *D. subobscura* i *D. guanche* utilitzada com a *outgroup* es va estimar com el nombre promig de diferències nucleotídiques per posició en totes les comparacions interespecífiques possibles.

$$K = \frac{k}{L}$$

sent  $k$  el nombre promig de diferències nucleotídiques en les comparacions interespecífiques i  $L$  el nombre de posicions. La divergència nucleotídica estimada es corregeix per Jukes i Cantor (1969) donada la problemàtica de les possibles mutacions a la mateixa posició.

### Desequilibri de lligament entre posicions

El grau de desequilibri de lligament global de les diferents regions en estudi es va analitzar mitjançant els estadístic  $Z_A$  (Rozas et al. 2001),  $Z_{nS}$  (Kelly 1997) i  $B$  de Wall (Wall 1999).

#### Estadístic $Z_{nS}$ (Kelly 1997, equació 3)

Estadístic que mesura globalment l'associació entre els llocs polimòrfics, sent el promig de l'estadístic  $r^2$  (Hill i Robertson 1968) a través de totes les possibles comparacions dos a dos.

$$Z_{nS} = \frac{2}{S(S-1)} \sum_{i=1}^{s-1} \sum_{j=i+1}^s r_{ij}^2$$

on  $S$  és el nombre de llocs polimòrfics i  $r_{ij}$  és l'estimador  $r$  (Hill i Robertson 1968), sent aquest estadístic

$$r^2 = \frac{D^2}{p(1-p)q(1-q)}$$

Considerant A i B com a dos loci amb els tipus d'al·lels  $A_0, A_1, B_0$  i  $B_1$ , respectivament i sent  $p$  i  $q$  les freqüències al·lel·liques d' $A_0$  i  $B_0$  respectivament.  $D = f_{00} - pq$ , sent  $f_{00}$  la freqüència del tipus gamètic  $A_0B_0$ .

#### Estadístic $Z_A$ (Rozas et al. 2001)

L'estadístic  $Z_A$  és equivalent a l'estadístic  $Z_{nS}$ , però només tenint en compte les posicions polimòrfiques informatives adjacents. Una posició polimòrfica és informativa quan almenys hi ha dues variants cadascuna present almenys en dues seqüències.

$$Z_A = \frac{1}{(S-1)} \sum_{i=1}^{s-1} r_{i,i+1}^2$$

on  $S$  és el nombre de llocs polimòrfics i  $r^2$  és l'estadístic  $r^2$  (Hill i Robertson 1968).

Estadístic  $B$  de Wall (1999)

L'estadístic  $B$  de Wall utilitza la informació present en un parell de llocs segregants adjacents. Es considera que un parell de llocs segregants adjacents és congruent si el conjunt de les dades que configuren els dos llocs contenen només dos haplotips.

$$B = B'/(S-1)$$

on  $S$  és el nombre de llocs segregants de la mostra i  $B'$  és el nombre de parells de llocs segregants adjacents que són congruents. L'estadístic  $B$  de Wall pot variar de 0 a 1, sent un estadístic informatiu sobre el desequilibri de lligament. A més recombinació, s'esperarien valors d'aquest estadístic més propers a zero ja que la probabilitat de que els llocs segregants adjacents compartissin la mateixa genealogia disminuiria. Per contra, s'esperarien valors elevats d'aquest estadístic sota un escenari demogràfic de subdivisió poblacional.

**Test de Hudson-Kreitman-Aguadé (HKA)**

El test HKA compara la variabilitat nucleotídica observada dintre d'espècie (polimorfisme) amb la variabilitat entre espècies (divergència) ja que sota condicions de neutralitat s'espera que la relació entre polimorfisme i divergència sigui similar a diferents regions genòmiques. El test HKA mesura la bondat d'ajust dels valors observats respecte els esperats d'acord a una distribució  $\chi^2$ . El test HKA multilocus es va realitzar amb el programa HKA distribuït per Jody Hey i disponible a <http://genfaculty.rutgers.edu/hey>.

**M2.2 Anàlisi de la variació nucleotídica a les regions afectades per la inversió  $J_1$** 

Sis de les 24 regions en estudi estan distribuïdes al llarg de la inversió  $J_1$ : p127, p221, p174, p39, p137 i p222. Les anàlisis d'aquestes regions es van realitzar independentment per les 18 línies  $J_1$  i per les 9 línies  $J_{st}$ .

**Distància al punt de trencament**

La distància de cada regió al punt de trencament (Taula M2.1) es va estimar considerant una distribució homogènia del DNA al llarg de tota la inversió  $J_1$  i utilitzant com a

referència les dades publicades de l'espècie *D. pseudoobscura* i *D. melanogaster* (Adams et al. 2000; Schaeffer et al. 2008). Per calcular la longitud de la inversió  $J_1$  expressada en cM, es va realitzar una extrapolació dels cM d'aquesta inversió a partir de les dades disponibles a Cáceres et al. (1999b).

**Taula M2.1.** Estimes de la distància de cada regió en estudi al punt de trencament més proper de la inversió  $J_1$ .

		cM (centiMorgan)	Mb (Megabases) en funció de <i>D. melanogaster</i>	Mb (Megabases) en funció de <i>D. pseudoobscura</i>
<b>Element D de Muller</b>		192 cM (Cáceres et al. 1999b)	22,6 Mb (Adams et al. 2000)	30,5 Mb (Schaeffer et al. 2008)
<b>Longitud de la inversió <math>J_1</math></b>		30	3,5	4,8
<b>Distància al punt de trencament</b>	p222	0,8	0,1	0,1
	p127	5,5	0,6	0,8
	p137	7,9	0,9	1,3
	p221	9,5	1,1	1,6
	p39	10,3	1,2	1,7
	p174	11,8	1,4	1,9
	p148	5,5	0,7	0,9
	p32096	8,7	1,1	1,4

Les regions p148 i p32096, ombrejades en gris, són les regions fora de la inversió però més properes als punts de trencament en estudi.

### Diferenciació genètica entre les ordenacions $J_1$ i $J_{st}$

La diferenciació genètica entre les dues ordenacions analitzades  $J_1$  i  $J_{st}$  es va estimar mitjançant diferents estadístics.

Nombre promig de diferències nucleotídiques per posició entre ordenacions ( $D_{xy}$ ) (Nei 1987, equació 10.20)

El nombre promig de diferències nucleotídiques entre les dues ordenacions cromosòmiques es va estimar com a:

$$D_{xy} = \sum_{ij} x_i y_i d_{ij}$$

on  $x_i$  i  $y_j$  són les freqüències dels haplotips  $i$  i  $j$  de les ordenacions  $X$  i  $Y$ , respectivament, i  $d_{ij}$  és la proporció de diferències nucleotídiques entre l'haplotip  $i$  d' $X$  i l'haplotip  $j$  d' $Y$ .

Nombre net de substitucions nucleotídiques per posició entre ordenacions ( $D_a$ ) (Nei 1987, equació 10.21)

El nombre net de substitucions nucleotídiques entre les dues ordenacions es va estimar com a

$$D_a = D_{xy} - \frac{(\pi_x + \pi_y)}{2}$$

on  $\pi_x$  i  $\pi_y$  és la diversitat nucleotídica en l'ordenació  $X$  i  $Y$ , respectivament.

Test de permutacions (Hudson et al. 1992b)

Per contrastar si la diferenciació genètica entre les dues ordenacions era significativa es va realitzar el test de permutacions. Aquest test es basa en l'estadístic  $K_{st}^*$  que també mesura la diferenciació genètica entre ordenacions i la seva significació estadística es determina mitjançant simulacions de Monte Carlo.

$F_{ST}$  (Hudson et al. 1992a, equació 3)

El flux gènic entre les ordenacions cromosòmiques es va calcular mitjançant l'estadístic  $F_{ST}$ .

$$F_{ST} = 1 - \frac{H_w}{H_b}$$

sent  $H_w$  la mitjana del nombre de diferències nucleotídiques entre seqüències de la mateixa ordenació i  $H_b$  la mitjana del nombre de diferències nucleotídiques entre seqüències d'ordenacions diferents.

### **Conversió gènica**

Seguint l'algoritme proposat per Bertrán et al. (1997, equacions A1 i A4), es va estimar el nombre i la longitud dels tractes de conversió gènica entre ordenacions cromosòmiques. Malgrat que aquest algoritme va ser dissenyat per identificar tractes de conversió gènica, de fet detecta esdeveniments d'intercanvi genètic entre ordenacions.

### **Genealogia i xarxa de la mostra de seqüències**

Es va utilitzar el programa MEGA5 (Tamura et al. 2011) amb el mètode de *neighbor-joining* (Saitou i Nei 1987) per obtenir la genealogia dels al·lels del concatenat de les seqüències de les sis regions que es troben dintre de la inversió  $J_1$ . Es van utilitzar 21 línies homocariotípiques, 14 de les quals tenen l'ordenació cromosòmica  $J_1$  i 7 l'ordenació  $J_{st}$ . No es van utilitzar les 27 línies homocariotípiques obtingudes ja que en algunes de les línies amb ordenació  $J_1$  i  $J_{st}$  no va ser possible amplificar algunes de les regions i per tant aquestes línies no van ser considerades en el concatenat. Amb les regions en les quals no es van detectar tractes d'intercanvi genètic entre ordenacions es va contruir una xarxa d'haplotips mitjançant el programa *Network* 4.6.1.1 ([http://www.fluxus-engineering.com/sharenet\\_rn.htm](http://www.fluxus-engineering.com/sharenet_rn.htm)) seguint la metodologia *median-joining network* proposada per Bandelt et al. (1999).

### **Desequilibri de lligament amb la inversió**

El grau d'associació entre les variants nucleotídiques de les posicions polimòrfiques informatives i l'ordenació del cromosoma J ( $J_{st}$ ,  $J_1$ ) es va estimar mitjançant la mitjana del paràmetre  $r^2$  (Hill and Robertson 1968). A més a més, el grau de desequilibri global de les diferents regions en estudi es va analitzar mitjançant l'estadístic  $Z_{ns}$  (Kelly 1997).

### M3. INFERÈNCIES EVOLUTIVES

#### M3.1 Estimació del temps de divergència entre diversos llinatges

Es va estimar el temps de divergència entre el subgrup *subobscura* i el subgrup *pseudoobscura* i també el de divergència entre *D. subobscura* i *D. guanche* (subgrup *subobscura*) mitjançant una anàlisi d'inferència bayesiana. Posteriorment, es va utilitzar l'estima més probable del temps de divergència entre *D. subobscura* i *D. guanche*, per calcular la taxa de mutació de cada una de les 16 regions en estudi necessària per les anàlisis posteriors.

Es va utilitzar únicament una de les 27 línies seqüenciades de la població natural de *D. subobscura* i la línia de *D. guanche* utilitzada com a *outgroup* en les anàlisis de divergència. Es va realitzar una cerca BLAST de les seqüències homòlogues de les 16 regions en estudi en les espècies *D. melanogaster*, *D. pseudoobscura* i *D. persimilis* a través de [www.flybase.com](http://www.flybase.com). Posteriorment, es va utilitzar el programa MAFFT versió 7 (Katoh i Standley 2013) amb els paràmetres per defecte per construir els alineaments múltiples de les seqüències homòlogues. Aquests alineaments, que van ser visualitzats amb el programa BioEdit (Hall 1999), van ser tractats amb el programa Gblocks versió 0.91 (Castresana 2000) amb els paràmetres restrictius per defecte per descartar possibles posicions mal alineades, obtenint un total de 17 kb alineades. A continuació es va avaluar independentment per a cada regió el model de substitució que millor s'ajusta a les dades, mitjançant el criteri d'informació Bayesiana (BIC) implementat en el programa MEGA v.5.1. (Tamura et al. 2011). Finalment es van estimar els diferents temps de divergència mitjançant inferència bayesiana a través del programa BEAST v.1.6.2 (Drummond i Rambaut 2007). Es van realitzar 16 particions independents, una per a cada regió en estudi, en relació al model de substitució. Es va utilitzar un rellotge molecular relaxat no correlacionat. Es van calibrar els nodes mitjançant datacions recopilades d'estudis previs, utilitzant distribucions *a priori* uniformes, amb valors de 30 a 35 milions d'anys pel temps de divergència entre el grup *obscura* i el grup *melanogaster* (Throckmorton, 1975; Obbard et al. 2012) i un temps de

divergència entre *D. pseudoobscura* i *D. persimilis* de 0,18 a 0,22 milions d'anys (Kulathinal et al. 2009). Es va córrer una cadena de *Markov* de 100 milions de generacions, mostrejant paràmetres cada 1000 generacions. Es va examinar la convergència utilitzant el programa *Tracer* i es va comprovar que tots els paràmetres d'interès tinguessin un valor de grandària efectiva de mostreig (*effective sampling size*, ESS) major de 200. Es va realitzar una rèplica de l'anàlisi i es va combinar amb l'anterior mitjançant el programa *Logcombiner*. Es va eliminar el 10% inicial de les mostres com *burn-in* i es va obtenir l'arbre resum de màxima credibilitat dels clades (*maximum clade credibility tree*) utilitzant la mediana de les altures dels nodes, mitjançant el programa *TreeAnnotator*. L'arbre va ser editat amb el programa *Figtree v.1.4*.

### **M3.2 Mutacions recurrents i l'espectre de freqüències desdoblada**

Prèviament a les simulacions de coalescència, es va realitzar un estudi de l'espectre de freqüències desdoblada observat, obtingut per parsimònia mitjançant les dades de polimorfisme, concretament dels llocs segregants i utilitzant les seqüències de *D. guanche* per polaritzar les mutacions. Es va calcular per a cada regió l'espectre de freqüències desdoblada observat mitjançant el *script* SFS.pl disponible a la pàgina <http://www.rilab.org/> i posteriorment es va fer un espectre conjunt per totes les regions a excepció de la regió p32096 que no es considera en l'anàlisi ja que la grandària mostral d'aquesta és només de 13 individus. Al tenir regions amb una grandària mostral de 18 a 21 individus, es va aplicar un plantejament probabilístic basat en una distribució hipergeomètrica per no eliminar aquelles línies en aquelles regions amb més de 18 individus, mitjançant un *script* cedit per Pablo Librado. Inferir l'espectre de freqüències per parsimònia pot donar errors si no es té en compte la possibilitat de les mutacions recurrents. Per això, posteriorment, es va realitzar una anàlisi d'aquest espectre de freqüències desdoblada observat, aplicant una correcció estadística per corregir la possible identificació errònia de les variants ancestrals. La correcció estadística que es va realitzar, mitjançant un *script* cedit per Sebastián Ramos-Onsins, és similar a la proposada a Hernandez et al. (2007). Amb l'espectre de freqüències



desdoblament corregit es va calcular el tant per cent de guany o pèrdua de mutacions, respecte l'espectre de freqüències desdoblament observat i d'aquesta manera es va poder avaluar quina era la proporció de mutacions presents a la població a elevada freqüència que havien estat identificades erròniament. Així, es va poder avaluar si el no considerar la possibilitat de possibles mutacions recurrents podia afectar els valors d'aquells estadístics simulats basats en la distribució de les freqüències de les mutacions, concretament de l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i per tant afectar les inferències demogràfiques quan en les anàlisis d'ABC s'utilitza aquest estadístic. Una identificació errònia de les variants ancestrals en l'estadístic  $H$  normalitzat de Fay i Wu que utilitza una espècie *outgroup* per polaritzar les mutacions, podria classificar incorrectament les mutacions a baixa freqüència com a mutacions a elevada freqüència (Baudry i Depaulis 2003) i donar valors esbiaixats d'aquest estadístic si alhora de realitzar simulacions de coalescència no es tenen en compte les mutacions recurrents.

### **M3.3 Inferències demogràfiques (*Approximate Bayesian Computation*)**

Es va inferir la història evolutiva de l'espècie *D. subobscura* en la regió paleàrtica mitjançant una anàlisi d'*Approximate Bayesian Computation* (ABC; Tavaré et al. 1997; Pritchard et al. 1999; Beaumont et al. 2002). La metodologia emprada segueix la seqüència mostrada en la figura M3.1 i està en part basada en la metodologia proposada a Csilléry et al. (2012).

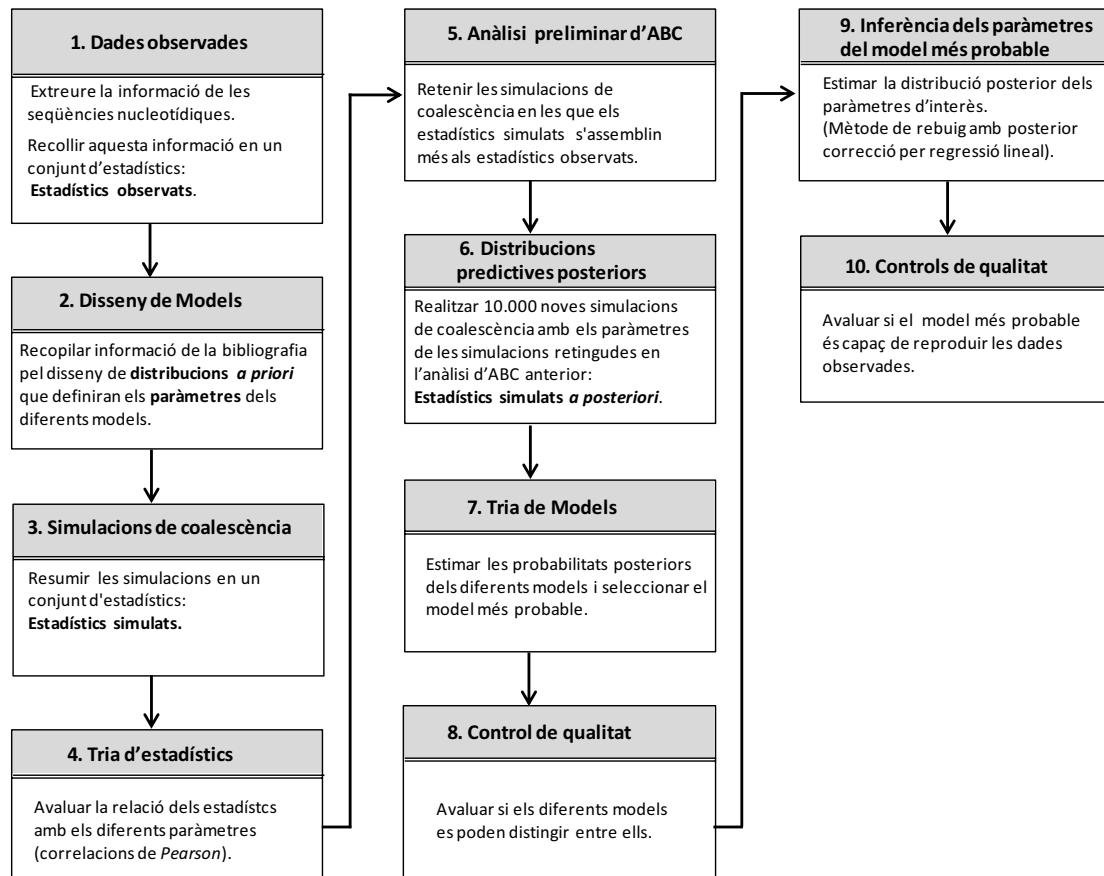


Figura M3.1. Diagrama de la metodologia emprada en les anàlisis d'ABC, dividida en 10 punts.

### Introducció al procés d'ABC

Les anàlisis d'*Approximate Bayesian Computation* (ABC) tenen les seves arrels en l'estadística bayesiana. L'objectiu de l'ABC és el d'inferir la distribució posterior d'un paràmetre ( $\theta$ ) condicionat a unes certes dades empíriques ( $D$ ). Aquesta distribució posterior ( $p(\theta|D)$ ) pot ser obtinguda com el producte de la distribució *a priori* ( $\pi(\theta)$ ) que representa el coneixement previ del paràmetre i la probabilitat d'observar uns valors particulars de les dades empíriques donats uns valors dels paràmetres, és a dir la versemblança ( $p(D|\theta)$ ).

El desenvolupament de la teoria neutralista de l'evolució molecular (Kimura 1968; 1980) i de la teoria de la coalescència (Kingman 1982; Hudson 1990) han revolucionat les bases dels càlculs de probabilitat en els models genealògics en dos punts clau: l'ús d'un conjunt d'estadístics per resumir les dades empíriques ( $S(x)$ ) i l'ús de simulacions de Monte Carlo, evitant la necessitat de calcular la funció de probabilitat. L'algorisme bàsic de l'ABC (Tavaré et al. 1997; Pritchard et al.

1999; Beaumont et al. 2002; Marjoram et al. 2003) és:  $p(\theta | S(x)) = \pi(\theta) * p(S(x) | \theta)$ . La metodologia de l'ABC es podria, per tant, dividir en aquests 4 punts.

1. Mostrejar el paràmetre  $\theta$  de la seva distribució *a priori*  $\longrightarrow \theta_i \sim \pi(\theta)$
2. Simular dades (D) donats uns certs paràmetres ( $\theta$ )  $\longrightarrow D_i \sim p(D | \theta)$
3. Resumir les dades ( $D_i$ ) en un conjunt d'estadístics  $\longrightarrow S(x)$
4. Acceptar aquelles simulacions on els estadístics  $\longrightarrow |S(x) - S(y)| < \delta$

simulats s'assemblin més als observats  $S(y)$ ,

en funció d'una distància euclidiana ( $\delta$ )

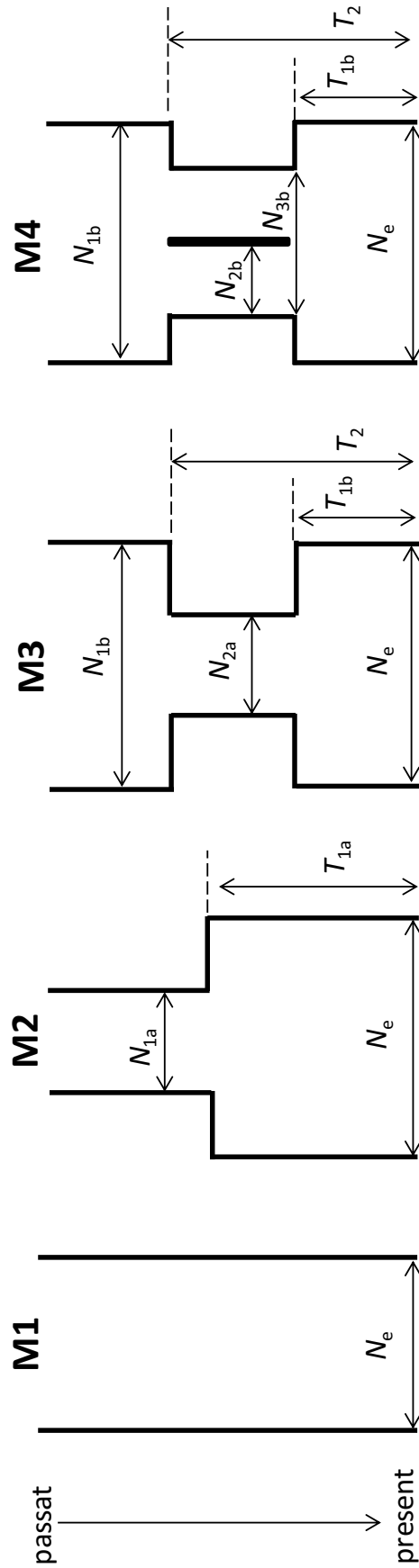
#### **Dades observades**

Es recull la informació de la variabilitat nucleotídica de les seqüències de DNA, en la mitjana i la variància d'un conjunt d'estadístics resum que s'anomenen estadístics observats.

#### **Disseny de models**

Per tal de dissenyar els models a contrastar, es va recopilar informació de la bibliografia tant pel què fa a hipòtesis prèvies de la història evolutiva de *D. subobscura* (Rozas et al. 1991; Menozzi i Krimbas 1992; Pascual et al. 2007) com de factors ambientals que podrien haver afectat la seva demografia (Van Andel i Tzedakis 1996; Hewitt 2000, 2004). Es van plantejar quatre models tenint en compte que aquests sempre seran una simplificació de la història evolutiva real de l'espècie. Els models evolutius estan caracteritzats per una sèrie de paràmetres definits per les distribucions *a priori* que són les que engloben la informació prèvia disponible. El primer model (M1) és un model neutre estàndard que s'utilitza com a hipòtesi nul·la, en el qual l'espècie no ha patit canvis demogràfics al llarg de la seva història. Els altres tres models (M2, M3, M4) descriuen una sèrie de canvis demogràfics en determinats períodes de temps. Aquests models es van dissenyar sota la premissa que els canvis climàtics produïts durant l'Holocè i una part del Pleistocè (Hewitt 2004) podrien haver modelat la variabilitat nucleotídica observada actualment a l'espècie. El model M2 planteja una expansió poblacional, escenari evolutiu que ja havia estat proposat com

el més plausible en estudis que analitzen la variabilitat nucleotídica en una única regió nuclear en relació al polimorfisme per inversions cromosòmiques (Rozas i Aguadé 1990; Sánchez-Gracia et al. 2011). El model M3 és un model de coll d'ampolla en el qual primer es produeix una reducció poblacional i posteriorment una expansió. De fet, Castro et al. (2010) proposen aquest model com el més plausible en relació al polimorfisme nucleotídic observat en l'estudi d'un fragment de 942 parells de bases d'un gen mitocondrial. L'últim model proposat (M4) és un model de barreja de dues subpoblacions en el qual primer es produeix un coll d'ampolla associat a un fenomen de subdivisió poblacional i posteriorment es produeix la barreja de les dues subpoblacions i la seva expansió. Aquest model, que anomenarem abreviadament model de barreja, es va dissenyar considerant la possible existència de múltiples refugis glacials a Europa durant el Pleistocè i l'Holocè (Van Andel et al. 1996; Hewitt 2000). En el cas de *D. subobscura*, Menozzi i Krimbas (1992), a través de l'anàlisi de diferents ordenacions cromosòmiques, descriuen com l'existència de refugis glacials a Europa hauria pogut perfilar la història evolutiva de l'espècie. En la figura M3.2 hi ha una representació dels diferents models en estudi.



**Figura M3.2:** Representació esquemàtica dels diferents models analitzats. M1, Model neutre estàndard. M2, Model d'expansió. M3, Model de coll d'ampolla. M4, Model de barreja.  $N_e$ , grandària efectiva actual de l'espècie;  $N_x$ , ràtio entre la grandària efectiva quan es produeix el canvi i la grandària efectiva actual.  $T_x$ , temps expressats en  $4N_e$  generacions en el qual es produeix el canvi poblacional (els subíndex  $x$ , representen els diferents temps en els quals es produeixen els canvis poblacionals).

Les diferents distribucions *a priori* que defineixen els paràmetres de cada model demogràfic, estan dissenyades en funció de la grandària efectiva actual de l'espècie ( $N_e$ ). A partir de dades recopilades de la bibliografia, es va dissenyar una distribució *a priori* uniforme de la grandària efectiva de  $1 \times 10^6$  (Pascual et al. 2007) a  $1 \times 10^7$  individus (Comeron 1997). En la Taula M3.1 hi ha un resum dels diferents paràmetres que conformen els models evolutius en estudi i de les distribucions *a priori* que defineixen aquests paràmetres.

Taula M3.1. Resum dels diferents paràmetres que conformen els models en l'estudi d'ABC.

Símbol	Definició	Paràmetre	Distribució <i>a priori</i>	Model
$\theta$	Paràmetre de mutació	$4N_e\mu$	U 0,01 – 0,1	M1, M2, M3, M4
R	Paràmetre de recombinació	$4N_e r^*$	U 0,001 – 0,4	M1, M2, M3, M4
$N_{1a}$	Grandària ancestral	$N_{1a}/N_e$	U 0,001 – 0,1	M2
$N_{1b}$	Grandària ancestral prèvia al coll d'ampolla	$N_{1b}/N_e$	U 0,1 – 2,0	M3, M4
$N_{2a}$	Reducció durant el coll d'ampolla	$N_{2a}/N_e$	U 0,001 – 0,1	M3
$N_{3b}$	Reducció durant la subdivisió poblacional	$N_{3b}/N_e$	U 0,001 – 0,1	M4
$N_{2b}$	Subpoblació	$N_{2b}/N_{3b}$	U 0,01 – 0,5	M4
$T_{1a}$	Temps d'inici de l'expansió (1.500 - 190.000 anys)	$T_{1a}/4N_e$	$\sim \exp(-)$ 0,0001 – 0,2	M2
$T_{1b}$	Temps d'inici de la recuperació del coll d'ampolla (1.500 - 130.000 anys)	$T_{1b}/4N_e$	$\sim \exp(-)$ 0,0001 – 0,16	M3, M4
<i>Dur</i>	Duració del coll d'ampolla (20.000 - 60.000 anys)	$Dur/4N_e$	$\sim \exp(-)$ 0,002 – 0,07	M3, M4
$T_2$	Temps d'inici del coll d'ampolla	$T_{1b} + Dur$	$\sim \exp(-)$ 0,002 – 0,2	M3, M4

U: distribució *a priori* uniforme.  $\exp(-)$ : distribució *a priori* exponencial negativa. Els models plantejats són: M1, Model neutre estàndard. M2, Model d'expansió. M3, Model de coll d'ampolla. M4, Model de barreja. Tots els paràmetres estan dissenyats en funció de la grandària efectiva actual de l'espècie i considerant cinc generacions per any.  $4N_e r^*$ : el paràmetre de recombinació en el cas de *Drosophila*, on no hi ha recombinació als mascles ve definit per l'expressió  $R = 2N_e r$ , essent  $r$  la recombinació a les femelles. La duració dels esdeveniments en aquells models que contempnen un coll d'ampolla (M3 i M4) reflecteix de forma aproximada la possible duració dels diferents períodes glacials i interglacials ocorreguts durant els últims 190.000 anys.

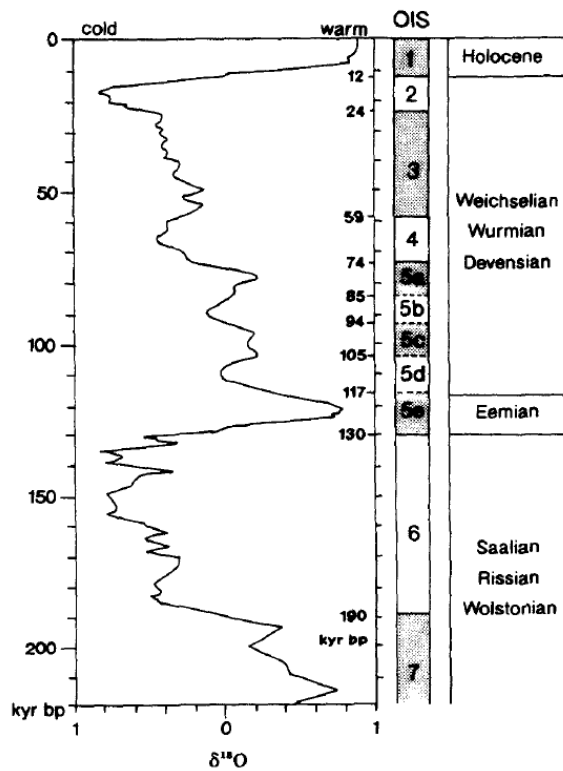
Els paràmetres poblacionals comuns a tots els models són: el paràmetre poblacional de mutació ( $\theta$ ), sent  $\theta = 4N_e\mu$  on  $\mu$  és la taxa de mutació per posició i per generació, i el paràmetre poblacional de recombinació ( $R$ ), sent en general  $R=4N_e r$  on  $r$  correspon a la taxa de recombinació per posició i per generació (Taula M3.1). La taxa de mutació per posició i per generació de cada regió ( $\mu$ ) es va estimar a partir de la divergència nucleotídica  $K$  (Nei, 1987) entre *D. subobscura* i *D. guanche* i del temps de divergència entre les dues espècies inferit prèviament en aquest treball considerant, en el cas d'espècies del grup *obscura* de *Drosophila*, cinc generacions per any. El paràmetre poblacional de recombinació ( $R$ ) per nucleòtid, va ser dissenyat mitjançant una distribució *a priori* uniforme amb un rang de valors ampli de manera que tingués en compte la possible variabilitat observada en les taxes de recombinació en espècies del grup *obscura* (Ortiz-Barrientos et al. 2006) i en les dades disponibles de l'espècie *D. subobscura* (Pegueroles et al. 2010). El paràmetre de recombinació es va incorporar a les simulacions de coalescència, però posteriorment no es va estimar la seva distribució posterior en el model més probable, ja que es va considerar com un paràmetre *nuisance*, és a dir un paràmetre que s'incorpora perquè els models siguin més reals però que no és d'interès de l'estudi (Lopes et al. 2010).

Els paràmetres demogràfics propis de cada model estan caracteritzats per fluctuacions en les grandàries poblacionals en determinats períodes de temps. Els canvis poblacionals ( $N_x$ ) dissenyats són instantanis i venen definits per distribucions *a priori* uniformes, corresponent a la ràtio entre la grandària poblacional quan es produeix el canvi i la grandària efectiva actual.

Els diferents paràmetres dels temps de coalescència en els quals es produeixen els canvis poblacionals ( $T_x$ ) s'expressen en unitats de  $4N_e$  generacions considerant cinc generacions per any. La metodologia que es va dur a terme per dissenyar les distribucions *a priori* pels paràmetres de temps ( $T_x$ ) segueix la següent seqüència. Primer es van dissenyar distribucions uniformes pels diferents temps en anys i es van passar a temps en generacions considerant cinc generacions per any ( $T_g$ ). Posteriorment, es van condicionar aquestes distribucions uniformes a la variació en la

grandària efectiva de l'espècie que també segueix una distribució *a priori* uniforme, aplicant la fórmula  $T_x = T_g / 4N_e$  (sent  $T_x$  el paràmetre de temps de coalescència expressat en  $4N_e$  generacions,  $T_g$  el temps expressat en generacions i  $N_e$  la grandària efectiva de l'espècie). Aquesta metodologia condueix a que aquestes distribucions *a priori* s'aproximin a distribucions exponencials negatives. Els temps abracen un període que en anys va des de 1.500 a 190.000, període del qual es tenen bons registres paleoambientals tant Europeus (Van Andel i Tzedakis 1996) com Nord Africans (Castañeda et al. 2009). Per definir els temps es va seguir l'escala bàsica de les corbes de *SPECMAN* (Imbrie et al. 1984) (Figura M3.3), una estratigrafia global dels isòtops d'oxigen calibrats amb els canvis periòdics de la geometria orbital de la Terra, que possibilita unes aproximacions raonables de les oscil·lacions climàtiques globals definides per períodes càlids (interglacials) i freds (glacials). Aquests canvis climàtics ocorreguts durant una part del Pleistocè i l'Holocè es correspondrien en l'escala temporal als períodes humits i àrids ocorreguts al Nord d'Àfrica (Castañeda et al. 2009). Així doncs, ara fa aproximadament 190.000 anys, en el Pleistocè mitjà, Europa es troba immersa en un període glacial conegut com a glaciació de Riss, corresponent a l'estadi isotòpic 6 (*OIS6*). La glaciació de Riss finalitza ara fa aproximadament 130.000 anys (inici del Pleistocè superior), començant el període interglacial Eemià, que coincideix amb la base de l'estadi isotòpic 5 (*OIS5*). Els estadis isotòpics 4, 3 i 2 (*OIS4*, *OIS3* i *OIS2*) corresponen al període glacial Würmià, que correspon a la última glaciació del Pleistocè que s'inicia ara fa aproximadament 74.000 anys i finalitza al voltant d'ara fa 10.000 anys. Durant aquesta darrera glaciació, en l'estadi *OIS4* es produeix el primer major avançament de gel (ara fa aproximadament 65.000 anys) i també una fase càlida (*OIS3* ara fa aproximadament 39.000-36.000 anys) que precedeix al màxim glacial d'aquest últim període glacial ara fa aproximadament 20.000 anys (*OIS2*). Fa aproximadament 10.000 anys, finalitza aquesta última glaciació iniciant el període conegut com a Holocè (*OIS1*), que constitueix el darrer període interglacial i s'estén fins l'actualitat (Van Andel et al. 1996).





**Figura M3.3.** Cronostratigrafia global dels isòtops d'oxigen (*SPECMAN*) dels últims 200.000 anys a Europa (modificat d'Imbrie et al. 1984). A l'eix de les Y s'indica els anys que es mostren en mil·lennis (Kyr); a les abscisses s'indica l'Oxigen bentònic normalitzat ( $\delta^{18}\text{O}$ ). Els nombres simbolitzen els límits en els diferents períodes dels isòtops d'oxigen, mostrant en ombrejat els períodes càlids (Martison et al. 1987).

### Simulacions de Monte Carlo

Per cada model evolutiu es van simular un milió de conjunts de dades condicionades als paràmetres que el defineixen ( $D_i \sim p(D|\theta)$ ). Es van resumir aquestes dades simulades en la mitjana i la variància d'un conjunt d'estadístics: els estadístics simulats ( $S(x)$ ). Les simulacions de coalescència es van realitzar amb el programa *mlcoalsim2* (Ramos-Onsins et al. 2007) que permet realitzar simulacions de coalescència multilocus, en les quals la grandària mostral i el nombre de nucleòtids pot variar d'una regió a una altra d'acord amb la seva variació a les dades observades. *mlcoalsim2* és una modificació del programa *ms* (Hudson 2002) que realitza simulacions de Monte Carlo les quals tenen els seus fonaments en la teoria de la coalescència (Kingman 1982; Hudson 1990). Les simulacions de coalescència es van realitzar considerant la possibilitat de mutacions a la mateixa posició (*multiple hits*), calculada a través de la relació entre la taxa de transicions i

transversions. Es va utilitzar *D. guanche* com a espècie *outgroup* en les simulacions, incloent-hi el temps de divergència amb *D. subobscura* que es va estimar a partir de les anàlisis filogenètiques prèvies realitzades en aquest treball amb el programa BEAST v.1.6.2 (Drummond i Rambaut 2007).

### **Tria d'estadístics**

Per avaluar la informació que podien aportar els estadístics simulats sobre els diferents paràmetres que conformen els models, es va calcular la correlació entre cadascun del milió d'estadístics simulats d'un model i els valors dels paràmetres que el defineixen de cadascuna de les simulacions, mitjançant el coeficient de correlació de *Pearson* que mesura la relació lineal entre dues variables aleatòries quantitatives. Com més elevada sigui la correlació observada entre l'estadístic i el paràmetre, major serà la informació aportada per aquest estadístic (Clotault et al. 2011). Una vegada es van haver finalitzat aquestes anàlisis, es va escollir el conjunt d'estadístics més adequat pel conjunt dels models plantejats. Aquests estadístics van ser utilitzats en les diferents anàlisis d'ABC.

### **Distribucions predictives posteriors**

Abans de comparar els models entre ells, és important verificar que aquests models siguin capaços de proporcionar una bona aproximació de les dades observades, doncs si no fos així, les estimes de les probabilitats posteriors dels diferents models podrien estar esbiaixades (Gelman et al. 2003). La metodologia emprada va ser la següent. Primerament, i amb el conjunt d'estadístics escollit per la tria de models, es va realitzar per cada model una anàlisi d'ABC seguint el mètode de rebuig (Pritchard et al. 1999) amb posterior correcció basada en regressió lineal local (Beaumont et al. 2002). Per cadascun dels models en estudi es van estimar les distribucions posteriors dels diferents paràmetres que els defineixen, retenint del milió de simulacions totals les 10.000 simulacions en les quals el seu conjunt d'estadístics simulats fos més similar al conjunt d'estadístics observats. Posteriorment, per cada model plantejat es van realitzar 10.000

simulacions predictives posteriors utilitzant com a distribucions *a priori* el conjunt de paràmetres associats a les simulacions retingudes en l'anàlisi anterior. Es va resumir la informació continguda en les simulacions predictives posteriors en el mateix conjunt d'estadístics utilitzat en les anàlisis anteriors: estadístics simulats *a posteriori*. Finalment, es va avaluar si els estadístics simulats *a posteriori* eren capaços de reproduir les dades observades. Es va calcular la significació estadística mitjançant la comparació dels estadístics observats i els estadístics simulats *a posteriori*, utilitzant un 5% del tipus d'error I pels dos talls de la distribució, mitjançant l'*script output\_selector* disponible a la pàgina web <http://bioinformatics.cragenomica.es/numgenomics/people/sebas/software/software.html>.

### Tria de models

Un dels punts clau dels mètodes basats en inferències bayesianes, és que aquests també poden ser aplicats a la tria de models, sent una metodologia en principi més senzilla que la utilitzada per la inferència dels paràmetres. El mecanisme directe de la tria de models va ser introduït per Weiss i Von Haessler (1998) i Pritchard et al. (1999). La idea bàsica de la tria de models seria la següent: agrupar totes les simulacions generades pels diferents models, retenir només aquelles simulacions on els estadístics simulats s'assemblin més als observats en funció d'una distància euclidiana ( $|S(x)-S(y)| < \delta$ ) i aproximar la probabilitat posterior de cada model en funció de la fracció de simulacions retingudes de cadascun d'ells. Així doncs, donats una sèrie de models  $\mu_1, \mu_2, \dots, \mu_\Omega$  amb les seves probabilitats *a priori*  $\sum_i \pi(\mu_i) = 1$ , es pot calcular la probabilitat posterior  $p(\mu_i | S(x))$  de cadascun dels models i també la ràtio de probabilitats o el factor Bayesià entre dos models  $i$  i  $j$  com:

$$\frac{p(\mu_i | S(x))}{p(\mu_j | S(x))} \div \frac{p(\mu_i)}{p(\mu_j)} = \frac{p(S(x) | \mu_i)}{p(S(x) | \mu_j)}$$

sent la part dreta de l'equació la ràtio entre les probabilitats marginals (per revisió, veure Beaumont 2010). Beaumont (2008) proposa el mètode logístic de regressió multinomial per estimar les probabilitats posteriors de forma directa. Una vegada es van haver obtingut les

probabilitats posteriors de cada model, es va avaluar el suport relatiu d'un model respecte a un altre mitjançant factors bayesians, és a dir calculant la ràtio de probabilitats entre dos models.

El mètode logístic de regressió multinomial (Beaumont et al. 2008) es va aplicar mitjançant la funció *postpr* disponible en la biblioteca *abc* del paquet informàtic d'R (R Development Core Team 2011). El nombre de simulacions que es va utilitzar en la tria de models va ser d'un milió i el llindar de tolerància escollit va ser de 0,01, és a dir, del milió total de simulacions es van retenir les 10.000 simulacions en les quals el seu conjunt d'estadístics simulats fos més similar al conjunt d'estadístics observats.

#### **Control de qualitat en la tria de models**

Es va realitzar un control de qualitat per avaluar si els diferents models es podien distingir entre sí, a través d'una anàlisi de validació creuada (Hastie et al. 2009). S'escull una simulació a l'atzar d'un dels models proposats, anomenada simulació de validació i el conjunt d'estadístics associats a aquesta simulació són utilitzats com a conjunt d'estadístics *pseudo-observats*. Aquest conjunt d'estadístics *pseudo-observats* es compara amb el total dels estadístics simulats dels altres models i també amb la resta d'estadístics del model que en aquell moment genera el conjunt d'estadístics *pseudo-observats*. Aquest procés es repeteix 1000 vegades per cadascun dels models en estudi. S'esperaria que si la tria de models hagués estat correcta, aquesta anàlisi assignaria una major probabilitat al model que en aquell moment estigués generant els estadístics *pseudo-observats*. Aquesta anàlisi proporciona una matriu de confusió (Hastie et al. 2009) on a cada fila de la matriu es veu representat el nombre de simulacions assignades al model correcte i el nombre de simulacions assignades incorrectament. Les anàlisis de validacions creuades es van realitzar mitjançant la funció *cv4postpr* disponible en la biblioteca *abc* del paquet informàtic d'R (R Development Core Team 2011).

### Inferència dels paràmetres del model més probable

Per estimar les distribucions posteriors dels diferents paràmetres [ $p(\theta|S(x))=\pi(\theta) p(S(x)|\theta)$ ], es va seguir el mètode de rebuig (Tavaré et al. 1997) amb correcció basada en regressió lineal local (Beaumont et al. 2002), transformant els paràmetres a escala logarítmica i amb correcció per heteroscedasticitat. El nombre de simulacions que es va utilitzar va ser d'un milió i el llindar de tolerància escollit ( $|S(x)-S(y)|<\delta$ ) va ser de 0,001. Es va utilitzar la funció *abc*, disponible en la biblioteca *abc* d'R (R Development Core Team 2011), mitjançant el mètode *loclinear*.

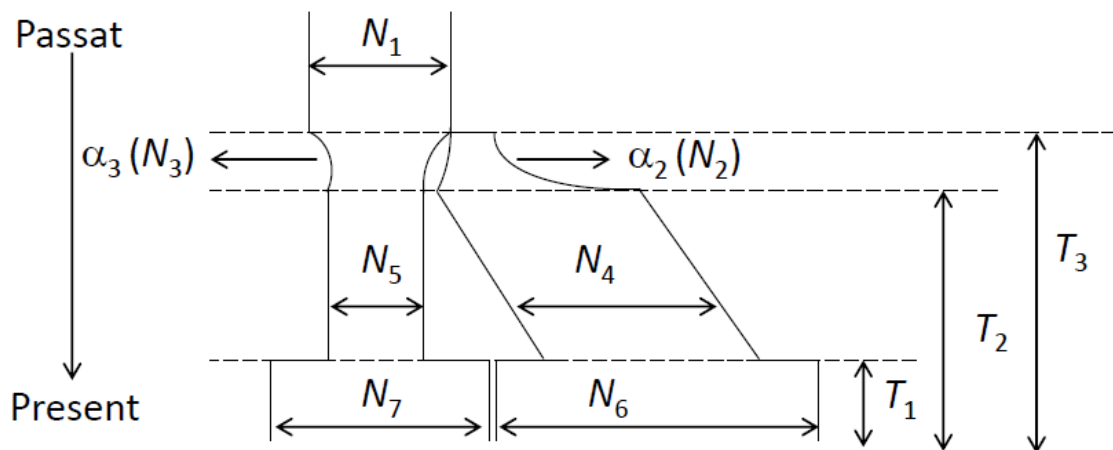
### Control de qualitat en la inferència dels paràmetres: espectre de freqüències desdoblament simulat *a posteriori*

Per avaluar si el model més probable era capaç de reproduir les dades observades, es va comparar l'espectre de freqüències desdoblament observat amb un espectre de freqüències simulat *a posteriori*. Per dissenyar l'espectre de freqüències simulat *a posteriori*, es van realitzar 10.000 noves simulacions *a posteriori* amb el programa *mlcoalsim 2* (Ramos-Onsins et al. 2007) utilitzant com a distribucions *a priori* el conjunt de paràmetres associats a les simulacions retingudes de l'anàlisi anterior, considerant que totes les regions en estudi tenien una grandària mostral de 18 individus (veure apartat M3.2).

### M3.4 Estima de l'edat de la inversió $J_1$ mitjançant una anàlisi d'ABC

Per tal d'estimar l'edat de la inversió  $J_1$ , es proposa un model senzill que té en compte l'origen únic d'una inversió i la rapidesa amb que la nova ordenació assoleix la freqüència d'equilibri, així com l'efecte reductor de la recombinació que tenen les inversions cromosòmiques. Es va fer una anàlisi d'*Approximate Bayesian Computation* (ABC), utilitzant el concatenat de les regions p222 i p127 (per un total de 2,2 kb, excloent les posicions amb *gaps*), sent aquestes regions les més properes als punts de trencament i les úniques que no presenten tractes d'intercanvi genètic entre ordenacions (veure els Resultats R4). Les simulacions de coalescència es van realitzar amb el programa *msABC* (Pavlidis et al. 2010), considerant absència

de recombinació ja que en les regions considerades aquesta es veu fortament reduïda en els individus heterocariotípics. Es va realitzar un milió de simulacions pel model demogràfic proposat en aquest treball (figura M3.4). Els estadístics emprats en l'anàlisi van ser la diversitat nucleotídica per locus ( $\pi$ , Tajima 1983), l'estadístic  $D$  de Tajima (1989), l'estadístic  $Z_{ns}$  (Kelly 1997) i l'estadístic  $F_{ST}$  (Hudson et al. 1992). Aquesta anàlisi es va realitzar mitjançant la biblioteca *abc* d'R (R Development Core Team 2011), aplicant l'algorisme de rebuig amb posterior correcció per regressió lineal (Beaumont et al. 2002) amb un llindar de tolerància de 0,001.



**Figura M3.4.** Representació esquemàtica del model dissenyat per estimar l'edat de la inversió  $J_1$ .  $N_e$ , grandària efectiva actual de la subpoblació amb l'ordenació  $J_1$ ;  $N_7$ , grandària efectiva actual de la subpoblació amb l'ordenació  $J_{st}$ .  $N_x$ , ràtio entre la grandària efectiva quan es produeix el canvi i la grandària efectiva actual de l'ordenació  $J_{st}$ .  $T_x$ , temps expressat en  $4N_e$  generacions en el qual es produeix el canvi poblacional (els subíndex  $x$  representen els diferents temps en els quals es produeixen els canvis poblacionals). Els diferents paràmetres estan calculats en funció de la grandària efectiva actual de l'ordenació  $J_{st}$  ( $N_7$ ) i considerant cinc generacions per any.

Es va dissenyar un model demogràfic en el que en un temps  $T_3$  s'origina la inversió  $J_1$  en un únic cromosoma, incrementant l'ordenació  $J_1$  ràpidament de freqüència comportant-se, per tant, com un procés d'escombrada selectiva parcial (Maynard-Smith i Haigh 1974) i dividint a la població en dues subpoblacions ( $J_1$  i  $J_{st}$ ) com a resultat de la manca d'intercanvi genètic entre elles en les regions properes als punts de trencament. La distribució *a priori* que defineix el paràmetre  $T_3$  avarca un rang de 600.000 a 200.000 anys. El llindar superior de la distribució *a priori* d'aquest paràmetre es va dissenyar en funció de l'estima del temps de divergència entre *D. subobscura* i *D.*

*madeirensis* inferida per Ramos-Onsins et al. (1998) de 600.000 anys. Les espècies *D. guanche* i *D. madeirensis* estretament relacionades amb *D. subobscura* presenten pel segment invertit l'ordenació estàndard ( $J_{st}$ ), la qual cosa fa pressuposar que la formació de la inversió  $J_1$  hauria de ser posterior a la divergència entre *D. subobscura* i *D. madeirensis*. El límit inferior d'aquesta distribució es va calcular seguint la metodologia proposada a Rozas et al. (1999).

Considerant que la inversió  $J_1$  proporciona una avantatge selectiva als individus portadors, la subpoblació  $J_1$  ( $N_2$ ) incrementa ràpidament de freqüència de forma exponencial, fins a la seva freqüència d'equilibri (Hartl i Clark 2007). L'expansió de l'ordenació o subpoblació  $J_1$  comporta la reducció exponencial complementària de l'ordenació  $J_{st}$ . Basant-nos en les dades publicades a Krimbas (1993) en relació a les freqüències observades de les dues ordenacions ( $J_1$  i  $J_{st}$ ) en l'àrea de mostreig, es va dissenyar una distribució *a priori* uniforme en què la freqüència de l'ordenació o subpoblació  $J_{st}$  varia de 0,3 a 0,4 sent complementària a la de l'ordenació o subpoblació  $J_1$ . Com a últim esdeveniment, s'incorpora una expansió poblacional inferida de les anàlisis d'ABC realitzades prèviament en aquest treball a partir de les 16 regions del cromosoma J no afectades per inversions (veure Resultats R5). Com a distribucions *a priori* pels paràmetres  $T_1$ ,  $N_6$  i  $N_7$ , es va utilitzar una aproximació del rang inferit dels paràmetres que descriuen el model d'expansió estimat com el més probable (M2).

En la Taula M3.2 hi ha un resum dels diferents paràmetres del model dissenyat que considera l'origen i manteniment d'una inversió i de les distribucions *a priori* que defineixen aquests paràmetres. Una vegada estimada la distribució posterior del paràmetre d'interès  $T_3$ , es va realitzar com a control de qualitat d'aquesta estima una anàlisi de validacions predictives posteriors seguint la mateixa metodologia utilitzada en les inferències del model demogràfic més probable per l'espècie *D. subobscura*.

Taula M3.2. Resum dels diferents paràmetres del model que considera l'origen i manteniment d'una inversió.

Símbol	Definició	Paràmetre	Distribució <i>a priori</i>
$\theta$	Paràmetre de mutació	$\theta = 4N_7\mu * nsites$	p222: 24,0 – 37,2 p127: 31,7 – 49,1
$N_e$	Grandària efectiva actual	$N_e$	$U 8,0 \times 10^6 - 9,6 \times 10^6$
freq	Freqüència actual de les ordenacions	freq $J_{st}$ freq $J_1$	$J_{st}$ : U 0,31 – 0,4 $J_1$ : U (1- freq $J_{st}$ )
$N_7$	Grandària efectiva actual de l'ordenació $J_{st}$	$N_e * \text{freq } J_{st}$	$2,4 \times 10^6 - 3,8 \times 10^6$
$N_6$	Grandària efectiva actual de l'ordenació $J_1$	$(N_e * \text{freq } J_1) / N_7$	1,5 – 2,2
$N_1$	Grandària ancestral	$(N_e * 0.08) / N_7$	0,2 – 0,26
$N_5$	Manteniment de l'ordenació $J_{st}$	$(N_1 * \text{freq } J_{st}) / N_7$	0,08
$N_4$	Manteniment de l'ordenació $J_1$	$(N_1 * \text{freq } J_1) / N_7$	0,12 – 0,18
$\alpha_2$	Formació i expansió de l'inversió $J_1$	$N_2 = N_7 * \exp^{-\alpha T_2}$ $N_2 = 1/N_7$	48,9 – 23,1
$\alpha_3$	Reducció de l'ordenació $J_{st}$ proporcional a l'expansió $J_1$	$(N_1 * \text{freq } J_{st}) = N_7 * \exp^{\alpha T_2}$	-38,6 – -8,4
$T_3$	Temps de formació de la inversió $J_1$ (200.000 - 600.000 anys)	$T_3/4N_7$	0,06535 – 0,300
$Dur$	Duració de l'expansió de l'ordenació $J_1$ (5-200 generacions)	$Dur/4N_7$	$7,9 \times 10^{-6} - 6,2 \times 10^{-6}$
$T_2$	Temps d'arribada a la freqüència d'equilibri de les ordenacions (199.800 - 599.995 anys)	$T_3 - Dur$	0,06 – 0,30
$T_1$	Temps d'inici de l'expansió (100.000 - 120.000 anys)	$T_1/4N_7$	0,039 – 0,060

U: distribució uniforme. En ombrejat es mostren les distribucions *a priori* que s'han utilitzat per dissenyar el model però que no s'inclouen com a paràmetre en les simulacions. Els temps s'expressen en unitats de  $4N_7$  generacions, sent  $N_7$  la grandària efectiva de la subpoblació  $J_{st}$  i considerant cinc generacions per any. La taxa de mutació per posició i per generació ( $\mu$ ) es considera de  $2,51 \times 10^{-9}$  (veure resultats, Taula R5.2). El paràmetre de mutació està dissenyat tenint en compte la longitud de la regió en estudi (*nsites*).  $N_e$ , grandària efectiva de *D. subobscura* estimada en aquest treball (veure Resultats, Taula R5.8).



### **M3.5 Detecció d'una regió candidata de ser diana de la selecció**

La disponibilitat del model demogràfic més probable, va permetre obtenir prediccions sobre els nivells i patrons de variabilitat en absència de selecció i, per tant, va ser possible detectar una regió la variabilitat de la qual no pot ser explicada únicament mitjançant la història demogràfica de la espècie. La significació estadística es va obtenir a partir de la distribució obtinguda mitjançant simulacions de coalescència (10.000 rèpliques), utilitzant com a distribucions *a priori* les simulacions retingudes en el model més probable. Les simulacions de coalescència es van realitzar amb el programa mlcoalsim2 (Ramos-Onsins et al. 2007), considerant la possibilitat de mutacions recurrents (*multiple hits*) i utilitzant *D. guanche* com a espècie *outgroup*.

MS4. MATERIAL SUPLEMENTARI

MS4.1 Encebadors d'amplificació i seqüenciació

Taula MS4.1.a. Encebadors d'amplificació i de seqüenciació en l'espècie *D. subobscura* i *D. guanche* corresponents a les 24 regions en estudi.

Regió genòmica	espècie	Amplificació		Seqüenciació	
		Forward/Reverse	Forward	Reverse	Reverse
p18	<i>D. subobscura</i>	p18AI 5' GACACACGAGGAAACACATC 3'	p18AI 5' GACACACGAGGAAACACATC 3'	p18AI 5' TGGGTACGGTAAGTGGACAC 3'	p18AI 5' TGGGTACGGTAAGTGGACAC 3'
		p18AII 5' TGGGTACGGTAAGTGGACAC 3'	p18s1 5' GCGAAACAGAGGGACTT 3'	p18s4a 5' CACTGGCTTTAGCGATCC 3'	p18s4a 5' CACTGGCTTTAGCGATCC 3'
			p18s2 5' GTCTATTGGGGCTTGAG 3'	p18s4b 5' TGGAAAGGGCAGACAATAG 3'	p18s4b 5' TGGAAAGGGCAGACAATAG 3'
	<i>D. guanche</i>	p18AI 5' GACACACGAGGAAACACATC 3'	p18AI 5' GACACACGAGGAAACACATC 3'	p18AI 5' TGGGTACGGTAAGTGGACAC 3'	p18AI 5' TGGGTACGGTAAGTGGACAC 3'
		p18AII 5' TGGGTACGGTAAGTGGACAC 3'	p18s1 5' GCGAAACAGAGGGACTT 3'	p18s4a 5' CACTGGCTTTAGCGATCC 3'	p18s4a 5' CACTGGCTTTAGCGATCC 3'
			p18s2 5' GTCTATTGGGGCTTGAG 3'	p18s4b 5' TGGAAAGGGCAGACAATAG 3'	p18s4b 5' TGGAAAGGGCAGACAATAG 3'
p39	<i>D. subobscura</i>	p39AI 5' GGTCTTGTGGTCACTTATG 3'	p39AI 5' GGTCTTGTGGTCACTTATG 3'	p39AI 5' CGAGCAATTACAGCAAAACAG 3'	p39AI 5' CGAGCAATTACAGCAAAACAG 3'
		p39AII 5' CGAGCAATTACAGCAAAACAG 3'	p39s1 5' CACTTCCCGATGCTTTTC 3'	p39s4 5' GGCAAGTTGAATCGTTAA 3'	p39s4 5' GGCAAGTTGAATCGTTAA 3'
			p39s2 5' CGTCGGATAGAGGGATAG 3'	p39s5 5' GCATGTCAAGCGGTTAC 3'	p39s5 5' GCATGTCAAGCGGTTAC 3'
	<i>D. guanche</i>	p39AI 5' GGTCTTGTGGTCACTTATG 3'	p39AI 5' GGTCTTGTGGTCACTTATG 3'	p39AI 5' CGAGCAATTACAGCAAAACAG 3'	p39AI 5' CGAGCAATTACAGCAAAACAG 3'
		p39AII 5' CGAGCAATTACAGCAAAACAG 3'	p39s1 5' CACTTCCCGATGCTTTTC 3'	p39s4 5' GGCAAGTTGAATCGTTAA 3'	p39s4 5' GGCAAGTTGAATCGTTAA 3'
			p39s2 5' CGTCGGATAGAGGGATAG 3'	p39s5 5' GCATGTCAAGCGGTTAC 3'	p39s5 5' GCATGTCAAGCGGTTAC 3'
p108	<i>D. subobscura</i>	p108AI 5' TCTGGGCTATTGGGTACA 3'	p108AI 5' TCTGGGCTATTGGGTACA 3'	p108AI 5' AATCGCTGAGACAATAATGG 3'	p108AI 5' AATCGCTGAGACAATAATGG 3'
		p108AII 5' AATCGCTGAGACAATAATGG 3'	p108s1 5' AGGCACACGACGACAATG 3'	p108s4a 5' GTACTCCCTATCCGACTC 3'	p108s4a 5' GTACTCCCTATCCGACTC 3'
			p108s2 5' ACCTGCTTCCGCTACTG 3'	p108s4b 5' AGCCCGTACTGAAACATC 3'	p108s4b 5' AGCCCGTACTGAAACATC 3'
	<i>D. guanche</i>	p108AI 5' TCTGGGCTATTGGGTACA 3'	p108AI 5' TCTGGGCTATTGGGTACA 3'	p108AI 5' AATCGCTGAGACAATAATGG 3'	p108AI 5' AATCGCTGAGACAATAATGG 3'
		p108AII 5' AATCGCTGAGACAATAATGG 3'	p108s1 5' AGGCACACGACGACAATG 3'	p108s4a 5' GTACTCCCTATCCGACTC 3'	p108s4a 5' GTACTCCCTATCCGACTC 3'
			p108s2 5' ACCTGCTTCCGCTACTG 3'	p108s4b 5' AGCCCGTACTGAAACATC 3'	p108s4b 5' AGCCCGTACTGAAACATC 3'
p127	<i>D. subobscura</i>	p127AI 5' GGACTAATCTCACCAACAC 3'	p127AI 5' GGACTAATCTCACCAACAC 3'	p127AI 5' CATCTCCGAGGTCAACAAC 3'	p127AI 5' CATCTCCGAGGTCAACAAC 3'
		p127AII 5' CATCTCCGAGGTCAACAAC 3'	p127s1 5' AGGTTTGGTTGGGTTTG 3'	p127s2 5' AGTGGATCTATTCGTGTG 3'	p127s2 5' AGTGGATCTATTCGTGTG 3'
	<i>D. guanche</i>	p127AI 5' GGACTAATCTCACCAACAC 3'	p127AI 5' GGACTAATCTCACCAACAC 3'	p127AI 5' CATCTCCGAGGTCAACAAC 3'	p127AI 5' CATCTCCGAGGTCAACAAC 3'
		p127AII 5' CATCTCCGAGGTCAACAAC 3'	p127s1 5' AGGTTTGGTTGGGTTTG 3'	p127s2 5' AGTGGATCTATTCGTGTG 3'	p127s2 5' AGTGGATCTATTCGTGTG 3'
p137	<i>D. subobscura</i>	p137AI 5' TGACTTTTTGGCTCCGACTC 3'	p137AI 5' TGACTTTTTGGCTCCGACTC 3'	p137AI 5' CGGTGTTTTCTGGGGTTCAT 3'	p137AI 5' CGGTGTTTTCTGGGGTTCAT 3'
		p137AII 5' CGGTGTTTTCTGGGGTTCAT 3'	p137s1 5' GGGAAATCGCTGGAATAT 3'	p137s5a 5' CGCCTAACTAAAGCATTG 3'	p137s5a 5' CGCCTAACTAAAGCATTG 3'
		(pensar que ni ha dos altres alternatius)	p137s2 5' CGGTGAGAGGAAGTAAAG 3'	p137s5b 5' AACCTTCCAGCAATAG 3'	p137s5b 5' AACCTTCCAGCAATAG 3'
	<i>D. guanche</i>	p137AI 5' TGACTTTTTGGCTCCGACTC 3'	p137AI 5' TGACTTTTTGGCTCCGACTC 3'	p137AI 5' CGGTGTTTTCTGGGGTTCAT 3'	p137AI 5' CGGTGTTTTCTGGGGTTCAT 3'
		p137AII 5' CGGTGTTTTCTGGGGTTCAT 3'	p137s1 5' GGGAAATCGCTGGAATAT 3'	p137s5a 5' CGCCTAACTAAAGCATTG 3'	p137s5a 5' CGCCTAACTAAAGCATTG 3'
			p137s2 5' CGGTGAGAGGAAGTAAAG 3'	p137s5b 5' AACCTTCCAGCAATAG 3'	p137s5b 5' AACCTTCCAGCAATAG 3'
p148	<i>D. subobscura</i>	p148AI 5' CCCCTGGCTCTTCTACTG 3'	p148AI 5' CCCCTGGCTCTTCTACTG 3'	p148AI 5' CGTTTGTGTGCTGCTTATGC 3'	p148AI 5' CGTTTGTGTGCTGCTTATGC 3'
		p148AII 5' CGTTTGTGTGCTGCTTATGC 3'	p148s1 5' ATGGTGTGGGTCGTATAG 3'	p148s5a 5' ACACAGACACACATAGC 3'	p148s5a 5' ACACAGACACACATAGC 3'
			p148s2a 5' ATGAAAGAGAGCGTAAAG 3'	p148s5b 5' TAGATAGAGAGGGTCAAG 3'	p148s5b 5' TAGATAGAGAGGGTCAAG 3'
	<i>D. guanche</i>	p148AI 5' CCCCTGGCTCTTCTACTG 3'	p148AI 5' CCCCTGGCTCTTCTACTG 3'	p148AI 5' CGTTTGTGTGCTGCTTATGC 3'	p148AI 5' CGTTTGTGTGCTGCTTATGC 3'
		p148AII 5' CGTTTGTGTGCTGCTTATGC 3'	p148s1 5' ATGGTGTGGGTCGTATAG 3'	p148s5a 5' ACACAGACACACATAGC 3'	p148s5a 5' ACACAGACACACATAGC 3'
			p148s2a 5' ATGAAAGAGAGCGTAAAG 3'	p148s5b 5' TAGATAGAGAGGGTCAAG 3'	p148s5b 5' TAGATAGAGAGGGTCAAG 3'

Taula MS4.1.b. Encebadors d'amplificació i de seqüenciació en l'espècie *D. subobscura* i *D. guanche* corresponents a les 24 regions en estudi.

Regió genòmica	espècie	Amplificació		Sequenciació						
		Forward/Reverse	Forward	Reverse	Reverse					
p169	<i>D. subobscura</i>	p169AI 5' ACTCATCACGGGCAGCACTC 3' p169AII 5' GGAACGAGCAGGACCAACAA 3'	p169AI 5' ACTCATCACGGGCAGCACTC 3'	p169AII 5' GGAACGAGCAGGACCAACAA 3'						
	<i>D. guanche</i>	p169AI 5' ACTCATCACGGGCAGCACTC 3' p169AII 5' GGAACGAGCAGGACCAACAA 3'	p169AI 5' ACTCATCACGGGCAGCACTC 3'	p169AII 5' GGAACGAGCAGGACCAACAA 3'						
p174	<i>D. subobscura</i>	p174AI 5' TGTGTGGCAAGTGCATATG 3' p174AI 5' TGTGTGGCAAGTGCATATG 3'	p174AI 5' TGTGTGGCAAGTGCATATG 3' p174s1 5' CGGACAACAACAGGCTAC 3' p174s2a 5' GGAAGGGGCTGCGTAACA 3' p174s2b 5' AAGTCAGCGGAAACATTG 3' p174s3 5' ACCTCAGCCAAGTTAATG 3'	p174AI 5' TGTGTGGCAAGTGCATATG 3' p174s4 5' CTACCGTGCATTATTG 3' p174s5 5' GCGATGGACGAGAGTAGC 3' p174s6 5' TTGTGGTTTGGCAATG 3'						
		<i>D. guanche</i>	p174GI 5' CAACAGCAGCCCCAGAACAG 3' p174GII 5' GCGTGAAGCCGTTGACATTG 3'	p174GI 5' CAACAGCAGCCCCAGAACAG 3' p174s1 5' CGGACAACAACAGGCTAC 3' p174s2a 5' GGAAGGGGCTGCGTAACA 3'	p174GII 5' GCGTGAAGCCGTTGACATTG 3' p174s4 5' CTACCGTGCATTATTG 3' p174s5 5' GCGATGGACGAGAGTAGC 3'					
			<i>D. subobscura</i>	p175AI 5' CACCACACAGCCAGAATGAG 3' p175AII 5' CACCACACAGCCAGAATGAG 3'	p175AI 5' CACCACACAGCCAGAATGAG 3' p175s1 5' GCAGCGTATAGAGCAGAC 3' p175s2a 5' ATCTGCGTTCCTTCATG 3' p175s2b 5' CATATTGTTCCGATCTCA 3'	p175AI 5' CACCACACAGCCAGAATGAG 3' p175s3 5' ATCGGGATGGCAGTTTAG 3' p175s4 5' GACTGCGTGTCTGTTG 3'				
	<i>D. guanche</i>			p175AI 5' CACCACACAGCCAGAATGAG 3' p175AII 5' CACCACACAGCCAGAATGAG 3'	p175AI 5' CACCACACAGCCAGAATGAG 3' p175s1 5' GCAGCGTATAGAGCAGAC 3' p175g1 5' CATATTGTTCCGATCTCA 3'	p175AI 5' CACCACACAGCCAGAATGAG 3' p175g2 5' TTCAGTATGCCGTAGTG 3' p175s4 5' GACTGCGTGTCTGTTG 3' p175g3 5' ATTCGCTCTGTTGACATC 3'				
		<i>D. subobscura</i>		p187AI 5' GAGCAACAACAGATCATAG 3' p187AII 5' CAGGGAGAGTGTTCTGTAAT 3'	p187AI 5' GAGCAACAACAGATCATAG 3' p187s1a 5' GGCTAGGCGATAAGTGAG 3' p187s1b 5' GTAAGCCAAAATTATAG 3' p187s1c 5' GCGTCATCTATGCGTTCT 3' p187s2 5' TTGTGTTTCCGTGTATCG 3' p187s3 5' CGAGGCGGATAAACAGA 3'	p187AII 5' CAGGGAGAGTGTTCTGTAAT 3' p187s4 5' TTGTGTTCCGTGTATCG 3' p187s5a 5' AAACGCACTACCACACAG 3' p187s5b 5' GCAGAACGCATAGATGAC 3' p187s6 5' GAATGCAGCACTGTCT 3'				
				<i>D. guanche</i>	p187AI 5' GAGCAACAACAGATCATAG 3' p187AII 5' CAGGGAGAGTGTTCTGTAAT 3'	p187g1 5' AGACAAGTGTGACATTC 3' p187g2 5' TTAGTTACACAGCCTTAG 3' p187g3 5' TGGTAGTGCGTTTCTAG 3'	p187s4 5' TTGTGTTCCGTGTATCG 3' p187s5a 5' AAACGCACTACCACACAG 3' p187s5b 5' GCAGAACGCATAGATGAC 3' p187s6 5' GAATGCAGCACTGTCT 3'			
			p196		<i>D. subobscura</i>	p196AI 5' CGCTGCTCGTCAAAATAG 3' p196AII 5' GCATTCAAACGCACTCAGC 3'	p196AI 5' CGCTGCTCGTCAAAATAG 3' p196s1a 5' GGGGATTAGTGTGAAAC 3' p196s1b 5' CAGTTCAACAGCCTCAAG 3' p196s2a 5' TCTCCACTTTTTATC 3' p196s2b 5' ACAGACTCTGCCTAACT 3' p196s2c 5' ACAGAGACAGTCCGATTCT 3' p196s2d 5' CACTGATTCTGGTTAGC 3' p196s3a 5' ACTTCTGTGAGCACTTC 3' p196s3b 5' CGCTAAACCAAAACAAGT 3'	p196AII 5' GCATTCAAACGCACTCAGC 3' p196s4a 5' TTTCCGACTTGACCTTGC 3' p196s4b 5' TCATCCACTAACCAACT 3' p196s4d 5' ACAATGTGGGGAACTG 3' p196s5a 5' CGTTCACTTCTCTTTG 3' p196s5b 5' AGGTCACTCACAATCTC 3' p196s6 5' GCTAACCAAGAATCAGT		
	<i>D. guanche</i>					p203AI 5' TAGGCTTGCAGAGAAACAT 3' p203AII 5' GTGCTCTGCTGCGATTCTC 3'	p203s1a 5' AAGGGAAACGCATACTAC 3' p203s1b 5' CTGCGAGAGCGAAACATG 3' p203s2 5' GATGATGTTCCGACTAG 3' p203s3a 5' GCGACGATTCTGTTCTCA 3' p203s3b 5' TTGCGTCTGCTTTGTTG 3' p203s4 5' ACACATAAACCAACAAC 3'	p203AII 5' GTGCTCTGCTGCGATTCTC 3' p203s5a 5' TATACGGTCAACCTGTTCT 3' p203s5b 5' CCCACAACCTAATGAGTC 3' p203s6 5' AGAGACAGCAGCGTAGA 3' p203s7a 5' CTGCCITTTGTTGACTTG 3' p203s7b 5' CAAATGTCTGGTAGTTCT 3'		
						<i>D. subobscura</i>	p204AI 5' TGCGGCAACGATGTGAAGAG 3' p204AII 5' TGCGGCAACGATGTGAAGAG 3'	p204AI 5' TGCGGCAACGATGTGAAGAG 3' p204s1a 5' TCCACCTGCTGATTCTC 3' p204s1b 5' TGCGGATTTTACAACAC 3' p204s2a 5' ATGCTATGTGGGCTCAAC 3' p204s2b 5' CAAACGCTGTGAACTGTG 3'	p204AI 5' TGCGGCAACGATGTGAAGAG 3' p204s3 5' ACACATCTGCTATTAG 3' p204s4 5' TTCTTAGCAGGCGATGTC 3'	
							<i>D. guanche</i>	p204AI 5' TGCGGCAACGATGTGAAGAG 3' p204AII 5' TGCGGCAACGATGTGAAGAG 3'	p204AI 5' TGCGGCAACGATGTGAAGAG 3' p204s1a 5' TCCACCTGCTGATTCTC 3' p204s2a 5' ATGCTATGTGGGCTCAAC 3'	p204AI 5' TGCGGCAACGATGTGAAGAG 3' p204s3 5' ACACATCTGCTATTAG 3' p204s4 5' TTCTTAGCAGGCGATGTC 3'

Taula MS4.1.c. Encebadors d'amplificació i de seqüenciació en l'espècie *D. subobscura* i *D. guanche* corresponents a les 24 regions en estudi.

Regió genòmica	espècie	Amplificació		Sequenciació	
		Forward/Reverse	Forward	Reverse	Reverse
p215	<i>D. subobscura</i>	p215AI 5' AACAGAGTCAGAGTCCAAGT 3'	p215AI 5' AACAGAGTCAGAGTCCAAGT 3'	p215AI 5' GGGAAAGAGAGGCATATAGG 3'	p215AI 5' GGGAAAGAGAGGCATATAGG 3'
		p215AII 5' GGGAAAGAGAGGCATATAGG 3'	p215s1a 5' TGAGAATGAGGGACACAC 3'	p215s4 5' CTGTGCTGCTGTTGTTCC 3'	p215s4 5' CTGTGCTGCTGTTGTTCC 3'
			p215s1b 5' TCCCGCTCAACTCCACTC 3'	p215s5a 5' TTCTTCTCCACTTCAC 3'	p215s5a 5' TTCTTCTCCACTTCAC 3'
			p215s1c 5' GACTAGCAGGATGACAAC 3'	p215s5b 5' AAGTAAACACAGCCAGAG 3'	p215s5b 5' AAGTAAACACAGCCAGAG 3'
			p215s1d 5' TCCCGCTCAACTCCACTC 3'	p215s6 5' GAAGTTTAGACGCCACAG 3'	p215s6 5' GAAGTTTAGACGCCACAG 3'
			p215s2 5' TTGAACTGTCTGCTCTC 3'	p215s7a 5' CGTGTGCCCTATTCTC 3'	p215s7a 5' CGTGTGCCCTATTCTC 3'
	<i>D. guanche</i>	p215AI 5' AACAGAGTCAGAGTCCAAGT 3'	p215AI 5' AACAGAGTCAGAGTCCAAGT 3'	p215AI 5' GGGAAAGAGAGGCATATAGG 3'	p215AI 5' GGGAAAGAGAGGCATATAGG 3'
		p215AII 5' GGGAAAGAGAGGCATATAGG 3'	p215s1a 5' TGAGAATGAGGGACACAC 3'	p215s4 5' CTGTGCTGCTGTTGTTCC 3'	p215s4 5' CTGTGCTGCTGTTGTTCC 3'
			p215s2 5' TTGAACTGTCTGCTCTC 3'	p215s5a 5' TTCTTCTCCACTTCAC 3'	p215s5a 5' TTCTTCTCCACTTCAC 3'
			p215s3a 5' TGAGGATAATGTACAAG 3'	p215s6 5' GAAGTTTAGACGCCACAG 3'	p215s6 5' GAAGTTTAGACGCCACAG 3'
			p215s3b 5' CCTCTGGTCTGTTTACT 3'	p215s7a 5' CGTGTGCCCTATTCTC 3'	p215s7a 5' CGTGTGCCCTATTCTC 3'
			p215s3c 5' TAAGGCTCACGAAATCAC 3'		
p220	<i>D. subobscura</i>	p220AI 5' GACTGGCTGGACACGGACAT 3'	p220AI 5' GACTGGCTGGACACGGACAT 3'	p220AI 5' ATGGATTTTCAGAGGCAACC 3'	p220AI 5' ATGGATTTTCAGAGGCAACC 3'
		p220AII 5' ATGGATTTTCAGAGGCAACC 3'	p220s1 5' ACTTTCTGCCCATAAATC 3'	p220s2 5' TAAGCGGCAATCGTAGAG 3'	p220s2 5' TAAGCGGCAATCGTAGAG 3'
	<i>D. guanche</i>	p220AI 5' GACTGGCTGGACACGGACAT 3'	p220AI 5' GACTGGCTGGACACGGACAT 3'	p220AI 5' ATGGATTTTCAGAGGCAACC 3'	p220AI 5' ATGGATTTTCAGAGGCAACC 3'
		p220AII 5' ATGGATTTTCAGAGGCAACC 3'	p220s1 5' ACTTTCTGCCCATAAATC 3'	p220s2 5' TAAGCGGCAATCGTAGAG 3'	p220s2 5' TAAGCGGCAATCGTAGAG 3'
p221	<i>D. subobscura</i>	p221AI 5' AACAGTTCAGGGTGGTAAGG 3'	p221AI 5' AACAGTTCAGGGTGGTAAGG 3'	p221AI 5' CTCTGTCTCTGTTGTGCC 3'	p221AI 5' CTCTGTCTCTGTTGTGCC 3'
		p221AII 5' CTCTGTCTCTGTTGTGCC 3'	p221s1 5' AGGGGAGGGGTCATAAAC 3'	p221s4 5' ACACCTGGCAGCAGCATC 3'	p221s4 5' ACACCTGGCAGCAGCATC 3'
			p221s2 5' AACTATGGCTCGGCACAG 3'	p221s5a 5' CGCTTCTCCATTACC 3'	p221s5a 5' CGCTTCTCCATTACC 3'
			p221s3 5' CACGACAAAAGCGAAAAG 3'	p221s5b 5' GGATGGTAGGAGCAAGAG 3'	p221s5b 5' GGATGGTAGGAGCAAGAG 3'
				p221s6 5' CGTAAGAGGGAGACAAC 3'	p221s6 5' CGTAAGAGGGAGACAAC 3'
	<i>D. guanche</i>	p221AI 5' AACAGTTCAGGGTGGTAAGG 3'	p221AI 5' AACAGTTCAGGGTGGTAAGG 3'	p221AI 5' CTCTGTCTCTGTTGTGCC 3'	p221AI 5' CTCTGTCTCTGTTGTGCC 3'
		p221AII 5' CTCTGTCTCTGTTGTGCC 3'	p221s1 5' AGGGGAGGGGTCATAAAC 3'	p221s4 5' ACACCTGGCAGCAGCATC 3'	p221s4 5' ACACCTGGCAGCAGCATC 3'
			p221s2 5' AACTATGGCTCGGCACAG 3'	p221s5a 5' CGCTTCTCCATTACC 3'	p221s5a 5' CGCTTCTCCATTACC 3'
			p221s3 5' CACGACAAAAGCGAAAAG 3'	p221s6 5' CGTAAGAGGGAGACAAC 3'	p221s6 5' CGTAAGAGGGAGACAAC 3'
p222	<i>D. subobscura</i>	p222AI 5' ATTCTTCTCTGGCGTCTG 3'	p222AI 5' ATTCTTCTCTGGCGTCTG 3'	p222AI 5' AGGCGACAGAAAAGGATGAG 3'	p222AI 5' AGGCGACAGAAAAGGATGAG 3'
		p222AII 5' AGGCGACAGAAAAGGATGAG 3'	p222s1 5' CTACACTACATCCGCTCG 3'	p222s2 5' CTGCTGCACAAACGGTAC 3'	p222s2 5' CTGCTGCACAAACGGTAC 3'
	<i>D. guanche</i>	p222AI 5' ATTCTTCTCTGGCGTCTG 3'	p222AI 5' ATTCTTCTCTGGCGTCTG 3'	p222AI 5' AGGCGACAGAAAAGGATGAG 3'	p222AI 5' AGGCGACAGAAAAGGATGAG 3'
		p222AII 5' AGGCGACAGAAAAGGATGAG 3'	p222s1 5' CTACACTACATCCGCTCG 3'	p222s2 5' CTGCTGCACAAACGGTAC 3'	p222s2 5' CTGCTGCACAAACGGTAC 3'
p227	<i>D. subobscura</i>	p227AI 5' GCATATCTGTGAGTCGTCTG 3'	p227AI 5' GCATATCTGTGAGTCGTCTG 3'	p227AI 5' CTAAGTATGTCGCCTCAA 3'	p227AI 5' CTAAGTATGTCGCCTCAA 3'
		p227AII 5' CTAAGTATGTCGCCTCAA 3'			
	<i>D. guanche</i>	p227AI 5' GCATATCTGTGAGTCGTCTG 3'	p227AI 5' GCATATCTGTGAGTCGTCTG 3'	p227AI 5' CTAAGTATGTCGCCTCAA 3'	p227AI 5' CTAAGTATGTCGCCTCAA 3'
		p227AII 5' CTAAGTATGTCGCCTCAA 3'	p227g1 5' AGTTTCAGAGTGGCTAAG 3'	p227g2 5' CACTGAGGCAAGGCTAAG 3'	p227g2 5' CACTGAGGCAAGGCTAAG 3'
p237	<i>D. subobscura</i>	p237AI 5' TCGGAGCGGAACATCAAGAC 3'	p237AI 5' TCGGAGCGGAACATCAAGAC 3'	p237AI 5' GGTTGACATTGGCAGTCT 3'	p237AI 5' GGTTGACATTGGCAGTCT 3'
		p237AII 5' GGTTGACATTGGCAGTCT 3'	p237s1 5' AGTCAGTCTAAGGAGTC 3'	p237s3 5' TTAGGCGATAGGTAGATG 3'	p237s3 5' TTAGGCGATAGGTAGATG 3'
			p237s2a 5' CGTTCGATCACAGACTT 3'	p237s4 5' ACAGACCACAAACAATAG 3'	p237s4 5' ACAGACCACAAACAATAG 3'
			p237s2b 5' TGTGGGTTTGGGAAGAG 3'		
	<i>D. guanche</i>	p237AI 5' TCGGAGCGGAACATCAAGAC 3'	p237AI 5' TCGGAGCGGAACATCAAGAC 3'	p237AI 5' GGTTGACATTGGCAGTCT 3'	p237AI 5' GGTTGACATTGGCAGTCT 3'
		p237AII 5' GGTTGACATTGGCAGTCT 3'	p237g1 5' GGAATGCCAAGGTTTTTA 3'	p237s3 5' TTAGGCGATAGGTAGATG 3'	p237s3 5' TTAGGCGATAGGTAGATG 3'
			p237g2 5' CAACGACCATAGGAGAG 3'	p237s4 5' ACAGACCACAAACAATAG 3'	p237s4 5' ACAGACCACAAACAATAG 3'
			p237g3 5' TTGACAATGCCAGAGAAC 3'	p237g4 5' ATAAAACCTTGGCATTCC 3'	p237g4 5' ATAAAACCTTGGCATTCC 3'
p248	<i>D. subobscura</i>	p248AI 5' TTCCTGTTCACTGGATAG 3'	p248AI 5' TTCCTGTTCACTGGATAG 3'	p248AI 5' TGAGAGCCATAGAATAACA 3'	p248AI 5' TGAGAGCCATAGAATAACA 3'
		p248AII 5' TGAGAGCCATAGAATAACA 3'	p248s1 5' CCTCCGATCATAAGAATA 3'	p248s4 5' CTCCGTTTTGCTGTTAC 3'	p248s4 5' CTCCGTTTTGCTGTTAC 3'
			p248s2 5' CACACAGATACGGGATAC 3'	p248s5 5' GTATCCCGTATCTGTATG 3'	p248s5 5' GTATCCCGTATCTGTATG 3'
	<i>D. guanche</i>	p248AI 5' TTCCTGTTCACTGGATAG 3'	p248AI 5' TTCCTGTTCACTGGATAG 3'	p248AI 5' TGAGAGCCATAGAATAACA 3'	p248AI 5' TGAGAGCCATAGAATAACA 3'
		p248AII 5' TGAGAGCCATAGAATAACA 3'	p248g1 5' TATAGTTCTGCGGGTAGC 3'	p248s4 5' CTCCGTTTTGCTGTTAC 3'	p248s4 5' CTCCGTTTTGCTGTTAC 3'
			p248g2 5' AGGACACTTTGGGACACA 3'	p248s5 5' GTATCCCGTATCTGTATG 3'	p248s5 5' GTATCCCGTATCTGTATG 3'
	p248g3 5' CGCTGTGCTGTTTACT 3'	p248g4 5' CAACGCCTCAATGGAATC 3'	p248g4 5' CAACGCCTCAATGGAATC 3'		
p271	<i>D. subobscura</i>	p271AI 5' CACATCCGCCAGCCATTAT 3'	p271AI 5' CACATCCGCCAGCCATTAT 3'	p271s3 5' GTGGAGCAGGAAGTGGAGTT 3'	p271s3 5' GTGGAGCAGGAAGTGGAGTT 3'
		p271AII 5' GTGGAGCAGGAAGTGGAGTT 3'	p271s1 5' AGGGTCTGTCTGCTAC 3'	p271s4a 5' GCAAGATTTCAGGGACTC 3'	p271s4a 5' GCAAGATTTCAGGGACTC 3'
			p271s2a 5' CCAAACAACGCAATC 3'	p271s4b 5' GCAAGATTTCAGGGACTC 3'	p271s4b 5' GCAAGATTTCAGGGACTC 3'
	<i>D. guanche</i>	p271AI 5' CACATCCGCCAGCCATTAT 3'	p271AI 5' CACATCCGCCAGCCATTAT 3'	p271s3 5' GTGGAGCAGGAAGTGGAGTT 3'	p271s3 5' GTGGAGCAGGAAGTGGAGTT 3'
		p271AII 5' GTGGAGCAGGAAGTGGAGTT 3'	p271g1 5' CAACCATGGCAATATC 3'	p271g2 5' AGACGGGAGGAATGAATC 3'	p271g2 5' AGACGGGAGGAATGAATC 3'

Taula MS4.1.d. Encebadors d'amplificació i de seqüenciació en l'espècie *D. subobscura* i *D. guanche* corresponents a les 24 regions en estudi.

Regió genòmica	espècie	Amplificació		Sequenciació	
		Forward/Reverse	Forward	Reverse	Reverse
p10849	<i>D. subobscura</i>	p10849AI 5' CCAGGATATTCTCGTTCTTC 3'	p10849AI 5' CCAGGATATTCTCGTTCTTC 3'	p10849s2 5' ATAGGTGTGGGATTGTAG 3'	p10849s3 5' CGAAGAGCCATTGTTTAC 3'
		p10849AII 5' GGTTTCAGGGCTTCTACATAG 3'	p10849s1 5' CGATTACCAGAAGCAGTT 3'		
	<i>D. guanche</i>	p10849AI 5' CCAGGATATTCTCGTTCTTC 3'	p10849AI 5' CCAGGATATTCTCGTTCTTC 3'	p10849s2 5' ATAGGTGTGGGATTGTAG 3'	p10849s3 5' CGAAGAGCCATTGTTTAC 3'
		p10849AII 5' GGTTTCAGGGCTTCTACATAG 3'	p10849s1 5' CGATTACCAGAAGCAGTT 3'		
p32096	<i>D. subobscura</i>	p32096AI 5' TGCTGCTGTTGTTCTGTGC 3'	p32096s1 5' TGCTACTGCTGCCTCATC 3'	p32096AI 5' AGGGTCTGTGTCGTCTAC 3'	p32096s4 5' TAAACAGTATGGCAACAC 3'
		p32096AII 5' AGGGTCTGTGTCGTCTAC 3'	p32096s2 5' CGTGGCATTGTTACCTAC 3'	p32096s5 5' ATTGGAAGTCCGGGTCAG 3'	p32096s6 5' CGCCATCAAGGTCAAAG 3'
			p32096s3 5' ACCATCACCTAAGAGTC 3'		
	<i>D. guanche</i>	p32096AI 5' TGCTGCTGTTGTTCTGTGC 3'	p32096s1 5' TGCTACTGCTGCCTCATC 3'	p32096AI 5' AGGGTCTGTGTCGTCTAC 3'	p32096s4 5' TAAACAGTATGGCAACAC 3'
		p32096AII 5' AGGGTCTGTGTCGTCTAC 3'	p32096s2 5' CGTGGCATTGTTACCTAC 3'	p32096s5 5' ATTGGAAGTCCGGGTCAG 3'	p32096s6 5' CGCCATCAAGGTCAAAG 3'
			p32096s3 5' ACCATCACCTAAGAGTC 3'		
p8593	<i>D. subobscura</i>	p8593AI 5' CGCCATCAGGAACATTGTAA 3'	p8593AI 5' CGCCATCAGGAACATTGTAA 3'	p8593AI 5' AGGACACTGTGAGCCATTG 3'	p8593s4 5' TCTGGTTGGGGATCTT 3'
		p8593AII 5' AGGACACTGTGAGCCATTG 3'	p8593s1 5' CACTATCGGTACGGCTAT 3'	p8593s5 5' CAAATGGAAGGGGGAATG 3'	
			p8593s2 5' TAACACGCACACGACAA 3'		
	<i>D. guanche</i>	p8593AI 5' CGCCATCAGGAACATTGTAA 3'	p8593AI 5' CGCCATCAGGAACATTGTAA 3'	p8593AI 5' AGGACACTGTGAGCCATTG 3'	p8593s4 5' TCTGGTTGGGGATCTT 3'
		p8593AII 5' AGGACACTGTGAGCCATTG 3'	p8593s1 5' CACTATCGGTACGGCTAT 3'	p8593s5 5' CAAATGGAAGGGGGAATG 3'	
			p8593s2 5' TAACACGCACACGACAA 3'		
		p8593s3 5' GCAACGCAACCGCAAGAG 3'			

## MS4.2 Condicions de PCR

Taula MS4.2.a. Condicions d'amplificació de les 24 regions en estudi en les espècies *D. subobscura* i *D. guanche*.

Regió Genòmica	PCR	<i>D. subobscura</i>	<i>D. guanche</i>
p18	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	56°C 2min 50 seg	56°C 2min
	Extensió final	65°C 5min	65°C 4min
	Conservació	4°C ∞	4°C ∞
p39	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	60°C 4 min	60°C 4 min
	Extensió final	65°C 8min	65°C 8min
	Conservació	4°C ∞	4°C ∞
p108	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	51°C 3 min 50 seg	51°C 3 min 50 seg
	Extensió final	65°C 6min	65°C 6min
	Conservació	4°C ∞	4°C ∞
p127	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	52°C 3 min 50 seg	52°C 3 min 50 seg
	Extensió final	65°C 6 min	65°C 6 min
	Conservació	4°C ∞	4°C ∞
p137	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	49°C 3 min 50 seg	49°C 3 min 50 seg
	Extensió final	65°C 6min	65°C 6min
	Conservació	4°C ∞	4°C ∞
p148	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	56°C 2 min 50 seg	54°C 3 min
	Extensió final	65°C 5 min	65°C 6min
	Conservació	4°C ∞	4°C ∞
p169	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	58°C 1 min	58°C 1 min
	Extensió final	65°C 2 min	65°C 2 min
	Conservació	4°C ∞	4°C ∞
p174	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	50°C 3 min	50°C 3 min
	Extensió final	65°C 5min	65°C 5min
	Conservació	4°C ∞	4°C ∞
p175	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	<i>Annealing</i>	46°C 10 seg	46°C 10 seg
	Extensió	56°C 2min 50 seg	56°C 2min
	Extensió final	65°C 5min	65°C 4min
	Conservació	4°C ∞	4°C ∞

Taula MS4.2.b. Condicions d'amplificació de les 24 regions en estudi en les espècies *D. subobscura* i *D. guanche*.

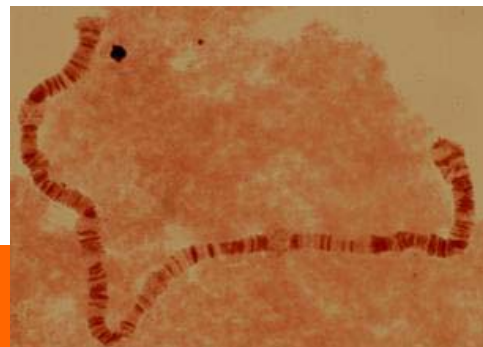
PCR		<i>D. subobscura</i>	<i>D. guanche</i>
<b>p187</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	54°C 2min 50 seg	54°C 3 min
	Extensió final	65°C 5min	65°C 6min
	Conservació	4°C ∞	4°C ∞
<b>p196</b>	Desnaturalització	94°C 3min	
	Desnaturalització	96°C 10 seg	
	Annealing	46°C 10 seg	
	Extensió	60°C 2min 50 seg	
	Extensió final	65°C 5 min	
	Conservació	4°C ∞	
<b>p203</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	60°C 4 min 50 seg	59°C 3min 30seg
	Extensió final	65°C 7 min	65°C 6min
	Conservació	4°C ∞	4°C ∞
<b>p204</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	56°C 2min 50 seg	50°C 2min 30seg
	Extensió final	65°C 5min	65°C 4min
	Conservació	4°C ∞	4°C ∞
<b>p215</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	56°C 2min 50 seg	54°C 3 min
	Extensió final	65°C 5min	65°C 6min
	Conservació	4°C ∞	4°C ∞
<b>p220</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	58°C 1 min	58°C 1 min
	Extensió final	65°C 2 min	65°C 2 min
	Conservació	4°C ∞	4°C ∞
<b>p221</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	61°C 2min	61°C 2min
	Extensió final	65°C 4min	65°C 4min
	Conservació	4°C ∞	4°C ∞
<b>p222</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	50°C 2 min	50°C 2 min
	Extensió final	65°C 4min	65°C 4min
	Conservació	4°C ∞	4°C ∞
<b>p227</b>	Desnaturalització	94°C 3min	94°C 3min
	Desnaturalització	96°C 10 seg	96°C 10 seg
	Annealing	46°C 10 seg	46°C 10 seg
	Extensió	54°C 2min 50 seg	54°C 2min 50 seg
	Extensió final	65°C 5 min	65°C 5 min
	Conservació	4°C ∞	4°C ∞

Taula MS4.2.c. Condicions d'amplificació de les 24 regions en estudi en les espècies *D. subobscura* i *D. guanche*.

PCR		<i>D. subobscura</i>	<i>D. guanche</i>	
<b>p237</b>	Desnaturalització	94°C 3min	94°C 3min	
	Desnaturalització	96°C 10 seg	96°C 10 seg	
	Annealing	30 cicles	46°C 10 seg	46°C 10 seg
	Extensió		58°C 2 min 50 seg	56°C 2min
	Extensió final	65°C 5min	65°C 4min	
	Conservació	4°C ∞	4°C ∞	
<b>p248</b>	Desnaturalització	94°C 3min	94°C 3min	
	Desnaturalització	96°C 10 seg	96°C 10 seg	
	Annealing	30 cicles	46°C 10 seg	46°C 10 seg
	Extensió		54°C 2min 50 seg mi 52°C 3min 30seg	
	Extensió final	65°C 5min	65°C 6min	
	Conservació	4°C ∞	4°C ∞	
<b>p271</b>	Desnaturalització	94°C 3min	94°C 3min	
	Desnaturalització	96°C 10 seg	96°C 10 seg	
	Annealing	30 cicles	46°C 10 seg	46°C 10 seg
	Extensió		53°C 1 min	53°C 1 min
	Extensió final	65°C 2min	65°C 2min	
	Conservació	4°C ∞	4°C ∞	
<b>p10849</b>	Desnaturalització	94°C 3min	94°C 3min	
	Desnaturalització	96°C 10 seg	96°C 10 seg	
	Annealing	30 cicles	46°C 10 seg	46°C 10 seg
	Extensió		50°C 2min	50°C 2min
	Extensió final	65°C 4min	65°C 4min	
	Conservació	4°C ∞	4°C ∞	
<b>p32096</b>	Desnaturalització	94°C 3min	94°C 3min	
	Desnaturalització	96°C 10 seg	96°C 10 seg	
	Annealing	30 cicles	46°C 10 seg	46°C 10 seg
	Extensió		61°C 4min 50 seg	61°C 4min 50 seg
	Extensió final	65°C 9min	65°C 9min	
	Conservació	4°C ∞	4°C ∞	
<b>p8593</b>	Desnaturalització	94°C 3min	94°C 3min	
	Desnaturalització	96°C 10 seg	96°C 10 seg	
	Annealing	30 cicles	46°C 10 seg	46°C 10 seg
	Extensió		50°C 2min	49°C 2 min
	Extensió final	65°C 4 min	65°C 4 min	
	Conservació	4°C ∞	4°C ∞	





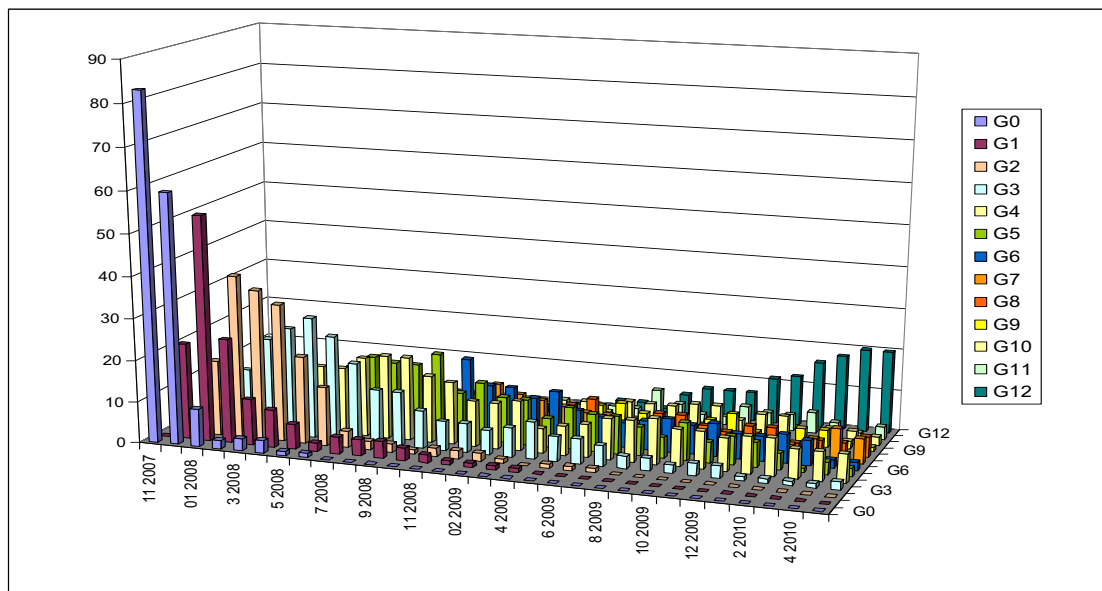


## RESULTATS



## R1. OBTENCIÓ DE LÍNIES HOMOCARIOTÍPIQUES PER AL CROMOSOMA J DE *D. SUBOBSCURA* A PARTIR D'UNA POBLACIÓ NATURAL DE BARCELONA

La mostra poblacional inicial va ser de 83 línies isomaternes a partir de les que es van realitzar almenys tretze encreuaments consecutius germà-germana per obtenir línies altament consanguínies. Aquests encreuaments es van efectuar durant un període de com a mínim dos anys en funció de la línia. En la Figura R1.1 s'observa la progressió temporal de la consanguinitat durant el període de Novembre de 2007 a Abril de 2010.

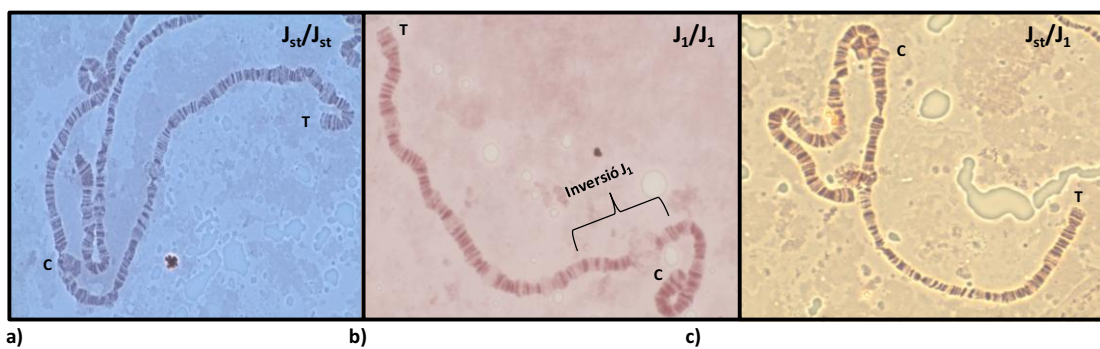


**Figura R1.1.** Progressió temporal del procés de consanguinitat. En l'eix de les X es mostra el període de temps expressat en mesos i anys (per ex: 11 2007 = Novembre de 2007) durant el qual s'efectuen els encreuaments germà-germana. En l'eix de les Y es mostra el nombre de línies isomaternes i en l'eix de les Z es mostra el nombre de generacions.

En la Figura R1.1 es veu reflectit com el procés d'obtenció de línies altament consanguínies en l'espècie *D. subobscura* és laboriós i costós. Durant el període representat, el nombre d'encreuaments els quals van tenir descendència viable va ser en promig d'un 54%. S'observa com a mesura que s'incrementa el nombre de generacions i, per tant l'homozigosi, la viabilitat de les línies disminueix indicant que hi ha nombrosos efectes deleteris associats amb la consanguinitat en l'espècie *D. subobscura*. Una vegada es van obtenir aquestes línies altament consanguínies es va comprovar que fossin homocariotípiques per al cromosoma J i es

va determinar la seva ordenació per aquest cromosoma. De les 83 línies isomaternes inicials es van obtenir un total de 27 línies que es diferencien per la presència o absència de la inversió  $J_1$ : 9 d'elles presenten l'ordenació  $J_{st}$  i 18 l'ordenació  $J_1$ . D'aquesta manera es van obtenir un 66% de línies amb l'ordenació  $J_1$  i un 33% de línies amb l'ordenació  $J_{st}$ . Aquestes freqüències observades concordarien amb les esperades en la seva àrea de mostreig d'acord a Krimbas (1993) i Orengo i Prevosti (1996).

En la Figura R1.2 es mostren tres preparacions de cromosomes politènics on s'observen les dues possibles ordenacions de les línies en estudi i els cromosomes d'un individu heterocariotip per al cromosoma J.



**Figura R1.2.** a) Cromosomes politènics d'un individu homocariotipic  $J_{st}/J_{st}$ . b) cromosomes politènics d'un individu homocariotipic  $J_1/J_1$ . c) Cromosomes politènics d'un individu heterocariotipic  $J_{st}/J_1$  on s'observa la nansa formada per la inversió  $J_1$ . Les preparacions mostrades provenen de la població natural de *D. subobscura* en estudi. C: centròmer, T: telòmer.

## R2. CARACTERITZACIÓ DE LES REGIONS LOCALITZADES AL LLARG DEL CROMOSOMA J DE *D. SUBOBSCURA*

El material inicial de partida va ser de 34 fags recombinants que havien estat aïllats aleatòriament d'una genoteca de l'espècie *D. subobscura* i localitzats per hibridació *in situ* en el cromosoma J d'aquesta espècie (Papaceit et al. 2006). Després de la subclonació de l'insert d'aquests 34 fags recombinants mitjançant el procés de *shotgun cloning*, es va realitzar la seqüenciació parcial d'un total de 80 subclons. Les cerques BLAST de les seqüències parcials van permetre caracteritzar un total de 19 regions no codificadores, quatre de les quals es localitzen dintre de la inversió  $J_1$ .

Per incrementar el nombre de regions en estudi dins de la inversió  $J_1$  o en regions fora de la inversió però properes al punt de trencament, es van seleccionar 3 gens que havien estat localitzats dintre o a la vora de la inversió  $J_1$  d'acord a Laayouni et al. (2007). Es van dissenyar nous encebadors sobre la seqüència genòmica de *D. pseudoobscura* en regions no codificadores properes a aquests gens i col·lineals entre les espècies *D. pseudoobscura* i *D. melanogaster*. Totes les sondes van ser posteriorment comprovades i localitzades citològicament per hibridació *in situ*. En la Figura R2.1 es mostren els resultats de les hibridacions *in situ* realitzades per comprovar la localització citològica d'aquestes noves regions.

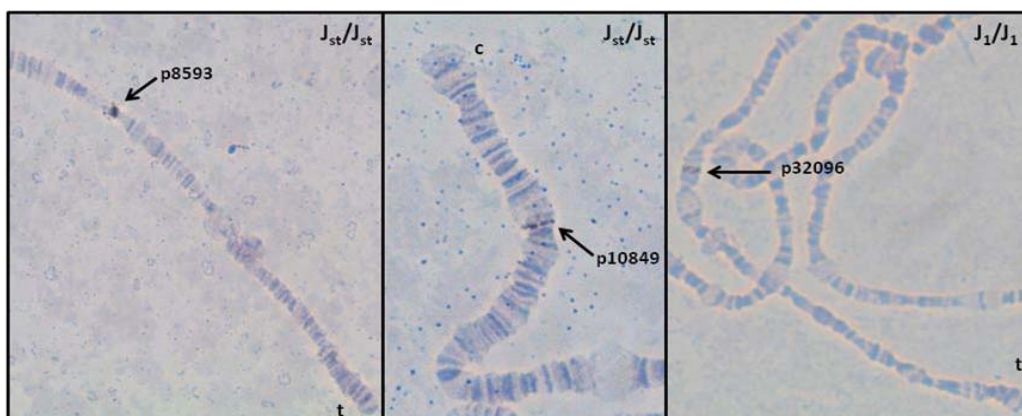


Figura R2.1. Hibridacions *in situ* de les sondes corresponents a les tres noves regions analitzades.

En la Taula R2.1 es mostren els tres gens seleccionats en l'espècie *D. melanogaster* i *D. pseudoobscura* per l'obtenció d'aquestes regions i en la Taula R2.2 la localització citològica de les tres noves regions.

Taula R2.1. Descripció de les tres noves regions en estudi.

Regió	Gen de <i>D. melanogaster</i> descrit a Laayouni et al. (2007)	Gen homòleg a <i>D. pseudoobscura</i>	Distància del gen a la regió analitzada	Gens flanquejants utilitzats per el disseny d'encebadors
<b>p32096</b>	<i>rols</i>	Dpse\GA16675	4,5 Kb	Dpse\GA16842 - Dpse\GA19744
<b>p10849</b>	<i>Sc2</i>	Dpse\GA10597	29 Kb	Dpse\GA12637 - Dpse\GA11341
<b>p8593</b>	<i>qm</i>	Dpse\GA21189	2,5Kb	Dpse\GA28324 - Dpse\GA13282

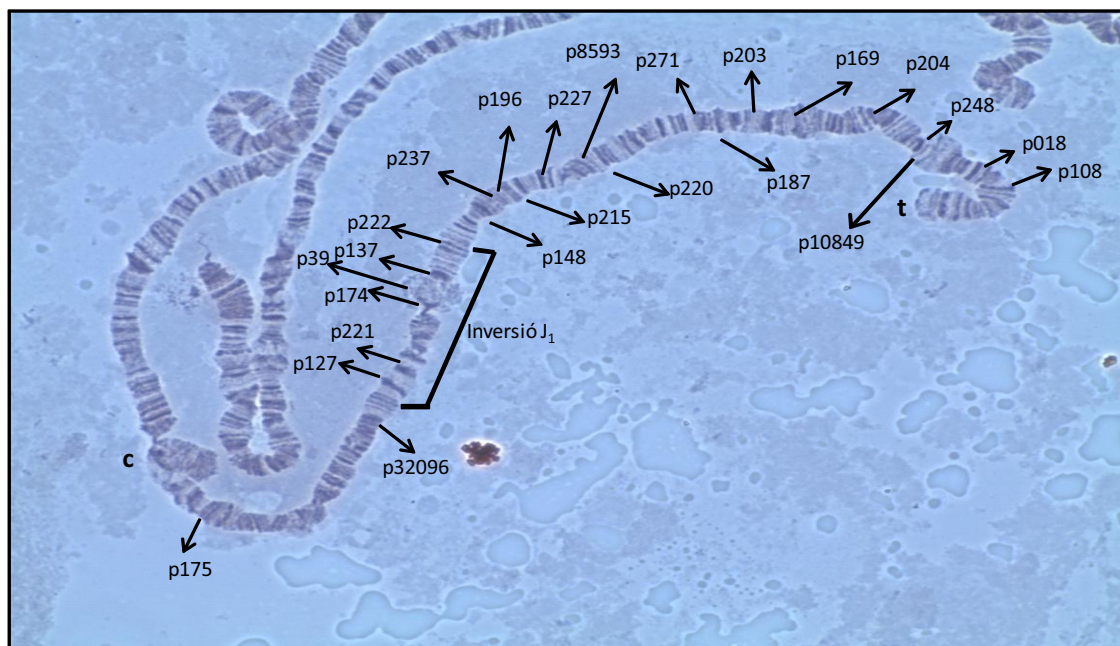
Taula R2.2. Localització citològica de les tres regions.

<i>D. subobscura</i>	Banda citològica descrita a Laayouni et al. (2007)	Banda citològica obtinguda en l'estudi
<b>p32096</b>	22C	21C
<b>p10849</b>	23B	33B
<b>p8593</b>	25D	27A

Tot i la proximitat de les tres noves regions analitzades als gens descrits a Laayouni et al. (2007; Taula R2.1), no es va poder confirmar, en cap cas, que les noves regions estiguessin localitzades dins de la inversió  $J_1$  ni a la vora dels punts de trencament. Per aquest motiu i per tal d'incrementar el nombre de regions analitzades dins de la inversió  $J_1$  es va utilitzar com a guia el genoma de *D. pseudoobscura* i com ancoratge les regions prèviament caracteritzades a partir dels inserits dels fags recombinants dins de la inversió  $J_1$ . Es van seleccionar noves regions allunyades de les d'ancoratge i es va comprovar la seva localització mitjançant la tècnica d'hibridació *in situ*. D'aquesta manera es van obtenir les regions p127 i p222, properes al punt de trencament de la inversió  $J_1$ .

En total es van poder obtenir 24 regions no codificadores, distribuïdes al llarg del cromosoma J de *D. subobscura*, 6 de les quals es troben dins de la inversió  $J_1$ . En la Figura R2.2

es veuen representades les diferents regions en estudi al llarg d'un cromosoma J amb l'ordenació estàndard. En la Taula R2.3 es mostren les diferents regions en estudi, la seva localització citològica en el cromosoma J, el tipus de regió no codificadora i les regions homòlogues en els genomes de *D. melanogaster* i *D. pseudoobscura* obtingudes de [www.flybase.com](http://www.flybase.com).



**Figura R2.2.** Localització de les diferents regions en estudi al llarg del cromosoma J de *D. subobscura*. c: centròmer; t: telòmer.



Taula R2.3. Localització citològica a *D. subobscura* de les regions en estudi, classificació i regions homòlogues a *D. pseudoobscura* i *D. melanogaster*.

Regions	Banda citològica	Classificació <sup>a</sup>	Regions homòlogues a <i>D. pseudoobscura</i> i <i>D. melanogaster</i>
<b>p175</b>	18D	intró	<i>D. pseudoobscura</i> : XR_group6:2470274..2476389 <i>D. melanogaster</i> : 3L:12138149..12142666
<b>p32096</b>	21C	pIGRs	<i>D. pseudoobscura</i> : XR_group8:9017539..9022931 <i>D. melanogaster</i> : 3L:8567913..8572755
<b>p127</b>	22E	pIGRs	<i>D. pseudoobscura</i> : XR_group8:2337468..2338467 <i>D. melanogaster</i> : 3L:9444041..9448093
<b>p221</b>	23C	intro	<i>D. pseudoobscura</i> : XR_group6:1297857..1301950 <i>D. melanogaster</i> : 3L:20840965..20845023
<b>p174</b>	24A	dIGRs	<i>D. pseudoobscura</i> : XR_group6:10218862..10223035 <i>D. melanogaster</i> : 3L:11457778..11461858
<b>p39</b>	24A	pIGRs	<i>D. pseudoobscura</i> : XR_group6:10615952..10622524 <i>D. melanogaster</i> : 3L:11149236..11151803
<b>p137</b>	24A	pIGRs	<i>D. pseudoobscura</i> : XR_group6:10641902..10645467 <i>D. melanogaster</i> : 3L:11126707..11128706
<b>p222</b>	24D	intro	<i>D. pseudoobscura</i> : XR_group6:1353071..1357973 <i>D. melanogaster</i> : 3L:7357941..7362018
<b>p148</b>	25D	dIGRs	<i>D. pseudoobscura</i> : XR_group8:8733422..8740390 <i>D. melanogaster</i> : 3L:8826730..8828663
<b>p237</b>	26A	intró	<i>D. pseudoobscura</i> : XR_group8:8843374..8849772 <i>D. melanogaster</i> : 3L:8728028..8730307
<b>p196</b>	26B	pIGRs	<i>D. pseudoobscura</i> : XR_group8:2587098...2593260 <i>D. melanogaster</i> : 3L:4490326..4496395
<b>p215</b>	26C	dIGRs	<i>D. pseudoobscura</i> : XR_group6:10863536..10870432 <i>D. melanogaster</i> : 3L:17992676..17998447
<b>p227</b>	27C	dIGRs	<i>D. pseudoobscura</i> : XR_group8:3509517..3514263 <i>D. melanogaster</i> : 3L:5853160..5857260
<b>p8593</b>	27A	pIGRs	<i>D. pseudoobscura</i> : XR_group8:3420271..3425263 <i>D. melanogaster</i> : 3L:11828708..11832731
<b>p220</b>	27E	pIGRs	<i>D. pseudoobscura</i> : XR_group6:2065758..2069861 <i>D. melanogaster</i> : 3L:5853160..5857260
<b>p271</b>	29A	pIGRs	<i>D. pseudoobscura</i> : XR_group6:5156548..5161970 <i>D. melanogaster</i> : 3L:1550925..1555577
<b>p187</b>	29D	pIGRs	<i>D. pseudoobscura</i> : XR_group6:6499483..6506756 <i>D. melanogaster</i> : 3L:891097..896952
<b>p203</b>	30C	pIGRs	<i>D. pseudoobscura</i> : XR_group6:13128243..13133983 <i>D. melanogaster</i> : 3L:20369535..20369734
<b>p169</b>	31B	intró	<i>D. pseudoobscura</i> : XR_group6:9890723..9895457 <i>D. melanogaster</i> : 3L:9782242..9782441
<b>p204</b>	32C	pIGRs	<i>D. pseudoobscura</i> : XR_group5:517979..520020 <i>D. melanogaster</i> : 3L:9085862..9089900
<b>p248</b>	33B	pIGRs	<i>D. pseudoobscura</i> : XR_group6:12877673..12883819 <i>D. melanogaster</i> : 3L:8387271..8389270
<b>p10849</b>	33B	pIGRs	<i>D. pseudoobscura</i> : XR_group8:516104..517103 <i>D. melanogaster</i> : 3L:1774120..1775119
<b>p18</b>	34D	dIGRs	<i>D. pseudoobscura</i> : XR_group6:3396500..3400845 <i>D. melanogaster</i> : 3L:12614862..12619096
<b>p108</b>	34E	dIGRs	<i>D. pseudoobscura</i> : XR_group8:1281396..1285638 <i>D. melanogaster</i> : 3L:3304393..3308441

En ombrejat es mostren les sis regions incloses dins de la inversió J<sub>1</sub>. <sup>a</sup> Les diferents regions es classifiquen en: intró, pIGRs (regions situades a menys de 2 kb del gen més proper) i dIGRs (regions situades a més de 4 kb del gen més proper).

### **R3. ESTUDI DELS NIVELLS I PATRONS DE VARIACIÓ A *D. SUBOBSCURA* EN REGIONS NO CODIFICADORES LOCALITZADES AL LLARG DEL CROMOSOMA J I NO AFECTADES PER INVERSIONS CROMOSÒMIQUES**

En aquest apartat es mostren els nivells i patrons de variació nucleotídica en una població natural de *D. subobscura* de Barcelona de les 18 regions no codificadores localitzades al llarg del cromosoma J i no afectades per inversions cromosòmiques. En el següent apartat (R4.) de resultats es mostraran els resultats per les 6 regions restants localitzades dins de la inversió  $J_1$ . Per l'obtenció de les dades de divergència es va utilitzar *D. guanche* com a espècie *outgroup*. Es va seqüenciar un individu d'aquesta espècie per cadascuna de les regions en estudi a excepció de la regió p196 on no va ser possible obtenir una amplificació correcta. La longitud de les 18 regions estudiades varia de 512 a 2935 nucleòtids amb una mitjana de 1877 nucleòtids per regió, el que correspon aproximadament a un total de 37 kb analitzades. Aquesta longitud és inferior quan s'inclou l'espècie *outgroup* degut a la presència d'insercions-deleccions (*indels*). El nombre de línies analitzades per la majoria de les regions va ser aproximadament de 20. L'absència de diferenciació genètica entre les línies  $J_{st}$  i  $J_1$  en aquestes regions va permetre l'anàlisi conjunta de totes les línies.

#### **R3.1 Nivells de variació nucleotídica**

Es van utilitzar varis estimadors per descriure els nivells de variació nucleotídica presents en aquestes 18 regions analitzades. En la Taula R3.1 es mostren els valors observats per cadascuna de les regions. Les taules suplementàries situades al final dels Resultats, mostren en detall el polimorfisme nucleotídic present en les regions estudiades.

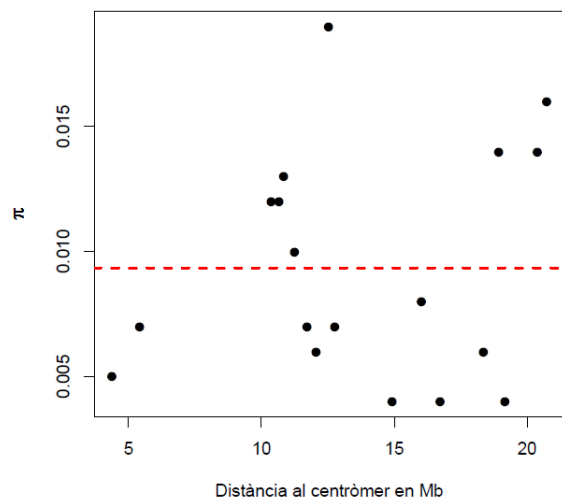
Taula R3.1. Estimes dels nivells de polimorfisme

Regions	<i>N</i>	<i>L</i>	<i>S</i>	$\theta$	$\pi$	<i>K</i>
p175	20	1914	47(26)	0,007	0,005	0,044
p32096	13	1600	45(28)	0,009	0,007	0,053
p148	20	2935	175(93)	0,017	0,012	0,045
p237	20	2187	140(90)	0,018	0,012	0,067
p196	20	2478	147(76)	0,017	0,013	n.a
p215	21	2469	146(90)	0,016	0,01	0,062
p8593	18	1643	64(40)	0,011	0,007	0,024
p227	21	718	18(9)	0,007	0,006	0,028
p220	20	1096	105(59)	0,027	0,019	0,089
p271	21	1388	53(29)	0,011	0,007	0,04
p187	20	2758	65(43)	0,007	0,004	0,035
p203	21	2275	103(64)	0,013	0,008	0,032
p169	20	515	17(14)	0,007	0,004	0,085
p204	21	2267	92(63)	0,011	0,006	0,035
p248	18	2779	185(112)	0,019	0,014	0,073
p10849	18	512	12(8)	0,007	0,004	0,078
p18	21	2781	194(98)	0,019	0,014	0,051
p108	21	1475	119(79)	0,022	0,016	0,065
Mitjana	20	1877	95,94	0,014	0,009	0,054
Variància	3,88	$6,38 \times 10^5$	$3,51 \times 10^3$	$3,62 \times 10^{-5}$	$2,08 \times 10^{-5}$	$4,15 \times 10^{-4}$

*N*, grandària de la mostra; *L*, longitud de la regió analitzada eliminant els *gaps*; *S*, nombre de llocs segregants (el nombre de llocs segregants amb variants úniques es mostra entre parèntesis);  $\theta$ , estimador de Watterson (1975);  $\pi$ , diversitat nucleotídica (Nei 1987) i *K*, divergència nucleotídica corregida d'acord a Jukes i Cantor (1969).

Totes les regions en estudi presenten un nombre de llocs polimòrfics (*S*) relativament elevat amb una considerable proporció de variants úniques. De fet en totes les regions el percentatge de variants úniques sobrepassa el 50% i s'observa un valor màxim del 82% en la regió p169. Els valors de l'estimador  $\theta$  de Watterson (1975) varien de 0,007 en la regió p175 a 0,027 en la regió p220, amb un valor promig de  $\theta = 0,014$ . La diversitat nucleotídica ( $\pi$ ) presenta un rang de valors més ampli, oscil·lant de 0,004 en la regió p187 a 0,019 en la regió p220 amb un valor promig de  $\pi = 0,009$ . Aquest valor promig de  $\pi$  és similar al valor promig de les estimes realitzades per posicions sinònimes en múltiples regions codificadores de *D. subobscura* (Papaceit, comunicació personal).

Es va estudiar la possible correlació entre la diversitat nucleotídica ( $\pi$ ) de cada regió i la seva distància al centròmer per tal d'avaluar si els nivells de variabilitat nucleotídica estaven distribuïts uniformement al llarg del cromosoma J. La distància de cada regió al centròmer es va estimar considerant una distribució homogènia del DNA al llarg de tot el cromosoma J i utilitzant com a referència de la grandària dels elements cromosòmics les dades publicades de l'espècie *D. pseudoobscura* (Schaeffer et al. 2008). En la Figura R3.1, hi ha representada la diversitat nucleotídica de cada regió en funció de la seva distància al centròmer. Els nivells de variació nucleotídica estan uniformement distribuïts al llarg del cromosoma J ja que no s'observa una correlació significativa entre la diversitat nucleotídica ( $\pi$ ) de cada regió i la seva distància al centròmer d'acord al test no paramètric de *Spearman* ( $P = 0,74$ ).



**Figura R3.1.** Relació entre la diversitat nucleotídica de cada regió i la seva distància al centròmer expressada en Mb. La línia discontinua és el valor promig de  $\pi$ .

### R3.2 Patrons de variació nucleotídica

Es van examinar els patrons de variació nucleotídica mitjançant diferents estadístics. En funció de la classificació proposada a Ramos-Onsins i Rozas (2002) aquests es van dividir en: estadístics de classe I basats en la distribució de freqüències de les mutacions i estadístics de classe II que utilitzen la informació de la distribució dels haplotips.

### Anàlisi dels patrons de polimorfisme mitjançant els estadístics de classe I

En la Taula R3.2 es mostren els valors estimats per diversos estadístics que analitzen el patró de polimorfisme i que estan basats en l'espectre de freqüències. Aquests estadístics són: *D*, estadístic *D* de Tajima (1989); *FD*, estadístic *D* de Fu i Li (1993) amb *outgroup*; *FF*, estadístic *F* de Fu i Li (1993) amb *outgroup*; *R<sub>2</sub>*, estadístic *R<sub>2</sub>* de Ramos-Onsins i Rozas (2002) i *Hnorm*, estadístic *H* normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006).

**Taula R3.2.** Estimes dels patrons de polimorfisme mitjançant estadístics basats en l'espectre de freqüències

Regions	<i>D</i>	<i>FD</i>	<i>FF</i>	<i>R<sub>2</sub></i>	<i>Hnorm</i>
<b>p175</b>	-1,28	-1,81	-2,00	0,073**	-0,01
<b>p32096</b>	-1,07	-1,88	-2,02	0,080**	0,19
<b>p148</b>	-1,18	-1,61	-1,83	0,076*	-0,34
<b>p237</b>	-1,49	-2,36*	-2,56*	0,060**	-0,20
<b>p196</b>	-1,01	n.a	n.a	0,083*	n.a
<b>p215</b>	-1,60#	-2,33*	-2,65*	0,060**	-0,78
<b>p8593</b>	-1,50	-2,20#	-2,44#	0,067**	0,40
<b>p227</b>	-0,29	-1,21	-1,10	0,116	0,11
<b>p220</b>	-1,27	-1,09	-1,36	0,076**	-1,70
<b>p271</b>	-1,32	-1,94#	-2,11#	0,076*	0,08
<b>p187</b>	-1,82*	-2,74**	-3,00**	0,050**	0,41
<b>p203</b>	-1,53	-2,37*	-2,54*	0,060**	-0,37
<b>p169</b>	-1,83*	-1,29	-1,70	0,061**	-3,49**
<b>p204</b>	-1,81#	-2,86**	-3,05**	0,050**	0,37
<b>p248</b>	-1,23	-2,06#	-2,20#	0,073**	-0,22
<b>p10849</b>	-1,26	-1,63	-1,83	0,089	0,31
<b>p18</b>	-1,19	-1,52	-1,73	0,077**	-0,43
<b>p108</b>	-1,22	-2,53*	-2,55*	0,070**	-0,16
<b>mitjana</b>	-1,33	-1,97	-2,16	0,07	-0,34
<b>variància</b>	0,13	0,28	0,29	2,37x10 <sup>-4</sup>	0,93

*D*, estadístic *D* de Tajima (1989); *FD*, estadístic *D* de Fu i Li (1993) amb *outgroup* (1993); *FF*, estadístic *F* de Fu i Li (1993) amb *outgroup* (1993); *R<sub>2</sub>*, estadístic *R<sub>2</sub>* de Ramos-Onsins i Rozas (2002) i *Hnorm*, estadístic *H* normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006). n.a, estimes no obtingudes en la regió p196 degut a la manca d'informació a l'espècie *outgroup*. Es mostren els nivells de significació obtinguts mitjançant 10.000 simulacions de coalescència sota el model neutre estàndard condicionades al nombre de llocs segregants i sota la premissa conservativa de no recombinació. La significació dels valors *P* segueix la següent nomenclatura: # 0,05 < *P* < 0,10; \* 0,02 < *P* < 0,05; \*\* *P* < 0,02.

L'estadístic *D* de Tajima (1989), que és sensible per detectar desviacions de l'espectre de freqüències en relació a les variants a freqüència baixa o intermèdia, mostra en totes les

regions en estudi valors negatius, amb un valor promig de  $D = -1,33$ . En 4 de les regions s'observen desviacions significatives ( $P < 0,05$ ) o marginalment significatives ( $0,05 < P < 0,1$ ) de l'espectre de freqüències en relació al que s'esperaria sota condicions de neutralitat. Els estadístics  $D$  i  $F$  de Fu i Li (1993) amb *outgroup*, que es fonamenten en la distribució de les mutacions esperades en les branques internes i externes d'una genealogia neutra suposant equilibri mutació-deriva, també presenten en totes les regions en estudi valors negatius amb un valor promig de  $FD$  i  $FF$  de  $-1,97$  i  $-2,16$ , respectivament. En 9 de les regions en estudi aquests dos estadístics mostren desviacions significatives ( $P < 0,05$ ) o marginalment significatives ( $0,05 < P < 0,1$ ) de l'espectre de freqüències. Les estimes dels estadístics  $D$  de Tajima, i  $D$  i  $F$  de Fu i Li en les regions en estudi indiquen desviacions de l'espectre de freqüències degut a un excés de variants a baixa freqüència, principalment variants úniques, posant de manifest que en l'actualitat la variació nucleotídica a *D. subobscura* no es troba en equilibri mutació-deriva. Aquest excés de variants a baixa freqüència podria indicar que l'espècie ha patit un procés d'expansió poblacional.

L'estadístic  $R_2$  (Ramos-Onsins i Rozas 2002), igual que l'estadístic  $F_s$  de Fu (1997), és un estadístic especialment sensible per la detecció d'expansions poblacionals però més potent quan es té en compte la recombinació intragènica i la grandària mostral és petita. S'observa en totes les regions en estudi, a excepció de la regió p227 i la regió p10849, que els valors de  $R_2$  es desvien significativament de l'esperat ( $P < 0,02$ ) sota condicions de neutralitat. Aquests valors significatius de l'estadístic  $R_2$ , juntament amb l'excés de variants a baixa freqüència detectat pels estadístics  $D$  de Tajima (1989), i  $D$  i  $F$  de Fu i Li (1993) amb *outgroup*, recolzarien que la història demogràfica de l'espècie pot haver estat marcada per una expansió poblacional recent.

L'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006), que és sensible per detectar desviacions de l'espectre de freqüències cap un excés de polimorfismes amb

variants derivades a elevada freqüència en relació a les variants a freqüència intermèdia, mostra en les regions en estudi valors tant negatius com positius però propers a zero, a excepció de la regió p169 i p220 amb valors per aquest estadístic de -3,49 i -1,70, respectivament. No s'observa en cap de les regions en estudi valors d'aquests estadístics que es desviïn significativament del que s'esperaria sota condicions de neutralitat, a excepció de la regió p169 en la que es detecta una desviació significativa de l'espectre de freqüències cap a variants derivades a elevada freqüència. El valor promig d'aquest estadístic és  $H_{norm} = -0,34$ .

#### **Anàlisi dels patrons de polimorfisme mitjançant els estadístics de classe II**

Es van examinar els patrons de polimorfisme mitjançant els estadístics basats en la distribució dels haplotips, sent aquests:  $H$ , nombre d'haplotips;  $H_w$ , diversitat haplotípica (Nei 1987);  $F_s$  de Fu (1997);  $Z_A$  de Rozas et al. (2001) i  $B$  de Wall (1999). En la Taula R3.3 es mostren els valors estimats per aquests estadístics.

Taula R3.3. Estimes dels patrons de polimorfisme mitjançant estadístics basats en la distribució dels haplotips

Regions	$H$	$H_w$	$F_s$	$Z_A$	$B$
<b>p175</b>	19	0,99	-10,22	0,16	0,13
<b>p32096</b>	13	1	-5,32	0,21	0,14
<b>p148</b>	19	0,99	-2,68	0,14	0,09
<b>p237</b>	20	1	-6,1	0,14	0,1
<b>p196</b>	20	1	-5,1	0,23	0,15
<b>p215</b>	21	1	-6,79	0,18	0,14
<b>p8593</b>	18	1	-9,09	0,14	0,11
<b>p227</b>	17	0,97	-10,24	0,21	0,18
<b>p220</b>	19	0,99	-4,92	0,22	0,16
<b>p271</b>	19	0,99	-8,17	0,16	0,12
<b>p187</b>	19	0,99	-9,31	0,16	0,11
<b>p203</b>	21	1	-8,81	0,12	0,07
<b>p169</b>	14	0,96	-10,11	0,07	0,06
<b>p204</b>	20	1	-7,77	0,13	0,1
<b>p248</b>	18	1	-3,48	0,21	0,16
<b>p10849</b>	10	0,88	-4,41	0,16	0,09
<b>p18</b>	20	1	-2,77	0,18	0,13
<b>p108</b>	21	1	-7,17	0,17	0,11
<b>mitjana</b>	18	0,99	-6,8	0,17	0,12
<b>variància</b>	9,01	$9,04 \times 10^{-4}$	6,59	$1,62 \times 10^{-3}$	$1,07 \times 10^{-3}$

$H$ , nombre d'haplotips;  $H_w$ , diversitat haplotípica (Nei 1987);  $F_s$ , estadístic  $F_s$  de Fu (1997);  $Z_A$ , estadístic  $Z_A$  (Rozas et al. 2001) i  $B$ , estadístic  $B$  de Wall (1999).

La diversitat haplotípica de les diferents regions en estudi és molt elevada, presentant valors superiors a 0,99 en totes les regions, a excepció de la regió p10849 on  $H_w = 0,88$ . Aquesta elevada diversitat haplotípica seria consistent amb l'elevada presència de variants úniques (*singletons*) observada en les regions en estudi i indicaria que aquestes variants no estan concentrades en línies particulars. L'estadístic  $F_s$  de Fu (Fu 1997) mostra en totes les regions en estudi valors negatius, amb un valor promig de  $F_s = -6,80$ .

En relació als estadístics que mesuren el grau de desequilibri de lligament o la no associació a l'atzar entre les variants nucleotídiques en diferents llocs polimòrfics, es va utilitzar l'estadístic  $Z_A$  (Rozas et al. 2001). El grau de desequilibri de lligament es va analitzar únicament per les posicions informatives i la seva significació es va estimar en funció del model neutre estàndard i sota la premissa conservativa de no recombinació. En cap de les regions en



estudi es va observar valors de l'estadístic  $Z_A$  significativament diferents del que s'esperaria sota condicions de neutralitat.

L'estadístic  $B$  de Wall (1999) mesura el desequilibri de lligament entre parells de llocs segregants adjacents i és sensible per detectar desviacions del model neutre estàndard produïdes per genealogies amb branques internes llargues en relació a les branques externes com podria succeir en un procés de subdivisió poblacional o de selecció balancejadora. No s'observa en cap de les regions en estudi valors d'aquest estadístic que es desviïn significativament del que s'esperaria sota condicions de neutralitat. Aquest resultat posa de manifest que la població de *D. subobscura* en estudi possiblement no està estructurada sinó que tendeix més a comportar-se com una població panmíctica tal i com planteja el model neutre estàndard.

No obstant això, la conclusió que s'extreu de les anàlisis dels nivells i patrons de polimorfisme és que, en l'actualitat, la variabilitat present a *D. subobscura* no es troba en equilibri mutació-deriva i que molt possiblement, factors demogràfics marcats per una expansió poblacional, poden haver perfilat la seva història evolutiva.

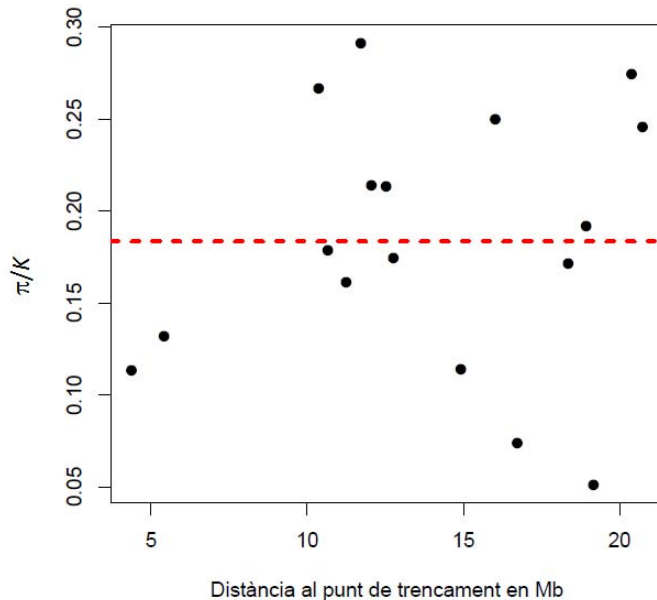
### **R3.3 Polimorfisme i divergència**

Es van estudiar 17 regions no codificadores, excloent de les anàlisis la regió p196 on no va ser possible obtenir la seqüència nucleotídica de l'espècie *outgroup*. Es va calcular la divergència nucleotídica ( $K$ , Nei 1987) corregida per Jukes i Cantor (1969) per cadascuna de les regions en estudi i utilitzant l'espècie *D. guanche* com a *outgroup*. La divergència nucleotídica promig de totes les regions en estudi és de  $K=0,05$  (Taula R3.1).

#### **$\pi/K$**

Es va utilitzar la relació  $\pi/K$  com a mesura relativa de la quantitat de variació nucleotídica d'una regió genòmica dins de l'espècie en relació a la divergència entre espècies, per corregir les possibles diferències entre regions de la taxa de mutació neutra. L'estima  $\pi/K$

es va analitzar en funció de la distància de cada regió al punt de trencament més proper. Les estimes de la diversitat nucleotídica ( $\pi$ ) i la divergència ( $K$ ) es mostren en la Taula R3.1. En la Figura R3.2 es mostra la relació  $\pi/K$  per a cada regió en estudi al llarg del cromosoma J.



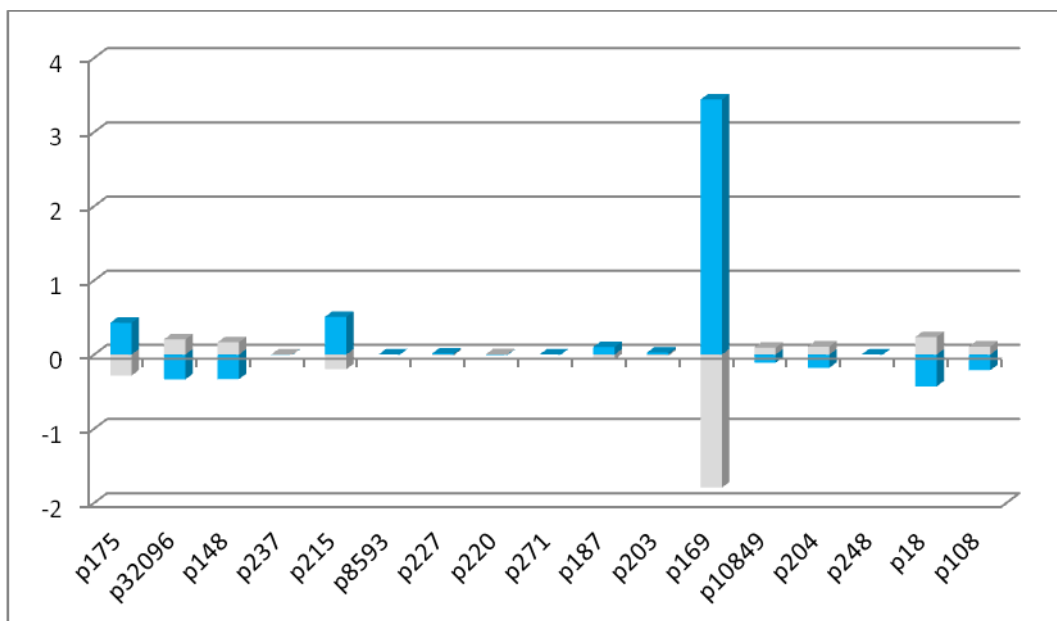
**Figura R3.2.** Estima de  $\pi/K$  per a cada regió en relació a la seva distància al centròmer. La línia horitzontal discontinua és el valor promig d'aquesta relació.

Els valors per la relació  $\pi/k$  estimats per les 17 regions en estudi són molt diversos, trobant-se valors d'aquesta relació des de 0,05 en la regió p10849 fins a 0,29 en la regió p8593, amb un valor promig de 0,18. Els valors de la ràtio  $\pi/K$  estan uniformement distribuïts al llarg del cromosoma J ja que no s'observa una correlació significativa entre la ràtio  $\pi/K$  de cada regió i la seva distància al centròmer d'acord al test no paramètric de *Spearman* ( $P = 0,75$ ).

#### Test d'HKA

Es va aplicar a nivell multilocus el test d'HKA (test de Hudson-Kreitman-Aguadé; Hudson et al. 1987) basat en la predicció de que els nivells de variació nucleotídica dins i entre espècies han de ser proporcionals en aquelles regions del genoma que evolucionen sota condicions de neutralitat (Hudson et al. 1987). El test multilocus no presenta resultats significatius, mostrant un valor de  $X^2 = 9,3851$  (16 g.II,  $P = 0,89$ ). En la Figura R3.3 es mostra la

contribució parcial del polimorfisme i la divergència de cada regió en estudi a l'estadístic  $\chi^2$ . Cal destacar que la regió p169, en la que s'observa un excés de divergència enfront del polimorfisme, és la que més contribueix a aquest estadístic. Aquest resultat, juntament amb el valor estimat d'aquesta regió per l'estadístic  $H$  normalitzat (Fay i Wu 2000; Zeng et al. 2006) el qual es desvia significativament del que s'esperaria sota condicions de neutralitat, podria indicar que aquesta regió hagi estat sotmesa a pressions selectives.

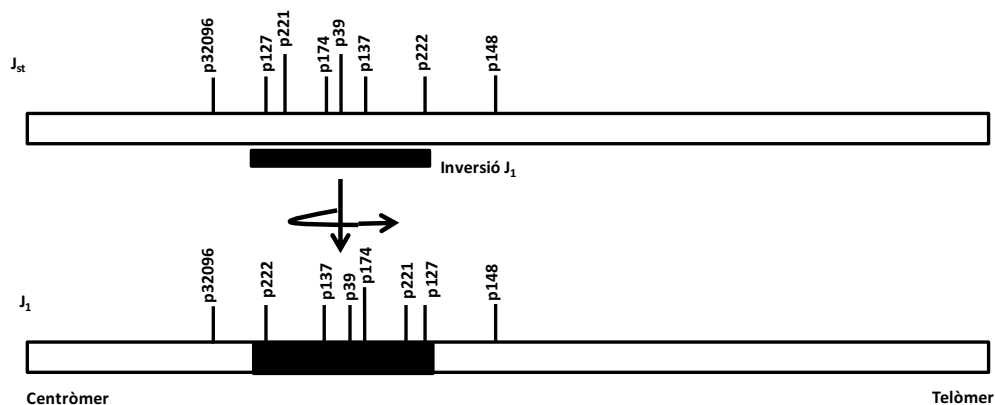


**Figura R3.3.** Test d'HKA. Contribució parcial del polimorfisme (en gris) i la divergència (en blau) de cada regió al valor total de l'estadístic  $\chi^2$  del test HKA multilocus.

#### R4. ANÀLISI DE LA VARIACIÓ EN REGIONS NO CODIFICADORES AFECTADES PER LA INVERSIÓ

##### $J_1$ DE *D. SUBOBSCURA*

En aquest apartat s'analitzen els nivells i patrons de variació nucleotídica en una població natural de *D. subobscura* de Barcelona en les sis regions estudiades localitzades dins de la inversió  $J_1$ . Les regions en estudi són les següents: p127, p221, p174, p39, p137 i p222. La longitud d'aquestes varia de 964 a 2572 nucleòtids (eliminant les posicions amb *gaps*) i corresponen a un total de 10,2 kb aproximadament. La grandària de la mostra és de 27 línies homocariotípiques que es diferencien per la presència o absència de la inversió  $J_1$ : 18 d'elles presenten l'ordenació  $J_1$  i 9 l'ordenació  $J_{st}$ . Les anàlisis es van realitzar per cada regió per separat i també amb les seqüències concatenades. En algunes de les anàlisis es van incloure també les regions p148 i p32096, amb una longitud aproximada de 4,5 kb, ja que aquestes són les regions més properes als punts de trencament entre les regions localitzades fora de la inversió  $J_1$ . D'aquesta manera es va poder avaluar l'efecte que la inversió podia tenir en la variabilitat genètica de les regions situades fora de la inversió però properes als punts de trencament. Per l'obtenció de les dades de divergència es va utilitzar l'espècie *D. guanche*. A la Figura R4.1 hi ha una representació esquemàtica de la localització de les diferents regions utilitzades en aquest apartat. Els polimorfismes nucleotídics presents en les regions en estudi es mostren en detall en el material suplementari del final dels Resultats.



**Figura R4.1.** Localització de les sis regions incloses dintre de la inversió  $J_1$  i de les dues regions fora de la inversió però properes als punts de trencament.

#### R4.1 Nivells de variació nucleotídica

Es van utilitzar varis estimadors per descriure els nivells de variació nucleotídica. En la

Taula R4.1 es mostren els resultats obtinguts.

**R4.1.** Nivells de variació nucleotídica i divergència en les diferents regions en estudi analitzats independentment per l'ordenació  $J_1$  i  $J_{st}$ .

Regions		<i>N</i>	<i>L</i>	<i>S</i>	$\theta$	$\pi$	<i>K</i>
<b>p222</b>	$J_1$	17	964	24 (14)	0,007	0,005	0,026
	$J_{st}$	9	964	12 (10)	0,004	0,003	0,029
<b>p127</b>	$J_1$	17	1257	24 (18)	0,006	0,003	0,035
	$J_{st}$	9	1257	13 (8)	0,004	0,003	0,035
<b>P137</b>	$J_1$	17	2560	157 (84)	0,018	0,014	0,048
	$J_{st}$	8	2560	96 (60)	0,014	0,012	0,047
<b>p221</b>	$J_1$	16	2046	103 (56)	0,015	0,012	0,061
	$J_{st}$	9	2046	66 (48)	0,011	0,009	0,058
<b>p39</b>	$J_1$	17	2572	187 (96)	0,021	0,017	0,070
	$J_{st}$	8	2572	89 (38)	0,013	0,014	0,071
<b>p174</b>	$J_1$	18	2040	141 (77)	0,020	0,014	0,065
	$J_{st}$	8	2040	87 (77)	0,015	0,015	0,064
Seqüències concatenades	$J_1$	14	10234	574 (299)	0,015	0,013	0,054
	$J_{st}$	7	10234	341 (241)	0,012	0,011	0,054

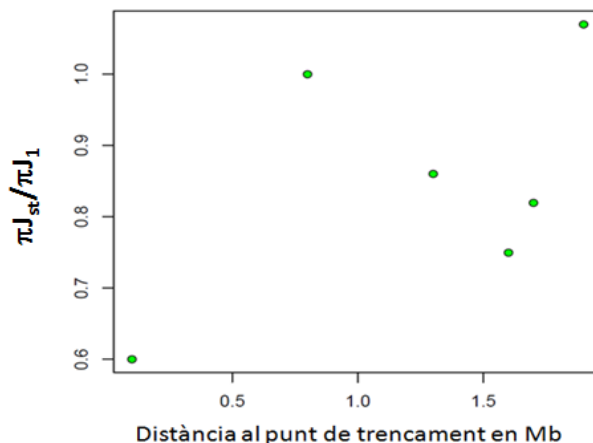
*N*, grandària de la mostra; *L*, longitud de la regió analitzada després d'eliminar els *gaps*; *S*, nombre de llocs segregants (el nombre de llocs segregants amb variants úniques es mostra entre parèntesis)  $\theta$ , estimador de Watterson (1975);  $\pi$ , diversitat nucleotídica (Nei 1987) i *K*: divergència nucleotídica corregida d'acord a Jukes i Cantor (1969).

Totes les regions en estudi presenten un nombre de llocs polimòrfics (*S*) elevat amb una important proporció de variants úniques (*singletons*). El percentatge de variants úniques sobrepasa el 50% en totes les regions, a excepció de la regió p39 en l'ordenació  $J_{st}$ . Tot i que el nombre de llocs polimòrfics i de variants úniques és, en general, inferior en les línies amb l'ordenació estàndard, s'ha de tenir en compte que el nombre de línies analitzades amb aquesta ordenació és inferior al nombre de línies  $J_1$ .

L'estimador  $\theta$  de Watterson (1975) presenta un ampli rang de valors que oscil·len de  $\theta = 0,004$  (p127  $J_{st}$ ) a  $\theta = 0,021$  (p39  $J_1$ ) mostrant un valor  $\theta = 0,015$  al concatenar les sis regions a

l'ordenació  $J_1$  i de  $\theta = 0,012$  a l'ordenació  $J_{st}$ . L'estimador  $\theta$  de Watterson mostra en tots els casos valors superiors en les línies que presenten l'ordenació  $J_1$ . No obstant, no s'observen diferències significatives entre les dues ordenacions (test no paramètric de *Wilcoxon*;  $P = 0,15$ ).

La diversitat nucleotídica ( $\pi$ ) mostra un rang de valors que varia de  $\pi = 0,003$  (p222  $J_{st}$ ) a  $\pi = 0,017$  (p39  $J_1$ ). El valor de  $\pi$  al concatenar les sis regions en estudi és de  $\pi = 0,013$  i  $\pi = 0,011$  per les ordenacions  $J_1$  i  $J_{st}$ , respectivament. Per analitzar si el nivell de polimorfisme dins de cada ordenació varia al llarg de la inversió, s'estudia la relació de  $\pi$  entre l'ordenació  $J_{st}$  i l'ordenació  $J_1$  ( $\pi_{J_{st}}/\pi_{J_1}$ ) en funció de la distància al punt de trencament més proper. No s'observen diferències significatives (test no paramètric d'*Spearman*;  $P = 0,33$ ) d'aquesta relació entre les diferents regions. En la Figura R4.2 hi ha representada la relació  $\pi_{J_{st}}/\pi_{J_1}$  en funció de la distància de cada regió al punt de trencament més proper.

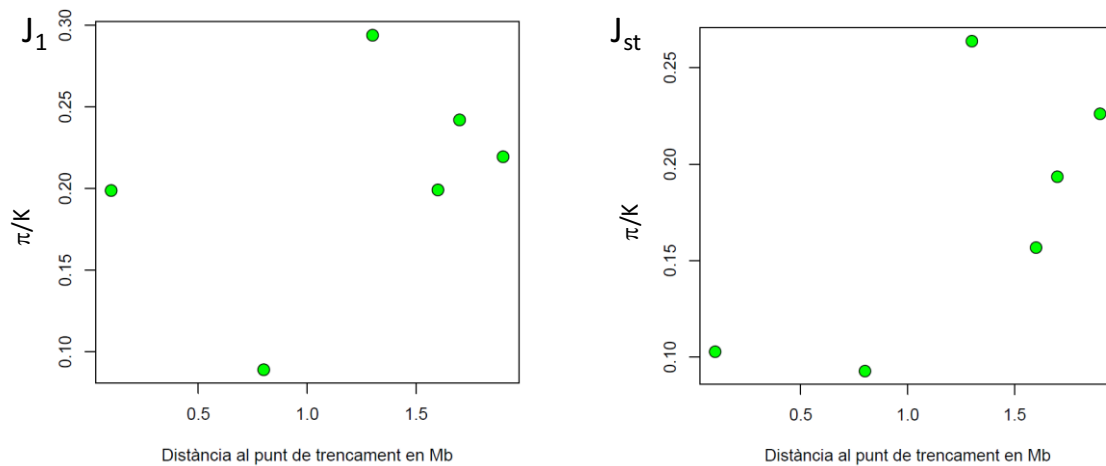


**Figura R4.2.** Relació  $\pi_{J_{st}}/\pi_{J_1}$  en funció de la distància de cada regió al punt de trencament més proper. Aquesta distància es va estimar considerant una distribució homogènia del DNA al llarg de tota la inversió  $J_1$  i utilitzant com a referència del contingut de DNA de l'element cromosòmic les dades publicades per l'espècie *D. pseudoobscura* (Schaeffer et al. 2008).

En general, les estimes de  $\theta$  són majors que les de  $\pi$  en les diferents classes cromosòmiques, diferència causada per un excés de polimorfismes amb variants que segreguen a baixa freqüència ja que  $\theta$  considera únicament el nombre de polimorfismes però no té en compte la freqüència dels diferents polimorfismes a la mostra.

$\pi/K$ 

No s'observen correlacions significatives entre l'estima de  $\pi/K$  i la distància al punt de trencament de l'ordenació  $J_1$  (coeficient de correlació d'*Spearman*;  $P = 0,26$ ), ni per l'ordenació  $J_{st}$  (coeficient de correlació d'*Spearman*;  $P = 0,21$ ), tot i que s'observa una disminució dels valors de  $\pi/K$  en les dues regions més properes als punts de trencament en l'ordenació  $J_{st}$ . Els resultats es mostren en la Figura R4.3.



**Figura R4.3.** Estima de  $\pi/K$  en l'ordenació  $J_1$  i l'ordenació  $J_{st}$  de cada regió en estudi en funció de la distància al punt de trencament més proper. Aquesta distància es va estimar considerant una distribució homogènia del DNA al llarg de tota la inversió  $J_1$  i utilitzant com a referència del contingut de DNA de l'element cromosòmic les dades publicades per l'espècie *D. pseudoobscura* (Schaeffer et al. 2008).

#### R4.2 Patrons de variació nucleotídica

Per estudiar els patrons de variació nucleotídica, es van utilitzar els estadístics basats en l'espectre de freqüències, sent aquests: estadístic  $D$  de Tajima (1989);  $FD$ , estadístic  $D$  de Fu i Li (1993) amb *outgroup*;  $FF$ , estadístic  $F$  de Fu i Li (1993) amb *outgroup* i estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000, Zeng et al. 2006). També es van analitzar els estadístics basats en la distribució dels haplotips:  $H$ , nombre d'haplotips i  $H_w$ , diversitat haplotípica (Nei 1987). La significació estadística per tots els tests es va realitzar mitjançant simulacions de coalescència (10.000 rèpliques) condicionades a  $\theta$  i amb la premissa conservativa de no recombinació intragènica. En els tests en els quals es requereix *outgroup*,

es va utilitzar la seqüència de *D. guanche*. Els resultats d'aquestes anàlisis es mostren en la Taula R4.2.

**Taula R4.2.** Patrons de variació en les diferents regions en estudi analitzades independentment per l'ordenació  $J_1$  i  $J_{st}$ .

Regions		<i>D</i>	<i>FD</i>	<i>FF</i>	<i>Hnorm</i>	<i>H</i>	<i>H<sub>w</sub></i>
<b>p221i</b>	$J_1$	-1,23	-1,14	-1,40	-0,90	13	0,96
	$J_{st}$	-1,52	-1,65	-1,93	0,24	8	0,97
<b>p127</b>	$J_1$	-1,80*	-2,67**	-2,89**	0,77	14	0,96
	$J_{st}$	-0,73	-1,04	-1,18	0,97	8	0,97
<b>P137</b>	$J_1$	-1,03	-1,20	-1,33	0,08	16	0,99
	$J_{st}$	-0,76	-1,16	-1,30	0,53	8	1,00
<b>p221</b>	$J_1$	-1,00	-1,55	-1,70	0,12	15	0,99
	$J_{st}$	-1,54	-1,47	-1,71	-0,25	9	1,00
<b>p39</b>	$J_1$	-1,05	-1,48	-1,67	-0,14	17	1,00
	$J_{st}$	0,11	0,49	0,55	-0,55	8	1,00
<b>p174</b>	$J_1$	-1,25	-1,43	-1,66	-0,24	17	0,99
	$J_{st}$	-0,35	-0,26	-0,37	-0,39	8	1,00
<b>Seqüències concatenades</b>	$J_1$	-0,92	-1,18	-1,34	0,07	14	0,98
	$J_{st}$	-0,54	-0,68	-0,79	0,01	7	0,99

*D*, estadístic *D* de Tajima (1989); *FD*, estadístic *D* de Fu i Li (1993) amb *outgroup*; *FF*, estadístic *F* de Fu i Li (1993) amb *outgroup*; *Hnorm*, estadístic *H* normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006); *H*, nombre d'haplotips; *H<sub>w</sub>*, diversitat haplotípica (Nei 1987). Es mostren els nivells de significació obtinguts mitjançant 10.000 simulacions de coalescència condicionades a  $\theta$  i amb la premissa conservativa de no recombinació. La significació dels valors *P* segueix la següent nomenclatura: #  $0,05 < P < 0,10$ ; \* $0,02 < P < 0,05$ ; \*\*  $P < 0,02$ .

L'estadístic *D* de Tajima presenta en tots els casos valors negatius, sent  $D = -0,92$  i  $D = -0,54$  els valors observats al concatenar les sis regions en estudi en les línies amb l'ordenació  $J_1$  i  $J_{st}$ , respectivament. Les línies amb l'ordenació  $J_{st}$  presenten en general valors de *D* de Tajima més propers a zero que els observats en l'ordenació  $J_1$ , fet que podria indicar que  $J_{st}$  és l'ordenació ancestral. En cap cas, s'observen valors de *D* de Tajima que es desviïn significativament del que s'esperaria sota condicions de neutralitat a excepció de la regió p127 en l'ordenació  $J_1$  ( $P < 0,05$ ). Els estadístics *D* i *F* de Fu i Li mostren un patró similar a l'observat per l'estadístic *D* de Tajima i únicament la regió p127 en l'ordenació  $J_1$  presenta valors que es



desvien significativament del que s'esperaria sota el model neutre estàndard. Els resultats observats d'aquests estadístics indiquen que l'espectre de freqüències està desviat, de forma general, cap a un excés de polimorfismes amb variants a baixa freqüència, principalment *singletons*.

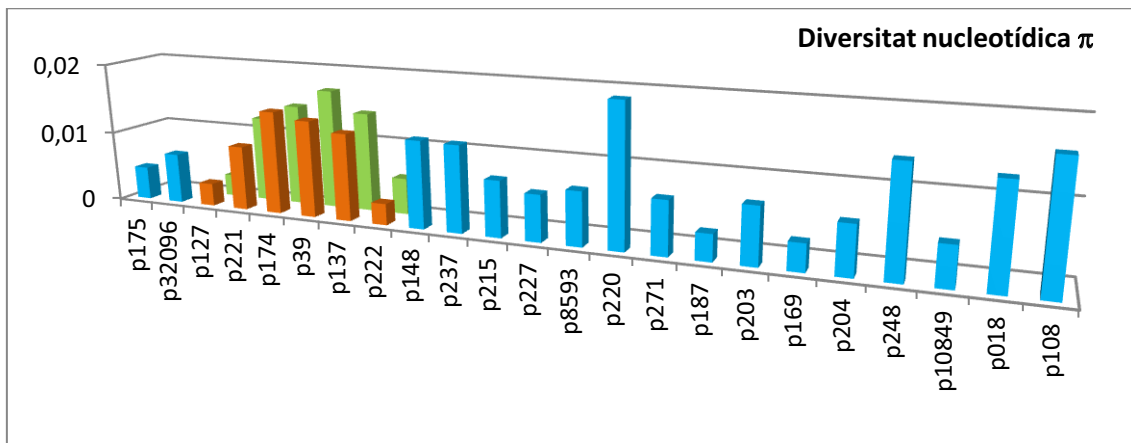
L'estadístic  $H$  normalitzat de Fay i Wu no mostra un patró de variació nucleotídica homogeni al llarg de totes les regions en estudi i els seus valors oscil·len de -0,90 en la regió p221 en l'ordenació  $J_1$  a 0,97 en la regió p127 en l'ordenació  $J_{st}$ . El valor de l' $H$  normalitzat de Fay i Wu és de 0,07 i de 0,01 al concatenar les sis regions en estudi per l'ordenació  $J_1$  i  $J_{st}$ , respectivament. En cap cas, aquest estadístic es desvia significativament del que s'esperaria sota condicions de neutralitat.

La diversitat haplotípica ( $H_w$ ) de les diferents regions en estudi és molt elevada, presentant valors superiors a 0,96 en totes les regions. Aquesta elevada diversitat haplotípica indicaria que les variants úniques (*singletons*) no estan concentrades en línies particulars.

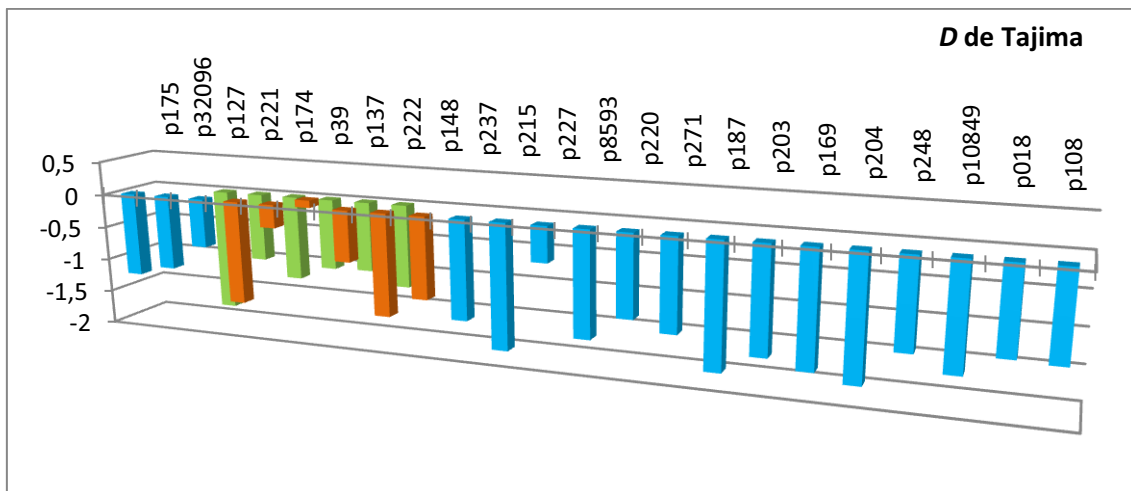
### **R4.3 Comparació dels nivells i patrons de polimorfisme en regions localitzades dins i fora de la inversió $J_1$**

Es va analitzar si els nivells i patrons de polimorfisme eren significativament diferents en les regions localitzades dins i fora de la inversió  $J_1$ . Les regions dins de la inversió es van analitzar segons l'ordenació i es van comparar independentment amb la resta de regions. Per l'estudi dels nivells de variació es va utilitzar l'estadístic diversitat nucleotídica (Figura R4.4) i pels patrons de variació l'estadístic  $D$  de Tajima (Figura R4.5). No s'observen diferències significatives ni en relació als nivells de diversitat nucleotídica mitjançant l'estadístic  $\pi$  (test no paramètric de *Wilcoxon*,  $P = 0,97$ , comparant les 6 regions localitzades dins del segment invertit en les línies  $J_{st}$  amb les 18 regions no afectades per la inversió; test no paramètric de *Wilcoxon*,  $P = 0,48$ , comparant les regions localitzades dins del segment invertit en les línies  $J_1$  amb les 18 regions no afectades per la inversió) ni en relació als patrons de variació analitzats

amb l'estadístic  $D$  de Tajima (test no paramètric de *Wilcoxon*,  $P = 0,12$ , comparant les 6 regions localitzades dins el segment invertit en les línies  $J_{st}$  amb les 18 regions no afectades per la inversió; test no paramètric de *Wilcoxon*,  $P = 0,18$ , comparant les regions localitzades dins el segment invertit en les línies  $J_1$  amb les 18 regions no afectades per la inversió).



**Figura R4.4.** Diversitat nucleotídica ( $\pi$ , Nei 1987) en les 24 regions en estudi. Per les regions localitzades dins de la inversió  $J_1$  les estimes de  $\pi$  s'indiquen independentment per les línies  $J_1$  (color verd) i les línies  $J_{st}$  (color taronja).



**Figura R4.5.**  $D$  de Tajima (1989) en les 24 regions en estudi. Per les regions localitzades dins de la inversió  $J_1$  les estimes de l'estadístic  $D$  de Tajima s'indiquen independentment per les línies  $J_1$  (color verd) i les línies  $J_{st}$  (color taronja).

Els resultats obtinguts de les anàlisis dels nivells i patrons de polimorfisme en les regions localitzades dins de la inversió  $J_1$ , mostren valors similars als obtinguts en les regions localitzades fora de la inversió, corroborant que possiblement els nivells i patrons observats en les regions localitzades dins de la inversió també estiguin reflectint, almenys en part, la història

demogràfica recent de l'espècie. No obstant això, degut al caràcter adaptatiu de les inversions cromosòmiques a *D. subobscura*, alhora d'analitzar la demografia d'aquesta espècie, no es consideraran les sis regions incloses dins de la inversió ja que tot i que no s'observin nivells i patrons de variació diferents, les regions incloses dins de la inversió  $J_1$  no serien *a priori* bones candidates alhora de realitzar inferències demogràfiques ja que podrien reflectir fonamentalment la pròpia història evolutiva de la inversió.

#### R4.4 Diferenciació genètica entre les ordenacions $J_1$ i $J_{st}$

La diferenciació genètica entre les dues ordenacions analitzades  $J_1$  i  $J_{st}$  es va estimar mitjançant els estadístics:  $D_{xy}$ , nombre promig de diferències nucleotídiques per posició entre poblacions (Nei 1987) i  $D_a$ , nombre net de substitucions nucleotídiques per posició entre poblacions (Nei 1987). Els resultats obtinguts es mostren en la Taula R4.3.

Taula R4.3. Diferenciació genètica entre les ordenacions  $J_1$  i  $J_{st}$

Regions	Compartits	Fixats	$S_{x1}$	$S_{x2}$	$D_{xy}$	$D_a$
<b>p222</b>	0	4	24	10	0,0100	0,0062
<b>p127</b>	0	2	24	12	0,0053	0,0022
<b>P137</b>	21	3	134	71	0,0209	0,0080
<b>p221</b>	10	0	93	52	0,0139	0,0037
<b>p39</b>	36	0	154	53	0,0201	0,0047
<b>p174</b>	26	0	111	54	0,0178	0,0037
<b>Seqüències concatenades</b>	<b>82</b>	<b>9</b>	<b>498</b>	<b>244</b>	<b>0,0167</b>	<b>0,0052</b>

Compartits, nombre de llocs polimòrfics segregants per les mateixes dues variants en les dues ordenacions; Fixats, nombre de llocs polimòrfics amb variants fixades diferencialment;  $S_{x1}$ , nombre de llocs polimòrfics en l'ordenació  $J_1$  i monomòrfics en l'ordenació  $J_{st}$ ;  $S_{x2}$ , nombre de llocs polimòrfics en l'ordenació  $J_{st}$  i monomòrfics en l'ordenació  $J_1$ ;  $D_{xy}$ , promig del nombre de diferències nucleotídiques per posició entre les dues ordenacions corregida per Jukes i Cantor (1969);  $D_a$ , nombre net de diferències nucleotídiques per posició entre ordenacions corregida per Jukes i Cantor (1969).

Les regions p222 i p127, presenten respectivament 4 i 2 diferències fixades i cap polimorfisme compartit. Aquestes regions són les més properes als punts de trencament. Per contra, les regions localitzades més al centre de la inversió  $J_1$  (p137, p221, p39 i p174), presenten un elevat nombre de polimorfismes compartits, probablement degut a l'intercanvi genètic entre ordenacions. Les estimes de diferenciació genètica ( $D_{xy}$ ) entre les ordenacions  $J_1$  i

$J_{st}$  varien de 0,0053 (p127) a 0,209 (p137). Quan l'estadístic  $D_{xy}$  és corregit pel nivell de variació en cada regió, la regió p127 continua presentant els nivells menors de diferenciació genètica i la regió p137 els majors.

La diferenciació genètica també es va estimar mitjançant l'estadístic  $F_{ST}$  (Hudson et al. 1992b). Per contrastar si la diferenciació genètica entre les dues ordenacions en estudi era significativa, es va aplicar el test de les permutacions proposat per Hudson et al. (1992b) basat en l'estadístic  $K_s^*$ . La significació estadística de l'estadístic  $K_s^*$ , es va determinar per simulacions de Monte Carlo en realitzar 1000 permutacions aleatòries de la mostra total. Per avaluar el possible efecte de la inversió en regions properes al punt de trencament però localitzades fora d'aquesta, es van incloure en l'anàlisi les regions p148 i p32096. Els resultats obtinguts es mostren en la Taula R4.4.

**Taula R4.4.** Diferenciació genètica entre les dues ordenacions i test de les permutacions

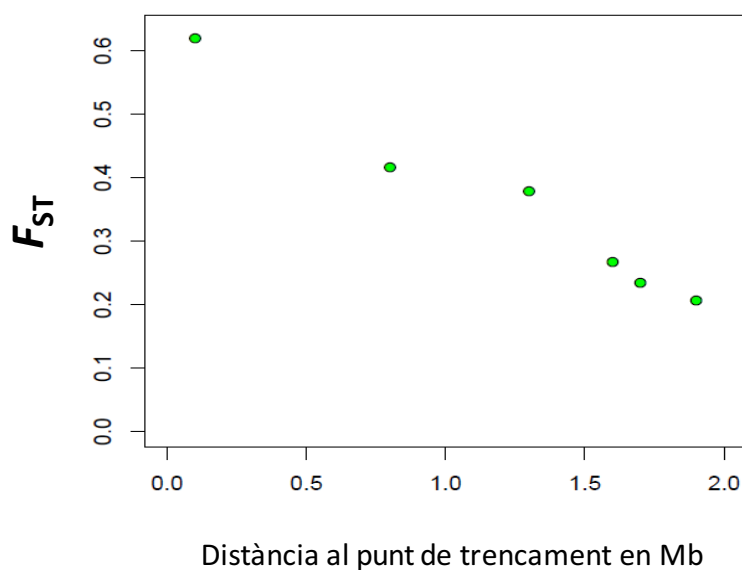
Regions	$F_{ST}$	$(k_s^*)$
<b>p222</b>	0,619	< 0,001
<b>p127</b>	0,415	< 0,001
<b>P137</b>	0,378	< 0,001
<b>p221</b>	0,266	< 0,001
<b>p39</b>	0,233	< 0,001
<b>p174</b>	0,206	< 0,001
Seqüències concatenades	0,307	< 0,001
<b>p148</b>	0,003	0,645
<b>p32096</b>	< 0,001	0,140

$F_{ST}$ , estima de la diferenciació genètica (Hudson 1992b);  $k_s^*$ , probabilitat del test de les permutacions basat en l'estadístic  $k_s^*$ . Es mostren les regions ordenades en funció de la seva proximitat al punt de trencament, així com també els resultats obtinguts al concatenar-les. En gris es mostren les regions p148 i p32096, les dues regions analitzades fora de la inversió  $J_1$  però més properes als punts de trencament en estudi.

S'observa una diferenciació genètica significativa entre les ordenacions  $J_1$  i  $J_{st}$  en totes les regions incloses dintre de la inversió  $J_1$  (test de les permutacions basat en l'estadístic  $K_s^*$ ;  $P < 0,001$ ). Per altra banda, es detecta una major diferenciació genètica en les regions p222 i p127 que són les regions més properes als punts de trencament i les que presenten diferències fixades. Tot i que en les tres regions centrals de la inversió no s'observin diferències fixades, la

diferenciació genètica observada continua sent altament significativa. Aquesta diferenciació genètica deixa de ser significativa en les dues regions analitzades més properes als punts de trencament però localitzades fora de la inversió  $J_1$  (p148 i p237). Aquestes regions es troben aproximadament a 0,9 i 1,4 Mb d'un dels punts de trencament de la inversió  $J_1$ .

Es va analitzar la diferenciació genètica de cada regió ( $F_{ST}$ ) localitzada dins de la inversió en funció de la distància al punt de trencament més proper. Els resultats obtinguts es mostren en la Figura R4.6.



**Figura R4.6.** Diferenciació genètica ( $F_{ST}$ ) entre ordenacions cromosòmiques per a cada regió en estudi en funció de la distància al punt de trencament més proper de la inversió  $J_1$ . Aquesta distància es va estimar considerant una distribució homogènia del DNA al llarg de tota la inversió  $J_1$  i utilitzant com a referència del contingut de DNA de l'element cromosòmic les dades publicades per l'espècie *D. pseudoobscura* (Schaeffer et al. 2008).

La diferenciació genètica en funció de la distància de cada regió al punt de trencament és significativa (coeficient de correlació d'*Spearman*;  $P < 0,005$ ) i s'observa una major diferenciació com més a la vora del punt de trencament es localitzen les regions.

#### R4.5 Conversió gènica

Tot i la clara diferenciació genètica entre les dues ordenacions, es van detectar polimorfismes compartits. La presència d'aquests polimorfismes s'hauria d'explicar per mutacions recurrents o alternativament a l'intercanvi genètic degut a conversió gènica o a

dobles entrecreuaments. Seguint l'algoritme proposat per Bertrán et al. (1997), es van analitzar el nombre d'esdeveniments d'intercanvi genètic entre ordenacions cromosòmiques.

Els resultats obtinguts es mostren en la Taula R4.5.

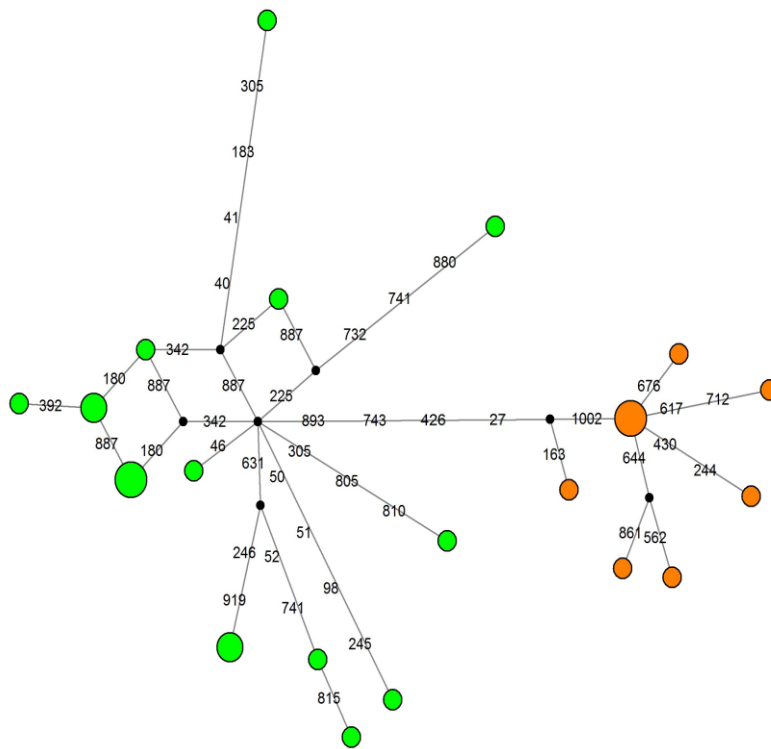
**Taula R4.5.** Tractes de conversió gènica identificats en les diferents regions en estudi

Regions	Nombre de tractes de conversió gènica	Llargada dels tractes
<b>p222</b>	0	
<b>p127</b>	0	
<b>P137</b>	4	8, 45, 22, 9
<b>p221</b>	3	2, 97, 191
<b>p39</b>	2	4, 140
<b>p174</b>	3	219, 136, 17, 455

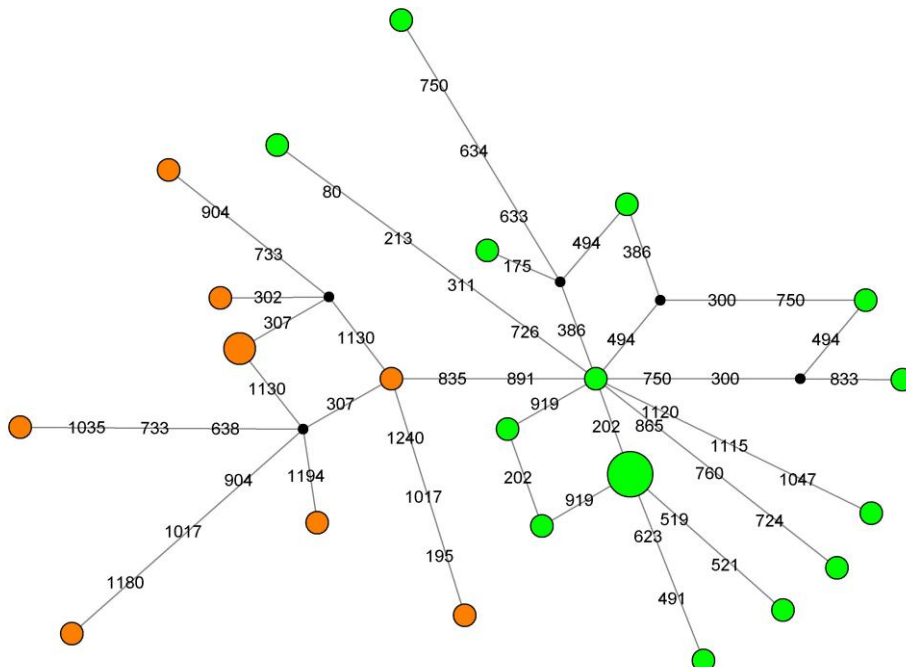
L'ordre de les diferents regions en estudi que es mostra a la taula es dona en funció de la seva proximitat al punt de trencament.

Es detecta intercanvi genètic en les regions localitzades més al centre de la inversió (p137, p221, p39 and p174), regions que, com s'ha comentat anteriorment, també presenten un elevat nombre de polimorfismes compartits. En les dues regions més properes als punts de trencament (p222 i p127) no es detecta intercanvi genètic. Aquests resultats evidencien la presència d'un major intercanvi genètic entre ordenacions en les regions centrals de la inversió, respecte a les més properes als punts de trencament. Si es considera que la taxa de conversió gènica és uniforme al llarg de tota la inversió, aquests resultats podrien indicar que els dobles entrecreuaments haurien contribuït a l'intercanvi genètic en regions centrals de la inversió  $J_1$ .

Per les dues regions més properes als punts de trencament (p222 i p127) i en les que no es detecta intercanvi genètic, es va reconstruir una xarxa d'haplotips, mitjançant el programa *Network* basat en mètodes de màxima versemblança i a través de la metodologia *median-joining network*. Els resultats es mostren en les Figures R4.7 i R4.8.



**Figura R4.7.** Xarxa d'haplotips de la regió p222. Les línies amb l'ordenació  $J_1$  estan representades en verd i les línies amb l'ordenació  $J_{st}$  estan representades en taronja. Quan més gran és el diàmetre de les circumferències més línies comparteixen aquest haplotip.

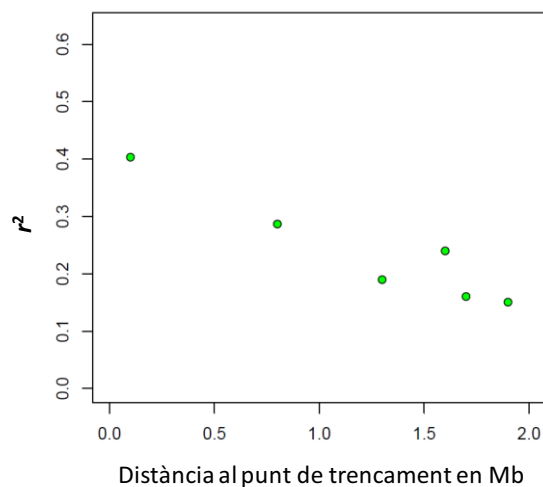


**Figura R4.8.** Xarxa d'haplotips de la regió p127. Les línies amb l'ordenació  $J_1$  estan representades en verd i les línies amb l'ordenació  $J_{st}$  estan representades en taronja. Quan més gran és el diàmetre de les circumferències més línies comparteixen aquest haplotip.

Tot i que les dues regions properes als punts de trencament presentin una elevada diferenciació genètica en relació a les dues ordenacions cromosòmiques, aquesta diferenciació és major en la regió p222, la regió més propera al punt de trencament, fet que s'observa per un nombre de mutacions més elevat en la branca que diferencia les dues ordenacions.

#### R4.6 Desequilibri de lligament

Per estudiar el desequilibri de lligament, es va analitzar l'associació entre les ordenacions cromosòmiques ( $J_1$  i  $J_{st}$ ) i les variants presents a les posicions polimòrfiques informatives (una posició és informativa quan almenys presenta dues variants cadascuna d'elles presents almenys en dues seqüències) mitjançant l'estadístic  $r^2$  (Hill i Robertson 1968). Es va estudiar el desequilibri de lligament en funció de la distància al punt de trencament més proper. Els resultats es mostren en la Figura R4.9.



**Figura R4.9.** Associació entre les dues ordenacions cromosòmiques i les variants presents a les posicions informatives per cada regió en funció de la distància al punt de trencament més proper. Aquesta distància es va estimar considerant una distribució homogènia del DNA al llarg de tota la inversió  $J_1$  i utilitzant com a referència del contingut de DNA de l'element cromosòmic les dades publicades per l'espècie *D. pseudoobscura* (Schaeffer et al. 2008).

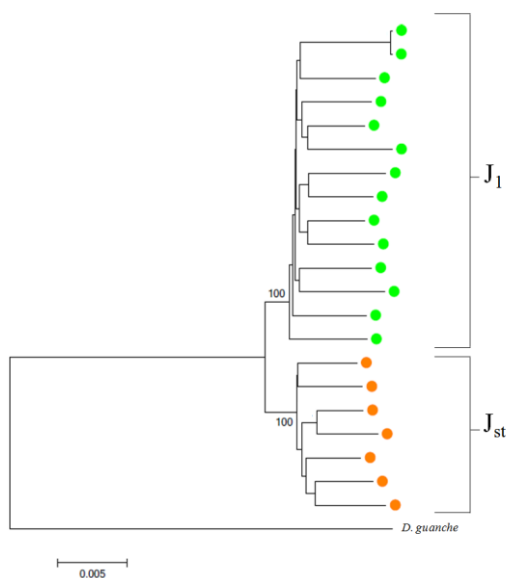
La regió més propera al punt de trencament (p222) és la que mostra una associació més elevada entre les diferents ordenacions cromosòmiques i la variació nucleotídica. S'observa una correlació positiva entre el desequilibri de lligament i la distància al punt de trencament (coeficient de correlació d'*Spearman*;  $P < 0,005$ ). Tot i que aquesta associació



decreix en les altres regions, si s'elimina la regió p222 de l'anàlisi es continua observant una correlació positiva entre el desequilibri de lligament i la distància al punt de trencament (coeficient de correlació d'*Spearman*;  $P < 0,037$ ). Dades teòriques prediuen que s'esperaria un elevat desequilibri de lligament en regions properes als punts de trencament i en regions on la selecció podria mantenir complexos de gens coadaptats (Navarro et al 1997, 2000).

#### R4.7 Genealogia de la mostra de seqüències

Es va utilitzar el programa MEGA5 (Tamura et al. 2011) amb el mètode de *neighbor-joining* (Saitou y Nei 1987) per obtenir la genealogia dels al·lels del concatenat de les sis regions localitzades dins de la inversió  $J_1$ . Per aquesta anàlisi es van utilitzar 21 línies homocariotípiques, 14 de les quals tenen l'ordenació cromosòmica  $J_1$  i 7 l'ordenació  $J_{st}$ . No es van utilitzar les 27 línies homocariotípiques obtingudes ja que en algunes de les línies amb ordenació  $J_1$  no va ser possible amplificar alguna de les regions estudiades. Els resultats de l'arbre genealògic es mostren a la Figura R4.10.



**Figura R4.10.** Genealogia obtinguda pel mètode *neighbor-joining* basat en la variació nucleotídica del concatenat de les seqüències de les regions incloses dintre de la inversió  $J_1$ . Els valors que es mostren en l'arbre genealògic són els percentatges de *bootstrap* (10.000 rèpliques). Les circumferències taronges indiquen aquelles línies que presenten l'ordenació  $J_{st}$  mentre que les circumferències verdes indiquen aquelles línies que tenen ordenació  $J_1$ . *D. guanche* s'utilitza com a *outgroup*.

La genealogia del concatenat de les sis regions en estudi mostra una clara agrupació de les línies segons la seva ordenació cromosòmica. A més, els dos clústers que agrupen, respectivament, les línies  $J_{st}$  i  $J_1$  presenten un elevat valor de *bootstrap*. Aquesta agrupació diferencial preval tot i la presència d'intercanvi genètic en quatre de les sis regions en estudi localitzades en les parts més centrals de la inversió (p137,p221, p174, p39). La genealogia de la mostra indica també que les dues ordenacions són antigues i és consistent amb l'origen monofilètic de les inversions cromosòmiques. La topologia de l'arbre és en forma d'estrella, caracteritzada per les branques dels nodes interns més curtes respecte a les dels nodes externs.

## R5 INFERÈNCIES DEMOGRÀFIQUES EN L'ESPÈCIE *D. SUBOBSCURA*

En els apartats R5.1 i R5.2 es presenten els resultats de les inferències demogràfiques realitzades en la població natural de *D. subobscura* de Barcelona. Per aquestes anàlisis es van utilitzar 16 de les 18 regions no codificadores distribuïdes al llarg del cromosoma J no afectades per inversions cromosòmiques ni per processos de selecció natural (veure Resultats R3). Donat que per una banda en la regió p196 no es va poder seqüenciar l'espècie *outgroup* i que, per una altra, les anàlisis efectuades en la regió p169 semblaven indicar que aquesta podia estar sotmesa a pressions selectives, aquestes dues regions no van ser considerades per aquestes anàlisis. Les regions que es van utilitzar són: p8593, p32096, p148, p237, p215, p227, p10849, p220, p175, p271, p187, p203, p204, p248, p018 i p108. Es van analitzar aproximadament 30 kb (excloent les posicions amb *gaps*) amb una grandària mostral que varia de 18 a 21 línies homocariotípiques (a excepció de la regió p32096 per la que només es van poder amplificar 13 línies) i es va utilitzar *D. guanche* com a *outgroup*. Una vegada es va identificar el model demogràfic més probable que pot explicar la variabilitat nucleotídica observada en l'actualitat a *D. subobscura* es van realitzar altres inferències utilitzant aquest model en les anàlisis tant en regions incloses dintre de la inversió J<sub>1</sub> (apartat R5.3) com en possibles regions afectades per la selecció natural (apartat R5.4).

### R5.1 Estimació del temps de divergència entre diversos llinatges

Es va estimar el temps de divergència entre *D. subobscura* i *D. guanche*, informació necessària per les anàlisis d'ABC posteriors, i el temps de separació entre els subgrups *pseudoobscura* i *subobscura*.

#### Model de substitució

El model de substitució nucleotídica més probable de totes les regions en estudi és el model HKY, amb taxes evolutives uniformes en la major part d'elles, a excepció de les regions p175, p215 i p18 amb taxes no uniformes modelades per una distribució Gamma i les regions

p237 i p148 amb taxes no uniformes, considerant que una proporció de les posicions són evolutivament invariables.

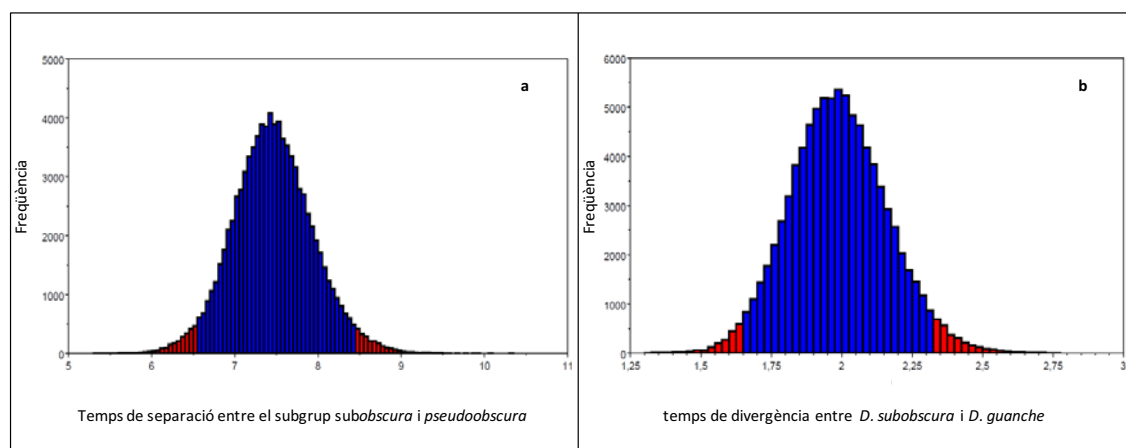
### Temps de divergència

Els resultats obtinguts amb el programa BEAST v.1.6.2 (Drummond i Rambaut 2007) es mostren en la Taula R5.1 on hi ha diversos estadístics descriptius de la distribució posterior dels paràmetres d'interès, en les Figures R5.1-a i R5.1-b on hi ha representades les distribucions posteriors d'aquests paràmetres i en la Figura R5.2 on es mostra l'arbre filogenètic més probable.

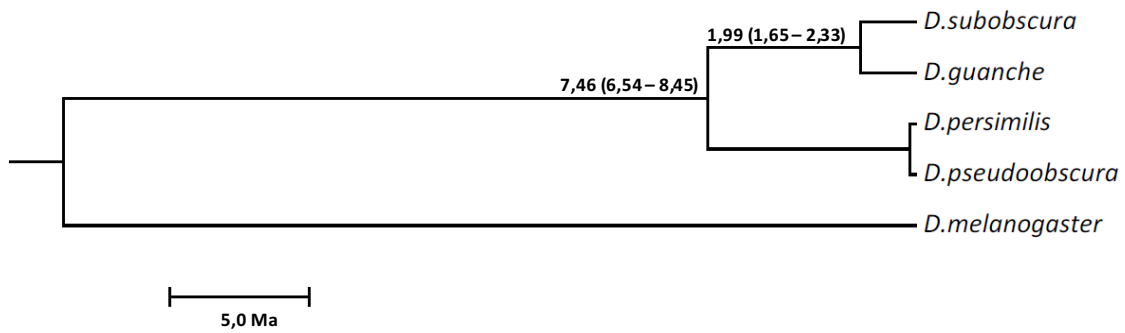
**Taula R5.1.** Estadístics descriptius de la distribució posterior dels paràmetres temps de separació entre el subgrups *subobscura* i *pseudoobscura* i temps de separació entre *D. subobscura* i *D. guanche*.

Temps de separació	grup <i>obscura</i> i grup <i>pseudoobscura</i>	<i>D. subobscura</i> i <i>D. guanche</i>
Mitjana	7,46	1,99
Desviació estàndard	$1,06 \times 10^{-02}$	$4,5 \times 10^{-03}$
mediana	7,45	1,99
límit inferior del 95% de la distribució	6,54	1,65
límit superior del 95% de la distribució	8,45	2,33

Els estadístics descriptius de la distribució posterior dels dos paràmetres en estudi venen expressats en milions d'anys. Es presenta la mitjana, la desviació estàndard i la mediana, així com també els límits inferior i superior del 95% de la distribució posterior.



**Figura R5.1-a.** Distribució posterior del temps de separació entre el grup *subobscura* i el grup *pseudoobscura*, expressat en milions d'anys. **b.** Distribució posterior del temps de divergència entre *D. subobscura* i *D. guanche*, expressat en milions d'anys. En color blau es mostra el 95% de la distribució.



**Figura R5.2.** Arbre filogenètic més probable inferit amb el programa BEAST. Al costat dels nodes es mostra el temps de divergència més probable entre els diferents llinatges expressat en milions d'anys (Ma). Entre parèntesis es mostra els límits inferior i superior del 95% de la distribució posterior dels temps de divergència dels diferents llinatges.

La mitjana estimada del temps de separació dels subgrups *subobscura* i *pseudoobscura* és de 7,5 milions d'anys (6,54 - 8,45 com a límits inferior i superior del 95% de la distribució). Així mateix, la mitjana estimada del temps de divergència entre *D. subobscura* i *D. guanche* és de 2 milions d'anys (1,65 - 2,33 com a límits inferior i superior del 95% de la distribució).

## R5.2 Història demogràfica recent de *D. subobscura* estimada a través de metodologies d'Approximate Bayesian Computation

### Disseny de la distribució *a priori* pel paràmetre poblacional de mutació ( $\theta$ )

Per dissenyar la distribució *a priori* del paràmetre poblacional de mutació ( $\theta$ ), sent  $\theta = 4N_e\mu$  (on  $N_e$  és la grandària efectiva de l'espècie i  $\mu$  la taxa de mutació per posició i per generació), es van considerar constants tant el temps de divergència entre *D. subobscura* i *D. guanche* com el nombre de generacions per any. Es va condicionar el paràmetre poblacional de mutació a la possible variació en la grandària efectiva de l'espècie, dissenyant per aquesta grandària una distribució *a priori* uniforme amb valors que oscil·len entre  $1 \times 10^6$  (Pascual et al. 2007) i  $1 \times 10^7$  (Comeron et al. 1997) individus. Es va calcular la taxa de mutació per posició i per generació ( $\mu$ ), de cadascuna de les 16 regions en estudi sota la premissa de neutralitat d'acord amb l'expressió  $\mu = K/2t$ . Es va utilitzar la divergència nucleotídica  $K$  (Nei 1987) corregida d'acord a Jukes i Cantor (1969) de cadascuna d'elles i el temps de divergència expressat en generacions entre *D. subobscura* i *D. guanche* que es va fixar en dos milions d'anys i

considerant cinc generacions per any. El valor promig de  $\mu$  que es va obtenir va ser de  $2,51 \times 10^{-9}$  mutacions per posició i per generació. Posteriorment es va dissenyar la distribució *a priori* del paràmetre poblacional de mutació, a partir d'aquest valor promig de la taxa de mutació i de la distribució *a priori* de la grandària efectiva de l'espècie ( $N_e$ ), obtenint una distribució *a priori* uniforme amb valors que oscil·len de  $\theta = 0,01$  a  $\theta = 0,1$ . En la Taula R5.2 es mostra la divergència estimada per les diferents regions i les corresponents estimes de la taxa de mutació.

**Taula R5.2.** Divergència nucleotídica i taxa de mutació per posició i per generació de les 16 regions en estudi.

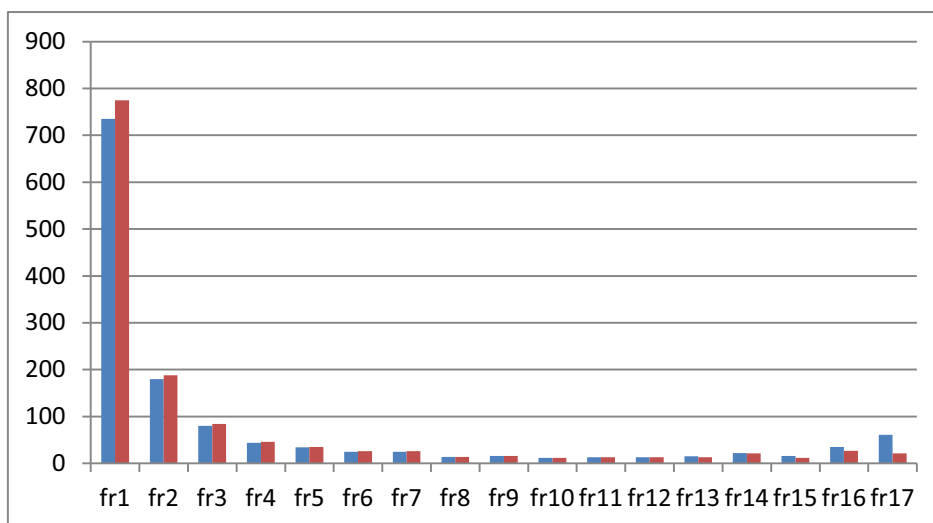
regions	<i>K</i>	$\mu$
p175	0,04	$2,00 \times 10^{-9}$
p32096	0,05	$2,70 \times 10^{-9}$
p148	0,05	$2,30 \times 10^{-9}$
p237	0,07	$3,40 \times 10^{-9}$
p215	0,06	$3,10 \times 10^{-9}$
p227	0,03	$1,40 \times 10^{-9}$
p8593	0,04	$2,20 \times 10^{-9}$
p220	0,09	$4,50 \times 10^{-9}$
p271	0,04	$1,75 \times 10^{-9}$
p187	0,03	$1,60 \times 10^{-9}$
p203	0,06	$2,75 \times 10^{-9}$
p10849	0,02	$1,20 \times 10^{-9}$
p204	0,04	$1,75 \times 10^{-9}$
p248	0,07	$3,70 \times 10^{-9}$
p018	0,05	$2,55 \times 10^{-9}$
p108	0,07	$3,25 \times 10^{-9}$
mitjana	0,05	$2,51 \times 10^{-9}$

*K*, divergència nucleotídica (Nei, 1987) corregida d'acord a Jukes i Cantor (1969);  $\mu$ , taxa de mutació per posició i per generació ( $\mu = K/2t$ , sent *t* el temps de divergència entre *D. subobscura* i *D. guanche* estimat en dos milions d'anys i considerant cinc generacions per any).

### Efecte de les mutacions recurrents en l'espectre de freqüències desdoblada

En la Figura R5.3 es mostra l'espectre de freqüències desdoblada observat, és a dir en el qual els polimorfismes es polaritzen per parsimònia tenint en compte la seqüència de l'*outgroup*, així com també es mostra aquest espectre corregit utilitzant la correcció proposada a Hernandez et al. (2007) que té en compte la possibilitat de mutacions recurrents. En la Taula

R5.3 es mostra el tant per cent de guany o pèrdua de polimorfismes en l'espectre de freqüències desdoblada corregida respecte l'espectre de freqüències desdoblada observada.



**Figura R5.3.** Espectre de freqüències desdoblada observada, polaritzant els polimorfismes per parsimònia tenint en compte la seqüència de l'espècie *outgroup D. guanche* (en blau) i espectre de freqüències desdoblada corregida (en vermell). En l'eix de les X es mostra la freqüència dels diferents polimorfismes i en l'eix de les Y el seu nombre.

**Taula R5.3.** Percentatge de guany o pèrdua de polimorfismes en l'espectre de freqüències desdoblada corregida, respecte l'espectre de freqüències desdoblada observada.

Freqüència dels polimorfismes	fr1	fr2	fr3	fr4	fr5	fr6	fr7	fr8	fr9	fr10	fr11	fr12	fr13	fr14	fr15	fr16	fr17
% guany o pèrdua	5,3	4,7	4,7	2,9	3,3	2,7	2,7	0	0	0	-5,2	-5,1	-7,8	-5,8	-23,7	-24	-64,8

Els valors del tant per cent de guany es mostren en positiu i els valors de tant per cent de pèrdua es mostren en negatiu.

L'espectre de freqüències desdoblada observada en les 15 regions en estudi en les que la grandària mostral és igual o superior a 18 individus (no es considera la regió p32096 en aquesta anàlisi degut a que la seva grandària mostral és de 13 individus) està caracteritzat per una elevada proporció de polimorfismes amb variants a baixa freqüència. Quan s'aplica la correcció proposada a Hernandez et al. (2007), el tant per cent de guany de polimorfismes amb variants a baixa freqüència no és elevat, ja que la població en estudi ja presenta inicialment un nombre elevat d'aquests polimorfismes. Amb els polimorfismes amb variants a elevada freqüència, s'observa el patró contrari: en l'espectre de freqüències desdoblada

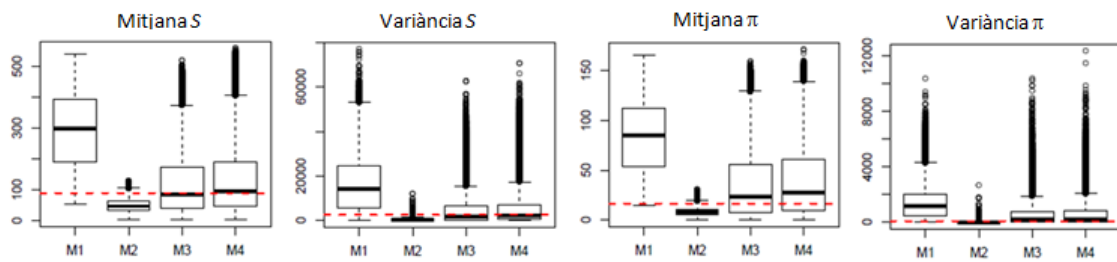
observat hi ha relativament pocs polimorfismes amb variants a elevada freqüència i quan es corregeix l'espectre de freqüències desdoblant hi ha una pèrdua important d'aquests polimorfismes.

En una espècie com *D. subobscura* que mostra uns nivells elevats de polimorfisme és important, per tant, tenir en compte la possibilitat de mutacions recurrents quan s'efectuen simulacions de coalescència. La inclusió d'aquesta possibilitat reflecteix d'una manera més acurada el procés mutacional i, per tant, permet utilitzar l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) que és informatiu per les variants derivades a elevada freqüència sense els problemes associats amb la identificació errònia de les variants ancestrals (Baudry i Depaulis 2003).

#### **Estadístics simulats**

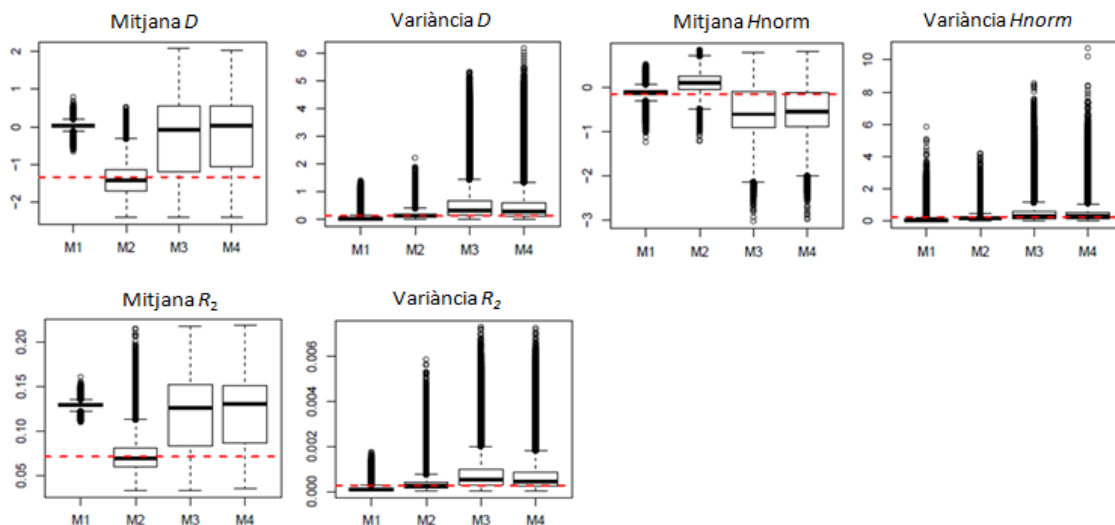
Per cadascun dels 4 models evolutius considerats (veure Figura M3.2), es va realitzar un milió de simulacions, condicionades als paràmetres que els defineixen. Es va resumir la informació continguda en cada simulació en la mitjana i la variància d'un conjunt d'estadístics, els estadístics simulats. Els estadístics simulats es van classificar en diferents categories: a) estadístics que aporten informació sobre els nivells de variació:  $S$ , nombre de llocs segregants i  $\pi$ , diversitat nucleotídica per locus (Tajima 1983); b) estadístics de classe I (Ramos-Onsins i Rozas, 2002) que detecten desviacions de l'espectre de freqüències:  $D$ , estadístic  $D$  de Tajima (1989);  $R_2$ , estadístic  $R_2$  de Ramos-Onsins i Rozas (2002) i  $H_{norm}$ , estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000, Zeng et al. 2006); c) estadístics de classe II (Ramos-Onsins i Rozas, 2002) que utilitzen la informació de la distribució dels haplotips i per tant del desequilibri de lligament i la recombinació:  $H_w$ , diversitat haplotípica (Nei 1987);  $F_s$  de Fu (1997) i  $Z_A$  de Rozas et al. (2001). En les Figures R5.4-R5.6 es mostra en diferents diagrames de caixa, la distribució de la mitjana i la variància del milió d'estadístics simulats pels 4 models demogràfics plantejats, així com també els valor promig observat d'aquests estadístics.





**Figura R5.4.** Diagrama de caixa del milió d'estadístics simulats que aporten informació dels nivells de variació sota els quatre models demogràfics en estudi. M1, model neutre estàndard; M2, model d'expansió; M3, model de coll d'ampolla i M4 model de barreja. La línia discontinua vermella és el valor promig de la mitjana o la variància dels estadístics observats, sent aquests estadístics:  $S$ , nombre de llocs segregants i  $\pi$ , diversitat nucleotídica per locus (Tajima 1983).

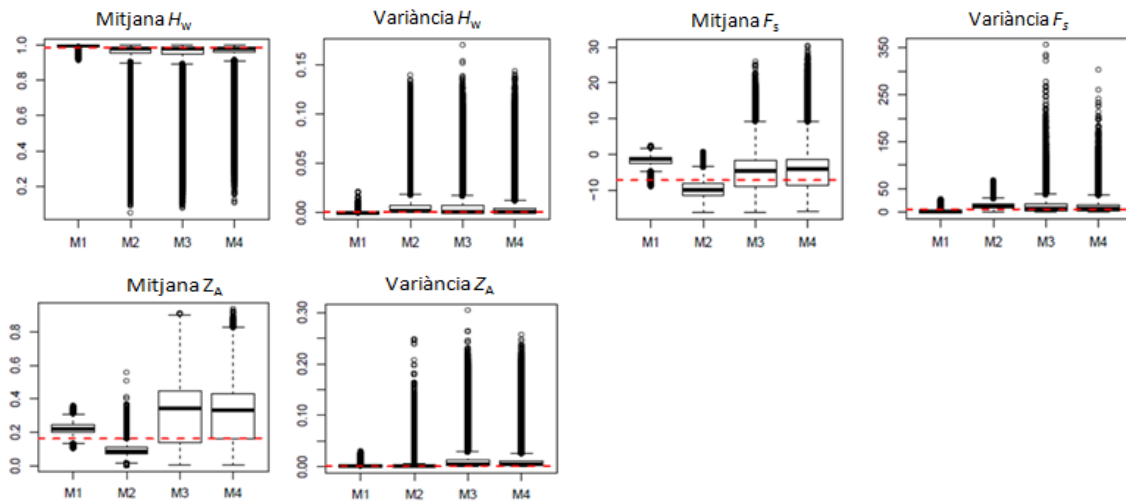
En els estadístics simulats que aporten informació sobre els nivells de variació, s'observa que els tres models que plantegen algun canvi demogràfic al llarg de la història evolutiva de *D. subobscura* (M2, M3 i M4), presenten en general valors que s'ajusten més a les dades observades respecte el model M1, model neutre estàndard.



**Figura R5.5.** Diagrama de caixa del milió d'estadístics simulats de classe I sota els quatre models demogràfics en estudi: M1, model neutre estàndard; M2, model d'expansió; M3, model de coll d'ampolla i M4 model de barreja. La línia discontinua vermella és el valor promig de la mitjana o la variància dels estadístics observats, sent aquests estadístics:  $D$ , estadístic  $D$  de Tajima (1989);  $H_{norm}$ , estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i  $R_2$ , estadístic  $R_2$  de Ramos-Onsins i Rozas (2002).

En els models que plantegen un coll d'ampolla (M3 i M4) s'observa una variància més elevada en els valors simulats de l'estadístic  $D$  de Tajima, respecte el model d'expansió M2. Aquest patró és similar al de l'estadístic  $R_2$ , sensible per detectar expansions poblacionals.

L'elevat nombre de polimorfismes amb variants a baixa freqüència observat en les anàlisis de les seqüències de DNA podria reflectir, per tant, una expansió en el passat de l'espècie. En relació a l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006), s'observa com en els models M2, M3 i M4 els valors de la distribució d'aquest estadístic difereixen de l'esperat sota el model neutre estàndard.



**Figura R5.6.** Diagrama de caixa del milió d'estadístics simulats de classe II sota els quatre models demogràfics en estudi: M1, model neutre estàndard; M2, model d'expansió; M3, model de coll d'ampolla i M4 model de barreja. La línia discontinua vermella és el valor promig de la mitjana o la variància dels estadístics observats, sent aquests estadístics:  $H_w$ , diversitat haplotípica (Nei 1987);  $F_s$  de Fu (1997) i estadístic  $Z_A$  (Rozas et al. 2001).

Els quatre models en estudi presenten valors elevats de diversitat haplotípica ( $H_w$ ), podent reproduir en els 4 models els valors observats d'aquest estadístic. Per l'estadístic  $F_s$  de Fu (1997), sota el model neutre estàndard (M1) s'esperarien valors propers a zero, tal i com s'observa en les simulacions. En el model M2 els valors simulats de l'estadístic  $Z_A$ , tendeixen a presentar valors menors dels observats, tot i que estarien dintre del 95% de l'interval de confiança de la distribució.

En el model neutre estàndard M1, en el qual l'espècie no ha patit canvis en un període llarg de la seva història evolutiva i per tant la variabilitat nucleotídica està en equilibri derivat de mutació, els valors dels estadístics simulats  $D$  de Tajima i  $H$  normalitzat de Fay i Wu són propers a zero, tal i com s'esperaria sota aquestes condicions. Els valors observats en la

població en estudi estan allunyats dels valors simulats d'aquests estadístics sota el model neutre estàndard. Aquesta desviació, posa de manifest que la variació a *D. subobscura* no es troba en l'actualitat en equilibri deriva-mutació i que, per tant, la seva història evolutiva haurà estat influenciada per factors demogràfics. En les anàlisis posteriors no s'inclou, per tant, el model M1 en no poder explicar les dades observades.

### Tria d'estadístics

Es va realitzar una primera anàlisi per avaluar quins estadístics podien aportar més informació per estimar els diferents paràmetres en estudi, a través de la correlació de *Pearson* entre els estadístics i els paràmetres. En la Taula R5.4 es mostren els resultats obtinguts d'aquestes correlacions.

**Taula R5.4.** Coeficient de correlació de *Pearson* entre cadascun del milió d'estadístics simulats d'un model i els valors dels paràmetres que el defineixen de cadascuna de les simulacions en els 4 models en estudi. M1, model neutre estàndard; M2, model d'expansió; M3, model de coll d'ampolla i M4 model de barreja.

Models/estadístics	paràmetres comuns						paràmetres propis de cada model										
	$\theta$			R			$N_{1a}$	$T_{1a}$	$N_{2a}$	$T_{1b}$	$N_{1b}$	$T_2$	$N_{1b}$	$T_2$	$N_{3b}$	$T_{1b}$	$N_{2b}$
	M2	M3	M4	M2	M3	M4	M2	M2	M3	M3	M3	M3	M4	M4	M4	M4	M4
mitjana <i>S</i>	0,5	0,6	0,6	<0,05	<0,05	<0,05	0,4	<0,05	0,5	0,4	0,3	0,5	0,3	0,5	0,3	0,5	0,2
variància <i>S</i>	0,5	0,6	0,6	<0,05	0,1	0,1	0,4	<0,05	0,3	0,4	0,3	0,4	0,3	0,4	0,2	0,4	0,1
mitjana $\pi$ per locus	0,5	0,6	0,6	<0,05	<0,05	<0,05	0,6	0,1	0,4	0,4	0,3	0,5	0,3	0,5	0,3	0,5	0,2
variància $\pi$ per locus	0,5	0,6	0,6	0,1	0,1	0,1	0,5	0,1	0,3	0,4	0,3	0,4	0,3	0,4	0,2	0,4	0,1
mitjana <i>D</i>	0,1	0,4	0,4	<0,05	<0,05	<0,05	0,6	0,2	0,5	0,5	0,2	0,5	0,2	0,5	0,4	0,5	0,1
variància <i>D</i>	<0,05	0,1	0,1	0,2	0,3	0,3	0,1	0,3	0,2	0,2	0,1	0,2	0,1	0,2	0,3	0,2	0,1
mitjana <i>H</i> norm	0,6	0,4	0,4	<0,05	<0,05	<0,05	0,2	0,7	0,3	0,5	0,2	0,5	0,1	0,5	0,1	0,5	<0,05
variància <i>H</i> norm	0,1	0,1	0,1	0,1	0,3	0,3	<0,05	0,3	0,2	0,2	0,1	0,2	0,1	0,2	0,2	0,2	0,1
mitjana <i>R</i> <sub>2</sub>	<0,05	0,4	0,4	<0,05	<0,05	<0,05	0,4	0,2	0,5	0,5	0,2	0,5	0,2	0,5	0,4	0,6	0,1
variància <i>R</i> <sub>2</sub>	0,1	0,1	0,1	0,1	0,3	0,3	0,1	0,3	0,3	0,2	0,1	0,2	0,1	0,1	0,3	0,1	0,1
mitjana <i>H</i> <sub>w</sub>	0,2	0,2	0,2	0,1	0,1	0,2	0,3	0,2	0,4	0,1	<0,05	<0,05	<0,05	0,1	0,4	<0,05	0,1
variància <i>H</i> <sub>w</sub>	0,2	0,3	0,3	<0,05	0,1	0,1	0,3	0,2	0,4	<0,05	<0,05	0,1	<0,05	0,1	0,4	0,1	0,1
mitjana <i>F</i> <sub>s</sub>	0,2	0,5	0,5	0,4	0,2	0,2	0,3	0,3	0,4	0,6	0,2	0,6	0,2	0,6	0,3	0,6	0,1
variància <i>F</i> <sub>s</sub>	0,3	0,2	0,2	0,1	0,2	0,2	0,3	<0,05	0,5	0,2	<0,05	0,2	<0,05	0,3	0,4	0,3	0,1
mitjana <i>Z</i> <sub>A</sub>	0,3	0,5	0,5	0,1	0,1	0,1	0,5	0,3	0,3	0,5	0,2	0,5	0,2	0,5	0,1	0,5	<0,05
variància <i>Z</i> <sub>A</sub>	<0,05	0,1	0,1	0,1	0,2	0,3	0,1	0,2	0,3	0,2	0,1	0,2	0,1	0,1	0,3	0,1	0,1

Els diferents colors mostren els valors del rang de correlació: negre ( $0,8 \leq |r|$ ), vermell ( $0,6 \leq |r| < 0,8$ ), taronja ( $0,4 \leq |r| < 0,6$ ), groc ( $0,2 \leq |r| < 0,4$ ) i no color ( $|r| < 0,2$ ). *S*, nombre de llocs segregants;  $\pi$ , diversitat nucleotídica per locus (Tajima 1983); *D*, estadístic *D* de Tajima (1989); *H*norm, estadístic *H* normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006); *R*<sub>2</sub>, estadístic *R*<sub>2</sub> de Ramos-Onsins i Rozas (2002); *H*<sub>w</sub>, diversitat haplotípica (Nei 1987); *F*<sub>s</sub>, estadístic *F*<sub>s</sub> de Fu (1997); *Z*<sub>A</sub>, estadístic *Z*<sub>A</sub> (Rozas et al. 2001).

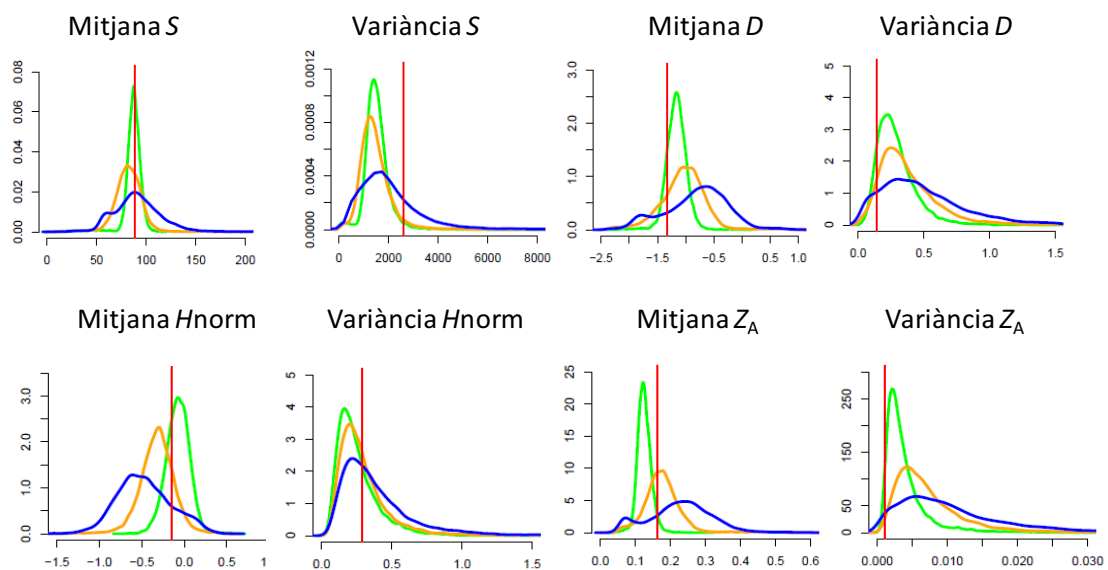
Els estadístics simulats que mostren una major correlació amb el paràmetre poblacional de mutació són els estadístics que aporten informació sobre els nivells de variabilitat, presentant ambdós estadístics nivells similars de correlació amb aquest paràmetre. En relació al paràmetre poblacional de recombinació, no s'observa ni per la mitjana ni per la variància de cap dels estadístics proposats, correlacions iguals o superiors a 0,5. Aquests valors posen de manifest la poca informació prèvia disponible en relació a la taxa de recombinació de l'espècie que conduí a una distribució *a priori* molt àmplia. Cal afegir que es va valorar la informació que poden aportar els estadístics de classe I en quant als paràmetres propis de cada model. Altrament, en el model M4 s'observa com pel paràmetre  $N_{2b}$ , paràmetre que defineix la proporció d'individus menor de les dues subpoblacions, cap de les correlacions presenta valors superiors a 0,2.

El conjunt d'estadístics utilitzat en la tria de models va ser el següent: en relació als estadístics que aporten informació sobre els nivells de variabilitat nucleotídica es va utilitzar el nombre de llocs segregants ( $S$ ), en relació als estadístics que aporten informació sobre els patrons de variabilitat nucleotídica es van utilitzar l'estadístic  $D$  de Tajima (1989) i l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006), i finalment en relació als estadístics que poden aportar informació sobre el desequilibri de lligament es va utilitzar l'estadístic  $Z_A$  (Rozas et al. 2001).

### **Distribucions predictives posteriors**

Prèviament a la tria de models, es va avaluar si aquests podien reproduir les dades observades mitjançant distribucions predictives posteriors, ja que si no fos així comparar-los entre ells podria donar estimes de les probabilitats posteriors esbiaixades. Amb el conjunt d'estadístics triat, es va realitzar una anàlisi d'ABC per cadascun dels tres models demogràfics plantejats (M2, M3 i M4), utilitzant un llindar de tolerància del 0,01, és a dir del milió de simulacions total, i es van escollir les 10.000 simulacions en les que els estadístics simulats

s'assemblaven més als estadístics observats, mitjançant el mètode de correcció per regressió lineal (Beaumont et al. 2002). Amb els 10.000 valors de les distribucions posteriors dels paràmetres de cada model en estudi, es van construir les distribucions predictives posteriors. En la Figura R5.7 es mostren les distribucions dels 10.000 estadístics simulats *a posteriori* en els diferents models en estudi i en la Taula R5.5 la probabilitat dels diferents estadístics resum.



**Figura R5.7.** Distribucions predictives posteriors de cadascun dels estadístics simulats *a posteriori* escollits per la tria de models. Els models M2, M3 i M4 estan representats pels colors verd, taronja i blau respectivament. La línia vertical vermella indica el valor promig de la mitjana o la variància dels estadístics observats utilitzats en la tria de models, sent aquests estadístics:  $S$ , nombre de llocs segregants;  $D$ , estadístic  $D$  de Tajima (1989);  $H_{norm}$ , estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i  $Z_A$ , estadístic  $Z_A$  (Rozas et al. 2001). En l'eix de les  $X$  es mostren els valors obtinguts pels diferents estadístics i en l'eix de les  $Y$  la freqüència d'aquests valors.

**Taula R5.5.** Probabilitat dels diferents estadístics resum observats obtinguda a partir de les seves distribucions predictives posteriors, considerant les dues cues de la distribució i un error de tipus I del 5%.

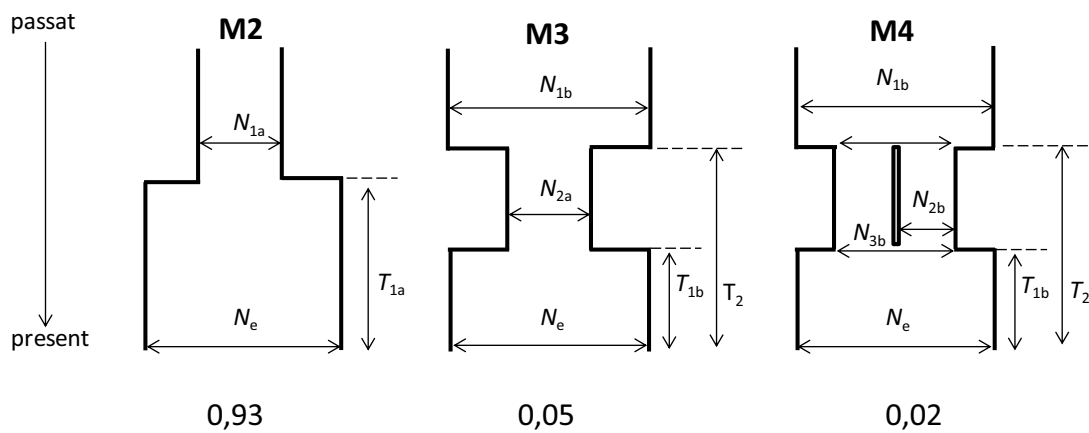
<b>Estadístics/Models</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>
<b>mitjana S</b>	0,86	0,70	0,64
<b>variància S</b>	0,27	0,75	0,95
<b>mitjana D</b>	0,24	0,22	0,28
<b>variància D</b>	0,24	0,11	0,14
<b>mitjana Hnorm</b>	0,54	0,24	0,25
<b>variància Hnorm</b>	0,54	0,90	0,90
<b>mitjana Z<sub>A</sub></b>	0,09	0,62	0,38
<b>variància Z<sub>A</sub></b>	0,08	0,01	0,03

Estadístics: S, nombre de llocs segregants; D, estadístic D de Tajima (1989); Hnorm, estadístic H normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i Z<sub>A</sub>, estadístic Z<sub>A</sub> (Rozas et al. 2001). Models: M2, model d'expansió; M3, model de coll d'ampolla; M4, model de barreja.

Les distribucions del conjunt d'estadístics simulats *a posteriori*, posen de manifest que els tres models que es volen contrastar en les anàlisis d'ABC són capaços de reproduir els estadístics observats. De fet, els valors observats es situen sempre dintre del 95% de les distribucions predictives posteriors, a excepció de la variància de l'estadístic Z<sub>A</sub> en els model M3 i M4 ( $P < 0,05$ ). En les Taules RS6.25-RS6.28 i les Figures RS6.1 i RS6.2 disponibles al final de l'apartat dels resultats, es mostren les estimes de les distribucions posteriors dels diferents paràmetres que defineixen els models M3 i M4, respectivament.

### **Tria de models**

Es van estimar les probabilitats posteriors dels tres models en estudi mitjançant el mètode de rebuig amb posterior correcció per regressió lineal (Beaumont et al. 2008), utilitzant un milió de simulacions i un llindar de tolerància del 0,01. Com s'ha comentat anteriorment, el conjunt d'estadístics escollit en la tria de models va ser el següent: nombre de llocs segregants (S), estadístic D de Tajima (1989), estadístic H normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i estadístic Z<sub>A</sub> (Rozas et al. 2003). En la Figura R5.8 es veu representat un esquema dels tres models contrastats amb la probabilitat posterior de cadascun d'ells.



**Figura R5.8.** Representació esquemàtica dels diferents models en estudi. M2, model d'expansió; M3, model de coll d'ampolla i M4 model de barreja. Les probabilitats posteriors s'indiquen sota l'esquema de cadascun d'ells.  $N_e$ , grandària efectiva actual.  $N_x$ , grandària efectiva en el període  $x$ .  $T_x$ , temps en els quals es produeixen els diferents canvis poblacionals.

La tria de models dóna un elevat suport al model M2, sent aquest el model més probable dels tres proposats, amb una probabilitat posterior del 0,93. Per tant, un escenari d'expansió poblacional és el que millor descriu la variabilitat nucleotídica observada en l'actualitat a *D. subobscura* i per tant la història demogràfica recent de l'espècie. El model de coll d'ampolla M3 presenta una probabilitat posterior de 0,05 i finalment el model de barreja poblacional M4 presenta una probabilitat posterior de 0,02. En la Taula R5.6 hi ha els quocients de les probabilitats entre dos models, és a dir els seus factors bayesians.

**Taula R5.6.** Factors bayesians obtinguts a partir de la ràtio de probabilitats en comparacions dos a dos dels tres models en estudi.

Models	M2	M3	M4
M2	1	18,85	52,74
M3	0,05	1	2,80
M4	0,02	0,36	1

Els valors ombrejats de les diagonals de la matriu, són el resultat de calcular la ràtio de probabilitats del mateix model, per tant s'obté en tots els casos un valor d'1.

El model M2 presenta respecte els models M3 i M4 uns factors bayesians de 18,9 i 52,7, respectivament. En funció de la classificació proposada a Kass i Raftery (1995), aquests factors bayesians proporcionen un elevat suport de què el model M2 és el més probable enfront dels altres dos models. El model M3, presenta respecte el model M4 un factor bayesià de 2,8. El model M4 presenta en les dues comparacions factors bayesians inferiors a 0,5 fet

que posa de manifest que és un model, respecte els altres dos, molt poc probable. Cal destacar que si s'utilitza la diversitat nucleotídica dintre del conjunt d'estadístics resum, en lloc del nombre de llocs segregants, el model més probable continua essent el model d'expansió amb una probabilitat posterior de 0,78.

### Control de qualitat en la tria de models

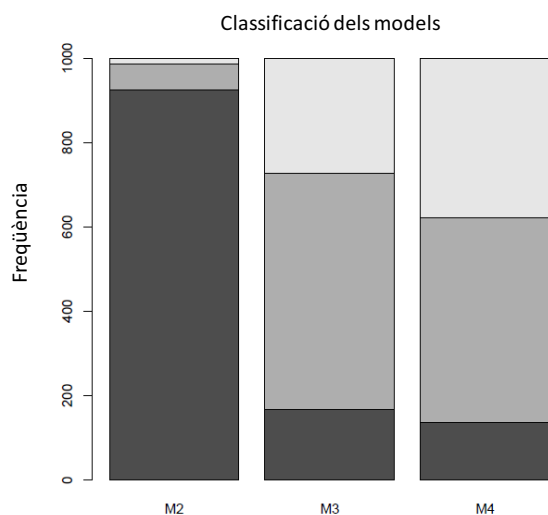
El control de qualitat en la tria de models basat en realitzar validacions creuades per avaluar si els diferents models poden distingir-se entre sí (veure Materials i Mètodes M3), s'il·lustra en la Taula R5.7 i la Figura R5.9.

**Taula R5.7.** Validacions creuades dels tres models demogràfics en estudi, resumides en una matriu de confusió.

Models	M2	M3	M4
M2	923	60	17
M3	153	565	282
M4	149	503	348

En cada fila de la matriu s'observa el nombre de simulacions assignades correctament a cada model.

S'observa com el model M2, el més probable, es pot classificar correctament en un 92% dels casos, mentre que el model M3 en un 56%. El model M4 no es pot classificar correctament ja que en un 65% dels casos s'assigna erròniament.

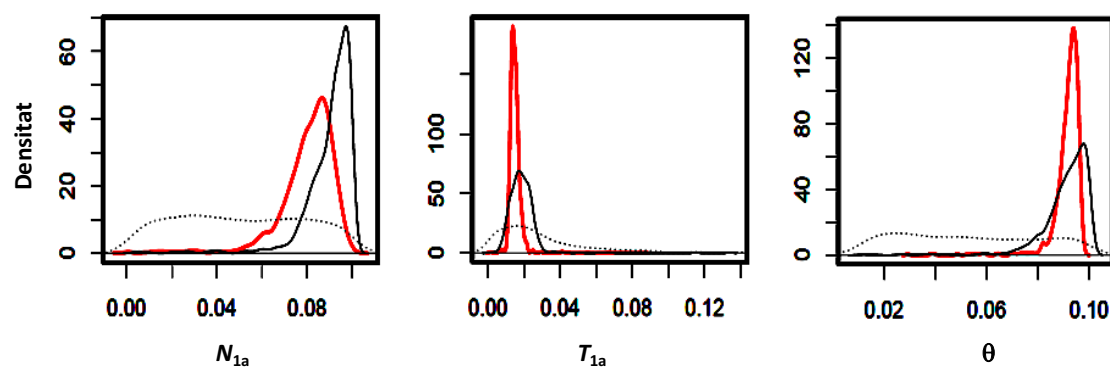


**Figura R5.9.** Diagrama de *Tukey* en el qual es visualitzen els resultats de les validacions creuades obtingudes en els controls de qualitat de la tria de models. Les diferents dades *pseudo-observades* de cada model es visualitzen amb una graduació de colors grisos diferents anant del model M2 de tonalitat més forta al model M4 amb la tonalitat més feble.



### Inferència dels paràmetres del model més probable

En aquest apartat s'estimen els paràmetres que caracteritzen el model d'expansió M2, model estimat com el més probable dels tres en estudi. El model M2 està caracteritzat per tres paràmetres:  $\theta$  (paràmetre poblacional de mutació),  $N_{1a}$  (grandària efectiva prèvia a l'expansió poblacional) i  $T_{1a}$  (temps en el qual es produeix aquesta expansió poblacional). Les estimes dels diferents paràmetres que caracteritzen el model M2, es van realitzar mitjançant una anàlisi d'ABC aplicant el mètode de rebuig amb posterior correcció per regressió lineal (Beaumont et al. 2002), utilitzant un milió de simulacions i un llindar de tolerància del 0,001. Per aquesta anàlisi es va utilitzar el mateix conjunt d'estadístics que en la tria de models a excepció del nombre de llocs segregants que va ser reemplaçat per la diversitat nucleotídica per locus (Tajima 1983), ja que la diversitat nucleotídica presenta una major correlació que el nombre de llocs segregants amb els paràmetres  $N_{1a}$  i  $T_{1a}$  que defineixen el model M2 (veure correlacions de *Pearson*, Taula R5.4). En la Figura R5.10 es mostren les distribucions posteriors dels diferents paràmetres en estudi i en la Taula R5.8 els diferents estadístics descriptius de la distribució posterior dels paràmetres d'interès del model M2. Les distribucions posteriors obtingudes pels tres paràmetres que caracteritzen el model M2, són capaces de discriminar entre el total de valors de les distribucions *a priori*.



**Figura R5.10.** Distribució posterior dels diferents paràmetres que defineixen el model M2, amb un llindar de tolerància del 0,001. Les distribucions amb línia discontinua representen el valor de cada paràmetre en les simulacions realitzades, en tant que les distribucions en negre i vermell representen el valor corresponent pel conjunt de 1.000 simulacions retinudes mitjançant el mètode de rebuig (Tavaré et al. 1997) i el mètode de correcció per regressió lineal (Beaumont et al. 2002) respectivament. Els

diferents paràmetres són:  $N_{1a}$ , quocient entre la grandària efectiva prèvia a l'expansió i la grandària efectiva actual;  $T_{1a}$ , temps d'expansió poblacional expressat en  $4N_e$  generacions;  $\theta$ , paràmetre poblacional de mutació ( $\theta = 4N_e \mu$ , sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació).

**Taula R5.8.** Estadístics descriptius de la distribució posterior dels paràmetres del model M2, sent aquests la mitjana, mediana i la moda així com també els límits inferior i superior del 95% la distribució, escollint un llinar de tolerància del 0,001 i mitjançant el mètode de correcció per regressió lineal (Beaumont et al. 2002).

<b>Estadístics descriptius/ Paràmetres</b>	<b><math>N_{1a}</math></b>	<b><math>T_{1a}</math></b>	<b><math>\theta</math></b>
<b>límit inferior del 95% de la distribució</b>	0,060	0,012	0,084
<b>Mediana</b>	0,083	0,015	0,093
<b>Mitjana</b>	0,081	0,015	0,092
<b>Moda</b>	0,087	0,014	0,094
<b>límit superior del 95% de la distribució</b>	0,094	0,019	0,096

Els diferents paràmetres són:  $N_{1a}$ , quocient entre la grandària efectiva prèvia a l'expansió i la grandària efectiva actual;  $T_{1a}$ , temps d'expansió poblacional expressat en  $4N_e$  generacions;  $\theta$ , paràmetre poblacional de mutació ( $\theta = 4N_e \mu$ , sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació).

Per cadascun dels 1000 valors retinguts de  $\theta$  ( $\theta = 4 N_e \mu$ ) i en funció d'una taxa de mutació per posició i per generació de  $2,51 \times 10^{-9}$  (Taula R5.2), mitjana estimada per les regions en estudi a partir de la divergència nucleotídica, es va calcular la grandària efectiva actual de la població ( $N_e$ ). A partir dels valors obtinguts d'aquesta grandària efectiva es van estimar els valors dels altres dos paràmetres demogràfics del model M2, sent aquests el temps en anys en el qual es produeix l'expansió poblacional ( $T_{1a}$ ) i la grandària efectiva prèvia a aquesta expansió ( $N_{1a}$ ). En el cas del paràmetre  $T_{1a}$ , que havia estat condicionat a la grandària efectiva, els diferents valors es van passar a anys prèviament a estimar la distribució posterior d'aquest paràmetre. En la Taula R5.9 es mostren diferents estadístics descriptius de les distribucions posteriors dels paràmetres del model M2.

**Taula R5.9.** Estadístics descriptius de la distribució posterior dels paràmetres del model M2 calculats en funció de la grandària efectiva actual estimada a partir dels valors obtinguts de  $\theta$ . Es presenta la mitjana, mediana i la moda així com també els límits inferior i superior del 95% la distribució, obtinguda amb un llindar de tolerància del 0,001 i mitjançant el mètode de correcció per regressió lineal.

<b>Estadístics descriptius/ Paràmetres</b>	<b><math>N_{1a}</math></b>	<b><math>T_{1a}</math></b>	<b><math>N_e</math></b>
<b>límit inferior del 95% de la distribució</b>	$5,03 \times 10^5$	77871	$8,40 \times 10^6$
<b>Mediana</b>	$7,66 \times 10^5$	111507	$9,24 \times 10^6$
<b>Mitjana</b>	$7,34 \times 10^5$	112038	$9,12 \times 10^6$
<b>Moda</b>	$8,16 \times 10^5$	110425	$9,34 \times 10^6$
<b>límit superior del 95% de la distribució</b>	$8,98 \times 10^5$	145803	$9,58 \times 10^6$

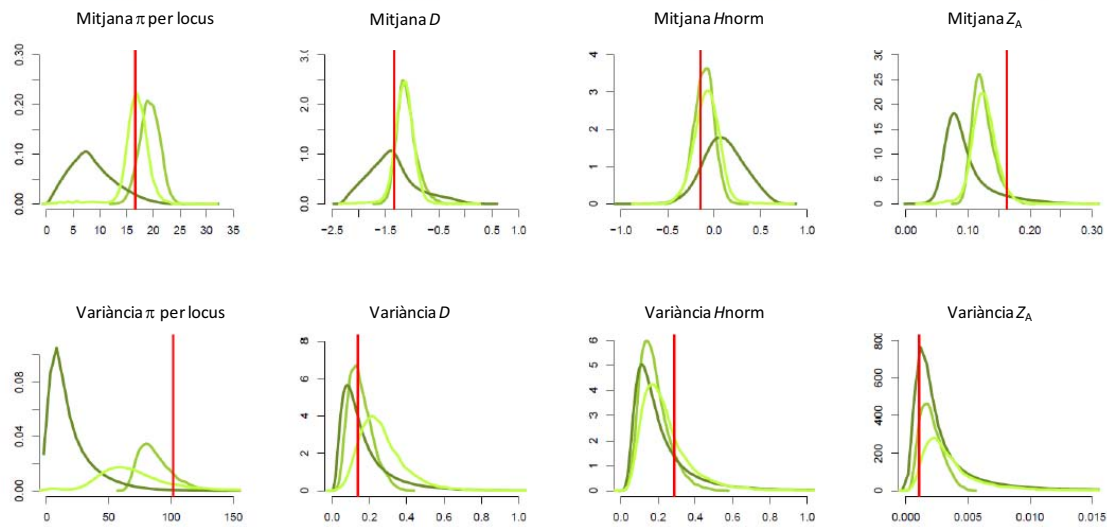
Els diferents paràmetres són:  $N_{1a}$ , grandària efectiva prèvia a l'expansió expressada en nombre d'individus;  $T_{1a}$ , temps en el que es produeix l'expansió, expressat en anys, considerant cinc generacions per any;  $N_e$ , grandària efectiva actual de l'espècie.

Les distribucions posteriors dels paràmetres que conformen el model M2 mostren com la grandària efectiva actual més probable de l'espècie és d'aproximadament 9,3 milions ( $8,4 \times 10^6$  -  $9,6 \times 10^6$  milions com a límits inferior i superior del 95% de la distribució). La grandària efectiva de la població prèvia a l'expansió poblacional, és d'aproximadament un ordre de magnitud inferior respecte a la grandària efectiva actual, sent l'estima més probable d'aquest paràmetre de  $8,1 \times 10^5$  individus ( $5,3 \times 10^5$  -  $8,9 \times 10^5$  com a límits inferior i superior del 95% de la distribució). Les estimes més probables de quan es produeix aquesta expansió és d'ara fa aproximadament 110.000 anys (78.000 - 145.000 anys com a límits inferior i superior del 95% la distribució).

### **Controls de qualitat en les inferències dels paràmetres del model més probable**

#### Distribucions predictives posteriors

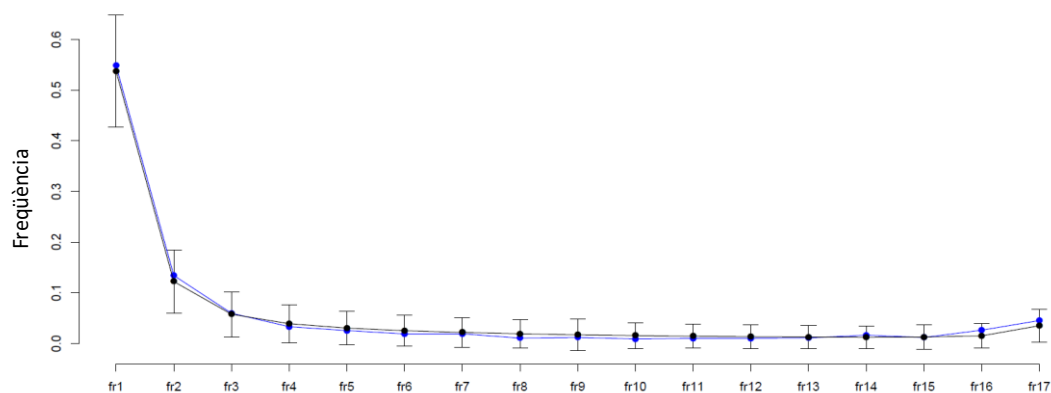
En la Figura R5.11 es mostren en relació al model M2 les distribucions del conjunt d'estadístics resum triats corresponents al milió total de simulacions, les 1000 simulacions retingudes en les anàlisis d'ABC per inferir les distribucions posteriors dels diferents paràmetres i les distribucions predictives posteriors d'aquest model. Les distribucions predictives posteriors corroboren que el model M2 és capaç d'explicar la variabilitat genètica observada en l'actualitat a *D. subobscura*.



**Figura R5.11.** Distribucions del conjunt d'estadístics resum triat corresponents al model M2. La graduació de colors verds del més fosc al més clar indica el següent: milió total de simulacions, les 1000 simulacions retingudes en les anàlisis d'ABC per inferir les distribucions posteriors dels diferents paràmetres i les distribucions predictives posteriors d'aquest. La línia vertical vermella indica el valor promig de la mitjana o la variància dels estadístics observats, sent aquests estadístics:  $\pi$ , diversitat nucleotídica per locus (Tajima 1983);  $D$ , estadístic  $D$  de Tajima (1989);  $H_{norm}$ , estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i  $Z_A$ , estadístic  $Z_A$  (Rozas et al. 2001). En l'eix de les  $X$  es mostren els valors obtinguts pels diferents estadístics i en l'eix de les  $Y$  la freqüència d'aquests estadístics.

### Espectre de freqüències desdoblament simulat *a posteriori*

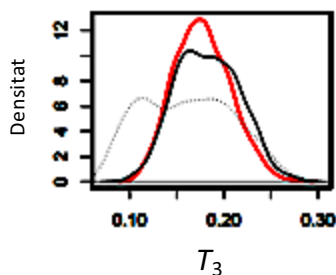
L'espectre de freqüències desdoblament simulat *a posteriori* corrobora que el model M2 és capaç d'explicar la variabilitat genètica observada en l'actualitat en l'espècie. En la Figura R5.12 es mostren els resultats obtinguts.



**Figura R5.12.** Espectre de freqüències desdoblament corresponent a les dades observades (color blau); espectre de freqüències desdoblament simulat *a posteriori* (color negre) obtingut a partir de la realització de 10.000 noves simulacions *a posteriori*, utilitzant com a distribucions *a priori* per a cada paràmetre el conjunt dels paràmetres retinguts de l'anàlisi prèvia d'ABC.

### R5.3 Estima de l'edat de la inversió $J_1$ mitjançant una anàlisi d'ABC

Es va inferir l'edat de la inversió  $J_1$  a través d'una anàlisi d'ABC, mitjançant el concatenat de les dues regions seqüenciades més properes als punts de trencament i localitzades dintre de la inversió  $J_1$ , sent aquestes les regions p127 i p222. La grandària de la mostra va ser d'un total de 26 línies homocariotípiques. Els estadístics que es van utilitzar per aquesta anàlisi són: la diversitat nucleotídica per locus (Tajima 1983) que aporta informació sobre els nivells de variació, l'estadístic  $D$  de Tajima (1989) que aporta informació sobre els patrons de variació, l'estadístic  $Z_{ns}$  (Kelly 1997) que aporta informació sobre el desequilibri de lligament i l'estadístic  $F_{ST}$  (Hudson et al. 1992) que aporta informació sobre la diferenciació genètica entre les ordenacions  $J_1$  i  $J_{st}$ . En relació als estadístics  $\pi$ ,  $D$  i  $Z_{ns}$  es van utilitzar com a estadístics observats els valors obtinguts de concatenar les dues regions en estudi en relació a les línies que presenten l'ordenació  $J_1$ . En la Figura R5.13 es mostra la distribució posterior del paràmetre d'interès  $T_3$  (edat de la inversió  $J_1$ ).



**Figura R5.13.** Distribució posterior del paràmetre  $T_3$  (edat de la inversió  $J_1$ ), amb un llindar de tolerància del 0,001. Les distribucions amb línia discontinua representen el valor de cada paràmetre en les simulacions realitzades, en tant que les distribucions en negre i vermell representen el valor corresponent pel conjunt de 1000 simulacions retingudes mitjançant el mètode de rebuig (Tavaré et al. 1997) i el mètode de correcció per regressió lineal (Beaumont et al. 2002), respectivament.

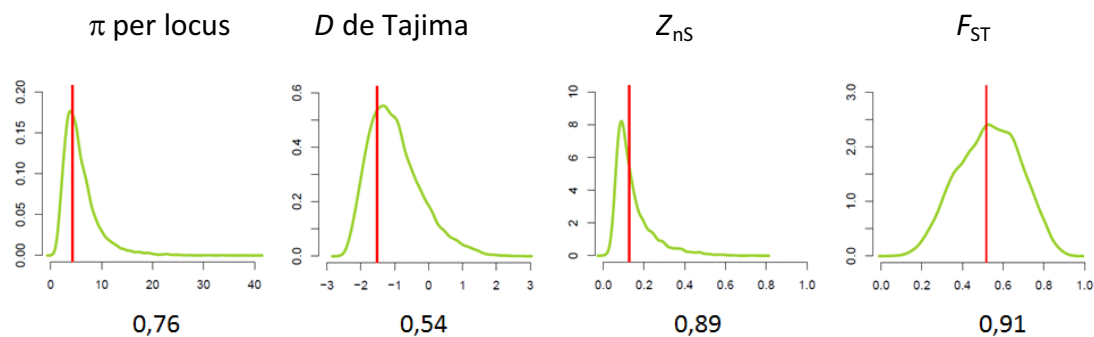
Les anàlisis d'ABC efectuades indiquen que l'edat més probable de la inversió  $J_1$  és d'ara fa aproximadament 466.000 anys (335.334 – 574.555 anys com a límits inferior i superior del 95% la distribució). En la Taula R5.10 es mostren diferents estadístics descriptius de la distribució posterior d'aquest paràmetre.

**Taula R5.10.** Estadístics descriptius de la distribució posterior del paràmetre  $T_3$  (edat de la inversió  $J_1$ ) sent aquests la mitjana, mediana i la moda així com també els límits inferior i superior del 95% la distribució, escollint un llindar de tolerància del 0,001 i mitjançant el mètode de correcció per regressió lineal (Beaumont et al. 2002).

Estadístics descriptius/Paràmetres	$T_3$	Edat de la inversió $J_1$
<b>Límit inferior del 95% de la distribució</b>	0,133	335334
<b>Mediana</b>	0,181	459959
<b>Mitjana</b>	0,18	456917
<b>Moda</b>	0,189	466402
<b>Límit superior del 95% de la distribució</b>	0,235	574555

### Control de qualitat en la inferència de l'edat de la inversió $J_1$ a través de validacions predictives posteriors

En la Figura R5.14 es mostren els resultats dels estadístics simulats *a posteriori* mitjançant 10.000 simulacions predictives posteriors les quals s'efectuen utilitzant com a distribucions *a priori* el conjunt de paràmetres associats a les simulacions retingudes en l'anàlisi anterior.



**Figura R5.14.** Distribucions predictives posteriors de cadascun dels estadístics *simulats a posteriori*. La línia vertical és el valor dels estadístics observats utilitzats en l'anàlisi, sent aquests estadístics: la diversitat nucleotídica per locus (Tajima 1983), l'estadístic  $D$  de Tajima (1989), l'estadístic  $Z_{nS}$  (Kelly 1997) i l'estadístic  $F_{ST}$  (Hudson et al. 1992b). En l'eix de les  $X$  es mostren els valors obtinguts pels diferents estadístics i en l'eix de les  $Y$  la freqüència d'aquests valors. Sota les distribucions predictives posteriors es mostra la probabilitat dels diferents estadístics resum observats obtinguda a partir de les seves distribucions predictives posteriors, considerant les dues cues de la distribució i un error de tipus I del 5%.

Les validacions predictives posteriors, corroboren que el model proposat per estimar l'edat de la inversió  $J_1$ , pot reproduir la variabilitat observada en les regions properes al punt de trencament.

#### R5.4 Detecció d'una regió candidata de ser diana de la selecció

La regió p169 no va ser considerada en l'apartat de les inferències demogràfiques ja que les anàlisis dels nivells i patrons de polimorfisme semblaven indicar que la variació en aquesta regió podia reflectir un canvi selectiu recent. Aquesta regió és homòloga a un intró del gen anomenat GA12801 a *D. pseudoobscura* i *Zasp67* (CG14168) a *D. melanogaster*. En la Figura R5.15 es mostra aquest gen a les dues espècies, d'acord a la informació disponible a [www.flybase.com](http://www.flybase.com).

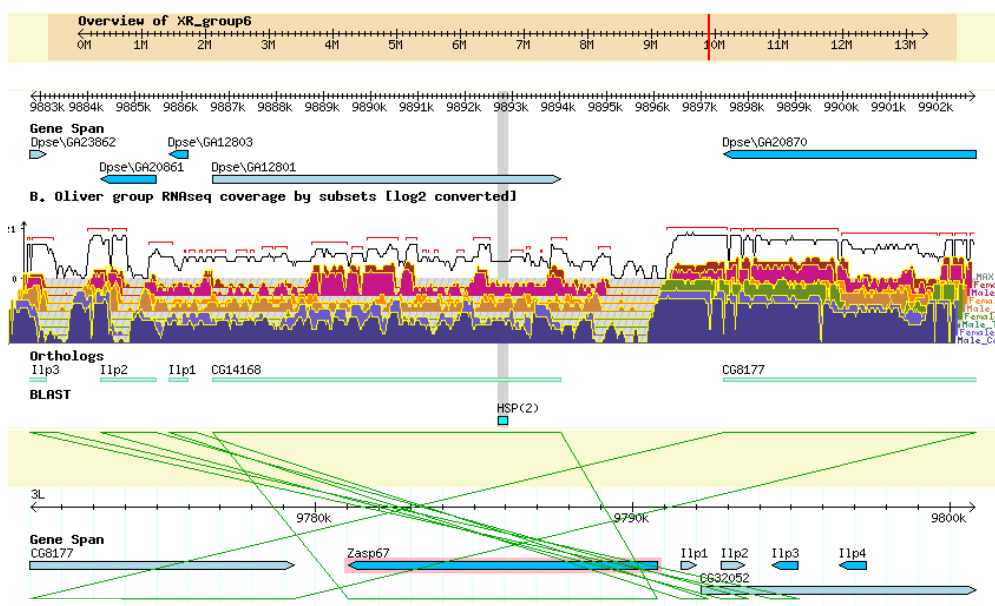
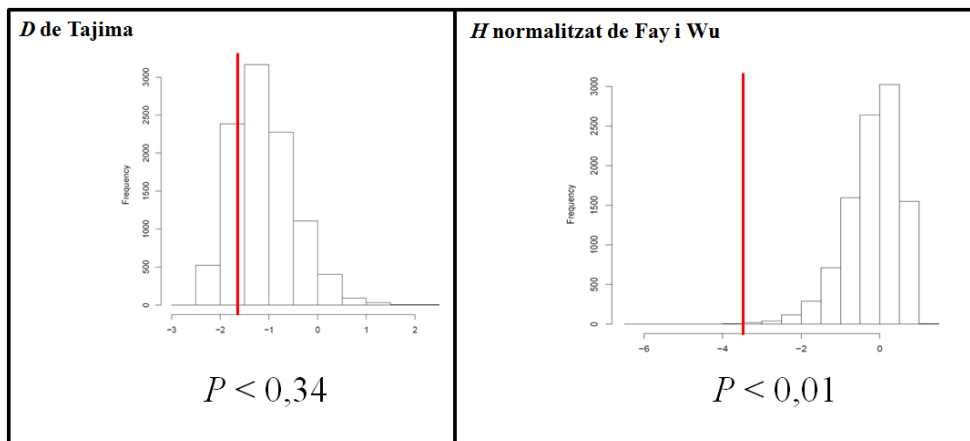


Figura R5.15. Homologia de la regió p169 en les espècies *D. pseudoobscura* i *D. melanogaster*.

La disponibilitat del model demogràfic d'expansió poblacional M2, va permetre obtenir prediccions sobre els patrons de variabilitat que s'esperarien en aquesta regió sota aquest model. En la Figura R5.16, es mostra la distribució empírica dels estadístics  $D$  de Tajima i  $H$  normalitzat de Fay i Wu construïda a partir de les 10.000 simulacions de coalescència, realitzades utilitzant com a distribucions *a priori* els valors corresponents de les simulacions retingudes en l'anàlisi d'ABC efectuada per obtenir la distribució posterior dels diferents paràmetres que conformen el model M2.



**Figura R5.16.** Distribució esperada en la regió p169 de l'estadístic  $D$  de Tajima (1989) i de l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) obtinguda a partir de simulacions de coalescència (10.000 rèpliques) utilitzant les distribucions posteriors dels paràmetres que defineixen el model d'expansió M2 com distribucions *a priori* i tenint en compte les possibles mutacions recurrents. La línia vertical vermella és el valor observat d'aquests dos estadístics en la regió p169. Sota les figures es mostra la significació estadística, considerant les dues cues de la distribució i un error de tipus I del 5%.

S'observa una desviació significativa ( $P < 0,01$ ) del que s'esperaria sota un model d'expansió poblacional en l'estadístic  $H$  normalitzat de Fay i Wu (Fay y Wu 2000; Zeng et al. 2006), indicant un excés significatiu de variants derivades a elevada freqüència no explicables per aquest model demogràfic.



**RS6. MATERIAL SUPLEMENTARI****RS6.1 Taules de polimorfisme de les regions localitzades al llarg dels cromosoma J i no afectades per la inversió J<sub>1</sub>**

Polimorfisme nucleotídic en cadascuna de les 18 regions no codificadores en estudi localitzades al llarg del cromosoma J de *D. subobscura* i no afectades per la inversió J<sub>1</sub>. En general, aquestes regions es van seqüenciar en 21 línies homocariotípiques per al cromosoma J obtingudes a partir d'una població natural de *D. subobscura* de Barcelona i una línia de *D. guanche* utilitzada com a *outgroup*. En la part superior de les taules, s'indica la numeració de les posicions polimòrfiques en funció de l'alineament múltiple. Els punts (·) indiquen el mateix nucleòtid que en la primera seqüència que s'utilitza de referència. En els *indels* (D<sub>x</sub>, deleció; I<sub>x</sub>, Inserció ) i els Microsatèl·lits (M<sub>x</sub>) s'indiquen el número de la primera i última posició afectada, així com la seva longitud en el subíndex x. En la última fila es mostra la informació que presenta la seqüència de l'espècie *D. guanche* utilitzada com a *outgroup* en aquelles posicions que són polimòrfiques a *D. subobscura*.



Taula RS6.2. Polimorfismes nucleotídics present en la regió p32096 en 13 línies de *D. subobscura* i una línia de *D. guanche*.

ID	Sequència	Mutacions	Posicions																		
			58	69	121	124	129	131	132	153	161	164	189	196	267	281	292	305	316	389	415
J10F07	GGATAC--GTACGCTAGGG	-	G	G	A	T	A	C	-	-	G	T	A	C	G	C	T	A	G	G	G
J10F15	..C..C..T..T..C..A	AAGGGGAC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F19	..C..C..T..T..C..A	AAGGGGAC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F26	..A..C..-..T..A..A..C..A	AACTGCGAC(AACGGCGAC) <sub>3</sub>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F28	..C..C..-..C..T..A..C..A	AACTGCGAC(AACGGCGAC) <sub>3</sub>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F27	..C..C..C..-..T..T..C..C..	(AACGGCGAC) <sub>2</sub>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F31	..G..C..-..T..T..A..C..C..	(AACGGCGAC) <sub>3</sub>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F74	..C..C..-..T..T..A..C..C..	AACGGCGAC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F78	..C..C..-..T..T..A..C..C..	(GAC) <sub>2</sub>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F79	..C..C..-..T..T..A..C..C..	AACGGCGAC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F16	..C..C..-..T..T..A..C..C..	AACGGCGAC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F18	..C..C..-..T..T..A..C..C..	AACGGCGAC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J10F51	..C..C..-..T..T..A..C..C..	AACGGC(GAC) <sub>5</sub>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>D. guanche</i>	A..C..-..T..T..A..C..C..	-	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.





Taula RS6.4. Polimorfismes nucleotídic present en la regió p237 en 20 línies de *D. subobscura* i una línia de *D. guanche*.

	2077	2078	2089	2120	2237	2242	2247	2251	2272	2276	2284	2293	2302	2308	2310	2327	2340	2343
P237	2077	2078	2089	2120	2237	2242	2247	2251	2272	2276	2284	2293	2302	2308	2310	2327	2340	2343
J10F07	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F14	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F15	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F19	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F26	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F27	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F28	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F31	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F32	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F44	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F58	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F74	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F77	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F78	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
J10F79	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
JstOF16	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
JstOF18	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
JstOF51	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
JstOF54	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
JstOF49	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>
<i>D. Guanche</i>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>	(AGC) <sub>3</sub>







Taula RS6.7. Polimorfismes nucleotídic present en la regió p227en 21 línies de *D. subobscura* i una línia de *D. guanche*.

	D <sub>52</sub>												D <sub>6</sub>											
	145	150	180	335	366	367	394	404	413	418	424	499	505	509	557	574	582	628	642	674	689	759		
p227	A	A	A	G	A	G	T	C	G	G	T	C	T	T	A	G	C	T	T	A	A	T		
J10F07	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	G	.	.	.	C		
J10F14	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.		
J10F15	.	.	.	C	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.		
J10F19	G	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.		
J10F26	G	.	.	.	.	.	C	.	C	.	.	.	.	.	.	.	A	.	.	.	.	C		
J10F27	G	.	.	.	.	.	G	C	-	A	.	G	C	C	.	.	.	G	A	.	.	.		
J10F28	G	G	.	.	.	.	G	C	-	A	.	G	C	C	.	.	.	G	.	.	.	A		
J10F31	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	A	G	.	.	.	.		
J10F32	G	.	.	.	.	.	C	G	.	C	.	.	.	.	.	.	A	.	.	.	.	.		
J10F44	G	.	.	.	.	.	C	G	.	C	.	.	.	.	.	.	A	.	.	.	.	.		
J10F58	.	.	.	.	.	.	G	.	.	.	.	.	G	C	.	.	.	.	.	.	.	.		
J10F74	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.		
J10F77	G	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	A	G	.	.	.	.		
J10F78	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	A	G	.	.	.	.		
J10F79	G	.	.	.	.	.	C	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.		
JstOF16	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	G	.	.	.		
JstOF18	G	.	.	.	.	.	-	.	.	.	.	.	G	C	.	.	.	G	.	.	.	.		
JstOF21	.	.	.	.	.	.	.	.	.	.	.	.	G	C	.	A	.	.	.	.	.	.		
JstOF51	G	.	.	.	.	.	G	C	-	A	.	G	C	C	.	.	.	G	.	.	.	.		
JstOF54	.	.	.	.	.	.	G	G	.	.	.	.	G	C	.	.	A	G	.	.	.	.		
JstOF49	G	.	T	.	C	.	G	.	.	.	.	G	C	C	.	.	A	.	.	T	.	.		
<i>D.guanche</i>	.	.	.	.	.	.	A	G	-	-	-	.	.	.	.	.	A	G	.	.	.	C		



Taula RS6.9. Polimorfismes nucleotídics present en la regió p220 en 20 línies de *D. subobscura* i una línia de *D. guanche*.

	860	878	879	882	903-904	919	924	927	931	932	945	948	966	973	974	975	978	980	987	1051	1125	1127	1158	1184	1187	1214	1227
	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	D <sub>6</sub>	
	I <sub>1</sub>	I <sub>2</sub>	I <sub>3</sub>	I <sub>4</sub>	I <sub>5</sub>	I <sub>6</sub>	I <sub>7</sub>	I <sub>8</sub>	I <sub>9</sub>	I <sub>10</sub>	I <sub>11</sub>	I <sub>12</sub>	I <sub>13</sub>	I <sub>14</sub>	I <sub>15</sub>	I <sub>16</sub>	I <sub>17</sub>	I <sub>18</sub>	I <sub>19</sub>	I <sub>20</sub>	I <sub>21</sub>	I <sub>22</sub>	I <sub>23</sub>	I <sub>24</sub>	I <sub>25</sub>	I <sub>26</sub>	
P220	T	T	G	-	A	C	T	G	G	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
J10F07	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F14	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F15	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F16	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F26	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F27	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F28	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F31	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F32	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F44	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F58	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F74	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F77	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F78	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F79	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F16	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F18	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F21	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F51	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F54	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
J10F49	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T
<i>D. Guanche</i>	A	T	G	A	A	G	G	T	T	G	G	A	A	G	G	T	T	G	G	A	A	G	A	A	G	G	T







Taula RS6.13. Polimorfismes nucleotídic present en la regió p169 en 20 línies de *D. subobscura* i una línia de *D. guanche*.

p169	D <sub>3</sub>	D <sub>3</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>7</sub>	D <sub>2</sub>	D <sub>15</sub>	D <sub>6</sub>																														
	257	261	274	276	279	287	396	397	409	434	435	443	451	460	461	468	473	481	482	483-484	493	496	506	600	601	615	616	634	663	665								
J10F07	T	T	C	-	A	C	T	-	C	C	T	C	T	C	T	A	T	G	C	-	C	T	T	A	A	C	T	G	T	T	A							
J10F14	.	.	.	.	.	.	C	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.						
J10F15	.	.	.	.	.	.	C	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.					
J10F26	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.					
J10F27	.	.	.	.	T	.	.	T	G	.	.	.	.	.	.	.	.	.	T	-	.	.	.	.	.	.	.	.	.	.	.	.	.					
J10F28	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.				
J10F31	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.			
J10F32	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.			
J10F44	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
J10F58	.	.	T	-	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
J10F74	.	.	.	.	.	.	.	T	T	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
J10F77	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
J10F78	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
J10F79	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
JstOF16	.	.	.	.	.	.	.	T	.	.	C	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
JstOF18	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
JstOF21	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
JstOF51	C	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
JstOF54	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
JstOF49	C	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
<i>D. Guanche</i>	C	-	T	A	T	.	.	T	T	G	C	-	-	G	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T





Taula RS6.15. Polimorfismes nucleotídics present en la regió p248 en 18 línies de *D. subobscura* i una línia de *D. guanche*.

Gen	5'	3'	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>	D <sub>6</sub>	D <sub>7</sub>	D <sub>8</sub>	D <sub>9</sub>	D <sub>10</sub>	D <sub>11</sub>	D <sub>12</sub>	D <sub>13</sub>	D <sub>14</sub>	D <sub>15</sub>	D <sub>16</sub>	D <sub>17</sub>	D <sub>18</sub>	D <sub>19</sub>	D <sub>20</sub>	D <sub>21</sub>	D <sub>22</sub>	D <sub>23</sub>	D <sub>24</sub>	D <sub>25</sub>	D <sub>26</sub>	D <sub>27</sub>	D <sub>28</sub>	D <sub>29</sub>	D <sub>30</sub>	D <sub>31</sub>	D <sub>32</sub>	D <sub>33</sub>	D <sub>34</sub>	D <sub>35</sub>	D <sub>36</sub>	D <sub>37</sub>	D <sub>38</sub>	D <sub>39</sub>	D <sub>40</sub>	D <sub>41</sub>	D <sub>42</sub>	D <sub>43</sub>	D <sub>44</sub>	D <sub>45</sub>	D <sub>46</sub>	D <sub>47</sub>	D <sub>48</sub>	D <sub>49</sub>	D <sub>50</sub>	D <sub>51</sub>	D <sub>52</sub>	D <sub>53</sub>	D <sub>54</sub>	D <sub>55</sub>	D <sub>56</sub>	D <sub>57</sub>	D <sub>58</sub>	D <sub>59</sub>	D <sub>60</sub>	D <sub>61</sub>	D <sub>62</sub>	D <sub>63</sub>	D <sub>64</sub>	D <sub>65</sub>	D <sub>66</sub>	D <sub>67</sub>	D <sub>68</sub>	D <sub>69</sub>	D <sub>70</sub>	D <sub>71</sub>	D <sub>72</sub>	D <sub>73</sub>	D <sub>74</sub>	D <sub>75</sub>	D <sub>76</sub>	D <sub>77</sub>	D <sub>78</sub>	D <sub>79</sub>	D <sub>80</sub>	D <sub>81</sub>	D <sub>82</sub>	D <sub>83</sub>	D <sub>84</sub>	D <sub>85</sub>	D <sub>86</sub>	D <sub>87</sub>	D <sub>88</sub>	D <sub>89</sub>	D <sub>90</sub>	D <sub>91</sub>	D <sub>92</sub>	D <sub>93</sub>	D <sub>94</sub>	D <sub>95</sub>	D <sub>96</sub>	D <sub>97</sub>	D <sub>98</sub>	D <sub>99</sub>	D <sub>100</sub>	D <sub>101</sub>	D <sub>102</sub>	D <sub>103</sub>	D <sub>104</sub>	D <sub>105</sub>	D <sub>106</sub>	D <sub>107</sub>	D <sub>108</sub>	D <sub>109</sub>	D <sub>110</sub>	D <sub>111</sub>	D <sub>112</sub>	D <sub>113</sub>	D <sub>114</sub>	D <sub>115</sub>	D <sub>116</sub>	D <sub>117</sub>	D <sub>118</sub>	D <sub>119</sub>	D <sub>120</sub>	D <sub>121</sub>	D <sub>122</sub>	D <sub>123</sub>	D <sub>124</sub>	D <sub>125</sub>	D <sub>126</sub>	D <sub>127</sub>	D <sub>128</sub>	D <sub>129</sub>	D <sub>130</sub>	D <sub>131</sub>	D <sub>132</sub>	D <sub>133</sub>	D <sub>134</sub>	D <sub>135</sub>	D <sub>136</sub>	D <sub>137</sub>	D <sub>138</sub>	D <sub>139</sub>	D <sub>140</sub>	D <sub>141</sub>	D <sub>142</sub>	D <sub>143</sub>	D <sub>144</sub>	D <sub>145</sub>	D <sub>146</sub>	D <sub>147</sub>	D <sub>148</sub>	D <sub>149</sub>	D <sub>150</sub>	D <sub>151</sub>	D <sub>152</sub>	D <sub>153</sub>	D <sub>154</sub>	D <sub>155</sub>	D <sub>156</sub>	D <sub>157</sub>	D <sub>158</sub>	D <sub>159</sub>	D <sub>160</sub>	D <sub>161</sub>	D <sub>162</sub>	D <sub>163</sub>	D <sub>164</sub>	D <sub>165</sub>	D <sub>166</sub>	D <sub>167</sub>	D <sub>168</sub>	D <sub>169</sub>	D <sub>170</sub>	D <sub>171</sub>	D <sub>172</sub>	D <sub>173</sub>	D <sub>174</sub>	D <sub>175</sub>	D <sub>176</sub>	D <sub>177</sub>	D <sub>178</sub>	D <sub>179</sub>	D <sub>180</sub>	D <sub>181</sub>	D <sub>182</sub>	D <sub>183</sub>	D <sub>184</sub>	D <sub>185</sub>	D <sub>186</sub>	D <sub>187</sub>	D <sub>188</sub>	D <sub>189</sub>	D <sub>190</sub>	D <sub>191</sub>	D <sub>192</sub>	D <sub>193</sub>	D <sub>194</sub>	D <sub>195</sub>	D <sub>196</sub>	D <sub>197</sub>	D <sub>198</sub>	D <sub>199</sub>	D <sub>200</sub>	D <sub>201</sub>	D <sub>202</sub>	D <sub>203</sub>	D <sub>204</sub>	D <sub>205</sub>	D <sub>206</sub>	D <sub>207</sub>	D <sub>208</sub>	D <sub>209</sub>	D <sub>210</sub>	D <sub>211</sub>	D <sub>212</sub>	D <sub>213</sub>	D <sub>214</sub>	D <sub>215</sub>	D <sub>216</sub>	D <sub>217</sub>	D <sub>218</sub>	D <sub>219</sub>	D <sub>220</sub>	D <sub>221</sub>	D <sub>222</sub>	D <sub>223</sub>	D <sub>224</sub>	D <sub>225</sub>	D <sub>226</sub>	D <sub>227</sub>	D <sub>228</sub>	D <sub>229</sub>	D <sub>230</sub>	D <sub>231</sub>	D <sub>232</sub>	D <sub>233</sub>	D <sub>234</sub>	D <sub>235</sub>	D <sub>236</sub>	D <sub>237</sub>	D <sub>238</sub>	D <sub>239</sub>	D <sub>240</sub>	D <sub>241</sub>	D <sub>242</sub>	D <sub>243</sub>	D <sub>244</sub>	D <sub>245</sub>	D <sub>246</sub>	D <sub>247</sub>	D <sub>248</sub>	D <sub>249</sub>	D <sub>250</sub>	D <sub>251</sub>	D <sub>252</sub>	D <sub>253</sub>	D <sub>254</sub>	D <sub>255</sub>	D <sub>256</sub>	D <sub>257</sub>	D <sub>258</sub>	D <sub>259</sub>	D <sub>260</sub>	D <sub>261</sub>	D <sub>262</sub>	D <sub>263</sub>	D <sub>264</sub>	D <sub>265</sub>	D <sub>266</sub>	D <sub>267</sub>	D <sub>268</sub>	D <sub>269</sub>	D <sub>270</sub>	D <sub>271</sub>	D <sub>272</sub>	D <sub>273</sub>	D <sub>274</sub>	D <sub>275</sub>	D <sub>276</sub>	D <sub>277</sub>	D <sub>278</sub>	D <sub>279</sub>	D <sub>280</sub>	D <sub>281</sub>	D <sub>282</sub>	D <sub>283</sub>	D <sub>284</sub>	D <sub>285</sub>	D <sub>286</sub>	D <sub>287</sub>	D <sub>288</sub>	D <sub>289</sub>	D <sub>290</sub>	D <sub>291</sub>	D <sub>292</sub>	D <sub>293</sub>	D <sub>294</sub>	D <sub>295</sub>	D <sub>296</sub>	D <sub>297</sub>	D <sub>298</sub>	D <sub>299</sub>	D <sub>300</sub>	D <sub>301</sub>	D <sub>302</sub>	D <sub>303</sub>	D <sub>304</sub>	D <sub>305</sub>	D <sub>306</sub>	D <sub>307</sub>	D <sub>308</sub>	D <sub>309</sub>	D <sub>310</sub>	D <sub>311</sub>	D <sub>312</sub>	D <sub>313</sub>	D <sub>314</sub>	D <sub>315</sub>	D <sub>316</sub>	D <sub>317</sub>	D <sub>318</sub>	D <sub>319</sub>	D <sub>320</sub>	D <sub>321</sub>	D <sub>322</sub>	D <sub>323</sub>	D <sub>324</sub>	D <sub>325</sub>	D <sub>326</sub>	D <sub>327</sub>	D <sub>328</sub>	D <sub>329</sub>	D <sub>330</sub>	D <sub>331</sub>	D <sub>332</sub>	D <sub>333</sub>	D <sub>334</sub>	D <sub>335</sub>	D <sub>336</sub>	D <sub>337</sub>	D <sub>338</sub>	D <sub>339</sub>	D <sub>340</sub>	D <sub>341</sub>	D <sub>342</sub>	D <sub>343</sub>	D <sub>344</sub>	D <sub>345</sub>	D <sub>346</sub>	D <sub>347</sub>	D <sub>348</sub>	D <sub>349</sub>	D <sub>350</sub>	D <sub>351</sub>	D <sub>352</sub>	D <sub>353</sub>	D <sub>354</sub>	D <sub>355</sub>	D <sub>356</sub>	D <sub>357</sub>	D <sub>358</sub>	D <sub>359</sub>	D <sub>360</sub>	D <sub>361</sub>	D <sub>362</sub>	D <sub>363</sub>	D <sub>364</sub>	D <sub>365</sub>	D <sub>366</sub>	D <sub>367</sub>	D <sub>368</sub>	D <sub>369</sub>	D <sub>370</sub>	D <sub>371</sub>	D <sub>372</sub>	D <sub>373</sub>	D <sub>374</sub>	D <sub>375</sub>	D <sub>376</sub>	D <sub>377</sub>	D <sub>378</sub>	D <sub>379</sub>	D <sub>380</sub>	D <sub>381</sub>	D <sub>382</sub>	D <sub>383</sub>	D <sub>384</sub>	D <sub>385</sub>	D <sub>386</sub>	D <sub>387</sub>	D <sub>388</sub>	D <sub>389</sub>	D <sub>390</sub>	D <sub>391</sub>	D <sub>392</sub>	D <sub>393</sub>	D <sub>394</sub>	D <sub>395</sub>	D <sub>396</sub>	D <sub>397</sub>	D <sub>398</sub>	D <sub>399</sub>	D <sub>400</sub>	D <sub>401</sub>	D <sub>402</sub>	D <sub>403</sub>	D <sub>404</sub>	D <sub>405</sub>	D <sub>406</sub>	D <sub>407</sub>	D <sub>408</sub>	D <sub>409</sub>	D <sub>410</sub>	D <sub>411</sub>	D <sub>412</sub>	D <sub>413</sub>	D <sub>414</sub>	D <sub>415</sub>	D <sub>416</sub>	D <sub>417</sub>	D <sub>418</sub>	D <sub>419</sub>	D <sub>420</sub>	D <sub>421</sub>	D <sub>422</sub>	D <sub>423</sub>	D <sub>424</sub>	D <sub>425</sub>	D <sub>426</sub>	D <sub>427</sub>	D <sub>428</sub>	D <sub>429</sub>	D <sub>430</sub>	D <sub>431</sub>	D <sub>432</sub>	D <sub>433</sub>	D <sub>434</sub>	D <sub>435</sub>	D <sub>436</sub>	D <sub>437</sub>	D <sub>438</sub>	D <sub>439</sub>	D <sub>440</sub>	D <sub>441</sub>	D <sub>442</sub>	D <sub>443</sub>	D <sub>444</sub>	D <sub>445</sub>	D <sub>446</sub>	D <sub>447</sub>	D <sub>448</sub>	D <sub>449</sub>	D <sub>450</sub>	D <sub>451</sub>	D <sub>452</sub>	D <sub>453</sub>	D <sub>454</sub>	D <sub>455</sub>	D <sub>456</sub>	D <sub>457</sub>	D <sub>458</sub>	D <sub>459</sub>	D <sub>460</sub>	D <sub>461</sub>	D <sub>462</sub>	D <sub>463</sub>	D <sub>464</sub>	D <sub>465</sub>	D <sub>466</sub>	D <sub>467</sub>	D <sub>468</sub>	D <sub>469</sub>	D <sub>470</sub>	D <sub>471</sub>	D <sub>472</sub>	D <sub>473</sub>	D <sub>474</sub>	D <sub>475</sub>	D <sub>476</sub>	D <sub>477</sub>	D <sub>478</sub>	D <sub>479</sub>	D <sub>480</sub>	D <sub>481</sub>	D <sub>482</sub>	D <sub>483</sub>	D <sub>484</sub>	D <sub>485</sub>	D <sub>486</sub>	D <sub>487</sub>	D <sub>488</sub>	D <sub>489</sub>	D <sub>490</sub>	D <sub>491</sub>	D <sub>492</sub>	D <sub>493</sub>	D <sub>494</sub>	D <sub>495</sub>	D <sub>496</sub>	D <sub>497</sub>	D <sub>498</sub>	D <sub>499</sub>	D <sub>500</sub>	D <sub>501</sub>	D <sub>502</sub>	D <sub>503</sub>	D <sub>504</sub>	D <sub>505</sub>	D <sub>506</sub>	D <sub>507</sub>	D <sub>508</sub>	D <sub>509</sub>	D <sub>510</sub>	D <sub>511</sub>	D <sub>512</sub>	D <sub>513</sub>	D <sub>514</sub>	D <sub>515</sub>	D <sub>516</sub>	D <sub>517</sub>	D <sub>518</sub>	D <sub>519</sub>	D <sub>520</sub>	D <sub>521</sub>	D <sub>522</sub>	D <sub>523</sub>	D <sub>524</sub>	D <sub>525</sub>	D <sub>526</sub>	D <sub>527</sub>	D <sub>528</sub>	D <sub>529</sub>	D <sub>530</sub>	D <sub>531</sub>	D <sub>532</sub>	D <sub>533</sub>	D <sub>534</sub>	D <sub>535</sub>	D <sub>536</sub>	D <sub>537</sub>	D <sub>538</sub>	D <sub>539</sub>	D <sub>540</sub>	D <sub>541</sub>	D <sub>542</sub>	D <sub>543</sub>	D <sub>544</sub>	D <sub>545</sub>	D <sub>546</sub>	D <sub>547</sub>	D <sub>548</sub>	D <sub>549</sub>	D <sub>550</sub>	D <sub>551</sub>	D <sub>552</sub>	D <sub>553</sub>	D <sub>554</sub>	D <sub>555</sub>	D <sub>556</sub>	D <sub>557</sub>	D <sub>558</sub>	D <sub>559</sub>	D <sub>560</sub>	D <sub>561</sub>	D <sub>562</sub>	D <sub>563</sub>	D <sub>564</sub>	D <sub>565</sub>	D <sub>566</sub>	D <sub>567</sub>	D <sub>568</sub>	D <sub>569</sub>	D <sub>570</sub>	D <sub>571</sub>	D <sub>572</sub>	D <sub>573</sub>	D <sub>574</sub>	D <sub>575</sub>	D <sub>576</sub>	D <sub>577</sub>	D <sub>578</sub>	D <sub>579</sub>	D <sub>580</sub>	D <sub>581</sub>	D <sub>582</sub>	D <sub>583</sub>	D <sub>584</sub>	D <sub>585</sub>	D <sub>586</sub>	D <sub>587</sub>	D <sub>588</sub>	D <sub>589</sub>	D <sub>590</sub>	D <sub>591</sub>	D <sub>592</sub>	D <sub>593</sub>	D <sub>594</sub>	D <sub>595</sub>	D <sub>596</sub>	D <sub>597</sub>	D <sub>598</sub>	D <sub>599</sub>	D <sub>600</sub>	D <sub>601</sub>	D <sub>602</sub>	D <sub>603</sub>	D <sub>604</sub>	D <sub>605</sub>	D <sub>606</sub>	D <sub>607</sub>	D <sub>608</sub>	D <sub>609</sub>	D <sub>610</sub>	D <sub>611</sub>	D <sub>612</sub>	D <sub>613</sub>	D <sub>614</sub>	D <sub>615</sub>	D <sub>616</sub>	D <sub>617</sub>	D <sub>618</sub>	D <sub>619</sub>	D <sub>620</sub>	D <sub>621</sub>	D <sub>622</sub>	D <sub>623</sub>	D <sub>624</sub>	D <sub>625</sub>	D <sub>626</sub>	D <sub>627</sub>	D <sub>628</sub>	D <sub>629</sub>	D <sub>630</sub>	D <sub>631</sub>	D <sub>632</sub>	D <sub>633</sub>	D <sub>634</sub>	D <sub>635</sub>	D <sub>636</sub>	D <sub>637</sub>	D <sub>638</sub>	D <sub>639</sub>	D <sub>640</sub>	D <sub>641</sub>	D <sub>642</sub>	D <sub>643</sub>	D <sub>644</sub>	D <sub>645</sub>	D <sub>646</sub>	D <sub>647</sub>	D <sub>648</sub>	D <sub>649</sub>	D <sub>650</sub>	D <sub>651</sub>	D <sub>652</sub>	D <sub>653</sub>	D <sub>654</sub>	D <sub>655</sub>	D <sub>656</sub>	D <sub>657</sub>	D <sub>658</sub>	D <sub>659</sub>	D <sub>660</sub>	D <sub>661</sub>	D <sub>662</sub>	D <sub>663</sub>	D <sub>664</sub>	D <sub>665</sub>	D <sub>666</sub>	D <sub>667</sub>	D <sub>668</sub>	D <sub>669</sub>	D <sub>670</sub>	D <sub>671</sub>	D <sub>672</sub>	D <sub>673</sub>	D <sub>674</sub>	D <sub>675</sub>	D <sub>676</sub>	D <sub>677</sub>	D <sub>678</sub>	D <sub>679</sub>	D <sub>680</sub>	D <sub>681</sub>	D <sub>682</sub>	D <sub>683</sub>	D <sub>684</sub>	D <sub>685</sub>	D <sub>686</sub>	D <sub>687</sub>	D <sub>688</sub>	D <sub>689</sub>	D <sub>690</sub>	D <sub>691</sub>	D <sub>692</sub>	D <sub>693</sub>	D <sub>694</sub>	D <sub>695</sub>	D <sub>696</sub>	D <sub>697</sub>	D <sub>698</sub>	D <sub>699</sub>	D <sub>700</sub>	D <sub>701</sub>	D <sub>702</sub>	D <sub>703</sub>	D <sub>704</sub>	D <sub>705</sub>	D <sub>706</sub>	D <sub>707</sub>	D <sub>708</sub>	D <sub>709</sub>	D <sub>710</sub>	D <sub>711</sub>	D <sub>712</sub>	D <sub>713</sub>	D <sub>714</sub>	D <sub>715</sub>	D <sub>716</sub>	D <sub>717</sub>	D <sub>718</sub>	D <sub>719</sub>	D <sub>720</sub>	D <sub>721</sub>	D <sub>722</sub>	D <sub>723</sub>	D <sub>724</sub>	D <sub>725</sub>	D <sub>726</sub>	D <sub>727</sub>	D <sub>728</sub>	D <sub>729</sub>	D <sub>730</sub>	D <sub>731</sub>	D <sub>732</sub>	D <sub>733</sub>	D <sub>734</sub>	D <sub>735</sub>	D <sub>736</sub>	D <sub>737</sub>	D <sub>738</sub>	D <sub>739</sub>	D <sub>740</sub>	D <sub>741</sub>	D <sub>742</sub>	D <sub>743</sub>	D <sub>744</sub>	D <sub>745</sub>	D <sub>746</sub>	D <sub>747</sub>	D <sub>748</sub>	D <sub>749</sub>	D <sub>750</sub>	D <sub>751</sub>	D <sub>752</sub>	D <sub>753</sub>	D <sub>754</sub>	D <sub>755</sub>	D <sub>756</sub>	D <sub>757</sub>	D <sub>758</sub>	D <sub>759</sub>	D <sub>760</sub>	D <sub>761</sub>	D <sub>762</sub>	D <sub>763</sub>	D <sub>764</sub>	D <sub>765</sub>	D <sub>766</sub>	D <sub>767</sub>	D <sub>768</sub>	D <sub>769</sub>	D <sub>770</sub>	D <sub>771</sub>	D <sub>772</sub>	D <sub>773</sub>	D <sub>774</sub>	D <sub>775</sub>	D <sub>776</sub>	D <sub>777</sub>	D <sub>778</sub>	D <sub>779</sub>	D <sub>780</sub>	D <sub>781</sub>	D <sub>782</sub>	D <sub>783</sub>	D <sub>784</sub>	D <sub>785</sub>	D <sub>786</sub>	D <sub>787</sub>	D <sub>788</sub>	D <sub>789</sub>	D <sub>790</sub>	D <sub>791</sub>	D <sub>792</sub>	D <sub>793</sub>	D <sub>794</sub>	D <sub>795</sub>	D <sub>796</sub>	D <sub>797</sub>	D <sub>798</sub>	D <sub>799</sub>	D <sub>800</sub>	D <sub>801</sub>	D <sub>802</sub>	D <sub>803</sub>	D <sub>804</sub>	D <sub>805</sub>	D <sub>806</sub>	D <sub>807</sub>	D <sub>808</sub>	D <sub>809</sub>	D <sub>810</sub>	D <sub>811</sub>	D <sub>812</sub>	D <sub>813</sub>	D <sub>814</sub>	D <sub>815</sub>	D <sub>816</sub>	D <sub>817</sub>	D <sub>818</sub>	D <sub>819</sub>	D <sub>820</sub>	D <sub>821</sub>	D <sub>822</sub>	D <sub>823</sub>	D <sub>824</sub>	D <sub>825</sub>	D <sub>826</sub>	D <sub>827</sub>	D <sub>828</sub>	D <sub>829</sub>	D <sub>830</sub>	D <sub>831</sub>	D <sub>832</sub>	D <sub>833</sub>	D <sub>834</sub>	D <sub>835</sub>	D <sub>836</sub>	D <sub>837</sub>	D <sub>838</sub>	D <sub>839</sub>	D <sub>840</sub>	D <sub>841</sub>	D <sub>842</sub>	D <sub>843</sub>	D <sub>844</sub>	D <sub>845</sub>	D <sub>846</sub>	D <sub>847</sub>	D <sub>848</sub>	D <sub>849</sub>	D <sub>850</sub>	D <sub>851</sub>	D <sub>852</sub>	D <sub>853</sub>	D <sub>854</sub>	D <sub>855</sub>	D <sub>856</sub>	D <sub>857</sub>	D <sub>858</sub>	D <sub>859</sub>	D <sub>860</sub>	D <sub>861</sub>	D <sub>862</sub>	D <sub>863</sub>	D <sub>864</sub>	D <sub>865</sub>	D <sub>866</sub>	D <sub>867</sub>	D <sub>868</sub>	D <sub>869</sub>	D <sub>870</sub>	D <sub>871</sub>	D <sub>872</sub>	D <sub>873</sub>	D <sub>874</sub>	D <sub>875</sub>	D <sub>876</sub>	D <sub>877</sub>	D <sub>878</sub>
-----	----	----	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------

Taula RS6.16. Polimorfismes nucleotídics present en la regió p10849 en 18 línies de *D. subobscura* i una línia de *D. guanche*.

	138	168	230	242	311	334	346	438	444	463	464	471	510	513
p10849	C	G	T	G	C	T	G	C	A	C	A	T	A	C
J1OF07	G	.	.	.	.	.	.	.	C	.	G	.	.	.
J1OF14	G	.	.	.	.	.	.	.	.	.	.	.	.	.
J1OF15	.	.	.	.	.	.	.	.	T	A	.	.	.	.
J1OF26	.	.	C	.	.	.	.	.	.	.	G	C	.	.
J1OF27	G	.	C	.	.	.	.	.	.	.	G	C	.	.
J1OF28	G	.	.	.	.	.	.	.	.	.	.	.	.	.
J1OF31	G	.	.	.	.	.	.	.	.	.	.	.	.	.
J1OF32	G	.	.	.	.	.	.	.	.	.	.	.	.	.
OF44	G	A	.	.	A	.	.	.	.	.	.	.	.	.
J1OF58	G	.	.	.	.	.	.	.	.	.	.	.	T	.
J1OF74	G	.	.	.	A	.	.	.	.	.	.	.	.	.
J1OF77	G	.	.	.	.	.	.	.	.	.	.	.	.	.
J1OF78	G	T	C	A	.	T	A	C	.	G	.	.	.	.
J1OF79	G	.	.	.	.	A	.	.	.	.	.	.	.	.
JsTOF16	G	.	.	.	.	A	.	.	.	.	.	.	.	.
JsTOF18	G	.	.	.	.	.	.	.	.	.	.	.	.	.
JsTOF51	G	.	.	.	.	.	.	.	.	.	.	.	.	.
JsTOF54	G	.	C	.	.	A	.	.	.	.	.	.	.	A
<i>D. guanche</i>	G	T	C	.	.	.	.	.	.	.	.	.	.	.





**RS6.2 Taules de polimorfisme de les regions localitzades dintre de la inversió J<sub>1</sub>**

Polimorfisme nucleotídic en cadascuna de les 6 regions no codificadores en estudi localitzades al llarg de la inversió J<sub>1</sub> de *D. subobscura*. En general, aquestes regions es van seqüenciar en 27 línies homocariotípiques per al cromosoma J obtingudes a partir d'una població natural de *D. subobscura* de Barcelona i una línia de *D. guanche* utilitzada com a *outgroup*. En la part superior de les taules, s'indica la numeració de les posicions polimòrfiques en funció de l'alineament múltiple. Els punts (·) indiquen el mateix nucleòtid que en la primera seqüència que s'utilitza de referència. En els *indels* (D<sub>x</sub>, deleció; I<sub>x</sub>, Inserció ) i els Microsatèl·lits (M<sub>x</sub>) s'indiquen el número de la primera i última posició afectada, així com la seva longitud en el subíndex x. En la última fila es mostra la informació que presenta la seqüència de l'espècie *D. guanche* utilitzada com a *outgroup* en aquelles posicions que són polimòrfiques a *D. subobscura*. En ombrejat es mostra les línies amb ordenació estàndard (J<sub>st</sub>).



Taula RS6.20. Polimorfisme nucleotidic present en la regió p127 en 26 línies de *D. subobscura* i una línia de *D. guanche*.

p127	l <sub>4</sub>	D <sub>7</sub>	l <sub>3</sub>	D <sub>4</sub>	D <sub>12</sub>	M	
8-11	-	G A C C A G G C G G T T A T T T G G - - G G G A C G T A - -	A T A - A G C A A C T G	T T C C	(CGCT) <sub>2</sub> CGCA	G C - G C	
80	-	C C	T	G	(CGCT) <sub>2</sub> CGCA	.	1280
81	-	C C	T	.	(CGCT) <sub>2</sub> CGCA	.	1241
175	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1201
195	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
175	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1195
202	-	C C	.	A T	(CGCT) <sub>2</sub> CGCA	.	1181
213	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
300	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
302	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
303	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
311	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
386	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
435	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
491	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
494	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
519	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
521	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
570-576	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
624	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
629	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
630	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
634	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
635	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
639	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
725	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
727	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
734	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
751	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
761	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
802	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
806	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
811	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
834	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
836	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
866	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
889-891	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
892	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
905	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
915	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
920	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1018	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1036	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1048	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1087-1101	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1101-1112	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1116	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1121	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1131	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
1154-1117	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F7	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F14	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F19	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F23	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F26	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F27	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F28	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F31	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F32	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F42	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F44	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F60	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F74	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F58	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F77	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F78	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
J10F79	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF1	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF13	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF16	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF18	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF21	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF47	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF49	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF51	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
JstOF54	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181
<i>D. guanche</i>	-	C C	.	.	(CGCT) <sub>2</sub> CGCA	.	1181

Taula RS6.21. Polimorfisme nucleotidic present en la regió p221 en 25 línies de *D. subobscura* i una línia de *D. guanche*.

p221	D <sub>5</sub>																									D <sub>1</sub>																									D <sub>1</sub> <sup>H</sup>																									D <sub>1</sub> <sup>S</sup>																									D <sub>1</sub>																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177		1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530	1531	1532	1533	1534	1535	1536	1537	1538	1539	1540	1541	1542	1543	1544	1545	1546	1547	1548	1549	1550	1551	1552	1553	1554	1555	1556	1557	1558	1559	1560	1561	1562	1563	1564	1565	1566	1567	1568	1569	1570	1571	1572	1573	1574	1575	1576	1577	1578	1579	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589	1590	1591	1592	1593	1594	1595	1596	1597	1598	1599	1600	1601	1602	1603	1604	1605	1606	1607	1608	1609	1610	1611	1612	1613	1614	1615	1616	1617	1618	1619	1620	1621	1622	1623	1624	1625	1626	1627	1628	1629	1630	1631	1632	1633	1634	1635	1636	1637	1638	1639	1640	1641	1642	1643	1644	1645	1646	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700	1701	1702	1703	1704	1705	1706	1707	1708	1709	1710	1711	1712	1713	1714	1715	1716	1717	1718	1719	1720	1721	1722	1723	1724	1725	1726	1727	1728	1729	1730	1731	1732	1733	1734	1735	1736	1737	1738	1739	1740	1741	1742	1743	1744	1745	1746	1747	1748	1749	1750	1751	1752	1753	1754	1755	1756	1757	1758	1759	1760	1761	1762	1763	1764	1765	1766	1767	1768	1769	1770	1771	1772	1773	1774	1775	1776	1777	1778	1779	1780	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811	1812	1813	1814	1815	1816	1817	1818	1819	1820	1821	1822	1823	1824	1825	1826	1827	1828	1829	1830	1831	1832	1833	1834	1835	1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847	1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419



Taula RS6.22.a. Polimorfisme nucleotídic present en la regió p137 en 25 línies de *D. subobscura* i una línia de *D. guanche*.

p137	D <sub>1</sub>		D <sub>2</sub>		D <sub>3</sub>		D <sub>4</sub>		D <sub>5</sub>		D <sub>6</sub>		D <sub>7</sub>		D <sub>8</sub>		D <sub>9</sub>		D <sub>10</sub>		D <sub>11</sub>		D <sub>12</sub>		D <sub>13</sub>				
	h <sub>18</sub>	D <sub>13</sub>	D <sub>11</sub>	D <sub>10</sub>	h <sub>17</sub>	D <sub>8</sub>	D <sub>9</sub>	D <sub>10</sub>	D <sub>11</sub>	D <sub>12</sub>	D <sub>13</sub>	D <sub>14</sub>	D <sub>15</sub>	D <sub>16</sub>	D <sub>17</sub>	D <sub>18</sub>	D <sub>19</sub>	D <sub>20</sub>	D <sub>21</sub>	D <sub>22</sub>	D <sub>23</sub>	D <sub>24</sub>	D <sub>25</sub>	D <sub>26</sub>	D <sub>27</sub>	D <sub>28</sub>	D <sub>29</sub>	D <sub>30</sub>	
J10F07	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
J10F14	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
J10F19	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F23	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F26	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F27	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F28	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F31	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F32	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F42	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F44	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F58	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F60	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F74	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F77	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F78	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F79	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F81	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F83	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F16	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F18	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F21	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F47	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F49	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F51	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
J10F51	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T







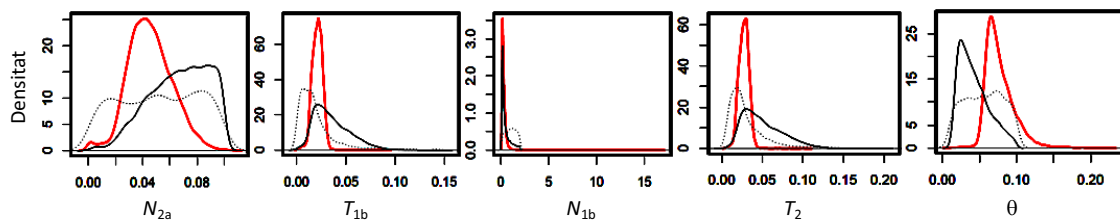




### RS6.3 Distribucions posteriors dels dos models menys probables

En les Figures RS6.1 i RS6.2 es mostren les distribucions posteriors dels diferents paràmetres que conformen els models M3 i M4. En les Taules RS6.25-RS6.28 es mostren diferents estadístics descriptius de la distribució posterior dels paràmetres d'interès dels models M3 i M4, respectivament. Aquestes distribucions posteriors es van analitzar prèviament a la tria de models, per corroborar que els models a contrastar podien ser comparats entre ells, de manera similar al descrit en el cas del model M2. Per obtenir aquestes distribucions posteriors es va seguir el mètode de rebuig (Pritchard et al. 1999) amb correcció basada en regressió lineal local (Beaumont et al. 2002) utilitzant un milió de simulacions i amb un llindar de tolerància del 0,01. Els estadístics que es van utilitzar per aquestes anàlisis són: nombre de llocs segregants ( $S$ ), estadístic  $D$  de Tajima (1989), estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) i estadístic  $Z_A$  (Rozas et al. 2003).

#### Model de coll d'ampolla (M3)



**Figura RS6.1.** Distribució posterior dels diferents paràmetres que defineixen el model M3, amb un llindar de tolerància del 0,01.

Les distribucions amb línia discontinua representen el valor de cada paràmetre en les simulacions realitzades, en tant que les distribucions en negre i vermell representen el valor corresponent pel conjunt de 10.000 simulacions retinudes mitjançant el mètode de rebuig (Tavaré et al. 1997) i el mètode de correcció per regressió lineal (Beaumont et al. 2002), respectivament. Els diferents paràmetres són:  $N_{2a}$ , grandària durant el coll d'ampolla;  $T_{1b}$ , temps en el qual es produeix la recuperació després del coll d'ampolla;  $N_{1b}$ , grandària prèvia al coll d'ampolla;  $T_2$ , temps d'inici del coll d'ampolla i  $\theta$ , paràmetre poblacional de mutació ( $\theta = 4N_e \mu$ , sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació).

**Taula RS6.25.** Estadístics descriptius de la distribució posterior dels paràmetres del model M3 sent aquests la mitjana, mediana i la moda així com també els límits inferior i superior del 95% de la distribució, escollint un llindar de tolerància del 0,01 i mitjançant el mètode de correcció per regressió lineal (Beaumont et al. 2002).

Estadístics descriptius/Paràmetres	$N_{2A}$	$T_{1b}$	$N_{1b}$	$T_2$	$\theta$
<b>límit inferior del 95% de la distribució</b>	0,023	0,013	0,112	0,018	0,055
<b>Mediana</b>	0,045	0,021	0,238	0,028	0,073
<b>Mitjana</b>	0,046	0,021	0,340	0,028	0,078
<b>Moda</b>	0,041	0,022	0,156	0,030	0,066
<b>límit superior del 95% de la distribució</b>	0,074	0,029	0,709	0,037	0,112

Els diferents paràmetres són:  $N_{2a}$ , grandària durant el coll d'ampolla;  $T_{1b}$ , temps en el qual es produeix la recuperació després del coll d'ampolla;  $N_{1b}$ , grandària prèvia al coll d'ampolla;  $T_2$ , temps d'inici del coll d'ampolla i  $\theta$ , paràmetre poblacional de mutació, ( $\theta = 4N_e\mu$ ) sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació. Els diferents temps s'expressen en  $4N_e$  generacions.

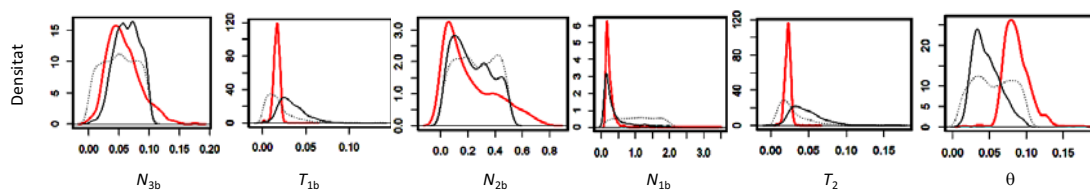
**Taula RS6.26.** Estadístics descriptius de la distribució posterior dels paràmetres del model M3 calculats en funció de la grandària efectiva actual estimada a partir dels valors obtinguts de  $\theta$ . Es presenta la mitjana, mediana i la moda així com també els límits inferior i superior del 95% de la distribució, amb un llindar de tolerància del 0,01 i mitjançant el mètode de correcció per regressió lineal.

Estadístics descriptius/Paràmetres	$N_{2a}$	$T_{1b}$	$N_{1b}$	$T_2$	$N_e$
<b>límit inferior del 95% de la distribució</b>	$1,26 \times 10^5$	56972	$6,14 \times 10^5$	78884	$5,48 \times 10^6$
<b>Mediana</b>	$3,27 \times 10^5$	122151	$1,73 \times 10^6$	162869	$7,27 \times 10^6$
<b>Mitjana</b>	$3,57 \times 10^5$	130518	$2,64 \times 10^6$	174024	$7,77 \times 10^6$
<b>Moda</b>	$2,70 \times 10^5$	115697	$1,03 \times 10^6$	157769	$6,57 \times 10^6$
<b>límit superior del 95% de la distribució</b>	$8,25 \times 10^5$	258805	$7,91 \times 10^6$	330199	$1,11 \times 10^6$

Els diferents paràmetres són:  $N_{2a}$ , grandària durant el coll d'ampolla;  $T_{1b}$ , temps en el qual es produeix la recuperació després del coll d'ampolla;  $N_{1b}$ , grandària prèvia al coll d'ampolla;  $T_2$ , temps d'inici del coll d'ampolla i  $\theta$  ( $\theta = 4 N_e\mu$ ), sent  $N_e$  la grandària efectiva actual,  $\mu$  la taxa de mutació per posició i per generació.



### Model de barreja (M4)



**Figura RS6.2.** Distribució posterior dels diferents paràmetres que defineixen el model M4, amb un llinar de tolerància del 0,01.

Les distribucions amb línia discontinua representen el valor de cada paràmetre en les simulacions realitzades, en tant que les distribucions en negre i vermell representen el valor corresponent pel conjunt de 10000 simulacions retingudes mitjançant el mètode de rebuig (Tavaré et al. 1997) i el mètode de correcció per regressió lineal (Beaumont et al. 2002), respectivament. Els diferents paràmetres són:  $N_{3b}$ , reducció durant la subdivisió poblacional;  $T_{1b}$ , temps d'inici de la recuperació després del coll d'ampolla;  $N_{2b}$ , subpoblació;  $N_{1b}$ , grandària ancestral prèvia al coll d'ampolla;  $T_2$ , temps d'inici del coll d'ampolla i  $\theta$ , paràmetre poblacional de mutació ( $\theta = 4N_e \mu$ , sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació).

**Taula RS6.27.** Estadístics descriptius de la distribució posterior dels paràmetres del model M4 sent aquests la mitjana, mediana i la moda així com també els límits inferior i superior del 95% de la distribució, escollint un llinar de tolerància del 0,01 i mitjançant el mètode de correcció per regressió lineal (Beaumont et al. 2002).

Paràmetres	$N_{3b}$	$T_{1b}$	$N_{2b}$	$N_{1b}$	$T_2$	$\theta$
límit inferior del 95% de la distribució	0,018	0,011	0,019	0,148	0,016	0,057
Mediana	0,045	0,019	0,163	0,285	0,026	0,075
Mitjana	0,048	0,019	0,214	0,358	0,026	0,079
Moda	0,042	0,019	0,044	0,187	0,027	0,068
límit superior del 95% de la distribució	0,089	0,027	0,551	0,744	0,036	0,116

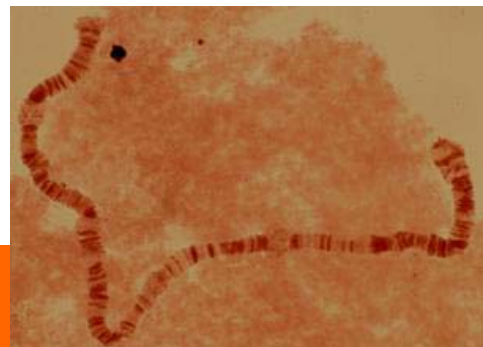
Els diferents paràmetres són:  $N_{3b}$ , reducció durant la subdivisió poblacional;  $T_{1b}$ , temps d'inici de la recuperació després del coll d'ampolla;  $N_{2b}$ , subpoblació;  $N_{1b}$ , grandària ancestral prèvia al coll d'ampolla;  $T_2$ , temps d'inici del coll d'ampolla i  $\theta$ , paràmetre poblacional de mutació, ( $\theta = 4N_e \mu$ ) sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació. Els diferents temps s'expressen en  $4N_e$  generacions.

**Taula RS6.28.** Estadístics descriptius de la distribució posterior dels paràmetres del model M4 calculats en funció de la grandària efectiva actual estimada a partir dels valors obtinguts de  $\theta$ . Es presenta la mitjana, mediana i la moda així com també els límits inferior i superior del 95% de la distribució, amb un llindar de tolerància del 0,01 i mitjançant el mètode de correcció per regressió lineal.

<b>Paràmetres</b>	<b><math>N_{3b}</math></b>	<b><math>T_{1b}</math></b>	<b><math>N_{2b}</math></b>	<b><math>N_{1b}</math></b>	<b><math>T_2</math></b>	<b><math>\theta</math></b>
<b>límit inferior del 95% de la distribució</b>	$1,02 \times 10^5$	49960	$1,94 \times 10^3$	$8,40 \times 10^5$	72669	$5,68 \times 10^6$
<b>Mediana</b>	$3,36 \times 10^5$	113546	$5,48 \times 10^4$	$2,13 \times 10^6$	155378	$7,47 \times 10^6$
<b>Mitjana</b>	$3,78 \times 10^5$	119602	$8,08 \times 10^4$	$2,82 \times 10^6$	163665	$7,87 \times 10^6$
<b>Moda</b>	$2,84 \times 10^5$	102948	$1,25 \times 10^4$	$1,27 \times 10^6$	146295	$6,77 \times 10^6$
<b>límit superior del 95% de la distribució</b>	$1,03 \times 10^6$	249562	$5,67 \times 10^5$	$8,60 \times 10^6$	332749	$1,16 \times 10^7$

Els diferents paràmetres són:  $N_{3b}$ , reducció durant la subdivisió poblacional;  $T_{1b}$ , temps d'inici de la recuperació després del coll d'ampolla;  $N_{2b}$ , subpoblació;  $N_{1b}$ , grandària ancestral prèvia al coll d'ampolla;  $T_2$ , temps d'inici del coll d'ampolla i  $\theta$  ( $\theta = 4 N_e \mu$ ),





## DISCUSSIÓ



## D1. OBTENCIÓ DE LÍNIES HOMOCARIOTÍPIQUES PER AL CROMOSOMA J DE *D. SUBOBSCURA*

Hi ha una llarga tradició d'estudis de genètica de poblacions els quals utilitzen l'espècie *D. subobscura* per analitzar la variació estructural cromosòmica, ja que l'elevada presència d'inversions cromosòmiques que presenta, juntament amb la fàcil visualització dels seus cromosomes politènics, fa que sigui una espècie ideal per aquests tipus d'estudis. El descobriment, entre altres, de clines latitudinals a *D. subobscura* relacionades amb les freqüències de les ordenacions cromosòmiques, va permetre establir el caràcter adaptatiu de les inversions cromosòmiques en aquesta espècie i discernir que aquestes clines no eren degudes a factors demogràfics (Prevosti et al. 1985, 1988; Krimbas i Powell 1992). S'han proposat diversos mecanismes en relació a l'augment de freqüència d'una determinada inversió cromosòmica i el manteniment posterior de les dues ordenacions en les poblacions naturals (Bauer 1936; Dobzhansky 1950; Van Valen i Levins 1968; Wasserman 1968; Kirkpatrick i Barton 2006). Aquests mecanismes estarien basats en l'efecte genètic de les inversions, és a dir en la reducció de la recombinació que es dona al llarg de la regió invertida entre individus portadors i no portadors de la inversió.

Aquesta elevada presència d'inversions cromosòmiques en l'espècie *D. subobscura* fa que el procés d'obtenció de línies homocariotípiques mitjançant encreuaments germà-germana sigui laboriós i costós. Krimbas (1993), comenta com en l'espècie *D. subobscura* és molt difícil obtenir línies homocariotípiques en tot el seu complement cromosòmic ni després de 12 generacions d'encreuaments germà-germana. En el nostre estudi, a través de l'observació visual dels cromosomes politènics d'almenys 8 larves de tercer estadi, es van obtenir diferents línies homocariotípiques en relació al cromosoma J, permetent d'aquesta manera la seqüenciació dels al·lels de les diferents regions en homozigosi. No obstant, després d'efectuar 13 generacions d'encreuaments germà-germana encara hi havia línies heterocariotípiques pel cromosoma J. En aquestes línies es van continuar efectuant encreuaments germà-germana fins

---

aconseguir que fossin homocariotípiques pel cromosoma J. Aquesta persistència de l'heterozigositat en relació al cromosoma J, podria reflectir una major eficàcia biològica dels individus heterozigots  $J_{st}/J_1$  en relació als homozigots  $J_{st}$  o  $J_1$ . S'ha de tenir en compte que la dificultat d'obtenció de línies homocariotípiques està estretament relacionada amb la presència d'al·lels deleteris recessius, fet que determina que a mesura que s'augmenta la consanguinitat, disminueix l'eficàcia biològica dels individus portadors d'aquests al·lels. S'ha de destacar que aquestes línies homocariotípiques constitueixen un recurs biològic que serà molt important per futurs estudis de genètica de poblacions no només en relació al cromosoma J sinó també en relació a altres cromosomes i a l'estudi de la variació estructural del genoma de *D. subobscura*.

## D2. ESTUDI DELS NIVELLS I PATRONS DE VARIACIÓ A *D. SUBOBSCURA* EN REGIONS NO CODIFICADORES LOCALITZADES AL LLARG DEL CROMOSOMA J I NO AFECTADES PER INVERSIONS CROMOSÒMIQUES

La variabilitat genètica de les poblacions naturals aporta informació sobre la seva història evolutiva ja que és aquesta variabilitat la que s'hereta d'una generació a una altra. Estudiar els nivells i patrons de variació a nivell de seqüències nucleotídiques hauria de permetre, per tant, analitzar tant la història demogràfica de l'espècie com detectar la petjada que la selecció positiva deixa en aquestes seqüències. Amb aquest objectiu, en aquesta part del treball es discuteixen els resultats obtinguts de l'estudi de diferents regions no codificadores localitzades al llarg del cromosoma J (element D de Muller) de *D. subobscura*. Les regions no codificadores són bones candidates per abordar aquests objectius ja que al ser regions subjectes a limitacions funcionals febles, en elles l'acció de la selecció purificadora no és tant acusada i per tant els nivells de variació haurien de ser més elevats que en regions codificadores. Les anàlisis dels nivells i patrons de polimorfisme en diferents regions no codificadores van permetre caracteritzar la variació genètica d'aquestes regions, realitzar inferències demogràfiques i analitzar si aquestes estaven afectades o no per esdeveniments adaptatius recents.

### D2.1 Nivells i patrons de variació

Les anàlisis dels nivells i patrons de variació en diferents regions no codificadores posen de manifest que la variació a la població natural de *D. subobscura* en estudi no es troba en equilibri deriva-mutació. Els estadístics basats en l'espectre de freqüències  $D$  de Tajima (1989) i  $D$  i  $F$  de Fu i Li amb *outgroup* (1993) mostren un patró de variació caracteritzat per un excés de polimorfismes amb variants a baixa freqüència. A més a més, s'observa una proporció molt elevada de polimorfismes amb variants úniques, arribant en algunes de les regions a més del 80%. S'ha d'afegir que en treballs anteriors en l'espècie *D. subobscura*, primer a nivell de



gens individuals (Rozas et al. 1995; Navarro-Sabaté et al. 2003; Sánchez-Gracia et al. 2011) i posteriorment a nivell multilocus (Munté et al. 2005; Nóbrega et al. 2008; Simões et al. 2012), s'observen patrons de variació similars als detectats en el nostre estudi.

Un espectre de freqüències desviat cap a polimorfismes amb variants a baixa freqüència en una regió determinada podria indicar que la història evolutiva de la població ha estat afectada per factors demogràfics o que aquesta regió ha estat sotmesa a pressions selectives. De fet, en determinades situacions els canvis adaptatius i no adaptatius poden afectar de forma similar al nivell i patró de polimorfisme en una regió del genoma determinada. En relació als factors selectius, un espectre de freqüències desviat cap a variants a baixa freqüència podria indicar l'acció recent de la selecció direccional positiva. La selecció positiva pot actuar sobre una mutació beneficiosa augmentant la seva freqüència en la població fins arribar a fixar-la. Associat a aquesta acció es produiria una escombrada selectiva de les mutacions neutres adjacents a la seleccionada que arribarien també finalment a fixar-se (Maynard i Haigh 1974). Després d'aquest procés d'escombrada selectiva sorgirien noves mutacions amb un patró de variació nucleotídica caracteritzat per un excés de variants a baixa freqüència. En relació als factors demogràfics que també podrien mostrar un espectre de freqüències similar al produït per un fenomen d'escombrada selectiva, es trobarien les expansions poblacionals o els colls d'ampolla amb una posterior recuperació. L'anàlisi a nivell multilocus va ser una bona estratègia per poder diferenciar si la variació detectada a les diferents regions havia estat modelada per factors selectius i/o demogràfics, ja que els canvis demogràfics, a diferència dels selectius, tenen un efecte sobre totes les regions del genoma. Els estadístics que detecten desviacions de l'espectre de freqüències, concretament la  $D$  de Tajima i la  $D$  i  $F$  de Fu i Li amb *outgroup* mostren un patró similar en totes les regions analitzades al llarg del cromosoma J. Aquests resultats posen de manifest que els nivells i patrons de variació observats en les diferents regions en estudi reflecteixen canvis demogràfics

recents de l'espècie més que processos selectius. Els resultats obtinguts pels estadístics  $R_2$  (Ramos-Onsins i Rozas 2002) i  $F_s$  (Fu 1997), sensibles per detectar expansions poblacionals, indiquen que molt possiblement la història demogràfica de l'espècie estaria caracteritzada per una expansió poblacional recent. A més a més, els elevats valors de diversitat haplotípica així com també els resultats obtinguts amb l'estadístic  $B$  de Wall (1999), indicarien que la població en estudi seria una població panmíctica en relació a la variació genètica. El test HKA (test de Hudson-Kreitman-Aguadé; Hudson et al. 1987) no detecta heterogeneïtat significativa en la relació polimorfisme/divergència entre les diferents regions en estudi. No obstant, s'observa una aportació majoritària al valor de  $\chi^2$  de la regió p169 (56%), que podria ser degut a que en aquesta regió o al seu voltant es trobés la possible diana d'un esdeveniment recent de selecció. S'ha d'afegir que la regió p169 és la única de les regions en estudi en la qual s'observa un valor de l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000; Zeng et al. 2006) que es desvia del que s'esperaria sota el model d'evolució neutre estàndard, essent aquest estadístic sensible per detectar variants derivades a elevada freqüència. Alhora de realitzar inferències demogràfiques es va decidir seguir un criteri conservatiu i per tant no es va incloure en les anàlisis la regió p169 pels indicis esmentats d'haver pogut estar afectada per un esdeveniment recent de selecció.

Així doncs i en concordança amb estudis anteriors efectuats en l'espècie *D. subobscura*, es pot afirmar que la història evolutiva recent d'aquesta espècie molt possiblement ha estat influenciada per processos demogràfics, posant de manifest la importància de disposar d'un model demogràfic que expliqui les dades observades i alhora permeti realitzar prediccions sobre els nivells i patrons de polimorfisme.

### D3. INFERÈNCIES DEMOGRÀFIQUES EN L'ESPÈCIE *D. SUBOBSCURA*

#### D3.1 Temps de divergència

En aquesta part del treball es discuteixen els resultats obtinguts de revisar les estimes del temps de divergència tant entre *D. subobscura* i *D. guanche* (espècies que juntament amb *D. madeirensis* formen el subgrup *subobscura*) com entre els subgrups *subobscura* i *pseudoobscura*, analitzant la variació nucleotídica en un total de 17 kb aproximadament. Les espècies que formen part del grup *obscura* estan distribuïdes tant en la regió neàrtica com paleàrtica. Hi ha varis estudis filogenètics que evidencien que els subgrups del grup *obscura* són monofilètics, proporcionant aquestes anàlisis un elevat suport als subgrups *affinis*, *sinobscura*, *pseudoobscura* i *subobscura* (Beckenbach 1993; O'Grady 1999). Gao et al. (2007) revisen la filogènia del grup *obscura* basant-se en la variació nucleotídica en sis regions, 4 d'elles localitzades en gens mitocondrials i dos en gens nuclears per un total de 4125 nucleòtids. Aquests autors arriben a la conclusió que els subgrups *subobscura* i *microlabis* són els nodes basals del grup *obscura* i que com a mínim la meitat dels llinatges del grup *obscura* s'haurien originat en el Miocè mitjà, aproximadament ara fa 15 milions d'anys.

Els resultats obtinguts en el present treball mostraren que el temps de divergència més probable entre *D. subobscura* i *D. guanche* és d'ara fa aproximadament 2 milions d'anys (1,65-2,33 milions d'anys com a límits inferior i superior del 95% de la distribució). Aquesta estima no difereix considerablement de l'obtinguda per Ramos-Onsins et al. (1998) a partir del gen *rp49*, que és de 1,8 milions d'anys. Les anàlisis efectuades en el present treball també posen de manifest que el temps de separació més probable dels subgrups *subobscura* i *pseudoobscura* és d'ara fa 7,5 milions d'anys (6,54 - 8,45 com a límits inferior i superior del 95% de la distribució), sent aquest temps menor a l'estimat tant per Ramos-Onsins et al. (1998) com per Segarra i Aguadé (1993): 8,1 i 9 milions d'anys, respectivament. Cal destacar que el temps de separació dels subgrups *subobscura* i *pseudoobscura* inferit en el present treball és molt

inferior a l'estimat per Tamura et al. (2004) de  $17,7 \pm 4,4$  milions d'anys en utilitzar les dades prèviament publicades dels gens *28S*, *COII* i *Adh* (O'Grady et al. 2002), i dels gens *COI* i *Gpdh* (Goto i Kimura 2001). Aquesta discrepància és, no obstant, només degut a que Tamura et al. (2004) prenen com a referència un temps de divergència entre els grups *D. melanogaster* i *obscura* d'aproximadament 60 milions d'anys, sent pràcticament el doble del temps de divergència considerat en aquest treball. Si es té en compte aquest fet les dues estimes són similars.

Disposar d'una estima robusta del temps de divergència entre *D. subobscura* i *D. guanche* és important per estimar la taxa de mutació per posició i per generació de cada regió a partir de la seva divergència interespecífica per tal d'obtenir la distribució *a priori* del paràmetre poblacional de mutació a utilitzar en les simulacions de coalescència. Així mateix és important quan en les simulacions de coalescència s'inclou *D. guanche* com a espècie *outgroup* per tal de polaritzar les mutacions en el llinatge de *D. subobscura*.

### **D3.2 Metodologies d'Approximate Bayesian Computation: pros i contres**

El desenvolupament durant les dues últimes dècades de les metodologies d'Approximate Bayesian Computation (Fu i Li 1997; Tavaré et al. 1997; Beaumont et al. 2002) ha estat clau per l'estudi de la història evolutiva de les poblacions. En aquesta part del treball es discuteixen les inferències realitzades sobre els canvis demogràfics recents en l'espècie *D. subobscura*. Els resultats d'ABC recolzen que, dels 3 models demogràfics que plantegen canvis en la grandària efectiva de l'espècie en determinats períodes de temps, el model més probable per explicar la variabilitat nucleotídica observada en la població natural de *D. subobscura* en estudi és el model d'expansió poblacional M2 (veure Figura R5.8).

### **Informació prèvia disponible en relació al paràmetre poblacional de recombinació**

Tot i que s'han realitzat diferents estudis per estimar la freqüència de recombinació a *D. subobscura* i obtenir mapes genètics (Christie 1939; Spurway 1945; Bird 1947; Koske i

Maynard Smith 1954; Loukas et al. 1979; Mestres et al. 1998), la informació disponible respecte al nivell de recombinació a l'espècie és encara escassa. En relació a altres espècies que pertanyen al grup *obscura*, Ortiz-Barrientos et al. (2006) estimen en 9 cM/Mb la taxa de recombinació en el braç cromosòmic XR de *D. pseudoobscura* (homòleg del cromosoma J de *D. subobscura*). Pegueroles et al. (2010) estudien la recombinació en diferents línies  $O_{st}$  i  $O_{3+4}$  de *D. subobscura* mitjançant l'estudi de 14 loci microsatèl·lits, observant taxes de recombinació al llarg del cromosoma O força uniformes tot i la presència d'alguns punts calents. Posteriorment, Pegueroles et al. (2013) estimen que la taxa de recombinació en línies homocariotípiques de *D. subobscura* pot variar de 2 a 11 cM/Mb amb una mitjana de 7 cM/Mb. No obstant aquestes dades disponibles en l'espècie *D. subobscura*, la manca de mapes genètics detallats del cromosoma J, així com també l'elevada variància observada en la taxa de recombinació en estudis previs, va fer que es dissenyés una distribució *a priori* pel paràmetre poblacional de recombinació amb un ampli rang de valors. L'amplitud d'aquesta distribució *a priori* podria ser un dels punts febles de l'anàlisi d'ABC efectuada, donat que el paràmetre de recombinació poblacional és un dels paràmetres importants a tenir en compte en aquesta metodologia. No obstant, l'elevat nombre de simulacions efectuades pels diferents models demogràfics plantejats, amb la selecció d'aquelles que millor expliquen les dades observades, hauria de pal·liar en major o menor mesura aquesta incertesa en el paràmetre poblacional de recombinació, paràmetre incorporat per obtenir models més reals tot i que no és objecte d'estudi directe en aquest treball.

### **Tria d'estadístics resum en la tria de models**

Un dels punts crítics de la metodologia d'ABC podria ser la tria d'estadístics efectuada ja que les probabilitats posteriors que s'obtenen poden ser molt dependents d'aquesta tria. En el present treball, es va escollir un conjunt d'estadístics que aporten informació complementària sobre els nivells i patrons de variació. Concretament, els estadístics triats per

ser utilitzats en la tria de models van ser la mitjana i la variància del nombre de llocs segregants, de l'estadístic  $D$  de Tajima (1989), de l'estadístic  $H$  normalitzat de Fay i Wu (Fay i Wu 2000, Zeng et al. 2006) i de l'estadístic  $Z_A$  (Rozas et al. 2001).

S'ha de destacar, que en la majoria d'estudis d'ABC efectuats a *Drosophila* (Pascual et al. 2007; Laurent et al. 2011; Duchon et al. 2013) únicament s'ha utilitzat l'estadístic  $D$  de Tajima com a estadístic resum de l'espectre de freqüències donada la sensibilitat de l'estadístic  $H$  normalitzat de Fay i Wu a l'assignació errònia de les variants derivades (Baudy and Depaulis 2003). En el nostre estudi es va adreçar aquest problema incorporant un *outgroup* en les simulacions, però el que és més important incorporant-hi també la possibilitat de mutacions recurrents. D'aquesta manera, es van comparar els estadístics observats amb els obtinguts a partir de simulacions que tenen en compte com s'originen realment les mutacions en les poblacions naturals. Es va valorar tanmateix si la incorporació de l'estadístic  $H$  normalitzat de Fay i Wu al recull d'estadístics emprat podia afectar la tria de models. A tal fi, es va estimar la probabilitat posterior de cadascun dels tres models considerats (M2, M3 i M4) exclouent del recull d'estadístics la mitjana i la variància de l'estadístic  $H$  normalitzat de Fay i Wu. Amb aquest nou recull d'estadístics, el model M2 continua essent el model amb una major probabilitat posterior (0,68 respecte 0,22 i 0,1 pels models M3 i M4, respectivament).

Finalment, es va constatar que quan en la tria de models s'utilitzen simulacions de coalescència que no incorporen la possibilitat de mutacions recurrents, la inclusió de l'estadístic  $H$  normalitzat de Fay i Wu en el recull d'estadístics emprat condueix a una inferència errònia del model més probable (en aquest cas seria el model M3). No obstant, si s'utilitzen simulacions de coalescència que no incorporen la possibilitat de mutacions recurrents però s'exclou del recull d'estadístics la mitjana i la variància de l' $H$  normalitzat de Fay i Wu, el model més probable continua essent el model M2. Aquesta observació mostra clarament com, al menys en espècies amb nivells elevats de variació com ho és *D. subobscura*,

la inclusió d'aquest estadístic en la metodologia d'ABC pot conduir a inferències errònies quan no es tenen en compte les mutacions recurrents.

#### **Validació de la tria de models**

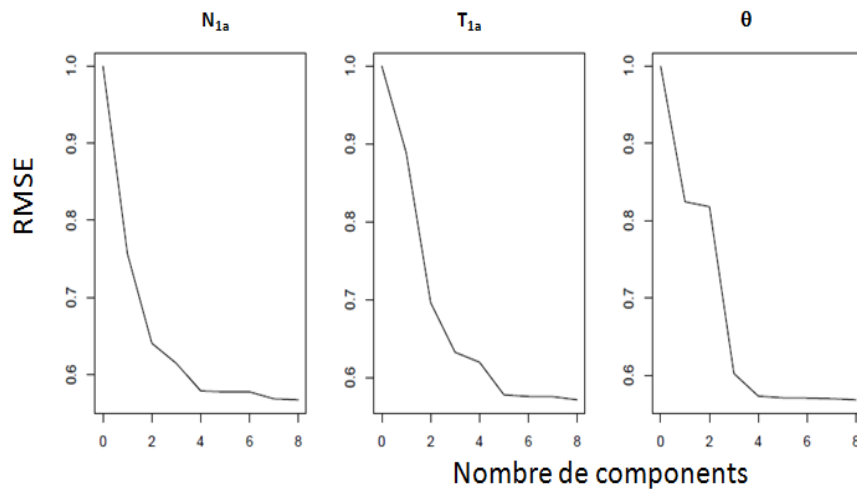
Les anàlisis d'ABC mostren com el model més probable és el model d'expansió M2 amb una probabilitat posterior de 0,93 (veure Figura R5.8). Un dels punts clau de la metodologia d'ABC és que permet analitzar si els models poden ser classificats correctament. El model d'expansió (M2) és classificat correctament en un 90% dels casos (veure Taula R5.7). Per contra el model de barreja (M4) no es pot classificar correctament ja que en un 65% dels casos s'assigna erròniament. Aquest resultat podria reflectir que el model M4 seria molt poc probable o que els estadístics utilitzats no són suficientment informatius en relació al paràmetre  $N_{2b}$  d'aquest model, sent aquest paràmetre la proporció d'individus de la subpoblació més petita (veure Taula R5.4). La primera possibilitat seria la més plausible ja que la distribució posterior obtinguda en relació al paràmetre  $N_{2b}$  (veure figura RS6.2) mostra com el nombre d'individus d'una de les dues subpoblacions és un nombre ínfim quan es compara amb la grandària de l'altra subpoblació. Per tant, les simulacions retingudes del model M4 fan que aquest es comporti com un model de coll d'ampolla sense subdivisió poblacional fet que produeix que quan es realitzen aquestes anàlisis de validació, el model M4 no es pugui distingir del model de coll d'ampolla (M3).

#### **Avaluació de diferents metodologies en les estimes de les distribucions posteriors dels paràmetres del model demogràfic més probable**

Els resultats de la tria de models van indicar que dels tres models comparats, el model demogràfic més probable per explicar la variació present a *D. subobscura*, és el model d'expansió (M2). Alhora d'obtenir les distribucions posteriors dels paràmetres que conformen aquest model d'expansió, es va seguir la metodologia proposada per Beaumont et al. (2002), basada en mètodes de correcció per regressió lineal local (veure Materials i Mètodes, apartat

M3). No obstant, tot i que les metodologies d'ABC hagin estat un gran avanç per fer inferències demogràfiques mitjançant l'ús d'un conjunt d'estadístics per resumir les dades empíriques (Tavaré 1997), un dels seus punts crítics seria precisament quin conjunt d'estadístics incloure, ja que en funció d'aquesta elecció, es podrien esbiaixar les estimes de les distribucions posteriors dels paràmetres. Així és com, triar pocs estadístics podria aportar una quantitat d'informació molt limitada sobre els paràmetres que conformen els models. Per contra, escollir un nombre massa elevat d'estadístics, podria produir "soroll" en les estimes de les distribucions posteriors dels paràmetres (Marjoram et al. 2003; Joyce i Marjoram 2008). Wegmann i Excoffier (2008) proposen reduir la dimensió del conjunt d'estadístics escollit aplicant una transformació d'aquests a través d'una anàlisi de regressió de mínims quadrats parcials (*Partial Least Square, PLS*) que seria un procediment similar a una anàlisi de components principals (PCA). Aquesta metodologia ha estat aplicada en diferents estudis tals com els de Laurent et al. (2011) i Veeramah et al. (2012). No obstant això, aquest procediment no pot ser aplicat en la tria de models, ja que la reducció dels estadístics en components *PLS* s'ha de construir independentment per cada model demogràfic, mentre que per la tria de models és necessari que el conjunt d'estadístics proposat sigui el mateix per tot el conjunt de models. Aquesta metodologia sí que es pot aplicar en les inferències dels paràmetres. Per avaluar si aquesta metodologia podia proporcionar una menor variància en les distribucions posteriors obtingudes dels paràmetres que conformen el model d'expansió, es va aplicar aquest procediment, construint els components *PLS* a partir de 10.000 simulacions tal i com es troba implementat en el paquet informàtic *ABCtoolbox* (Wegmann et al. 2010). Aquest paquet informàtic proporciona un error quadràtic mitjà de predicció (*root mean square error, RMSE*) amb el qual s'avalua el nombre adequat de components *PLS* (Figura D3.1).





**Figura D3.1.** Error quadràtic mitjà de predicció (RMSE) dels diferents paràmetres en estudi ( $N_{1a}$ ,  $T_{1a}$ ,  $\theta$ ) associat a la transformació dels 8 estadístics escollits: mitjana i variància de la diversitat nucleotídica per locus (Tajima 1983), de la  $D$  de Tajima (1989), de l' $H$  normalitzat de Fay i Wu (Fay i Wu 2000, Zeng et al. 2006) i del  $Z_a$ , (Rozas et al. 2001) en 8 components *PLS*. Prèviament a la reducció en components *PLS* es corregeix per la possibilitat de variàncies desiguals mitjançant una transformació *Box-Cox*.

En funció de l'avaluació de l'error quadràtic mitjà de predicció, es van escollir tres *PLS* i es va procedir a obtenir les distribucions posteriors dels diferents paràmetres mitjançant el paquet informàtic *ABCtoolbox* (Wegmann et al. 2010) que té implementat un procediment que es diferencia del proposat per Beaumont et al. (2002) ja que les estimes dels paràmetres es realitzen mitjançant models generals de regressió lineal ABC-GLM (Leuenberger i Wegmann 2009). Es va utilitzar el mateix milió de simulacions que en les anàlisis prèvies d'ABC, amb el mateix llindar de tolerància de 0,001. En la Taula D3.1 es mostren diferents estadístics descriptius en relació a la distribució posterior dels paràmetres d'interès del model M2.

**Taula D3.1.** Estadístics descriptius de la distribució posterior dels paràmetres del model M2, sent aquests la mitjana, mediana i la moda així com també els límits inferior i superior del 95% la distribució, escollint un llindar de tolerància del 0,001 i mitjançant el mètode de models generals de regressió lineal (Leuenerger i Wegmann 2009).

Estadístics descriptius/Paràmetres	$N_{1a}$	$T_{1a}$	$\theta$
<b>límit inferior del 95% de la distribució</b>	0,065	$2,0 \times 10^{-04}$	0,072
<b>Mediana</b>	0,085	0,020	0,090
<b>Mitjana</b>	0,850	0,089	0,098
<b>Moda</b>	0,090	0,019	0,094
<b>límit superior del 95% de la distribució</b>	0,100	0,039	0,100

Els diferents paràmetres són:  $N_{1a}$ , quocient entre la grandària efectiva prèvia a l'expansió i la grandària efectiva actual;  $T_{1a}$ , temps d'expansió poblacional expressat en  $4N_e$  generacions;  $\theta$ , paràmetre poblacional de mutació ( $\theta = 4N_e\mu$ , sent  $N_e$  la grandària efectiva actual i  $\mu$  la taxa de mutació per posició i per generació).

Les distribucions obtingudes mitjançant la metodologia proposada per Wegman i Excoffier (2008) són similars a les distribucions obtingudes en les nostres anàlisis en les quals no es van transformar els estadístics en components *PLS* (veure Taula R5.8). Aquests resultats corroboren primer que el conjunt d'estadístics triat ja seria d'entrada suficient, i segon que la reducció en components *PLS* no aportaria, almenys en el nostre estudi, uns resultats més robusts. I és que a l'efectuar la transformació dels estadístics en components *PLS*, no es redueix significativament la senyal no informativa dels estadístics i d'aquesta manera no només no es milloren les distribucions posteriors dels paràmetres sinó que a més a més es produeix un augment de la dispersió d'aquestes en relació al límit superior de la distribució. Per tant, tot i que la metodologia proposada per Wegmann i Excoffier (2008) proporcioni estimes correctes, no s'observa una millora significativa alhora de reduir la dimensió dels estadístics transformant-los en components *PLS*. Com indica Bertorelle et al. (2010), no s'ha demostrat que l'ús de components *PLS* produeixi estimes més robustes en les metodologies d'ABC.

### D3.3 Detecció d'una regió possiblement afectada per un esdeveniment selectiu recent

En les anàlisis dels nivells i patrons de polimorfisme, la regió p169 va destacar per presentar un defecte de polimorfisme enfront de la divergència (veure Figura R3.3) i un excés de variants derivades a elevada freqüència que es desvien significativament del que s'esperaria en funció del model d'expansió poblacional obtingut en les anàlisis d'ABC efectuades en aquest treball (veure Figura R5.16). Un excés de variants derivades a elevada freqüència és un dels patrons indicatius d'un procés d'escombrada selectiva recent (Fay i Wu 2000). En funció d'aquests resultats, per tant, seria possible considerar que s'ha produït la fixació d'una mutació adaptativa en aquesta regió o al seu voltant.

La regió p169 és homòloga a un dels introns del gen *Zasp67* recentment identificat a *D. melanogaster* per Katzemich et al. (2013). Aquest gen pertany a la família *Alp/Enigma*. Tot i que no es coneix amb claredat la funció de *Zasp67*, la proteïna codificada per aquest gen juntament amb les codificades pels gens *Zasp52* i *Zasp66* poden jugar un paper important en l'assemblatge de la miofibril·la (Katzemich et al. 2013). El gen *Zasp67* està inclòs en un bloc sintènic entre les espècies *D. melanogaster* i *D. pseudoobscura* d'aproximadament 123 Kb. Per tant, la regió p169 podria ser el punt d'inici per contrastar més detalladament la possible acció recent de la selecció positiva en aquesta regió en l'espècie *D. subobscura*, mitjançant l'anàlisi de la variació en regions d'aquest bloc sintènic a distàncies creixents de la regió p169. Afegir que en aquest bloc i únicament a unes 2,5 kb del gen *Zasp67* es troba el clúster format pels gens *ilp1-ilp4* (gens implicats en la via de la insulina). Esmentar, finalment, que Guirao-Rico i Aguadé (2013) observen que a *D. melanogaster* la variabilitat nucleotídica al clúster format pels gens *ilp1-ilp4* podria estar afectada per un possible esdeveniment selectiu recent.

### **D3.4 Història demogràfica de *D. subobscura***

Les poblacions naturals no són sistemes estàtics sinó que es veuen influenciades per les pressions selectives del medi natural que les envolta, pressions que poden deixar la seva empremta en la variabilitat genètica de les poblacions. En la regió Paleàrtica, àrea de distribució de *D. subobscura*, els últims 200.000 anys han estat marcats per grans oscil·lacions climàtiques, classificades en períodes glacials i interglacials. És aquest interval de temps el que es va prendre de referència en l'estudi de la història demogràfica de *D. subobscura* ja que si aquesta es veïés influenciada directament o indirectament per les condicions climàtiques, aquesta successió de períodes freds i càlids podrien haver produït canvis en la seva grandària i aquests canvis podrien veure's reflectits en la variabilitat nucleotídica observada en l'actualitat en les seqüències de DNA.

Les anàlisis d'*Approximate Bayesian Computation* realitzades en aquest estudi han recolzat que dels models considerats, que poden explicar la variabilitat nucleotídica observada en l'actualitat a *D. subobscura*, el més probable és un model d'expansió poblacional. Les estimes dels paràmetres que conformen aquest model indiquen que *D. subobscura* va patir una expansió poblacional ara fa aproximadament 110.000 anys (78.000 a 145.000 anys com a límits inferior i superior del 95% de la distribució) i que la grandària efectiva de la població prèvia a aquesta expansió era d'aproximadament 810.000 individus (500.000 a 890.000 individus com a límits inferior i superior del 95% de la distribució), pràcticament un ordre de magnitud inferior a la grandària efectiva actual, que s'estima en 9,3 milions d'individus (8,4 a 9,6 milions com a límits inferior i superior del 95% de la distribució).

#### **Influència dels canvis climàtics ocorreguts durant els últims 200.000 anys**

El penúltim període glacial (*OIS6*, glaciació de Riss), és considerat com el més vast de tot el Pleistocè (Van Andel i Tzedakis 1996; veure la Figura M3.3). Els marges dels gels de les regions del Nord d'Europa són els més extensos dels últims 200.000 anys i persisteixen durant

un temps més llarg que el de l'últim període glacial (glaciació de Würm). Els registres fòssils de pol·len mostren com durant la primera etapa de l'estadi isotòpic 6 (OIS6) a Europa hi van haver condicions climàtiques moderadament severes amb fluctuacions en l'abundància dels arbres, seguides per condicions més extremes, ara fa aproximadament 150.000 anys (Van Andel i Tzedakis 1996). Es considera que durant els dos períodes glacials ocorreguts durant els últims 200.000 anys (glaciació de Riss i glaciació de Würm) hi va haver tres refugis glacials de biodiversitat a Europa, localitzats en les tres penínsules mediterrànies: Ibèrica, Itàlica i Balcànica (Schmitt et al. 2007). Les anàlisis d'ABC efectuades, però, indiquen com durant el penúltim període glacial l'espècie no va patir una disminució en la seva grandària efectiva. Una de les consideracions que es podria tenir en compte del per què la grandària efectiva no es va veure afectada durant aquest període (OIS6), seria que l'espècie encara no hagués colonitzat les regions del Nord d'Europa. És a dir podria ser que *D. subobscura* no s'hagués vist afectada per la glaciació de Riss ja que en la seva àrea de distribució no s'haguessin donat unes condicions climàtiques tan extremes. I és que tot i que no es coneix amb claredat l'origen exacte de *D. subobscura*, es creu que aquest es localitza en regions del centre d'Àfrica (Krimbas et al. 1992). Seria possible considerar que des d'aquestes regions centrals Africanes paulatinament *D. subobscura* hagués colonitzat les regions del Nord d'Àfrica, com recolza que colonitzés Canàries i Madeira, i posteriorment del Sud d'Europa. De fet, les anàlisis d'ABC efectuades posen de manifest que l'espècie no ha patit una disminució significativa en la seva grandària efectiva al llarg de la seva història demogràfica recent, fet corroborat per la probabilitat tant baixa del model de coll d'ampolla (M3) respecte del model d'expansió (M2) (veure Figura R5.8). Així mateix, els resultats obtinguts tampoc recolzen el possible aïllament en dos o més refugis glacials associat a una reducció poblacional, en ser la probabilitat del model M4 (model d'aïllament i coll d'ampolla amb una posterior barreja de les dues subpoblacions) molt menor que la del model M2 (model d'expansió). Aquests resultats

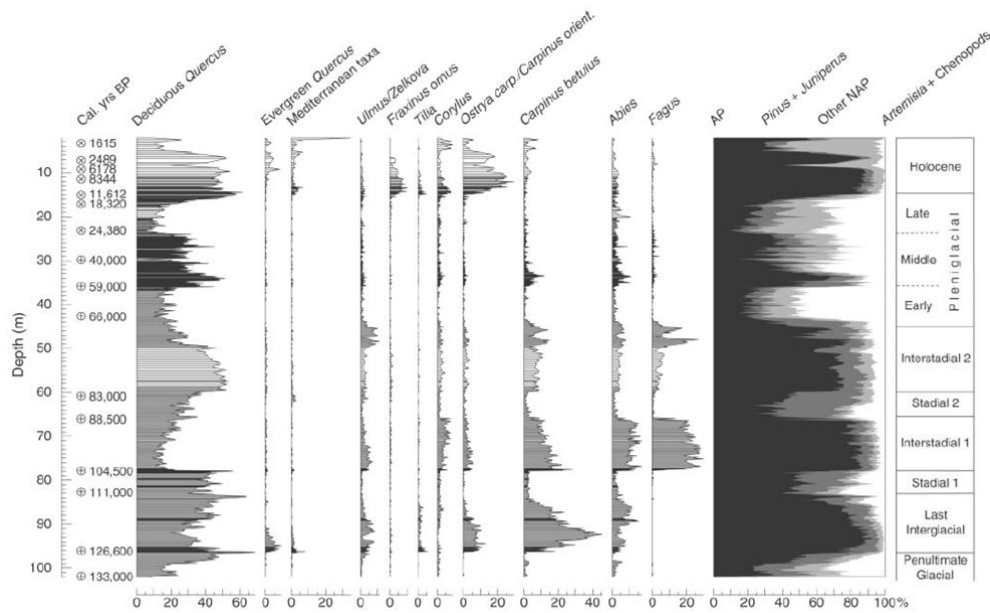
indiquen que tot i que s'hagués produït aquest aïllament, no hauria estat suficientment extrem perquè les poblacions evolucionessin independentment. Tal i com indiquen Rozas et al. (1990), *D. subobscura* es pot considerar una espècie panmíctica, en no detectar-se diferenciació genètica a nivell molecular entre poblacions de punts allunyats de la seva àrea de distribució. I és que a *D. subobscura* no s'observa estructuració geogràfica en relació a la variació present dins de les diferents ordenacions cromosòmiques a Europa (Rozas i Agudé, 1990, 1999; Rozas et al. 1995; Khadem et al. 1998; Pegueroles et al. 2013) ni tampoc entre Europa i regions del Nord d'Àfrica (Munté 1999), havent per tant evidències d'un considerable flux gènic entre les poblacions Europees i també entre les poblacions Europees i les Nord Africanes. Aquest elevat flux gènic, per tant, eradica la possible estructuració poblacional de l'espècie, inclús en el cas que aquesta hagués quedat aïllada en varis refugis glacials en algun moment de la seva història demogràfica recent.

Fa aproximadament 130.000 anys comença un procés ràpid de desglaciació dels gels acumulats durant la glaciació de Riss (Van Andel i Tzedakis 1996; veure la Figura M3.3), donant pas al període interglacial Eemià (OIS5), període que coincideix amb els valors més probables de l'inici de l'expansió poblacional de *D. subobscura*. Quan s'inicia aquest període interglacial a Europa, la vegetació boscosa comença a dispersar-se des dels diferents refugis glacials i diacrònicament colonitza diferents hàbitats, esdevenint el paisatge dominant. La vegetació europea durant aquest període mostra una successió de vegetació que s'inicia amb arbres caducifolis com alzines i oms, sent un dels tipus de boscos on es localitzen les poblacions de *D. subobscura* a la naturalesa. Durant l'interglacial Eemià (OIS5) les temperatures augmenten gradualment fins a 12°C aproximadament, respecte a les temperatures del penúltim període glacial. Durant la primera etapa de l'estadi isotòpic 5 (OIS5e) hi ha un interglacial òptim, ara fa aproximadament 125.000 anys, on les temperatures arriben a valors de 3°C més elevades respecte les temperatures observades en l'actualitat (Kukla et al. 2002), donant pas a densitats

poblacionals boscoses màximes. Les oscil·lacions temporals d'aquest interglacial òptim varien en regions del Nord i Sud d'Europa, amb una duració aproximada en el Nord de fins ara fa aproximadament 115.000 anys, respecte a les regions del Sud on s'estén fins ara fa aproximadament 110.000 anys (Tzedakis 2003). Després d'aquest interglacial òptim, es produeix una reducció progressiva de la biomassa d'arbres, formant-se paisatges boscosos més oberts. No obstant, dades disponibles semblen indicar que ara fa aproximadament 118.000-117.000 anys es produeix una segona expansió dels boscos de *Quercus* (Tzedakis et al. 2003), coincidint amb el valor més probable del temps en què es produí l'expansió de *D. subobscura*. S'ha de remarcar, que hi ha varis estudis tals com els de Hewitt et al. (2004) i Schmitt et al. (2007) que evidencien com els insectes herbívors, com per exemple aquells associats amb boscos d'alzines i pins en àrees mediterrànies, haurien sobreviscut en regions del Sud d'Europa i posteriorment haurien respost a l'augment de temperatura d'un període interglacial amb una ràpida expansió cap a regions del Nord. Per tant, és probable que durant l'interglacial Eemià, *D. subobscura* hagués colonitzat gran part del Nord de la seva àrea de distribució actual a la regió Paleàrtica.

El període interglacial Eemià va tenir una curta duració donant pas gradualment a l'últim període glacial, glaciació de Würm, que s'inicià ara fa aproximadament 74.000 anys. L'últim període glacial està caracteritzat per un pic mitjà de fred OIS4 (ara fa aproximadament 65.000 anys), seguit d'un període càlid parcial OIS3 (ara fa aproximadament 39.000 – 36.000 anys) fins arribar a l'últim màxim glacial ara fa aproximadament 18.000 anys (OIS2) (CLIMAP Project Members, 1984). No obstant, les anàlisis d'ABC indicarien com el darrer període glacial no va conduir a canvis significatius en la grandària poblacional de l'espècie. Hi ha dades del registre fòssil de pol·len que posen de manifest que després del penúltim període glacial (OIS6), l'amplitud de les oscil·lacions climàtiques no va afectar de manera tan significativa a l'abundància dels arbres tal i com s'havia pressuposat anteriorment, és a dir, seria possible

considerar que les poblacions d'arbres haguessin sobreviscut al l'últim període glacial (Tzedakis et al. 2004). En el cas d'espècies de *Quercus* i *Pinus*, Tzedakis et al. (2002) detecten a Grècia pol·len fòssil durant tota l'amplitud de l'últim període glacial, en tant que en d'altres espècies tals com *Abies* i *Ulmus* detecten una presència intermitent (Figura D3.2), sent Grècia part de l'àrea de distribució de l'espècie *D. subobscura* (Krimbas 1964).



**Figura D3.2.** Diagrama temporal d'abundància de pol·len en una profunditat màxima de 102m, de 1-284 seqüències mostrant els percentatges dels taxons seleccionats. Les edats es mostren en anys. Extret de Tzedakis et al. (2002).

Hi ha altres estudis, a més a més, tals com els de Willis et al. (2004), Magri et al. (2006), Varga (2009) i Stewart et al. (2010) que qüestionen la visió tradicional de que l'àrea de distribució dels boscos europeus s'hagués vist reduïda durant l'últim període glacial, en només els tres refugis mediterranis de biodiversitat. S'ha d'afegir que els boscos de *Quercus robur* marquen el límit nord de l'àrea de distribució europea de *D. subobscura* (Krimbas. 1993), per tant si aquest tipus de bosc o altres haguessin sobreviscut a l'últim període glacial, és molt probable que aquestes oscil·lacions climàtiques no haguessin afectat tant abruptament la grandària poblacional de l'espècie, ja que no s'hauria vist eradicat el seu hàbitat natural. Considerant també la pròpia biologia de l'espècie, s'ha de tenir en compte que *D. subobscura* és una espècie generalista que s'alimenta bàsicament de llevats i bacteris que fermenten



matèria orgànica vegetal (Krimbas 1993). Kerdelhué et al. (2009) indiquen com les espècies generalistes (en termes d'hàbitat, hoste o dieta) tenen un risc menys elevat de que desaparegui el seu nínxol ecològic. Per tant, no només seria la presència dels boscos temperats durant la última glaciació (Tzedakis et al. 2002) el que podria haver propiciat el manteniment de la grandària efectiva de *D. subobscura* sinó que també se li hauria de sumar el que és una espècie generalista. S'ha d'afegir que l'elevat polimorfisme per inversions cromosòmiques que presenta *D. subobscura*, moltes d'elles antigues (Munté et al. 2005; Nóbrega et al. 2008) i amb un clar caràcter adaptatiu, permet hipotetitzar que aquest polimorfisme contribueixi de forma important a la possiblement elevada capacitat d'adaptació i resistència als canvis climàtics de l'espècie. El fet que l'espècie ocupi gran part de tota la regió Paleàrtica, en la que es presenta en una gran varietat de condicions climàtiques i d'hàbitats (Monclús, 1964; Huey i Pascual, 2009) podria reflectir aquesta possibilitat.

Clàssicament, quan es parla de grans extincions durant el Quaternari, s'ha de pensar sobretot en els vertebrats com a model i no en espècies com els insectes. Russel et al. (1996) posen de manifest que els fòssils trobats d'insectes durant el Quaternari no permeten inferir grans taxes d'extinció durant aquest període. La constància d'espècies i de comunitats d'espècies podria ser la norma en el cas dels insectes durant almenys l'últim milió d'anys (Russel et al. 1996).

En el cas d'artròpodes, Porretta et al. (2012) analitzen la variabilitat nucleotídica de diferents marcadors mitocondrials i nuclears de diferents poblacions europees de l'espècie *Ixodes ricinus*. A partir dels seus resultats aquests autors consideren que durant l'últim període glacial, l'espècie no va quedar aïllada en refugis glacials separats sinó que es va mantenir en poblacions interconnectades a través del sud i centre d'Europa. Tot i que l'espècie *Ixodes ricinus* és un ectoparàsit i que per tant la seva biologia sigui diferent a la de *D. subobscura*, els resultats de Porretta et al. (2012) estarien en concordança amb els deduïts del nostre estudi.

### **Temps de generació i grandària efectiva**

Arenas et al. (2012) consideren que els nivells de diversitat que es preserven després d'un canvi climàtic dins i entre refugis, no només depenen de les habilitats de dispersió de les espècies sinó també de la velocitat del canvi. Durant un episodi de canvi climàtic, aquest canvi afectarà diferencialment a les espècies en funció del seus temps de generació i de la seva grandària efectiva (Charlesworth 2009; Arenas et al. 2012). La família *Drosophilidae* a la qual pertany *D. subobscura*, es caracteritza per temps de generació curts i elevades grandàries efectives, fet que els hi proporciona un potencial evolutiu elevat (Kellermann et al. 2012). Els resultats d'aquest treball indiquen que *D. subobscura* té una grandària efectiva molt elevada, d'aproximadament 9,3 milions d'individus, valor que està en concordança amb l'estima realitzada per Comerón (1997), que era de 10 milions d'individus. La grandària efectiva de *D. subobscura* es considera, per tant, similar a la grandària efectiva de les poblacions africanes de *D. melanogaster* que s'estima aproximadament en 8,6 milions d'individus (Li i Stephan 2006). El nombre de generacions per any a *D. subobscura*, a l'igual que a *D. pseudoobscura*, es considera que és de cinc generacions (Mestres et al. 2001), havent menys generacions per any respecte a altres espècies de la família *Drosophilidae* tals com *D. melanogaster* i *D. simulans* en que s'estimen 10 generacions per any (Sharp i Li 1989).

### **Termotolerància**

Les anàlisis d'ABC indiquen que durant el penúltim període glacial (OIS6) *D. subobscura* no hauria patit canvis en la seva grandària efectiva. Com s'ha comentat anteriorment, això podria ser degut a que l'espècie encara no hagués colonitzat les regions del Nord d'Europa afectades per unes condicions climàtiques més adverses, respecte a regions del Sud d'Europa o Nord d'Àfrica. Posteriorment, durant l'interglacial Eemià (OIS5), l'espècie hauria patit una expansió poblacional. Seria lògic pensar que en aquell període s'iniciés el canvi d'una sèrie de característiques climàtiques i ecològiques idònies per a l'espècie, podent ser la temperatura un

dels factors determinants. I és que les condicions climàtiques determinen els límits geogràfics per a moltes espècies, sent la temperatura una de les variables més importants que en delimiten la seva distribució i abundància (Parmesan, 2006). Durant el penúltim període interglacial (interglacial Eemià) la temperatura mitjana va pujar fins a 3<sup>o</sup> C respecte la temperatura actual. Aquest increment de la temperatura, podria haver facilitat unes condicions idònies per l'espècie fet que podria haver propiciat aquest procés d'expansió i també de dispersió en la regió Paleàrtica. Com a exemple, podríem posar la fertilitat de les femelles a *D. subobscura*. Buzzati-Traverso (1942) descriu la temperatura òptima per a la fertilitat de les femelles de *D. subobscura* en condicions de laboratori en 19-20<sup>o</sup>C, sent les temperatures que es podrien haver trobat durant l'interglacial òptim de l'OIS5. Sota aquestes temperatures el nombre d'ous que pot posar una femella és més gran (Clarke i Maynard Smith, 1961). A més a més, Walters et al. (2006) construeixen un model de predicció de com poden afectar els canvis climàtics a la dispersió dels insectes, arribant a la conclusió que l'augment de 3 a 5<sup>o</sup>C en la temperatura màxima del dia, pot incrementar la proporció de dispersions a gran distància, tal i com podria haver succeït a *D. subobscura* durant el període OIS5. Després del penúltim interglacial, els resultats obtinguts en el nostre treball indicarien que les fluctuacions climàtiques no van ser suficientment extremes per l'espècie perquè es produís una disminució considerable de la seva grandària efectiva, ni tampoc extremadament favorables perquè es produís un augment significatiu d'aquesta grandària efectiva. La termotolerància de l'espècie podria ser un factor determinant del per què la grandària efectiva de *D. subobscura* no s'hauria vist afectada durant l'últim període glacial (OIS4, OIS3 i OIS2). Estudis efectuats en condicions de laboratori demostren que *D. subobscura* té una termotolerància que avarca un rang de 3<sup>o</sup>C a 37<sup>o</sup>C (Krimbas 1993). Kellermann et al. (2012) realitzen un estudi amb 95 espècies de *drosofilids* i observen que *D. subobscura*, respecte a altres *drosofilids* és una espècie amb una elevada resistència al fred, fet que podria explicar que la seva grandària efectiva no s'hagués

vist tant afectada durant l'últim període glacial ni tan sols durant el seu pic màxim de fred ara fa aproximadament 20.000 anys (OIS2).

Finalment, s'ha d'afegir que els models demogràfics dissenyats són models senzills capaços d'explicar la variabilitat nucleotídica observada a *D. subobscura* però que no necessàriament descriuen el que realment ha succeït a l'espècie al llarg de la seva història evolutiva. No obstant, la disponibilitat d'un model demogràfic, ha permès i permetrà en estudis posteriors, la detecció de regions la variabilitat de les quals no pugui ser explicada per un model d'expansió poblacional i que siguin per tant candidates a haver estat dianes recents de la selecció adaptativa.

#### D4. ANÀLISI DE LA VARIACIÓ EN REGIONS NO CODIFICADORES AFECTADES PER LA INVERSIÓ

##### J<sub>1</sub> DE *D. SUBOBSCURA*

Des de que al 1917 Sturtevant descobreix l'existència de les inversions cromosòmiques, s'han realitzat multitud d'estudis amb l'objectiu de desentrellar quins mecanismes evolutius són els responsables de l'increment de freqüència i el manteniment de les inversions cromosòmiques i com la seva presència afecta a la variabilitat genètica de les poblacions. *D. subobscura* és una espècie ideal per aquests estudis ja que es caracteritza per presentar un elevat polimorfisme per inversions cromosòmiques. No tots els seus cromosomes presenten un nombre similar d'inversions, sent el cromosoma J el que presenta un menor polimorfisme per inversions cromosòmiques, si exceptuem el cromosoma puntiforme. En aquesta part del treball es discuteixen els resultats obtinguts de l'anàlisi de sis regions no codificadores distribuïdes al llarg de la inversió J<sub>1</sub> en 27 línies homocariotípiques obtingudes a partir d'una població natural de *D. subobscura* de Barcelona que es diferencien per la presència o absència de la inversió J<sub>1</sub>. Es van analitzar per separat les línies amb l'ordenació J<sub>st</sub> de les línies amb l'ordenació J<sub>1</sub>. La inversió J<sub>1</sub> és una inversió paracèntrica simple (és a dir, només hi ha un segment invertit), té una grandària mitjana-petita, segrega a freqüències relativament elevades en la població d'estudi i en l'àrea paleàrtica és una de les dues ordenacions majoritàries juntament amb l'ordenació J<sub>st</sub> (Krimbas i Powell 1992).

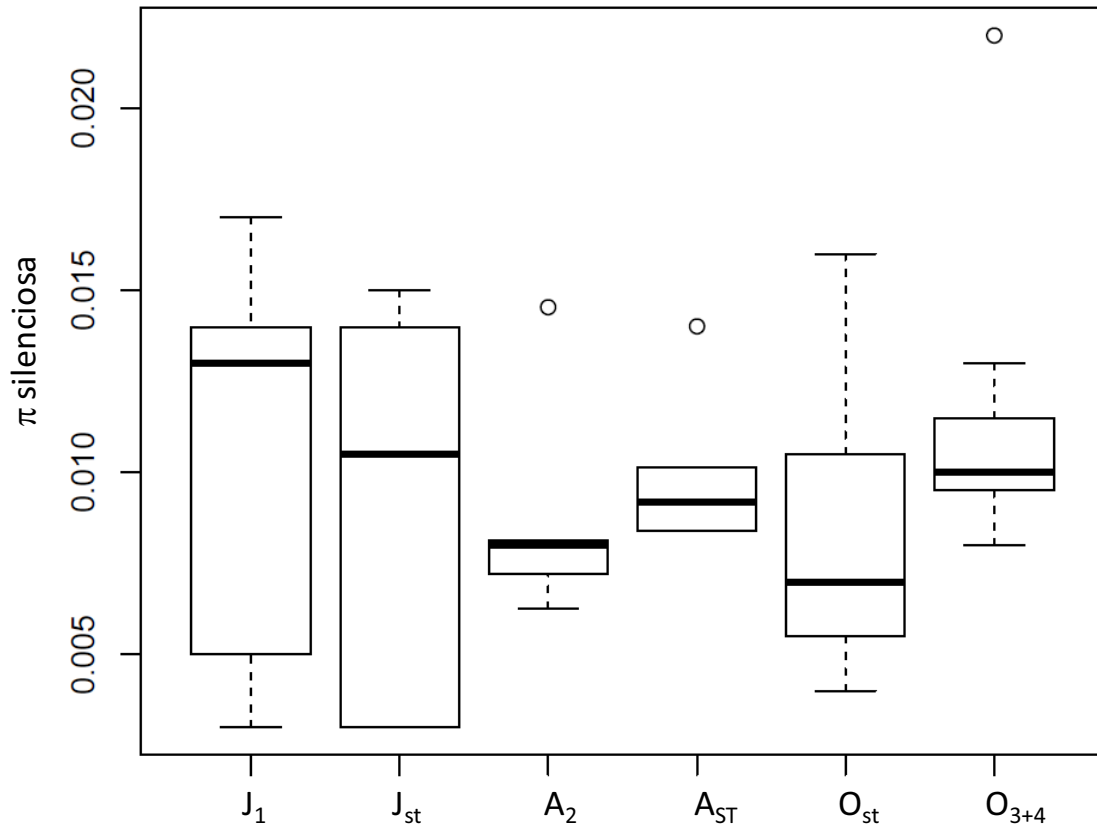
##### D4.1 Nivells i patrons de variació

La disponibilitat d'estudis previs a nivell multilocus, en els quals s'analitzen els nivells i patrons de polimorfisme en diferents ordenacions cromosòmiques de l'espècie *D. subobscura*, permet comparar els resultats del present treball amb els obtinguts per d'altres inversions polimòrfiques de l'espècie. Concretament es prenen de referència els treballs de Nóbrega et al. (2008) i Munté et al. (2005). Nóbrega et al. (2008) analitzen cinc regions localitzades al llarg de la inversió A<sub>2</sub> en 18 línies homocariotípiques provinents d'una població natural de *D.*

*subobscura* que es diferencien per la presència o absència d'aquesta inversió: 12 d'elles presenten l'ordenació  $A_2$  i 6 l'ordenació  $A_{st}$  que té una longitud aproximada de 41,3 cM segons estimen els autors. El treball de Nóbrega et al. (2008), esdevé adient per ser comparat amb el present treball ja que la inversió  $A_2$ , igual que la  $J_1$ , és una inversió paracèntrica simple. Munté et al. (2005) analitzen la variació nucleotídica en 8 regions, sis d'elles distribuïdes al llarg de la inversió  $O_3$  i les dues restants localitzades fora de la inversió però molt properes als punts de trencament, en 28 línies homocariotípiques d'una població natural de *D. subobscura*: 14 d'elles presenten l'ordenació  $O_{st}$  i les 14 restants l'ordenació  $O_{3+4}$ . Les ordenacions  $O_{st}$  i  $O_{3+4}$  es diferencien per dues inversions imbricades (inversió 3 i 4) que sorgeixen independentment de l'ordenació ancestral  $O_3$  extingida en l'actualitat (Ramos-Onsins et al. 1998). Els autors estimen la longitud de la inversió  $O_3$  en 27,4 cM, sent una inversió més petita que la  $J_1$  que té una longitud aproximada de 30 cM. Comparar el sistema cromosòmic  $O_{st}/O_{3+4}$  no és tan adequat en relació al nostre estudi degut a que l'ordenació  $O_{3+4}$  està formada per dues inversions imbricades.

La diversitat nucleotídica dels tres sistemes cromosòmics (Figura D4.1) es compara utilitzant com a referència en el cas dels estudis de Munté et al. (2005) i Nóbrega et al. (2008) la diversitat nucleotídica silenciosa ja que les regions en estudi del nostre treball són en tots els casos no codificadores. Els valors de diversitat nucleotídica silenciosa obtinguts per Nóbrega et al. (2008) es multipliquen per 4/3 ja que les diferents regions analitzades es troben localitzades al cromosoma X i d'aquesta manera poden ser comparades amb les altres regions localitzades en els autosomes. Es va aplicar el test no paramètric de *Kruskal-Wallis* per observar si existeixen diferències significatives entre les mitjanes dels valors de diversitat nucleotídica de les diferents ordenacions, no observant-se en cap cas diferències significatives (test no paramètric de *Kruskal-Wallis*;  $P = 0,19$ ). Sinó es corregeixen les regions incloses dintre del

cromosoma X, els resultats del test no paramètric de *Kruskal-Wallis* continuen essent no significatius (test no paramètric de *Kruskal-Wallis*;  $P=0,14$ ).



**Figura D4.1.** Diagrama de caixa corresponent a la diversitat nucleotídica silenciosa (Nei 1983) en relació a diferents ordenacions cromosòmiques en l'espècie *D. subobscura*. Els nivells de diversitat nucleotídica són extrems del nostre estudi, de Nóbrega et al. (2008) i Munté et al. (2005) en relació a les ordenacions  $J_1/J_{st}$ ,  $A_{st}/A_2$  i  $O_{st}/O_{3+4}$ , respectivament.

El criteri més acceptat alhora d'establir quines ordenacions cromosòmiques són les més primitives és aquell que té en compte la possible coexistència d'una ordenació en més d'una espècie, fet que provaria, per tant, que aquesta ordenació s'hagués originat abans de la separació d'ambdues espècies. S'observa com les espècies *D. guanche* i *D. madeirensis* que formen juntament amb *D. subobscura* el subgrup *subobscura* presenten tant pel cromosoma J com pel cromosoma A l'ordenació estàndard o una ordenació derivada de l'estàndard, fet que indicaria que  $J_{st}$  i  $A_{st}$  són les ancestrals. S'ha d'afegir, que generalment es considera que les ordenacions ancestrals haurien de presentar uns nivells de variació superiors respecte a les derivades, ja que s'espera que hauran tingut més temps per acumular variació. Aquesta

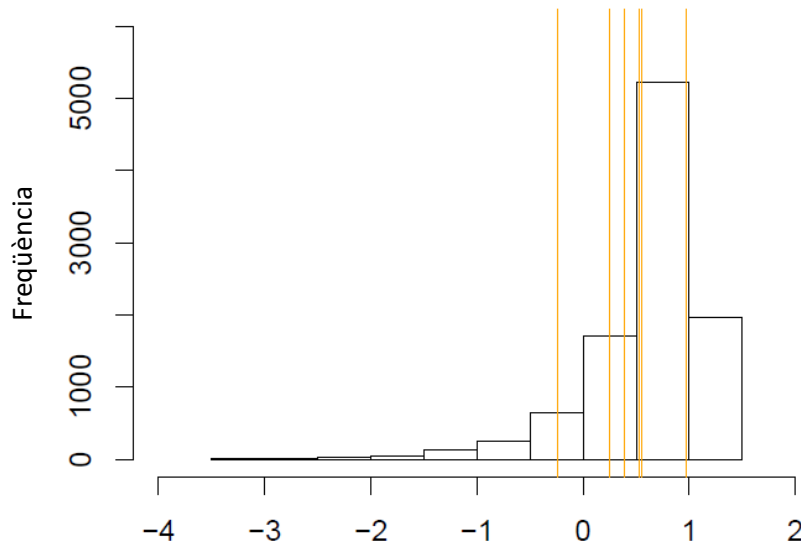
premissa es compleix en el cas del sistema cromosòmic  $A_{st}/A_2$ , on s'observa com l'ordenació  $A_{st}$ , considerada com a ancestral, presenta uns nivells de variació superiors a l'ordenació  $A_2$ , considerada com a derivada. En el cas del nostre treball, però, quan es comparen les dues ordenacions cromosòmiques, l'ordenació  $J_1$  presenta uns nivells de diversitat nucleotídica superiors als de l'ordenació  $J_{st}$ . Aquest resultat, a més a més, està en concordança amb un estudi previ efectuat per Simões et al. (2012), en el qual analitzen el sistema cromosòmic  $J_{st}/J_1$  a través de l'estudi de 4 marcadors microsatèl·lits, dos localitzats dintre de la inversió  $J_1$  i altres dos localitzats fora de la inversió però propers als punts de trencament. Per tant, malgrat s'hagi proposat que les ordenacions que presenten uns nivells de variació superiors són les més primitives, aquesta proposició no es pot confirmar en el cas de les ordenacions  $J_{st}/J_1$  degut a que l'ordenació  $J_{st}$  s'hauria de considerar com l'ancestral en funció de les ordenacions compartides amb espècies properes.

Per tant, s'hauria de pensar per què l'ordenació considerada com a ancestral presenta uns nivells de variació inferiors a la considerada com a derivada. Una de les possibles explicacions d'aquesta observació, i que es considera la més plausible, està relacionada amb la grandària efectiva de les dues ordenacions. Si l'ordenació  $J_{st}$  tingués una grandària efectiva considerablement inferior que l'ordenació  $J_1$ , el nivell de variació neutre que s'esperaria a l'equilibri deriva-mutació seria inferior i el nombre de generacions que haurien de transcorre per arribar a aquest equilibri també seria inferior. Una altra explicació que es podria considerar seria, per exemple, que aquests nivells de variació inferiors en l'ordenació  $J_{st}$  fossin deguts a factors demogràfics, tal i com podria succeir en el cas que s'hagués produït un coll d'ampolla en els individus portadors de l'ordenació  $J_{st}$ . Aquest coll d'ampolla, però, no podria ser recent, ja que els nivells de variació observats en les regions incloses dintre de la inversió  $J_1$  són relativament alts i, per tant, si hagués ocorregut un coll d'ampolla hauria passat un temps suficient per a que s'hagués pogut recuperar la variació. Per altra banda, la possibilitat d'un



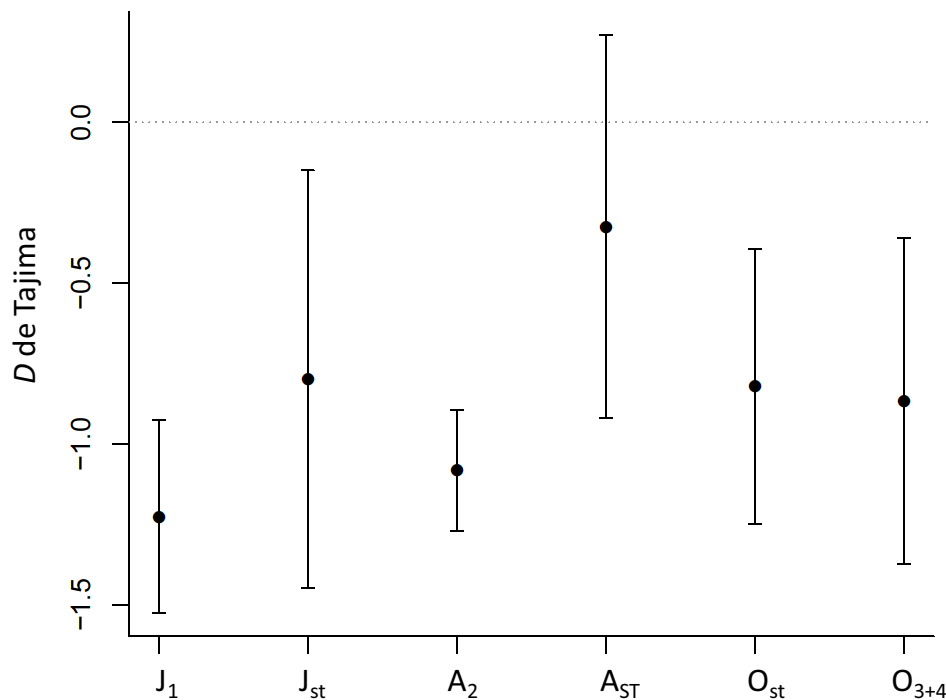
coll d'ampolla seria poc probable ja que si s'hagués produït aquest fenomen demogràfic, seria molt estrany que aquest hagués afectat a l'ordenació  $J_{st}$  i no a l'ordenació  $J_1$ . Una explicació alternativa que es podria considerar per interpretar els nivells de variació inferiors en els individus portadors de l'ordenació  $J_{st}$ , seria que en aquesta ordenació s'hagués produït un arrossegament selectiu que hagués eliminat variació prèvia. De fet, l'evolució adaptativa és un factor que pot distorsionar el nivell de polimorfisme neutre en les regions estretament lligades a les posicions sota selecció causant una reducció del polimorfisme. No obstant, aquesta explicació seria també poc probable ja que, a l'analitzar els nivells i patrons de variació a nivell multilocus en les diferents regions localitzades dintre de la inversió  $J_1$ , no s'observen en cap de les regions en estudi senyals indicatives de possibles esdeveniments selectius recents. Concretament, si s'analitza l'estadístic  $H$  normalitzat de Fay i Wu (2000) que és sensible per detectar variants derivades a elevada freqüència i per tant podria ser un estadístic sensible per detectar l'acció de la selecció positiva, no s'observa en cap de les regions en estudi valors d'aquest estadístic que es desviïn significativament del que s'esperaria sota un model d'evolució neutre estàndard. Per corroborar aquests resultats, a més a més, es va decidir obtenir la distribució empírica de l'estadístic  $H$  normalitzat de Fay i Wu a partir de simulacions de coalescència realitzades sota el model desenvolupat per inferir l'edat més probable de la inversió  $J_1$ . En les regions en estudi a l'ordenació  $J_{st}$ , els valors d'aquest estadístic no es desvien en cap cas significativament del que s'esperaria sota aquest model demogràfic (Figura D4.2). Per tant, aquests resultats indiquen que els nivells de variació inferiors en l'ordenació  $J_{st}$  no han estat produïts per esdeveniments de selecció positiva almenys relativament recents. Per tant, malgrat no poder afirmar rotundament en el cas del sistema cromosòmic  $J_{st}/J_1$  quina de les inversions seria l'ancestral i quina la derivada, s'ha considerat que l'ordenació  $J_{st}$  és l'ancestral degut a la seva presència a *D. madeirensis* i a la presència d'una ordenació derivada

de  $J_{st}$  a  $D. guanche$ , explicant els nivells de variació inferiors d'aquesta ordenació com una conseqüència de la seva grandària efectiva menor respecte l'ordenació  $J_1$ .



**Figura D4.2.** Distribució empírica de l'estadístic  $H$  normalitzat de Fay i Wu (2000) en l'ordenació  $J_{st}$ . Aquesta distribució es construeix després de realitzar 10.000 noves simulacions utilitzant com a distribucions *a priori* els paràmetres retinguts de l'anàlisi d'ABC efectuada per inferir l'edat més probable de la inversió  $J_1$  (veure Resultats R5.3). Les línies verticals de color taronja indiquen els valors observats de l'estadístic  $H$  normalitzat de Fay i Wu en les sis regions en estudi a l'ordenació  $J_{st}$ .

Els patrons de polimorfisme reflectits en l'estadístic  $D$  de Tajima (1989) d'aquest treball també s'han comparat amb els dels treballs realitzats per Munté et al. (2005) i Nóbrega et al. (2008) (Figura D4.3). S'observa com l'ordenació  $A_{st}$ , considerada com l'ordenació ancestral, presenta valors de l'estadístic  $D$  de Tajima més propers a zero respecte l'ordenació derivada  $A_2$ , sent aquest el patró que s'esperaria. En el cas del nostre estudi, els valors de  $D$  de Tajima de l'ordenació  $J_{st}$  tendeixen a ser més propers a zero respecte a l'ordenació  $J_1$ . Aquests valors més propers a l'equilibri deriva-mutació de l'ordenació  $J_{st}$ , respecte a l'ordenació  $J_1$ , podrien corroborar el seu caràcter ancestral i que la grandària efectiva de l'ordenació  $J_{st}$  fos inferior a la de l'ordenació  $J_1$ . No obstant aquestes consideracions, però, no es pot obviar l'elevada variància que presenten els valors de l'estadístic  $D$  de Tajima en les diferents ordenacions.



**Figura D4.3.** Mitjana de l'estadístic  $D$  de Tajima (1989) en relació a diferents ordenacions cromosòmiques en l'espècie *D. subobscura*. Els valors de l'estadístic són extrets de l'article de Nóbrega et al. (2008) i Munté et al. (2005), en relació a les ordenacions  $A_{st}/A_2$  i  $O_{st}/O_{3+4}$ , respectivament. En la figura es mostra la mitjana de l'estadístic  $D$  de Tajima en les diferents ordenacions així com també la seva desviació estàndard. La línia discontinua és el valor que s'esperaria de l'estadístic sota condicions de neutralitat.

En comparar els tres sistemes cromosòmics, s'observa com l'ordenació  $J_1$  és la que presenta uns valors negatius de l'estadístic  $D$  de Tajima més elevats, és a dir és la que presenta un excés més marcat de polimorfismes amb variants que segreguen a baixa freqüència. No obstant això, en els tres estudis s'observen uns patrons de variació caracteritzats per un excés de polimorfismes amb variants a baixa freqüència, tal i com indiquen els valors negatius de l'estadístic  $D$  de Tajima. Aquests resultats denoten que els tres sistemes cromosòmics presenten un espectre de freqüències que es desvia d'acord al que s'esperaria sota les prediccions del model neutre estàndard. Seria congruent considerar que molt possiblement sigui la història demogràfica recent de l'espècie, caracteritzada per una expansió poblacional, la que determini gran part dels patrons de variació observats en les regions localitzades dintre de les inversions cromosòmiques. Finalment, remarcar que en el cas de les inversions

cromosòmiques en l'espècie *D. subobscura*, s'haurà d'anar amb cura alhora d'interpretar els estudis dels nivells i patrons de variabilitat genètica. Observar un excés de polimorfismes amb variants a baixa freqüència, tests de neutralitat que es desvien significativament de l'esperat en condicions de neutralitat i genealogies amb forma d'estrella, no tindria per què reflectir un procés d'escombrada selectiva parcial associat a la formació d'una inversió, sinó que podria indicar l'efecte que la història demogràfica recent té sobre la variabilitat genètica de les poblacions naturals d'aquesta espècie.

#### **D4.2 Diferenciació genètica entre ordenacions**

Les inversions cromosòmiques són canvis estructurals dels cromosomes que modifiquen la recombinació, ja que aquesta es veu fortament suprimida en els individus heterocariotips, existint en el cas de *D. subobscura* una gran quantitat d'estudis a nivell molecular que ho corroboren (Rozas et al. 1999; Navarro-Sabaté et al. 2003; Munté et al. 2005; Nóbrega et al. 2008; Simões et al. 2012; Pegueroles et al. 2013). Degut a la baixa probabilitat de que es produeixin dobles entrecreuaments en les regions properes als punts de trencament (Navarro et al. 1997), la supressió de la recombinació no té per què ser igual al llarg de tota la regió afectada per la inversió. És aquesta elevada supressió de l'intercanvi genètic en les regions més properes als punts de trencament, el que faria esperar uns nivells de variació menors en aquestes regions, respecte a les regions més centrals de la inversió, ja que en elles la mutació seria pràcticament l'únic mecanisme generador de nova variació dins d'ordenació. Els resultats obtinguts en el nostre estudi segueixen aquesta predicció doncs els nivells de diversitat nucleotídica de les dues regions més properes als punts de trencament són menors respecte a les regions localitzades en punts més centrals de la inversió  $J_1$  (veure Taula R4.1). En comparar totes les regions en estudi localitzades al llarg del cromosoma J, s'observa com els valors de diversitat nucleotídica continuen sent generalment menors en les dues regions més properes als punts de trencament (veure Figura R4.4). Aquests nivells de diversitat

nucleotídica inferiors en les dues regions més properes als punts de trencament, sumats a que aquestes regions siguin les que presenten uns nivells més elevats de desequilibri de lligament i de diferenciació genètica entre ordenacions (veure Figures R4.6 i R4.9), corroboren la forta supressió de la recombinació en regions properes als punts de trencament.

La supressió de la recombinació en els individus heterocariotips actua, per tant, produint un fort bloqueig de l'intercanvi genètic entre les diferents ordenacions cromosòmiques. S'esperaria que la recuperació de la variació després del procés d'escombrada selectiva parcial associat a la formació i increment de freqüència d'una inversió, es veiés reflectit per una elevada diferenciació genètica entre les diferents ordenacions cromosòmiques, sempre i quan hagués passat un temps suficient des de la formació de la inversió per a què s'hagués pogut acumular variació diferencialment. Els resultats obtinguts en el present treball mostren com totes les regions localitzades dintre de la inversió  $J_1$  presenten una elevada diferenciació genètica entre ordenacions, estimada a partir de l'estadístic  $F_{st}$  (Hudson 1992) (veure Figura R4.6). Aquests resultats estan en concordança amb els observats en estudis previs en l'espècie *D. subobscura* on també s'observen uns elevats nivells de diferenciació genètica entre les diferents ordenacions cromosòmiques en estudi (Rozas et al. 1990; Munté et al. 2005; Nóbrega et al. 2008). Per tant, la forta supressió de la recombinació en les inversions cromosòmiques produeix una elevada estructuració de la variació genètica entre les diferents ordenacions en les regions invertides. Estudis previs descriuen, a més a més, que aquesta estructuració es podria estendre no només al llarg de tota la regió afectada per la inversió, sinó també en regions localitzades fora de la inversió però properes a aquesta. Així, en estudis efectuats en les espècies *D. pseudoobscura* i *D. persimilis*, es descriu com la supressió de la recombinació es pot estendre en regions localitzades fora de la inversió fins a 2,5 Mb (Machado et al. 2007; Kulathinal et al. 2009; McGaugh et al. 2012). Corbett-Detig et al. (2012) efectuen un estudi en l'espècie *D. melanogaster* i detecten que aquest efecte es pot

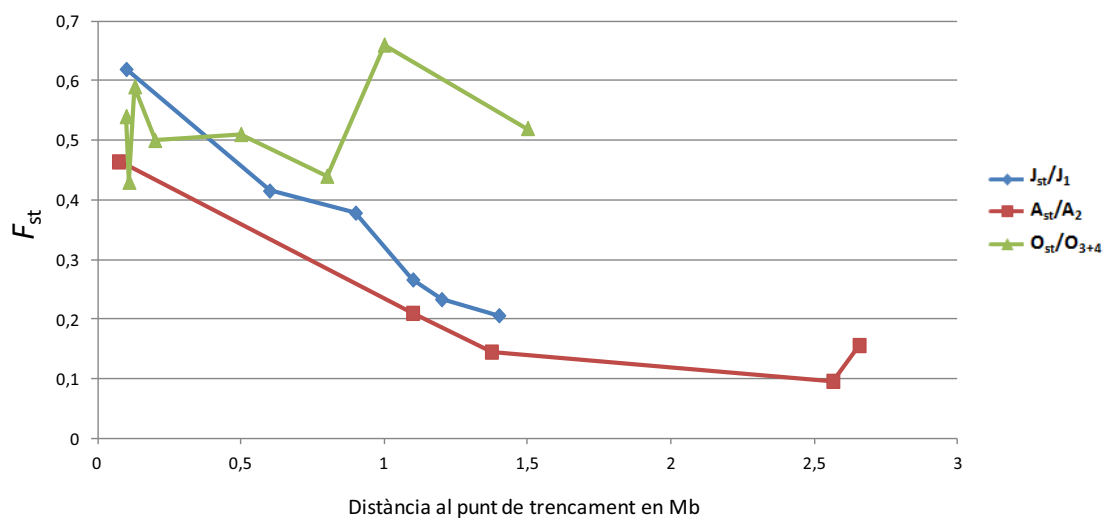
estendre fins a 4 Mb. Aquestes estimes són superiors a les observades en el nostre estudi, on les dues regions localitzades fora de la inversió però properes a un dels punts de trencament, aproximadament a 0,9 Mb i 1,4 Mb, ja no presenten una diferenciació genètica significativa entre ordenacions (veure Taula R4.4). Per tant, l'elevada presència d'inversions cromosòmiques en l'espècie *D. subobscura* podria produir una forta estructuració del seu genoma tot i que l'efecte produït per aquestes inversions, en el cas del nostre estudi, no semblaria que abracés una longitud molt distant fora de la inversió.

No obstant s'hagi observat aquest fort bloqueig de la recombinació en les inversions cromosòmiques, pot existir un cert grau d'intercanvi genètic entre ordenacions produït tant per dobles entrecreuaments com per conversió gènica. En el nostre estudi, precisament, els resultats obtinguts mostren com la diferenciació genètica no és homogènia al llarg de tota la inversió  $J_1$ , observant una major diferenciació com més a la vora dels punts de trencament es troben les diferents regions en estudi. Estudis teòrics prediuen que mentre l'intercanvi genètic produït per conversió gènica pot donar-se homogèniament al llarg de tota la inversió, els dobles entrecreuaments seran més probables en regions centrals de la inversió ja que aquests es troben fortament inhibits en regions properes als punts de trencament (Navarro et al. 1997, 2000). A l'estudiar el sistema cromosòmic  $J_{st}/J_1$  i aplicar l'algoritme desenvolupat per Bertrán et al. (1997), es van detectar tractes d'intercanvi genètic només en les regions centrals de la inversió  $J_1$  (veure Taula R4.5). Estudis teòrics (Navarro et al. 1997, 2000) prediuen, a més a més, que els dobles entrecreuaments només seran possibles en inversions cromosòmiques amb una longitud superior als 20 cM. Les estimes efectuades en el nostre treball, indiquen que la longitud de la inversió  $J_1$  és d'aproximadament 30 cM, fet que fa considerar que els dobles entrecreuaments siguin probables. Aquesta estima de la longitud en cM té en compte les dades recopilades per Cáceres et al. (1999) del mapa de lligament per tres elements cromosòmics i la longitud relativa de la inversió  $J_1$  en el mapa citològic. Per tant, és tan sols una

estima aproximada i és considerablement superior a la inferida indirectament per Loukas et al. (1979). S'ha d'afegir que les dues regions localitzades a la vora d'un dels punts de trencament no presenten polimorfismes compartits entre les ordenacions i sí polimorfismes fixats, sent el contrari del que s'observa en les regions més centrals de la inversió (veure Taula R4.3). Aquests polimorfismes compartits entre les ordenacions cromosòmiques en les regions més centrals de la inversió corroborarien l'existència d'intercanvi genètic en aquestes regions. Aquest intercanvi genètic podria ser degut principalment a dobles entrecreuaments malgrat que la conversió gènica també hi podria contribuir. Al contrari, prop dels punts de trencament només la conversió gènica permetria un intercanvi genètic entre ordenacions. De fet, un estudi efectuat recentment per Papaceit et al. (2013) identifiquen petits tractes de conversió gènica en les zones immediatament adjacents als mateixos punts de trencament de la inversió  $O_3$ . Aquest estudi reflectiria que l'intercanvi genètic per conversió gènica és possible inclús en els mateixos punts de trencament i que per tant també seria d'esperar en les regions properes.

Es van comparar els valors de diferenciació genètica d'aquest treball amb els valors obtinguts en els treballs de Nóbrega et al. (2008) i Munté et al. (2005) (Figura D4.4) que estudien el sistema cromosòmic  $A_{st}/A_2$  i  $O_{st}/O_{3+4}$ , respectivament. En els sistemes cromosòmics  $J_{st}/J_1$  i  $A_{st}/A_2$ , s'observa una disminució de la diferenciació genètica a l'allunyar-se del punt de trencament més proper. No obstant, quan es comparen aquests dos sistemes cromosòmics els valors de diferenciació genètica observats en les regions més centrals de la inversió  $A_2$  són inferiors respecte als de les regions centrals de la inversió  $J_1$ . Com que la inversió  $A_2$  té una longitud més gran respecte a la inversió  $J_1$ , la probabilitat de que es produeixin dobles entrecreuaments serà major i per tant és d'esperar que els nivells de diferenciació genètica d'aquesta inversió siguin menys elevats respecte a la inversió  $J_1$ . Per contra, en l'estudi de Munté et al. (2005) els quals analitzen el sistema cromosòmic  $O_{st}/O_{3+4}$ , observen uns nivells de diferenciació genètica distribuïts bastant homogèniament al llarg de tota la inversió  $O_3$ .

Aquests nivells de diferenciació genètica homogenis podrien ser deguts a que la inversió  $O_3$  té una longitud menor (27,4cM) respecte a la  $J_1$  o a la  $A_2$ , sent per tant, una inversió relativament petita per a què es puguin produir dobles entrecreuaments. No s'ha d'obviar però, que en l'estudi de Munté et al. (2005) la inversió  $O_3$  es troba superposada amb la inversió  $O_4$  formant el complex  $O_{3+4}$ . Aquesta doble nansa d'inversió en els heterocariotips  $O_{st}/O_{3+4}$  podria també dificultar la possibilitat de que es produeixin dobles entrecreuaments. I és que la presència d'inversions imbricades podria conduir a una major supressió dels dobles entrecreuaments respecte a altres inversions simples tals com la  $A_2$  o la  $J_1$ . Alternativament, si els dobles entrecreuaments es produïssin, aquesta diferenciació genètica relativament homogènia podria ser un efecte de l'acció de la selecció epistàtica operant en contra dels recombinants, podent ser un factor indicatiu de la presència de complexes gènics coadaptats en el manteniment d'aquestes ordenacions.



**Figura D4.4.** Valors de diferenciació genètica obtinguts en tres inversions cromosòmiques de l'espècie *D. subobscura*, en funció de la distància al punt de trencament de cada inversió. Les dades en relació al sistema cromosòmic  $A_{st}/A_2$  són obtingudes de Nóbrega et al. (2008) i les del sistema  $O_{st}/O_{3+4}$  de Munté et al. (2005).

En contrast amb els resultats obtinguts per Munté et al. (2005), en el cas de la inversió  $J_1$  la detecció d'intercanvi genètic en regions centrals de la inversió podria fer considerar que la hipòtesi de coadaptació de complexes poligènics no tindria per què ser la que millor explica el



manteniment d'aquesta inversió en poblacions naturals. Si la selecció estigués actuant en contra dels recombinants, s'hauria d'esperar observar valors de  $F_{st}$  elevats i distribuïts homogèniament al llarg de la inversió, o almenys detectar en alguna de les regions centrals de la inversió valors de  $F_{st}$  que haurien de ser més similars als observats en les regions més properes als punts de trencament. Per tant, es podria considerar en el cas de la inversió  $J_1$  que el manteniment d'aquesta inversió en poblacions naturals s'aproximés més a un mecanisme d'adaptació local tal i com proposen Kirkpatrick i Barton (2006). Aquests autors consideren que les inversions es troben afavorides en les poblacions naturals ja que la reducció de la recombinació en els heterocariotips manté una sèrie d'al·lels lligats a la inversió i localment adaptats, no sent necessàries les relacions epistàtiques entre ells.

#### **D4.3 Estimació de l'edat de la inversió $J_1$**

Està altament acceptat el caràcter monofilètic de les inversions cromosòmiques (Powell 1997), ja que seria molt improbable que en la formació d'una inversió cromosòmica es produïssin més d'una vegada exactament els mateixos trencaments involucrant la mateixa part del cromosoma. Concretament en el nostre estudi, els resultats obtinguts de la genealogia del concatenat de les sis regions incloses dintre de la inversió  $J_1$  seria consistent amb aquest caràcter monofilètic de les inversions cromosòmiques (veure Figura R4.10). A més a més, tot i l'intercanvi genètic detectat en regions centrals de la inversió  $J_1$ , aquest no és suficientment elevat per eradicar la forta diferenciació genètica entre les dues ordenacions tal i com denoten els dos clústers amb un elevat valor de *bootstrap* representats per les línies  $J_{st}$  i  $J_1$ . Cal remarcar que aquesta genealogia seria consistent amb que la inversió  $J_1$  fos una inversió antiga. Són precisament les anàlisis efectuades per establir l'edat més probable de la inversió  $J_1$  el que es discuteix en aquesta part del treball. Per realitzar aquestes anàlisis es va utilitzar el concatenat de les dues regions incloses dintre de la inversió  $J_1$  més properes als punts de trencament, sent aquestes les regions p127 i p222.

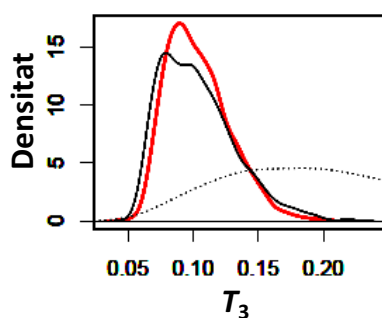
Com ja s'ha comentat anteriorment, la supressió de la recombinació dintre de les inversions cromosòmiques en individus heterocariotípics es veurà incrementada en regions properes als punts de trencament, podent ser informatives aquestes regions sobre l'origen i la història demogràfica d'una determinada inversió. Les anàlisis dutes a terme en les regions p222 i p127 van mostrar com aquestes són les úniques regions incloses dintre de la inversió  $J_1$  les quals no presenten cap polimorfisme compartit entre les dues ordenacions cromosòmiques  $J_{st}$  i  $J_1$ , així com tampoc es van identificar tractes d'intercanvi genètic. Aquests resultats indiquen que aquestes regions resulten adients per fer inferències demogràfiques sobre la història de la inversió  $J_1$ . S'ha d'afegir que la presència de 4 i 2 polimorfismes fixats en les regions p222 i p127, respectivament, és una dada indicativa de que l'origen de la inversió  $J_1$  no és recent.

Al llarg dels últims vint anys, s'han proposat diferents metodologies per estimar l'edat de les inversions cromosòmiques. Rozas i Aguadé (1994) proposen estimar l'edat de les inversions utilitzant la diversitat nucleotídica dels dos individus més divergents de la població en estudi. Posteriorment Rozas et al. (1999) proposen estimar l'edat de les inversions cromosòmiques, utilitzant la diversitat nucleotídica silenciosa i considerant que la inversió es troba en la fase de recuperació de variació i que tota la variació que s'observa dintre de la inversió des del seu origen hauria estat originada per mutació. Wallace et al. (2011) proposen inferir la història evolutiva de diferents ordenacions cromosòmiques del cromosoma 3 a *D. pseudoobscura* mitjançant inferències bayesianes, utilitzant el programa BEAST (Drummond i Rambaut 2007), estimant l'edat de diferents ordenacions cromosòmiques a partir de datacions externes de l'arbre filogenètic, tals com a partir del registre fòssil o dades prèvies disponibles. No obstant, aquesta metodologia no es considera la més adequada al no tenir en compte la història demogràfica de l'espècie.

Recentment, s'han desenvolupat diferents algorismes basats en simulacions de coalescència i aproximacions bayesianes, l'objectiu dels quals és estimar l'edat de les inversions cromosòmiques (Guerrero et al. 2012; Peischl et al. 2013). Corbett-Detig et al. (2013), desenvolupen un model d'expansió exponencial simple i utilitzen metodologies d'ABC per estimar l'edat mínima de formació de diferents inversions cromosòmiques en l'espècie *D. melanogaster*, utilitzant regions properes als punts de trencament. No obstant, aquest estudi podria tenir la limitació de que no engloba la resta de factors demogràfics que haurien pogut influenciar la història evolutiva de l'espècie i per tant també la història de les inversions. En el nostre estudi s'ha inferit l'edat de la inversió  $J_1$  mitjançant la metodologia d'ABC, utilitzant un model demogràfic senzill el qual engloba no només la història evolutiva de la inversió  $J_1$ , sinó també la història demogràfica recent de l'espècie (veure Materials i Mètodes M3.4). El sorgiment de la inversió  $J_1$  seria anàleg al d'un fenomen de subdivisió poblacional i la migració entre subpoblacions reflectiria la taxa d'intercanvi genètic. En el model demogràfic plantejat, no es va incloure migració ja que per les anàlisis només es van utilitzar les dues regions més properes als punts de trencament les quals no presenten intercanvi genètic. Les inferències realitzades van mostrar com l'edat més probable de formació de la inversió  $J_1$  seria d'ara fa aproximadament 466.000 anys (335.334 – 574.555 anys com a límits inferior i superior del 95% de la distribució).

La inversió  $J_1$  seria una inversió més antiga que altres inversions datades en l'espècie *D. subobscura* tals com diverses inversions del cromosoma O (Rozas et al. 1999) o la inversió  $A_2$  (Nóbrega et al. 2008). No obstant, la metodologia emprada per datar l'edat més probable de la inversió  $J_1$  difereix de la utilitzada per altres inversions de l'espècie *D. subobscura*, les quals no tenen en compte l'expansió poblacional recent ocorreguda a l'espècie. A tals propòsits i seguint la mateixa metodologia que la utilitzada per estimar l'edat de la inversió  $J_1$ , es va revisar l'edat més probable de la inversió  $A_2$ . El plantejament del model demogràfic varia

tenint en compte que la inversió  $A_2$  es troba localitzada en el cromosoma X, per tant la grandària efectiva emprada pel disseny de les distribucions *a priori* del model serà de  $\frac{3}{4}$  la grandària efectiva actual estimada de l'espècie. A més a més, per les freqüències de l'ordenació  $A_{st}$  es va dissenyar una distribució *a priori* uniforme amb un llindar de 0,2 a 0,5 sent complementària a la de l'ordenació  $A_2$ . Per fixar aquest llindar es van tenir en compte les freqüències observades d'aquesta inversió en l'àrea paleàrtica (Krimbas 1993). El límit inferior de la distribució *a priori* del paràmetre temps de formació de la inversió  $A_2$  es va calcular de la mateixa manera que per la inversió  $J_1$ , seguint la metodologia proposada a Rozas et al. (1999). Per revisar l'edat de la inversió  $A_2$  es van utilitzar les regions p150 i Sx1 disponibles a l'article de Nóbrega et al. (2008), sent aquestes les úniques dues regions de l'estudi en les quals no s'observen tractes de conversió gènica. Les anàlisis d'ABC van mostrar que l'edat més probable de la inversió  $A_2$  és d'ara fa aproximadament 165.000 anys (154.366 – 239.466 anys com a límits inferior i superior del 95% de la distribució) (Figura D4.5 i Taula D4.1). Aquesta estima és similar a l'obtinguda per Nóbrega et al. (2008). D'aquesta manera es van poder comparar les dues ordenacions, posant de manifest que la inversió  $J_1$  és una inversió més antiga que la inversió  $A_2$ .

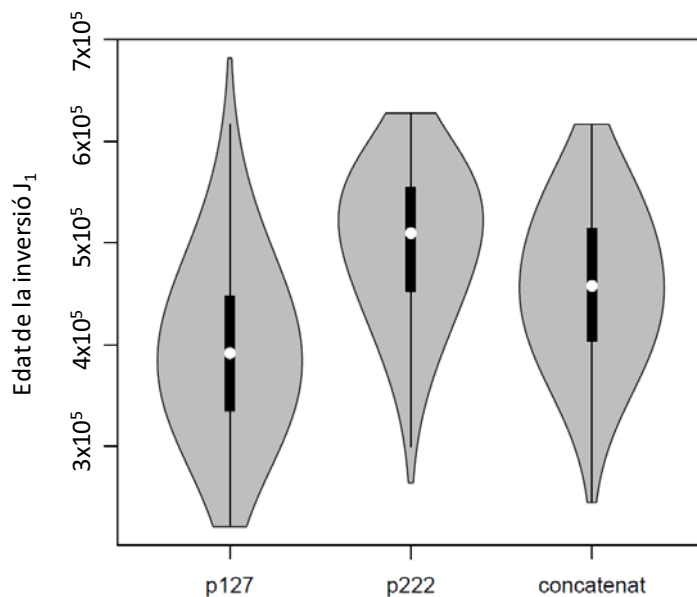


**Figura D4.5.** Distribució posterior del paràmetre  $T_3$  (edat de formació de la inversió  $A_2$ ), amb un llindar de tolerància del 0,001. Les distribucions amb línia discontinua representen el valor de cada paràmetre en les simulacions realitzades, en tant que les distribucions en negre i vermell representen el valor corresponent pel conjunt de 1000 simulacions retinudes mitjançant el mètode de rebuig (Tavaré et al. 1997) i el mètode de correcció per regressió lineal (Beaumont et al. 2002), respectivament.

**Taula D4.1.** Estadístics descriptius de la distribució posterior del paràmetre  $T_3$  (edat de formació de la inversió  $J_1$ ) sent aquests la mitjana, mediana i la moda així com també els límits inferior i superior del 95% de la distribució, escollint un llinar de tolerància del 0,001 i mitjançant el mètode de correcció per regressió lineal (Beaumont et al. 2002). Els estadístics utilitzats en l'anàlisi d'ABC obtinguts de Nóbrega et al. (2008) són la diversitat nucleotídica per locus (Tajima 1983), l'estadístic  $D$  de Tajima (1989) i l'estadístic  $Zns$  (Kelly 1997) per l'ordenació  $A_2$  i l'estadístic  $F_{ST}$  (Hudson et al. 1992).

Estadístics descriptius/ Paràmetres	$T_3$	Edat de formació de la inversió $A_2$
<b>límit inferior del 95% de la distribució</b>	0,0714	154385
<b>Mediana</b>	0,1001	179304
<b>Mitjana</b>	0,1036	185693
<b>Moda</b>	0,0916	165088
<b>límit superior del 95% de la distribució</b>	0,1480	239466

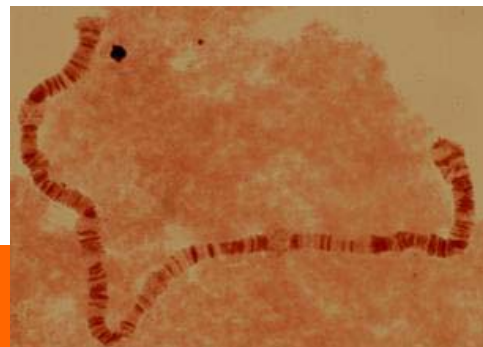
Les estimes de l'edat de la inversió  $J_1$  es van realitzar a partir del concatenat de les dues regions més properes al punt de trencament. No obstant, si s'analitzen les dues regions per separat, les estimes de la distribució posterior del paràmetre edat de la inversió  $J_1$  ( $T_3$ ) varien, sent l'estima més probable en el cas que s'utilitzi només la regió p127 de 410.000 anys (274.605-535.608 com a límits inferior i superior del 95% de la distribució) i l'estima més probable en el cas que s'utilitzi només la regió p222 de 560.000 anys (371.608-592.2247 com a límits inferior i superior del 95% de la distribució), tal i com s'observa en la Figura D4.6.



**Figura D.4.6.** Gràfic de violí, en el qual es combina un diagrama de caixa i un plot de densitat *Kernel*, que representa l'edat de la inversió  $J_1$ , en funció de si s'analitzen per separat les dues regions més properes als punts de trencament (p127 i p222) o les dades concatenades. L'escala de l'eix de les  $Y$  s'expressa en anys.

Per tant, la regió p222 que és la més propera a un dels punts de trencament indicaria que, respecte a la regió p127, l'ordenació  $J_1$  és més antiga. La regió p222 presenta uns nivells majors de diversitat nucleotídica per locus (Tajima 1983), uns valors de l'estadístic  $D$  de Tajima (1983) més propers a zero i uns nivells majors de diferenciació genètica, mesurada mitjançant l'estadístic  $F_{st}$  (Hudson 1992). Aquestes característiques de la variació genètica de la regió p222 en relació a p127 són consistents amb el resultat obtingut. Corbett-Detig et al. (2013), els quals estudien un conjunt d'inversions cromosòmiques en l'espècie *D. melanogaster*, analitzant per separat la variació en els extrems distals i proximals dels punts de trencament, també observen aquesta disparitat en les estimes d'ambdós punts de trencament encara que molt més acusades que en el cas de  $J_1$ . Seria interessant, doncs, en estudis posteriors incrementar l'anàlisi de les regions properes als punts de trencament, així com també dels mateixos punts de trencament, per poder confirmar aquestes diferències en els nivells i patrons de variació en els dos punts de trencament.





## CONCLUSIONS



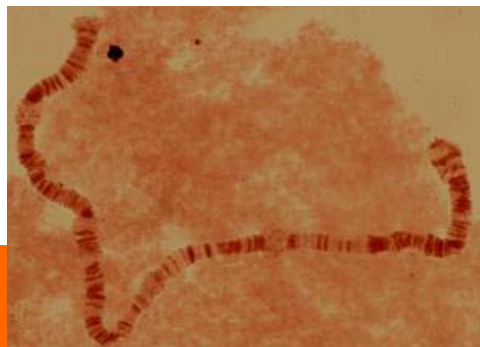


1. Els nivells i patrons de variació de les diferents regions en estudi estan caracteritzats per un excés de polimorfismes amb variants a baixa freqüència, indicant que la variació genètica en la població natural de *D. subobscura* estudiada no es troba en equilibri deriva-mutació.
2. Els patrons de variació relativament homogenis al llarg de totes les regions en estudi denoten que la variabilitat genètica d'aquestes regions està influenciada per factors demogràfics més que per factors selectius.
3. En espècies amb un elevat nivell de variació és important considerar la possibilitat de mutacions recurrents en les simulacions de coalescència per inferir la seva història evolutiva.
4. El model demogràfic més probable dels que poden explicar la variabilitat nucleotídica observada en la població natural de *D. subobscura*, és el model d'expansió poblacional recent amb una probabilitat posterior de 0,93 respecte als altres dos models analitzats.
5. L'edat més probable d'aquesta expansió poblacional és d'ara fa aproximadament 110.000 anys, coincidint amb el penúltim període interglacial i posant per tant de manifest que *D. subobscura* no ha vist alterada significativament la seva grandària efectiva durant períodes climàtics adversos posteriors, ni tan sols durant l'últim màxim glacial ara fa aproximadament 20.000 anys.
6. L'increment de la grandària efectiva ocorreguda durant l'expansió poblacional és pràcticament d'un ordre de magnitud, estimant-se la grandària efectiva actual en 9 milions d'individus aproximadament.
7. La disponibilitat d'un model demogràfic ha permès la detecció d'una possible diana de la selecció en la regió que inclou el gen *Zasp67*.

8. En les regions localitzades dintre de la inversió  $J_1$ , els nivells i patrons de variació (quan es consideren les ordenacions  $J_1$  i  $J_{st}$  independentment) no es diferencien significativament dels observats en les regions localitzades fora de la inversió, indicant que la variació a les regions localitzades dintre de la inversió  $J_1$  també es veu influenciada per la història demogràfica recent de l'espècie.
9. La menor variabilitat nucleotídica detectada a l'ordenació  $J_{st}$ , considerada com l'ancestral, pot ser deguda a una grandària efectiva menor d'aquesta ordenació respecte a l'ordenació derivada  $J_1$ .
10. La genealogia del concatenat de les sis regions incloses dintre de la inversió cromosòmica  $J_1$  posa de manifest el seu origen monofilètic i la forta diferenciació genètica entre ordenacions malgrat la presència d'intercanvi genètic.
11. La diferenciació genètica observada entre ordenacions en les diferents regions incloses dintre de la inversió  $J_1$  és altament significativa, observant-se una correlació també significativa tant entre la diferenciació genètica i la distància al punt de trencament més proper com entre aquesta distància i el desequilibri de lligament entre les variants de les posicions polimòrfiques informatives i l'ordenació cromosòmica ( $J_{st}$  o  $J_1$ ).
12. La disminució de la diferenciació genètica en regions centrals de la inversió  $J_1$  amb una longitud aproximada de 30 cM indicaria que aquesta longitud és suficient per què es donin dobles entrecreuaments dins de la nansa d'inversió.
13. L'elevada presència d'inversions cromosòmiques fa pensar que el genoma de *D. subobscura* pot estar altament estructurat. No obstant, no es detecta estructuració produïda per la inversió  $J_1$  a l'allunyar-se al voltant de 1 Mb fora de la inversió.
14. La utilització d'una metodologia per datar les inversions cromosòmiques que engloba tant els factors demogràfics com els factors associats a l'origen d'una nova inversió, ha

permès inferir l'edat més probable de la inversió  $J_1$  en ara fa aproximadament 460.000 anys, sent per tant una inversió relativament antiga.





## BIBLIOGRAFIA



- 
- Achaz G (2009) Frequency spectrum neutrality tests: one for all and all for one. *Genetics* **183**, 249-258.
  - Adams MD, Celniker SE, Holt RA, *et al* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.
  - Aguadé M (1999) Positive selection drives the evolution of the *Acp29AB* accessory gland protein in *Drosophila*. *Genetics* **152**, 543–551.
  - Aguadé M, Miyashita N, Langley CH (1992) Polymorphism and divergence in the *Mst26A* male accessory gland gene region in *Drosophila*. *Genetics* **132**, 755-770.
  - Andolfatto P (2005) Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**, 1149–1152.
  - Andolfatto P, Przeworski M (2001) Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics* **158**, 657-665.
  - Andolfatto P, JD Wall and M. Kreitman (1999) Unusual haplotype structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*. *Genetics* **153**, 1297-1311.
  - Arenas M, Ray N, Currat M, Excoffier L (2012) Consequences of range contractions and range shifts on molecular diversity. *Mol Biol Evol.* **29**, 207-218.
  - Ashburner M, Golic KG, Hawley RS (2005) *Drosophila: A laboratory handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
  - Ayala D, Fontaine MC, Cohuet A, Fontenille D, Vitalis R, Simard F (2011) Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. *Mol Biol Evol.* **28**, 745-758.
  - Ayala FJ, Serra L and Prevosti A (1989) A grand experiment in evolution: the *D. subobscura* colonization of the Americas. *Genome* **31**, 246-255.



- 
- Balanyà J, Oller JM, Huey RB, Gilchrist GW, Serra L (2006) Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science* **313**, 1773-1775.
  - Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* **16**, 37-48.
  - Bauer H (1936) Beitrage zur vergleichenden Morphologie der Speicheldriensenchromosomen. *Zool. Jahrb. Abt. Allg. Zool. Physiol. Tiere* **56**, 239-276.
  - Baudry E, Depaulis F (2003) Effect of misoriented sites on neutrality tests with outgroup. *Genetics* **165**, 1619-1622.
  - Beaumont, MA (2008) Joint determination of topology, divergence time and immigration in population trees. In Matsumura,S., Forster,P. and Renfrew,C. (eds) Simulations, Genetics and Human Prehistory, (McDonald Institute Monographs), McDonald Institute for Archaeological Research, Cambridge, pp 134–154.
  - Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035.
  - Beckenbach AT, Wei YW, Liu H (1993) Relationship in the *Drosophila obscura* group, inferred from mitochondrial cytochrome oxidase-II sequences. *Mol Biol Evol.* **10**, 619-634.
  - Beckenbach AT and Prevosti A (1986) Colonization of North America by the European species *D. subobscura* and *D. ambigua*. *Am Midl Nat.* **115**, 10-18.
  - Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP *et al* (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *Plos Biol.* **6**, e310.
  - Bender W, Akam M, Karch F, Beachy PA, Peifer M, Spierer P, Lewis EB, Hogness DS (1983) Molecular Genetics of the Bithorax Complex in *Drosophila melanogaster*. *Science* **221**, 23–29.

- 
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol.* **19** 2609-2625.
  - Bertrán E, Rozas J, Navarro A, Barbadilla A (1997) The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**, 89-99.
  - Bridges CB, Morgan TH (1919) Contributions to the genetics of *Drosophila melanogaster*. II. The second-chromosome group of mutant characters. *Carnegie Inst. Washington Publ.* **278**, 123-304.
  - Britten RJ, Kohne DE (1968) Repeated sequences in DNA. *Science* **161**, 529–540.
  - Brncic D, Prevosti A, Budnik M, Monclús M, Ocana J (1981) Colonization of *Drosophila subobscura* in Chile I. First population and cytogenetic studies. *Genetica* **56**, 3-9.
  - Buzzati-Traverso AA, Scossioli RE (1955) The *obscura* group of the genus *Drosophila*. *Adv. Genet.* **7**, 47.
  - Buzzati-Traverso A (1942) Genetica di popolazioni in *Drosophila*: I. Eterozigosi in *Drosophila subobscura* Collins. *Scientia Genetica* **2**, 190-223.
  - *C. elegans* Sequencing Consortium (2012) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **282**, 2012-2018.
  - Cáceres M, Puig M, Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Research* **11**,1353-1364.
  - Cáceres M, Ranz JM, Barbadilla A, Long M, Ruiz A (1999a) Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**, 415-418.
  - Cáceres M, Barbadilla A, Ruiz A (1999b) Recombination rate predicts inversion size in Diptera. *Genetics* **153**,251-259.

- 
- Casals F, Cáceres M, Manfrin MH, González J, Ruiz A (2005) Molecular characterization and chromosomal distribution of *Galileo*, *Kepler* and *Newton*, three foldback transposable elements of the *Drosophila buzzatii* species complex. *Genetics* **169**, 2047-2059.
  - Castañeda IS, Mulitza S, Schefuss E, Lopes dos Santos RA, Sinninghe Damsté JS, Schouten S (2009) Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proc Natl Acad Sci USA* **106**, 20159-20163.
  - Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540-552.
  - Castro JA, Barrio E, González A, Picornell A, Ramon MM, Moya A (2010) Nucleotide diversity of a *ND5* fragment confirms that population expansion is the most suitable explanation for the mt DNA haplotype polymorphism of *Drosophila subobscura*. *Genetica* **138**, 819-829.
  - Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* **10**, 195-205
  - Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ (2012) Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics* **190**, 1417-1432.
  - Christie ALM (1939) The effect of X-rays on sex in *Drosophila subobscura* and an account of some sex-linked characters. *Journal of Genetics* **75**, 123-131.
  - Clark AG, Eisen MB, Smith DR, *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–18.
  - Clark A, Gibson G, Kaufman T, Myers E, O'Grady P (2003) "Proposal for *Drosophila* as a model system for comparative genomics." *White Paper to NHGRI*.
  - Clarke JM, Maynard Smith J, Sondhi KC (1961) Asymmetrical response to selection for rate of development in *Drosophila subobscura*. *Genet. Research* **2**, 70-81.

- 
- CLIMAP Project Members (1984) The last interglacial ocean. *Quaternary Research* **21**, 123-224.
  - Clotault J, Thuillet AC, Buiron M, De Mita S, Couderc M, Hausmann BI, Mariac C, Vigouroux Y (2011) Evolutionary history of pearl millet (*Pennisetum glaucum* [L.] R. Br.) and selection on flowering genes since its domestication. *Mol Biol Evol.* **29**, 1199-1212.
  - Collin JE (1936) Note *Drosophila subobscura* sp.n. males, female. Addendum p. 60 to Gordon, *J. Genet.* **33**, 25-60.
  - Comerón JM (1997) Estudi de la variabilitat nucleotídica a *Drosophila*: Regió *Xdh* a *D. subobscura*. Ph.D. Thesis. Universitat de Barcelona.
  - Conceição IC, Aguadé M (2010) Odorant receptor (Or) genes: polymorphism and divergence in the *D. melanogaster* and *D. pseudoobscura* lineages. *PLoS One* **5**, e13389.
  - Corbett-Detig RB, Kolaczkowski B, Fang S, Nista PM, Holloway AK, Kern AD, Dewey CN, Song YS, Hahn MW and Begun DJ (2012) Genomic Variation in Natural Populations of *Drosophila melanogaster*. *Genetics* **192**, 533-598.
  - Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol.* **25**, 410-418.
  - Depaulis F, Veuille M (1998) Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Mol Biol Evol.* **15**, 1788–1790.
  - Dobzhansky TH (1950) Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. *Genetics* **35**, 288–302.
  - Dobzhansky TH (1937) Genetics and the Origin of Species. *Columbia Univ. Press*, New York; 2nd Ed., 1941; 3rd Ed., 1951.
  - Dobzhansky TH (1933) On the sterility of interracial hybrids in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences USA* **19**, 397-403.

- 
- Drummond A, Rambaut A (2007) BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* **7**, 214.
  - Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S (2013) Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* **193**, 291-301.
  - Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
  - Eanes, WF, Kirchner M, Yoon J, Biermann C, Wang IN *et al.* (1996) Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans*. *Genetics* **144**, 1027–1041.
  - ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia of dna elements) project. *Science* **306**, 636–640.
  - Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol.* **25**, 660-669.
  - Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413.
  - Fontdevila A, Zapata C, Álvarez G, Sánchez L, Méndez J, Enríquez I (1983) Genetic coadaptation in the chromosomal polymorphism of *Drosophila subobscura*. I. Seasonal changes of gametic disequilibrium in a natural population. *Genetics* **105**, 935-955.
  - Frutos R, Prevosti A (1984) Temporal change of chromosomal polymorphism in natural populations of *D. subobscura*. *Genetica* **63**, 181-187.
  - Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.
  - Fu YX, Li WH (1997) Estimating the age of the common ancestor of a sample of DNA sequences. *Mol Biol Evol.* **14**, 195–199.

- 
- Fu YX Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.
  - Gaffney DJ, Keightley PD (2006) Genomic Selective Constraints in Murid Noncoding DNA. *PLoS Genet* **2**, e204.
  - Gao JJ, Watabe HA, Aotsuka T, Pang JF, Zhang YP (2007) Molecular phylogeny of the *Drosophila obscura* species group, with emphasis on the Old World species. *BMC Evol Biol.* **7**, 87.
  - Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian Data Analysis, 2nd edn. Chapman & Hall, London.
  - Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M *et al.* (2003) A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736.
  - Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D (2003) Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**, 1269–1278.
  - Goto SG, and MT Kimura (2001) Phylogenetic utility of mitochondrial COI and nuclear Gpdh genes in *Drosophila*. *Mol Phylogenet Evol.* **18**, 404–422.
  - Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E (2013) On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol.* **5**, 578-590.
  - Guerrero RF, Rousset F, Kirkpatrick M (2012) Coalescent patterns for chromosomal inversions in divergent populations. *Philos Trans R Soc Lond B Biol Sci.* **367**, 430-438.
  - Guillén Y, Ruiz A (2012). Gene alterations at *Drosophila* inversion breakpoints provide prima facie evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics* **13**, 53.

- 
- Guirao-Rico S and M Aguadé (2013) Patterns of nucleotide diversity at the regions encompassing the *Drosophila* insulin-like peptide (dilp) genes: demography vs. positive selection in *Drosophila melanogaster*. *PLoS One* **8**, e53593.
  - Guirao-Rico S, Aguadé M (2009) Positive selection has driven the evolution of the *Drosophila* insulin-like receptor (*InR*) at different timescales. *Mol Biol Evol.* **26**, 1723-1732.
  - Haddrill PR, Bachtrog D, Andolfatto P (2008) Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol Biol Evol.* **25**, 1825–1834.
  - Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* **15**, 790–799.
  - Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp.* **41**, 95-98.
  - Harris H (1966) Enzyme polymorphism in man. *Proc. R. Soc. Lond. Ser. B.* **164**, 298-310.
  - Hartl DL, and AG Clark (2007) Principles of Population Genetics, Ed. 4. Sinauer Associates, Sunderland, MA.
  - Hastie T, Tibshirani R, Friedman JH (2009) The Elements of Statistical Learning. Springer, 2nd edition.
  - Hernandez RD, Williamson SH, Bustamante CD (2007) Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol Biol Evol.* **24**, 1792-1800.
  - Hewitt GM (2004). Genetic consequences of climatic changes in the Quaternary. *Phil. Trans. R. Soc. Lond. B* **359**, 183–195.
  - Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
  - Hill WG i Robsetson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet.* **38**, 226-231.

- 
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-338.
  - Hudson RR, Slatkin M and Maddison WP (1992a) Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583-589.
  - Hudson RR, Boos DD. and Kaplan NL. (1992b) A statistical test for detecting population subdivision. *Mol Biol Evol.* **9**, 138-151.
  - Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1–44.
  - Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-159.
  - Huey RB i Pascual M (2009) Partial thermoregulatory compensation by a rapid evolving invasive species along a latitudinal cline. *Ecology* **90**, 1715-1720.
  - Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science* **287**, 308-309.
  - Imbrie J, Hays JD, Martinson DG, McIntyre AC, Mix AC, Morley JJ, Pisias NG, Prell WL, Shackleton NJ (1984) The orbital theory of Pleistocene climate: Support from a revised chronology of the marine  $\delta^{18}\text{O}$  record. In: Berger AL, Imbrie J, Hays JD, Kukla GJ and Saltzman B. (eds), *Milankovitch and Climate*, pp. 269-306. Reidel, Dordrecht.
  - Joyce P, Marjoram P (2008) Approximately sufficient statistics and bayesian computation. *Stat Appl Genet Mol Biol.* **7**, 26.
  - Jukes TH and CR Cantor (1969) Evolution of protein molecules, p. 21-120. In *Mammalian Protein Metabolism*, HM Munro (ed.), Academic Press, New York.
  - Kass R and Raftery A (1995) Bayes factors and model uncertainty. *J Am Statist Ass.* **90**, 773–795.



- 
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**, 772-780.
  - Katzemich A, Liao KA, Czerniecki S, Schöck F (2013) *Alp*/Enigma family proteins cooperate in z-disc formation and myofibril assembly. *PLoS Genet.* **9**, e1003342.
  - Kellermann V, Loeschcke V, Hoffmann AA, Kristensen TN, Fløjgaard C, David JR, Svenning JC, Overgaard J (2012) Phylogenetic constraints in key functional traits behind species climate niches: patterns of desiccation and cold resistance across 95 *Drosophila* species. *Evolution* **66**, 3377-3389.
  - Kelly JK (1997). A test of neutrality based on interlocus associations. *Genetics* **146**, 1197-1206.
  - Kerdelhué C, Zane L, Simonato M, Salvato P, Rousset J, Roques A, Battisti A (2009) Quaternary history and contemporary patterns in a currently expanding species. *BMC Evol Biol.* **9**, 220.
  - Khadem M, Rozas J, Segarra C, Aguadé M (1998) Tracing the colonization of Madeira and the Canary Islands by *Drosophila subobscura* ancestors through the study of the *rp49* gene region. *J Evol Biol.* **11**, 439-452.
  - Kimura M (1983) Rare variant alleles in the light of the neutral theory. *Mol Biol Evol.* **1**, 84-93.
  - Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.
  - Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624-626.
  - Kingman JFC (1982) On the Genealogy of Large Populations. *Journal of Applied Probability* **19A**, 27-43.

- 
- Kirkpatrick M, and Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434.
  - Kohn MH, Fang S, Wu CI (2004) Inference of positive and negative selection on the 5' regulatory regions of *Drosophila* genes. *Mol Biol Evol.* **21**, 374–383.
  - Kolaczowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics* **187**, 245–260.
  - Koske T and J. Maynard Smith (1954) Genetics and cytology of *Drosophila subobscura*. X. The fifth linkage group. *Journal of Genetics* **52**, 521-541.
  - Krimbas CB (1993) *Drosophila subobscura*: Biology, Genetics, and Inversion Polymorphism. Hamburg: Verlag Dr Kovac.
  - Krimbas CB and JR Powell editors (1992) *Drosophila* inversion polymorphism. CRC Press. Boca Raton, Florida, USA.
  - Krimbas CB, and Loukas M (1980) The inversion polymorphism of *Drosophila subobscura*. *Evolutionary Biology* **12**, 163-234.
  - Krimbas CB (1964). The genetics of *Drosophila subobscura* populations. I. Inversion polymorphism in populations of southern Greece. *Evolution* **18**, 541–552.
  - Kreitman M, Aguadé M (1986) Excess Polymorphism at the *Adh* Locus in *Drosophila melanogaster*. *Genetics* **114**, 93–110.
  - Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412-417.
  - Kukla GJ, et al. (2002) Last interglacial climates. *Quat Res.* **58**, 2–13.
  - Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet.* **5**, e1000550.

- 
- Kunze-Mühl E, Müller E, Sperlich D (1958) Qualitative, quantitative und jahreszeitliche untersuchungen über den chromosomalen polymorphismus natürlicher populationen von *Drosophila subobscura* Coll. in der umgebung vo wien. *Zeitchsrift Induktive Abstammungs- u. Vererbungslehre* **89**, 636-646.
  - Laayouni H, García-Franco F, Chávez-Sandoval BE, Trotta V, Beltran S, Corominas M, Santos M (2007) Thermal evolution of gene expression profiles in *Drosophila subobscura*. *BMC Evol Biol.* **19**, 7-42.
  - Laayouni H, Hasson E, Santos M, Fontdevila A (2003) The evolutionary history of *Drosophila buzzatii*. XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. *Mol Biol Evol.* **20**, 931–944.
  - Langley CH, Montgomery E, Quattlebaum WF (1982) Restriction map variation in the *Adh* region of *Drosophila*. *Proc Natl Acad Sci USA* **79**, 5631–5635.
  - Langley CH, Stevens K, Cardeno C, Lee YC, Schridder DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, Fang S, Nista PM, Holloway AK, Kern AD, Dewey CN, Song YS, Hahn MW, Begun DJ (2012) Genomic Variation in Natural Populations of *Drosophila melanogaster*. *Genetics* **192**, 533-98.
  - Laurent SJ, Werzner A, Excoffier L, Stephan W (2011) Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. *Mol Biol Evol.* **28**, 2041-2051.
  - Leuenberger C, and D Wegmann (2009) Bayesian Computation and Model Selection without likelihoods. *Genetics* **184**, 243-252.
  - Lewin R (1986) Proposal to sequence the human genome stirs debate. *Science* **232**, 1598–1600.

- 
- Lewontin RC, Hubby IL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595-609.
  - Li H, Stephan W (2006) Inferring the Demographic History and Rate of Adaptive Substitution in *Drosophila*. *PLoS Genet* **2**, e166.
  - Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
  - Lim JK, Simmons MJ (1994) Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *BioEssays* **16**, 269–275.
  - Lindsey DL, Zimm GG (1992) The genome of *Drosophila melanogaster*. *Academic Press, Inc., San Diego, California*.
  - Llopart A, Aguadé M (2000) Nucleotide polymorphism at the *Rp11215* gene in *Drosophila subobscura*. Weak selection on synonymous mutations. *Genetics* **155**, 1245-1252.
  - Lopes JS, Beaumont MA (2010) ABC: a useful Bayesian tool for the analysis of population data. *Infect Genet Evol.* **10**, 826-833.
  - Loukas M, Krimbas CB, Mavragani-Tsipidou P i Kastritsis CD (1979) Genetics of *Drosophila subobscura* populations. VIII. Allozyme loci and their chromosome maps. *The Journal of Heredity* **70**, 17-26.
  - Lyon MF (2003) Transmission ratio distortion in mice. *Annu Rev Genet.* **37**, 393–408.
  - Machado CA, Haselkorn TS, Noor MAF (2007) Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **175**, 1289-12306.
  - Maddison WP, Maddison DR (2000) MACCLADE 4: Analysis of Phylogeny and Character Evolution. In: *Sinauer & Associates, Sunderland, Massachusetts*.

- 
- Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, Latafowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Petit RJ, de Beaulieu JL (2006) A new scenario for the quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.* **171**, 199-221.
  - Mainx F, Koske TH, Smithal E (1953) *Z. Indukt. Abstamm u. VererbLehre* **85**,354.
  - Maniatis T, Fritsch EF. and Sambrook J. in *Molecular Cloning: A Laboratory Manual*, 23–233 (Cold Spring Harbor Laboratory, New York, 1983).
  - Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *PNAS* **100**, 15324-15328.
  - Martinson D, Pias NG, Hays JD, Imbrie J, Moore TC, Shackleton NJ (1987) Age dating and the orbital theory of the Ice Ages: Development of a high-resolution 0-300,000 Year chronostratigraphy. *Quaternar3' Research* **27**, 1-29.
  - Marzo M, Puig M, Ruiz A (2008) The Foldback-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the *Drosophila* genus. *Proc Natl Acad Sci USA* **105**, 2957-2962.
  - Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35.
  - McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652-654.
  - McGaugh SE, Noor MAF (2012) Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philos Trans R Soc Lond B Biol Sci* **367**, 422-429.
  - Menozzi P, and Krimbas CB (1992) The inversion polymorphism of *D. subobscura* revisited: Synthetic maps of gene arrangement frequencies and their interpretation. *J Evol Biol.* **5**, 625–641.

- 
- Mestres F, Balanyà J, Arenas C, Solé E, Serra L (2001) Colonization of America by *Drosophila subobscura*: heterotic effect of chromosomal arrangements revealed by the persistence of lethal genes. *Proc Natl Acad Sci USA* **98**, 9167-70.
  - Mestres F, Sanz J, Serra L (1998) Chromosomal structure and recombination between inversions in *Drosophila subobscura*. *Hereditas* **128**, 105-113.
  - modENCODE Consortium *et al.* (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787.
  - Monclús M (1964) Distribución y ecología de Drosófilidos en España. I. Especies de *Drosophila* en la región catalana. *Genética Ibérica* **16**, 143-165.
  - Muller HJ (1940) Bearings of the *Drosophila* work on problems of systematic, p. 185- 268. In *The New Systematics*, J. Huxley (ed.), Oxford Univ. Press (Clarendon), London and New York.
  - Munté A, Rozas J, Aguadé M, Segarra C (2005) Chromosomal inversion polymorphism leads to extensive genetic structure: A multilocus survey in *Drosophila subobscura* . *Genetics* **169**, 1573–1578.
  - Munté A, Aguadé M, Segarra C (2000) Nucleotide variation at the yellow gene region is not reduced in *Drosophila subobscura*: a study in relation to chromosomal polymorphism. *Mol Biol Evol.* **17**, 1942-1955.
  - Munté A (1999) Evolució molecular del gen *yellow*: Polimorfisme nucleotídic a *Drosophila subobscura* i divergència interespecífica en els grups *obscura* i *melanogaster*. Ph.D. Thesis. Universitat de Barcelona.
  - Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence-- accelerated evolution in rearranged chromosomes. *Science* **11**, 321-324.
  - Navarro A, Barbadilla A, Ruiz A (2000) Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics* **155**, 685–698.

- 
- Navarro A, Betran E, Barbadilla A, Ruiz A (1997) Recombination and gene flux caused by gene conversion and crossing. *Genetics* **146**, 695-709.
  - Navarro-Sabaté A, Aguadé M, Segarra C (2003) Excess of nonsynonymous polymorphism at *AcpH-1* in different gene arrangements of *Drosophila subobscura*. *Mol Biol Evol.* **20**, 1833-1843.
  - Nei M (1987) Molecular evolutionary genetics *Columbia University Press, New York*.
  - Nóbrega C, Khadem M, Aguadé M, Segarra C (2008) Genetic exchange versus genetic differentiation in a medium-sized inversion of *Drosophila*: the  $A_2/A_{st}$  arrangements of *Drosophila subobscura*. *Mol Biol Evol.* **25**, 1534-1543.
  - Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A.* **98**, 12084–12088.
  - Obbard DJ, Maclennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM (2012) Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol Biol Evol.* **29**, 3459-3473.
  - O'Grady PM, Kidwell MG (2002) Phylogeny of the subgenus *sophophora* (*Diptera: drosophilidae*) based on combined analysis of nuclear and mitochondrial sequences. *Mol Phylogenet Evol.* **22**, 442-453.
  - O'Grady PM (1999) Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. *Mol Phylogenet Evol.* **12**, 124-139.
  - Ohno S (1972) So much 'junk' DNA in our genome. *Brookhaven Symp. Biol.* **23**, 366–370.
  - Ometto L, Glinka S, De Lorenzo D, Stephan W (2005) Inferring the effects of demography and selection on *Drosophila melanogaster* from a chromosome-wide DNA polymorphism study. *Mol Biol Evol.* **22**, 2119–2130.

- 
- Orengo DJ, Aguadé M (2004) Detecting the footprint of positive selection in a european population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. *Genetics* **167**,1759-1766.
  - Orengo DJ, Prevosti A (1999) Wing-size heritability in a natural population of *Drosophila subobscura*. *Heredity* **82**, 100-106.
  - OrengoDJ, and Prevosti A (1996) Temporal changes in chromosomal polymorphisms of *Drosophila subobscura* related to climate changes. *Evolution* **50**, 1346-1350.
  - Orengo DJ, Prevosti A (1994) Preadult competition between *Drosophila subobscura* and *Drosophila pseudoobscura*. *Zeitschrift fuer Zoologische Systematik und Evolutionsforschung* **32**, 44–50.
  - Ortiz-Barrientos D, Chang AS, Noor MAF (2006) A recombinational portrait of the *Drosophila pseudoobscura* genome. *Genet Res.* **87**, 23-31.
  - Painter TS (1933) A new method for the study of chromosome rearrangements and the plotting of chromosome maps. *Science* **78**, 585-586.
  - Papaceit M, Segarra C, Aguadé M (2013) Structure and population genetics of the breakpoints of a polymorphic inversion in *Drosophila subobscura*. *Evolution* **67**, 66-79.
  - Papaceit M, Aguadé M, Segarra C (2006) Chromosomal evolution of elements B and C in the *Sophophora* subgenus of *Drosophila*: Evolutionary rate and polymorphism. *Evolution* **60**, 768–781.
  - Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annu Rev Ecol Evol Syst.* **37**, 637–669.
  - Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P (2010) On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol Biol Evol.* **27**, 1226–1234.



- 
- Pascual M, Chapuis MP, Mestres F, Balanyà J, Huey RB, Gilchrist GW, Serra L, Estoup A (2007) Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Mol Ecol.* **16**, 3069-3083.
  - Pavlidis P, Metzler D, Stephan W (2012) Selective sweeps in multilocus models of quantitative traits. *Genetics* **192**, 225-239.
  - Pavlidis P, Laurent S, Stephan W (2010) msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Mol Ecol Resour.* **10**, 723–727.
  - Pegueroles C, Aquadro CF, Mestres F i Pascual M (2013) Gene flow and gene flux shape evolutionary patterns of variation in *Drosophila subobscura*. *Heredity* **110**, 520–529.
  - Pegueroles C, Araúz PA, Pascual M, Mestres F (2010) A recombination survey using microsatellites: the O chromosome of *Drosophila subobscura*. *Genetica* **138**, 795-804.
  - Peischl S, Koch E, Guerrero RF, Kirkpatrick M (2013) A sequential coalescent algorithm for chromosomal inversions. *Heredity* **111**, 200-209.
  - Porretta D, Mastrantonio V, Mona S, Epis S, Montagna M, Sassera D, Bandi C, Urbanelli S (2013) The integration of multiple independent data reveals an unusual response to Pleistocene climatic changes in the hard tick *Ixodes ricinus*. *Mol Ecol.* **22**, 1666-1682.
  - Powell JR, Dion K, Papaceit M, Aguadé M, Vicario S, Garrick RC (2010) Non-recombining genes in a recombination environment: the *Drosophila* "dot" chromosome. *Mol Biol Evol.* **28**, 825-833.
  - Powell JR (1997) Progress and prospects in evolutionary biology: the *Drosophila* model. *Oxford University Press, New York.*
  - Prevosti A, Serra LL, Segarra C, Aguadé M, Ribó G, Monclús M (1990) Clines of chromosomal arrangements of *Drosophila subobscura* in South America evolve closer to Old World patterns. *Evolution* **44**, 218-221.

- 
- Prevosti A, Ribó G, Serra L, Agudé M, Balanyà J, Monclús M, Mestres F (1988) «Colonization of America by *Drosophila subobscura*: Experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism». *Proc. Natl. Acad. Sci. USA* **85**, 5597-5600.
  - Prevosti A, Serra L, Ribó G, Agudé M, Sagarra E, Monclús M and García MP (1985) The colonization of *D. subobscura* in Chile. II. Clines in the chromosomal arrangements. *Evolution* **39**, 838-844.
  - Prevosti A, Garcia MP, Serra L, Agudé M, Ribó G, Sagarra E (1983) Association between allelic isozyme alleles and chromosomal arrangements in European populations and Chilean colonizers of *Drosophila subobscura*. *Isozymes* **10**, 171-191.
  - Pritchard JK, Stephens P, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
  - Pritchard JK, Seielstad MT, *et al.* (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol.* **16**, 1791-1798.
  - Ramirez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**, 555–567.
  - Ramos-Onsins SE, Mitchell-Olds T (2007) Mlcoalsim: multilocus coalescent simulations. *Evol Bioinform Online* **3**, 41-44.
  - Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**, 2092–2100.
  - Ramos-Onsins S, Segarra C, Rozas J and Agudé M (1998) Molecular and chromosomal phylogeny of the *obscura* group of *Drosophila* inferred from sequences of the *rp49* gene region. *Mol Phyl Evol.* **9**, 33-41.

- 
- Ranz JM, Maurin D, Chan YS, von Grotthuss M, Hillier LW, Roote J, Ashburner M, Bergman CM (2007) Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol.* **5**, e152.
  - Ranz JM, González J, Casals F, Ruiz A (2003) Low occurrence of gene transposition events during the evolution of the genus *Drosophila*. *Evolution* **57**, 1325–1335.
  - Ranz JM, Casals F, Ruiz A (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Res.* **11**, 230–239.
  - Ranz JM, Segarra C, Ruiz A (1997) Chromosomal homology and molecular organization of Muller's elements *D* and *E* in the *Drosophila repleta* species group. *Genetics* **145**, 281–295.
  - Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R *et al.* (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1-18.
  - Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol.* **16**, 351–358.
  - Roberts PA (1976) The genetics of chromosomal aberration. In: Ashburner M, Novitski E, editors. *The genetics and biology of Drosophila*. London: Academic Press. p. 67–184.
  - Robertson M (1986) The proper study of mankind. *Nature* **322**, 11.
  - Rodríguez-Trelles F, Rodríguez MA (2007) Comment on "Global genetic change tracks global climate warming in *Drosophila subobscura*". *Science* **315**, 1497.
  - Rottschaefer SM, Riehle MM, Coulibaly B, Sacko M, Niaré O, Morlais I, Traoré SF, Vernick KD, Lazzaro BP (2011) Exceptional diversity, maintenance of polymorphism, and recent directional selection on the *APL1* malaria resistance genes of *Anopheles gambiae*. *PLoS Biol.* **9**, e1000600.

- 
- Rozas J, Guillaud M, Blandin G, Aguadé M (2001) DNA variation at the *rp49* gene region of *Drosophila simulans*: Evolutionary inferences from an unusual haplotype structure. *Genetics* **158**, 1147-1155.
  - Rozas J, Segarra C, Ribó G, Aguadé M (1999) Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* **151**, 189-202.
  - Rozas J, Segarra C, Zapata C, Alvarez G, Aguadé, M (1995) Nucleotide polymorphism at the *rp49* region of *Drosophila subobscura*: Lack of geographic subdivision within chromosomal arrangements in Europe. *J Evol Biol* **8**, 355-367.
  - Rozas J and M. Aguadé (1994) Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *Proc. Natl Acad. Sci. USA* **91**, 11517-11521.
  - Rozas J, and Aguadé M (1991) Using restriction-map analysis to characterize the colonization process of *Drosophila subobscura* on the American continent.I. *rp49* region. *Mol. Biol. Evol.* **8**, 447-457.
  - Rozas J, Hernández M, Cabrera VM and Prevosti A (1990) Colonization of America by *Drosophila subobscura*. Effect of the founder event on the mitochondrial DNA polymorphism. *Mol Biol Evol.* **7**, 103-109.
  - Rozas J, and Aguadé M (1990) Evidence of Extensive Genetic Exchange in the *rp49* Region Among Polymorphic Chromosome Inversions in *Drosophila subobscura*. *Genetics* **126**, 417-26.
  - Russel G (1996) The response of insect faunas to glacial-Interglacial climatic fluctuations. *Philosophical Transactions: Biological Sciences* **344**, 1307.
  - Rychlik W. OLIGO 7 primer analysis software.( 2007) *Methods Mol Biol.* **402**, 35-60.

- 
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* **4**, 406-425.
  - Sánchez-Gracia A, Rozas J (2011) Molecular population genetics of the OBP83 genomic region in *Drosophila subobscura* and *D. guanche*: contrasting the effects of natural selection and gene arrangement expansion in the patterns of nucleotide variation. *Heredity* **106**, 191-201.
  - Santos M (2009) Recombination load in a chromosomal inversion polymorphism of *Drosophila subobscura*. *Genetics* **181**, 803-809.
  - Santos M, Céspedes W, Balanyà J, Trotta V, Calboli FC, Fontdevila A, Serra L (2005) Temperature-related genetic changes in laboratory populations of *Drosophila subobscura*: evidence against simple climatic-based explanations for latitudinal clines. *Am Nat.* **165**, 258-273.
  - Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, *et al.* (2008) Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179**, 1601–1655.
  - Schaeffer SW, Anderson WW (2005) Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* **171**, 1729-1739.
  - Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T (2005) A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from the standard neutral model of DNA sequence polymorphism. *Genetics* **169**, 1601–1615.
  - Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front Zool.* **17**, 4-11.
  - Segarra C, Ribó G, Agudé M (1996) Differentiation of Muller's chromosomal elements D and E in the *obscura* group of *Drosophila*. *Genetics* **144**, 139-146.

- 
- Segarra C, Lozovskaya ER, Ribó G, Aguadé M, Hartl DL (1995) P1 clones from *Drosophila melanogaster* as markers to study the chromosomal evolution of Muller's A element in two species of the *obscura* group of *Drosophila*. *Chromosoma* **104**, 129–136.
  - Segarra C and M Aguadé (1993) Nucleotide divergence of the *rp49* gene region between *Drosophila melanogaster* and two species of the *Obscura* group of *Drosophila*. *Journal of Molecular Evolution* **36**, 243-248.
  - Segarra C, Aguadé M (1992) Molecular organization of the X chromosome in different species of the *obscura* group of *Drosophila*. *Genetics* **130**, 513–521.
  - Sella G, Petrov DA, Przeworski M, Andolfatto P (2009) Pervasive natural selection in the *Drosophila* genome?. *PLoS Genetics* **5**, e1000495.
  - Sharp PM, Li WH (1989) On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol.* **28**, 398-402.
  - Simões P, Calabria G, Picão-Osório J, Balanyà J, Pascual M (2012) The genetic content of chromosomal inversions across a wide latitudinal gradient. *PLoS One* **7**, e51625.
  - Spurway H (1945) The genetics and cytology of *Drosophila subobscura*. I. Element A. Sex-linked mutants and their standard order. *Journal of Genetics* **46**, 278-286.
  - Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, et al. (2005) A common inversion under selection in Europeans. *Nat Genet.* **37**, 129–137.
  - Steinberg KM, Antonacci F, Sudmant PH, Kidd JM, Campbell CD, Vives L, Malig M, Scheinfeldt L, Beggs W, Ibrahim M, Lema G, Nyambo TB, Omar SA, Bodo JM, Froment A, Donnelly MP, Kidd KK, Tishkoff SA, Eichler EE (2012) Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat Genet.* **44**, 872-880.
  - Stevison LS, Hoehn KB, Noor MAF (2011) Effects of inversions on within- and between-species recombination and divergence. *Genome Biol Evol.* **3**, 830-841.

- 
- Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences* **277**, 661–671.
  - Stumm-Zollinger E (1953) Vergleichende Untersuchung über die Inversionshäufigkeit bei *Drosophila subobscura* in Populationen der Schweiz und Südwesteuropas. Zeitschrift Induktive Abstammungs- u. Vererbungslehre **85**, 382-407.
  - Sturtevant AH (1942) The classification of the genus *Drosophila* with the description of nine new species. *Univ. Texas Publ.* **4213**, 5.
  - Sturtevant AH, Novitski E (1941) The homologies of the chromosome elements in the genus *Drosophila*. *Genetics* **26**, 517-541.
  - Sturtevant H, Dobzhansky TH (1936a) Inversions in the third chromosome of wild races of *Drosophila pseudoobscura*, and their use in the study of the history of the species. *Proc. Nat. Acad. Sci.* **22**, 448-450.
  - Sturtevant H, Dobzhansky TH (1936b) Geographical distribution and cytology of “sex-ratio” in *Drosophila pseudoobscura* and related species. *Genetics* **21**, 473-490.
  - Sturtevant AH (1917) Genetics factors affecting the strength of linkage in *Drosophila*, *Proc. Natl. Acad. Sci. USA.* **3**, 555-558.
  - Sullivan W, Ashburner M, Hawley RS (2000) *Drosophila* Protocols. *Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.*
  - Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
  - Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.

- 
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* **28**, 2731-2739.
  - Tamura K, Subramanian S, Kumar S (2004) Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks. *Mol Biol Evol.* **21**, 36-44.
  - Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505–518.
  - Thornton KR, Andolfatto P (2006) Approximate bayesian inference reveals evidence for a recent, severe, bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**, 1607-1619.
  - Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In “Handbook of Genetics, Vol. 3” (R. C. King, Ed.), pp. 421–469, Plenum, New York.
  - Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* **179**, 455–473.
  - Tzedakis PC, Roucoux KH, de Abreu L, Shackleton NJ (2004) The duration of forest stages in southern Europe and interglacial climate variability. *Science* **24**, 2231-2235.
  - Tzedakis PC (2003) Timing and duration of Last Interglacial conditions in Europe: a chronicle of a changing chronology. *Quaternary Science Reviews* **22**, 763–768.
  - Tzedakis PC, Frogley MR, Heaton THE (2003) Last Interglacial conditions in southern Europe: evidence from Ioannina, northwest Greece. *Global and Planetary Change* **36**, 157–170.
  - Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC (2002). Buffered tree population changes in a quaternary refugium: evolutionary implications. *Science* **20**, 2044-2047.
  - Van Andel TH, Tzedakis PC (1996) Palaeolithic landscapes of Europe and environs, 150,000–25,000 years ago: an overview. *Quat Sci Rev.* **15**, 481–500.



- 
- Van Valen L and R Levins (1968) "The origins of inversion polymorphisms,". *American Naturalist* **102**, 5-24.
  - Varga Z (2009) Extra-Mediterranean, post-glacial vegetation history and area dynamics in Eastern Central European refugia. In: *Relict Species*, Vol. 1 (eds Habel JC & Assmann T), pp. 57–87. Springer Berlin Heidelberg, Berlin.
  - Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G, Soodyall H, Louie L, Hammer MF (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol Biol Evol.* **29**, 617-630.
  - Wall JD (1999) Recombination and the power of statistical tests of neutrality. *Genet Res.* **74**, 65–79.
  - Wallace AG, Detweiler D, Schaeffer SW (2011) Evolutionary history of the third chromosome gene arrangements of *Drosophila pseudoobscura* inferred from inversion breakpoints. *Mol Biol Evol.* **28**, 2219-2229.
  - Walters RJ, Hassall M, Telfer MG, Hewitt GM, Palutikof JP (2006) Modelling dispersal of a temperate insect in a changing climate. *Proc Biol Sci.* **273**, 2017-2023.
  - Wasserman M (1968) Recombination-induced chromosomal heterosis. *Genetics* **58**, 125–139.
  - Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* **7**, 256-276.
  - Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatic* **4**, 11-116.
  - Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* **182**, 1207-1218.

- Weiss G, von Haeseler A (1998) Inference of population history using a likelihood approach. *Genetics* **149**, 1539–1546.
- Willis K, Van Andel T (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* **23**, 2369-2387.
- Zeng K, Fu YX, Shi S, Wu CI (2006) Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**, 1431-1439.

