Escuela Técnica Superior de Ingenierías Agrarias Palencia

UVa

Population genetic structure of Iberian white poplar (*Populus alba* L.): the role of mating system, hybridization and demographical history

TESIS DOCTORAL

DAVID MACAYA SANZ
Ingeniero de Montes
Palencia, 2014

Universidad de Valladolid

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS AGRARIAS

DEPARTAMENTO DE PRODUCCIÓN VEGETAL Y RECURSOS FORESTALES

TESIS DOCTORAL:

Population genetic structure of Iberian white poplar (*Populus alba* L.): the role of mating system, hybridization and demographical history

Presentada por David Macaya Sanz para optar al grado de doctor por la Universidad de Valladolid

Dirigida por:
Dr. Santiago César González Martínez
Dr. Christian Lexer

Prólogo

Esta tesis doctoral se realizó en el Centro de Investigación Forestal (CIFOR) perteneciente al Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), bajo la codirección del Dr. Santiago César González Martínez y del Dr. Christian Lexer. Fue financiada mediante el programa de beca+contrato de Formación del Personal Investigador (FPI; referencia de la ayuda: BES-2006-13132) del Ministerio de Ciencia e Innovación del Gobierno de España, asociada al proyecto del Plan Nacional de I+D+i "Sistemas de reproducción, flujo genético efectivo y efectos parentales: aplicación a la gestión y conservación de recursos genéticos forestales" (AGL2005-07440-C02-01/FOR), proyecto con el cual se financiaron las investigaciones.

Los estudios realizados durante este periodo han dado lugar a la realización de cinco documentos científicos de los cuales dos han sido ya publicados en revistas indexadas en el *Science Citation Index* (SCI), y otros tres están en distintas fases de elaboración, y cuyos resultados se incluyen aquí como capítulos:

MACAYA-SANZ, D., HEUERTZ, M., LOPEZ-DE-HEREDIA, U., DE-LUCAS, A. I., HIDALGO, E., MAESTRO, C., PRADA, A., ALÍA, R., GONZÁLEZ-MARTÍNEZ, S. C. The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars. 2012. Molecular Ecology **21**: 3593-3609.

MACAYA-SANZ, D., SUTER, L., JOSEPH, J., BARBARÁ, T., ALBA, N., GONZÁLEZ-MARTÍNEZ, S. C., WIDMER, A., LEXER, C. Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species.2011. Heredity **107**: 478-486.

MACAYA-SANZ, D., HEUERTZ, M., LINDTKE, D., LEXER, C., GONZÁLEZ-MARTÍNEZ, S. C, 2014. Causes and consequences of large clonal assemblies in a poplar hybrid zone. Manuscrito.

MACAYA-SANZ, D., CLIMENT, J.M., GONZÁLEZ-MARTÍNEZ, S. C. Hermaphroditism and pollen gene flow in white poplar. Borrador.

MACAYA-SANZ, D., WEGRZYN, J., NEALE, D., GONZÁLEZ-MARTÍNEZ, S. C. Local adaptation imprints on wood formation genes. Borrador.

Así mismo, los trabajos en esta tesis han contribuido de forma menor a la publicación de otro documento indexado en el SCI:

DE CARVALHO, D., INGVARSSON, P. K., JOSEPH, J., SUTER, L., SEDIVY, C., **MACAYA-SANZ, D.**, COTTRELL, J., HEINZE, B., SCHANZER, I, LEXER, C. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree.2010. Molecular Ecology **19**: 1638-1650.

Prologue

This PhD thesis was conducted at the Forest Research Centre (CIFOR) of the National Institute for Agricultural and Food Research and Technology (INIA), and supervised by Dr Santiago César González Martínez and Dr Christian Lexer. It was carried out thanks to a PhD scholarship of the FPI programme (scholarship reference code: BES-2006-13132) funded by the Science and Innovation Ministry, Spanish Government, and linked to the National Plan of I+D+i project "Reproduction systems, effective gene flow and parental effects: application to management and conservation of forest genetic resources" (AGL2005-07440-C02-01/FOR). The conducted research was funded by this project.

The studies carried out during this thesis have produced five papers, of which two have already been published in SCI (*Science Citation Index*) journals, and three other are in different steps of elaboration, whose results are included herein as chapters:

MACAYA-SANZ, D., HEUERTZ, M., LOPEZ-DE-HEREDIA, U., DE-LUCAS, A. I., HIDALGO, E., MAESTRO, C., PRADA, A., ALÍA, R., GONZÁLEZ-MARTÍNEZ, S. C. The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars. 2012. Molecular Ecology **21**: 3593-3609.

MACAYA-SANZ, D., SUTER, L., JOSEPH, J., BARBARÁ, T., ALBA, N., GONZÁLEZ-MARTÍNEZ, S. C., WIDMER, A., LEXER, C. Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species.2011. Heredity **107**: 478-486.

MACAYA-SANZ, D., HEUERTZ, M., LINDTKE, D., LEXER, C., GONZÁLEZ-MARTÍNEZ, S. C, 2014. Causes and consequences of large clonal assemblies in a poplar hybrid zone. Manuscript.

MACAYA-SANZ, D., CLIMENT, J.M., GONZÁLEZ-MARTÍNEZ, S. C. Hermaphroditism and pollen gene flow in white poplar. First draft.

MACAYA-SANZ, D., WEGRZYN, J., NEALE, D., GONZÁLEZ-MARTÍNEZ, S. C. Local adaptation imprints on wood formation genes. First draft.

Moreover, this PhD research has contributed in a minor way to the publication of another SCI indexed paper:

DE CARVALHO, D., INGVARSSON, P. K., JOSEPH, J., SUTER, L., SEDIVY, C., **MACAYA-SANZ, D.**, COTTRELL, J., HEINZE, B., SCHANZER, I, LEXER, C. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree.2010. Molecular Ecology **19**: 1638-1650.

Contents

Resumen

El género *Populus* representa un caso interesante para el estudio de los sistemas reproductivos y de la evolución. Este género que combina la dioecia con la clonalidad como modos reproductivos, posee un cromosoma sexual incipiente (XIX), pero, al mismo tiempo, muestra casos de inconstancia sexual, e hibrida frecuentemente, si bien no entre ciertas Secciones. Además, muchas de sus especies crecen exclusivamente en áreas riparias y humedales (a menudo en poblaciones aparentemente aisladas). Si bien, estas especies riparias presentan diferencias sutiles en sus condicionantes ecológicos. Finalmente, posee un genoma relativamente pequeño, con un elevado grado de sintenia, que además ha sido secuenciado con éxito.

Los álamos blanco y negro (*Populus alba* y *P. nigra*, respectivamente) crecen acompañando muchos cursos de agua de la Península Ibérica, a veces en simpatría, mientras que el álamo temblón (*P. tremula*) habita en laderas montanas en parapatría. El álamo blanco no es capaz de hibridar con el negro de forma natural, pero puede hacerlo ocasionalmente con el álamo temblón, dando lugar al llamado álamo cano (*Populus* × *canescens*). Por otro lado, algunas poblaciones ibéricas de álamo blanco presentan altas tasas de clonalidad, así como individuos hermafroditas.

Con el objetivo de profundizar en el significado evolutivo y ecológico del sistema reproductivo de los álamos ibéricos, esta tesis doctoral busca (i) estudiar la estructura genética espacial de los álamos ibéricos a escala local y regional; (ii) evaluar el grado de adaptación local en álamo blanco mediante diferentes aproximaciones; (iii) investigar el grado de expansión, las causas y las implicaciones de la clonalidad; (iv) evaluar los aspectos ecológicos y evolutivos de la hibridación entre el álamo blanco y el temblón; y (v) mejorar el conocimiento del incipiente cromosoma sexual del género *Populus*.

Para alcanzar estas metas, se caracterizó el ADN del cloroplasto de múltiples poblaciones ibéricas de álamo y, en algunas de ellas, se evaluó la diferenciación genética cuantitativa de varios caracteres adaptativos. El nivel de clonalidad fue investigado en varias poblaciones ibéricas de álamos, de las cuales una se caracterizó más intensamente (la zona híbrida situada en las proximidades de Aranda de Duero), usando microsatélites nucleares. Estos marcadores se usaron también para genotipar varias poblaciones europeas de álamo temblón, una de ellas procedente del centro de España (Sierra de Gredos), así como

individuos de varios ensayos de procedencias – para los que también se midieron caracteres fenotípicos – y un retrocruzamiento artificial de *P.* × *canescens* hacia *P. alba* (BC_1), en este caso, con un mayor énfasis en el cromosmoma XIX. Asimismo fueron genotipadas varias progenies producto de polinización abierta de una población andaluza de álamo blanco subdioica (es decir, que posee individuos masculinos, femeninos y algunos hermafroditas). Finalmente, se secuenciaron un gran número de genes involucrados en la formación de madera en poblaciones ibéricas que representaban grupos genéticos diferenciados.

El álamo blanco y el negro mostraron patrones filogeográficos contrastados en la Península Ibérica., reflejando que las diferencias en los requerimientos ecológicos (como la tolerancia al frío) imprimen diferentes estructuras genéticas regionales. En particular, el confinamiento del álamo blanco a la cuenca hidrográfica y a la vertiente marítima fue mayor que el del álamo negro. Las poblaciones sureñas de álamo blanco también resultaron más diversas genéticamente y hubo mayor migración entre las regiones del Norte y del Sur que entre las del Este y el Oeste. Estos patrones sugieren un efecto de las glaciaciones del Holoceno sobre los álamos, a pesar de su vinculación a los cursos de agua. La presencia de numerosos haplotipos privados mostró que no tuvieron lugar extinciones regionales durante las glaciaciones en ninguna de las cuencas hidrográficas estudiadas.

En el álamo blanco se detectó adaptación local para dos caracteres cuantitativos (altura total y forma del tronco a los 3 años), pero a una escala espacial mayor que en otras especies forestales (entre cuencas hidrográficas pero no entre poblaciones de la misma cuenca). Los resultados sobre señales de adaptación genética molecular en los genes de formación de la madera fueron controvertidos; usando un método se detectaron varios loci con valores atípicos (*outliers*), sin embargo, no se encontró ninguno empleando un método más estricto, siendo probablemente fruto de cuestiones asociadas al tamaño muestral. En cuanto al álamo temblón, también se detectó adaptación local en caracteres cuantitativos. Asimismo en esta especie se descubrieron señales claras de patrones de variación genética, posiblemente generadas por la mezcla genética de diferentes linajes.

En álamo blanco, los niveles de clonalidad promediaron 4,2 ramets por genet, en poblaciones con 32-49 ramets muestreados, alcanzando tamaños de genet cercanos a los 100 m. Sin embargo, se encontraron clones mucho mayores (>10 km) en dos poblaciones. Una de estas poblaciones, situada en Aranda de Duero, en la zona septentrional del centro peninsular, se estudió más intensamente. Allí se encontraron dos clones extensos y antiguos, uno de ellos híbrido con álamo temblón, que se extendían alrededor de cien

kilómetros. Aunque la riqueza y la uniformidad genotípica fueron menores que en otras poblaciones europeas, la diversidad genética fue similar y no se observaron signos de empobrecimiento genético. Mediante modelos de coalescencia se apreció que el tamaño efectivo poblacional total está en declive, aunque probablemente los grandes clones están incrementado su número de ramets. Además, los genets de mayor extensión solían ser los progenitores de más individuos. Por último, la introgresión puntual de *P. tremula* y la heterozigosidad podrían estar involucrados en la capacidad de propagación clonal de esta población.

Se profundizó en el rol de la heterozigosidad interespecífica en la adaptación mediante el estudio del cruce BC_1. Se encontró distorsión en la segregación hacia *P. tremula*, probablemente debido a interacciones citonucleares o a la superioridad de los heterozigotos, apoyando la idea de que la heterozigosidad podría jugar un papel importante en la adaptación, en particular cuando ha sido potenciada mediante hibridación.

En cuanto al cromosoma XIX, si bien algunas observaciones coincidieron con lo esperado si éste fuera un cromosoma sexual, otras sin embargo fueron opuestas. No se detectó recombinación alguna en la progenie del retrocruzamiento entre cuatro marcadores del cromosoma XIX, como es de esperar para un cromosoma sexual. No obstante, la distribución de la diversidad genética en este cromosoma, para grupos separados de machos y hembras, y su reducida divergencia entre poblaciones naturales fueron en contra de la hipótesis del cromosoma sexual. En conjunto, mi investigación apoya la hipótesis de que el cromosoma XIX está en un estado muy inicial de degeneración hacia un cromosoma sexual maduro.

Finalmente, los análisis de paternidad realizados en la población de Jimena revelaron que el número efectivo de padres involucrados en la polinización osciló entre 1,58 y 8,33. El tamaño clonal y la distancia estuvieron aparentemente involucrados en la tasa de polinización. En cuanto a los hermafroditas, uno de ellos fue el padre efectivo de varios brinzales. Además, se apreció una correlación entre la feminidad (proporción media de flores femeninas por amento, considerando las flores hermafroditas como media flor femenina) y la tasa de germinación.

Basándose en los resultados de esta tesis doctoral, en concreto la presencia de adaptación local y estructura genética espacial regional entre las poblaciones ibéricas de álamos, es recomendable tener en cuenta la región de procedencia en los programas de restauración e

implementar planes de conservación *in situ* en base a la cuenca hidrográfica, ya que entre las poblaciones ibéricas de álamos tienen lugar procesos de adaptación local y de estructuración genética espacial a nivel regional. La reseñable extensión de algunos conjuntos clonales hace necesaria la consideración de la replicación clonal en cualquier gestión relacionada con estos árboles. Sin embargo, es igualmente importante remarcar que ese grado de clonalidad no está produciendo ningún efecto genético negativo en la población estudiada. La hibridación y la mezcla genética de linajes han mostrado que son formas valiosas de enriquecimiento genético, con repercusión en la *fitness* y en la adaptación local. Los gestores forestales podrían aprovechar este hecho en cuanto a la adopción de medidas contra las futuras amenazas relacionadas con el Cambio Climático.

Finalmente, se ha observado que el cromosoma XIX opera como un cromosoma sexual en etapas iniciales de desarrollo, lo que permite la ocurrencia ocasional de hermafroditas. Una profundización en la interacción entre el hermafroditismo, los cromosomas sexuales y la clonalidad permitiría entender mejor su significación ecológica relativa y las potenciales rutas evolutivas del álamo blanco, una importante especie riparia y dioica.

Abstract

Populus represents an interesting tree genus for breeding system and evolutionary studies. It combines dioecy and clonality as reproductive modes, possesses an incipient sex chromosome (XIX) but, at the same time, shows occasional sex inconstancy and hybridises frequently, albeit not among some Sections. Moreover, many of its species grow exclusively in riparian areas and wetlands (often in apparently isolated populations), but with subtly different ecological optima. Finally, it has a relatively small genome with elevated synteny that has been successfully sequenced.

White and black poplars (*Populus alba* and *P. nigra*, respectively) grow along many water courses in the Iberian Peninsula, sometimes in sympatry, whereas European aspen (*P. tremula*) inhabits upland hillsides in parapatry. Neither poplar can interbreed naturally, although, in occasions, white poplar and European aspen do (*Populus* × *canescens*). Furthermore, some Iberian white poplar populations possess high clonality rates or hermaphroditic individuals.

In order to gain insight into the evolutionary and ecological significance of the reproductive system in Iberian *Populus*, this PhD work aimed at (i) studying its spatial genetic structure at the local and regional scale; (ii) evaluating the degree of local adaptation in white poplar through different approaches; (iii) researching the extent, causes and implications of clonality; (iv) appraising the ecological and evolutionary issues of hybridisation between white poplar and European aspen; and (v) enhancing the knowledge of the incipient *Populus* sex chromosome.

To achieve these goals, chloroplast DNA of numerous Iberian populations of poplars was characterised and quantitative genetic differentiation of several adaptive traits was evaluated among contrasting populations. Clonality was investigated in several Iberian populations, with a higher focus in one of them (the hybrid zone located in the proximity of Aranda de Duero), using nuclear microsatellites. These markers were also used to genotype European populations of *P. tremula*, including one from central Spain (Sierra de Gredos), as well as individuals from common garden trials – for which some phenotypic traits were also measured –, and an artificial backcross from *P.* × *canescens* to *P. alba* (BC_1), in the latter case with a higher focus on chromosome XIX. Additionally, several open-pollinated progenies from an Andalusian subdioecious (i.e. comprising males, females and some

hermaphrodites) white poplar population were genotyped. Finally, several genes involved in wood formation were sequenced in Iberian poplar populations that were representative of different gene pools.

White and black poplars displayed contrasting phylogeographic patterns in the Iberian Peninsula, reflecting that differences in ecological requirements (e.g. tolerance to cold temperatures) impose different regional genetic structures. In particular, white poplar confinement to river and drainage basins was higher than black poplar's. Southern populations of white poplar were more genetically diverse, and migration between the northern and southern regions was higher than between the eastern and western ones. These patterns suggest the effect of Holocene glaciations on poplars, despite their dependence of water courses. The presence of numerous private haplotypes indicated that no regional extinctions have taken place during glaciations in any of the studied river basins.

In white poplar, local adaptation was detected for two quantitative traits (total height and stem form at age 3), but at a wider spatial scale than in other forest trees (among river basins but not among populations within river basins). Molecular genetic adaptation imprints in wood formation genes were controversial: several outlier loci were found using one method, but none was detected using a more stringent analysis. This probably reflects sample size issues. In European aspen, local adaptation was also detected in quantitative traits. In addition, clear signatures on patterns of standing genetic variation, probably created by the admixture of two European distinct lineages, were uncovered for this species.

In white poplar, levels of clonality averaged 4.2 ramets per genet, in populations with 32-49 sampled ramets, genets reaching sizes of ca. 100 m. However, in two populations much larger clones (>10 km) were found. One of these populations, Aranda de Duero in north-central Spain, was more intensively studied. In Aranda de Duero, I found two large and ancient clones, one of them a hybrid with aspen, which spread for over a hundred kilometres. While genotypic richness and evenness was lower than in other European populations, genetic diversity was similar in Aranda de Duero and there was no apparent risk of genetic depauperation. Coalescence modelling showed that overall effective population size was declining, although large genets were probably increasing in ramet number. Besides, larger genets were generally parents of more individuals. Punctual

genomic introgression from *P. tremula* and heterozygosity could be involved in the accelerated clonal spread in this population.

Further insides on the role of interspecific heterozygosity in adaptation were obtained from the BC_1 cross. I found segregation distortion towards *P. tremula*, probably due to cyto-nuclear interactions or heterozygote advantage. This supports the idea that heterozygosity could play an important role in adaptation, in particular when enhanced through hybridisation.

Some findings agreed with expectations for a sex chromosome in chromosome XIX, while other disagreed. No recombination was detected in backcrossed progeny among four markers of chromosome XIX, as expected for sex chromosomes. However, the distribution of genetic diversity at this chromosome of separated male and female groups and its reduced divergence in natural populations were against sex chromosome expectations. Overall, my research supports the hypothesis that chromosome XIX is at a very early step of degeneration into a mature sex chromosome.

Finally, the paternity analyses done in the Jimena population revealed that the effective number of fathers involved in pollination ranged from 1.58 to 8.33. Clone size and distance were apparently involved in pollination rate. Regarding hermaphrodites, one of them effectively sired offspring. In addition, a correlation between feminity (average proportion of female flowers per catkin, considering hermaphrodites flowers as half-female flowers) and germination rate was detected.

Based on the results of this PhD thesis – that local adaptation and regional spatial genetic structure occur among Iberian poplar populations –, it is advisable to take provenance origin into account in restoration programmes and to consider a river-basin basis when implementing *in situ* conservation plans. The conspicuous extent of clonality makes it necessary to consider natural clone replicates in any management measure concerning these trees, although it is equally important to remark that such degree of clonality is not producing detrimental genetic effects in the studied population. Hybridisation and lineage admixture have proven to be valuable ways of genetic enrichment, with repercussions on fitness and local adaptation. Forest managers could put this fact to good use by adopting simple mitigation measures in response to future climate change threats.

Finally, chromosome XIX operates like a sex chromosome at an initial stage of development, which makes room for the occasional existence of hermaphrodites. A deeper

view on the interplay among hermaphroditism, sex chromosomes and clonality would allow a better understanding of chromosome XIX's relative ecological significance, and of white poplar's potential evolutionary pathways, as the species is an important dioecious riparian tree.

CHAPTER 1: General introduction

Section I. Evolution and Ecology of *Populus*

Concepts on speciation and hybridization

The concept of species, though intuitive, is quite difficult to demarcate, in particular for hybridising species. This is probably because it is a somewhat simplified categorical concept, derived from a gradual and complex process. Classically, the concept of species was linked to the development of Reproductive Isolation (RI) between conspecific populations (Biological Species Concept by MAYR 1963), according to the assumption that speciation was a whole-genome phenomenon and that genomes were highly cohesive. RI was considered crucial in the development of a new species. However, RI alone cannot explain species demarcation as, for instance, two species may hybridise, i.e. reproduce effectively, and still be considered as different.

Recently, empirical observations using higher resolution genetic tools have helped to understand the genetics of speciation, and a new concept of species has emerged. This new theory, known as the "genic view of speciation" (LEXER and WIDMER 2008; WU 2001), places the basis of speciation on "speciation genes" that drive initial divergence by differential adaptation rather than on whole-genome differentiation. According to this theory, two groups of individuals belong to two distinct species when they are differentiated and such differentiation would not be lost upon contact, i.e. they would be able to continue diverging. Within this new paradigm, RI is considered a mere by-product of differential adaptation.

Under the "genic view of speciation", differential adaptation at the early stages of speciation induces divergence in some genes whose varied alleles produce fitness differences in contrasting populations. With time, the speciation process continues by expanding the array of genes involved in differential adaptation and, eventually, with a diverging genomic architecture, blocking exchange of some genomic regions and increasing RI. Subsequently, many types of RI (intrinsic and extrinsic; see below) reach significant levels, including extensive genomic incompatibility and despite possible low levels of introgression (Figure 1.1). In the final stage, RI is complete and gene flow completely precluded (WU 2001).

Speciation can take place in sympatry, parapatry or allopatry (COYNE and ORR 2004). It can also occur after hybridization (Figure 1.1), as it has been demonstrated in landmark studies on hybrid sunflowers (UNGERER *et al.* 1998). Before those, hybridization was seen as a phenomenon that hampered "evolutionary noise", i.e. speciation and adaptation. Supporting the studies on hybrid sunflowers, recent discoveries have shown that hybridization can effectively contribute to the evolution of life (NOLTE and TAUTZ 2010), as two species can bear hybrids that, because of particular genetic features (BAACK and RIESEBERG 2007), are adapted to niches where none of the parental species can live (BURKE and ARNOLD 2001). If this "hybrid swarm" persists long enough, it can ultimately lead to a new species.

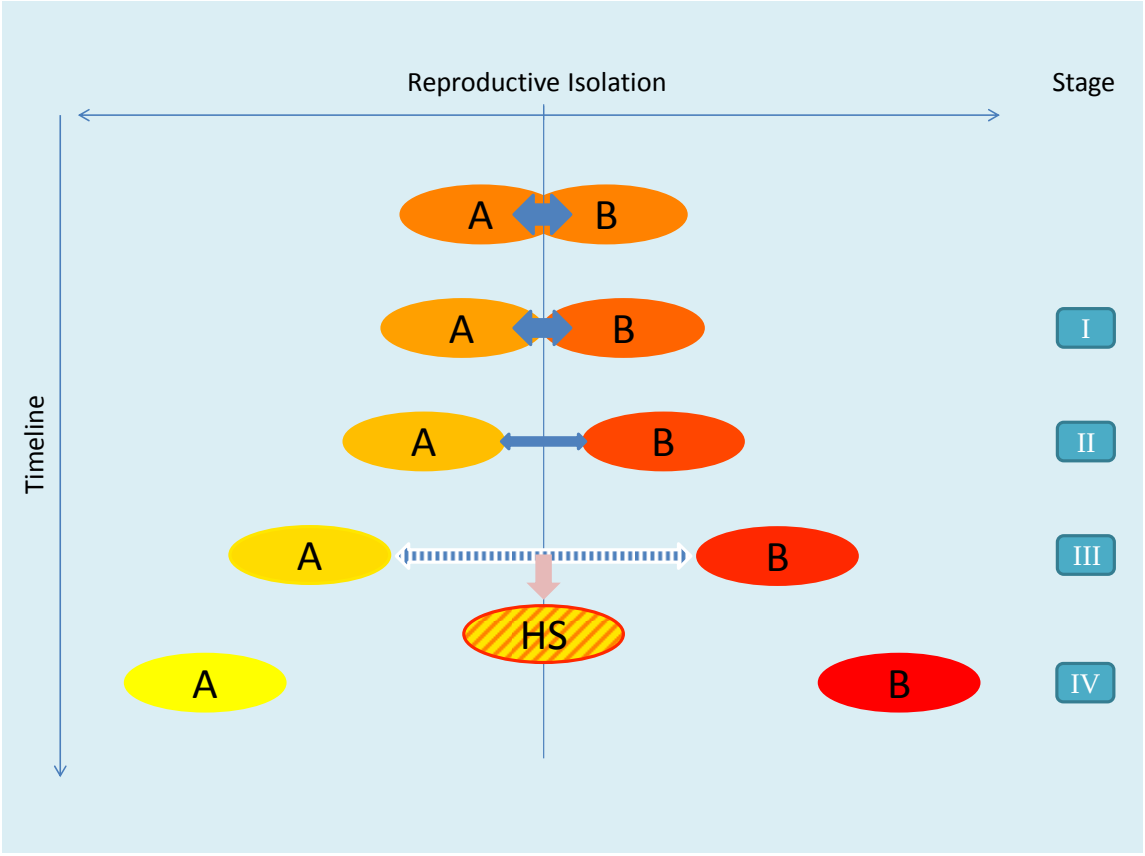


Figure 1.1 Schematic diagram of the process of speciation, representing the four stages described by WU (2001). Blue arrows represent gene flow between groups A and B. The wider the arrow, the more gene flow occurs. Dashed arrow indicates impeded gene flow. Similarity of colours indicates similarity between groups. Groups A and B begin as two populations of the same species and finish as two species unable to interbreed. HS: hybrid swarm.

RI can be prezygotic or postzygotic, depending on the stage in which reproduction is precluded, with the former playing a stronger role in speciation (RIESEBERG and WILLIS 2007; WIDMER *et al.* 2009). Postzygotic isolation can be driven by extrinsic and intrinsic mechanisms (COYNE and ORR 2004). Extrinsic isolation is linked to the possible reduction of fitness in hybrids, for instance by maladaptation to available niches or by increased difficulties to mate. Intrinsic isolation accounts for possible genomic incompatibilities and is usually related to genetic divergence (WIDMER *et al.* 2009), requiring a longer time to arise, although some genomic events, like polyploidy, can suddenly build intrinsic isolation (RIESEBERG and WILLIS 2007).

Sex chromosomes and sex-linked genes sometimes play an important role in the process of speciation. Even though, autosomal genes are responsible of divergence in its first stages, the particular characteristics of sexual chromosomes may render them a relevant factor at posterior speciation phases. In particular, they can promote the development of sexual isolation related to genes involved in sex-specific fitness (QVARNSTROM and BAILEY 2009).

Phylogeny and origin

Populus is a group of broadleaved trees (i.e. angiosperms) of the Malphigiales order, and *Salicaceae* family. It has 29 species, currently divided into six sections: *Populus* (formerly *Leuce*), *Tacamahaca*, *Aigeiros*, *Leucoides*, *Turanga*, and *Abaso* (ECKENWALDER 1996). This classification was based on morphological traits and crossability. Later, genetic divergence has revealed that genus *Populus* is not monophyletic, with section *Abaso* being less close than genus *Salix* to the rest of sections, and sections *Aigeiros*, *Leucoides* and *Tacamahaca* being intermingled (CERVERA *et al.* 2005). Sections *Populus* and *Turanga* are monophyletic.

In English, *Populus* trees are commonly referred to as poplars, cottonwoods and aspens, names that do not correspond to their phylogenetic arrangement. White poplars (*P. alba*), black poplars (*P. nigra*), and European aspens (*P. tremula*) are the native European members of the genus, but *P. alba* and *P. tremula* belong to section *Populus*, and *P. nigra* to section *Aigeiros*. Both poplar species are riverine trees, whereas aspens grow in more montane habitats, although in Northern Europe it also grows in the lowlands. White poplar has a wide distribution in Eurasia, reaching the Far East, but in Europe appears mainly across the Mediterranean basin. Black poplar occupies more oceanic climate stands. The three species are naturally present in the Iberian Peninsula. Some doubts about the native status

of *P. alba* have been dispelled by fossil remains and verified its autochthony (ROIRON *et al.* 2004).

The origin of the genus *Populu*s has been a controverted topic, with authors dating it from late Miocene (5-10 Mya; YIN *et al.* 2008) to the late Paleocene or early Eocene (50-60 Mya; KERSTEN *et al.* 2013). BELL *et al.* (2010) dated the separation between genus *Populus* and Salix at 31 Mya, although unambiguous fossil records place the origin no later than the Eocene, or probably during the late Paleocene (60 Mya) (SLAVOV and ZHELEV 2010; and references therein).

Phylogenetic analyses point at the following sequence of divergence of sections from the main trunk: *Abaso*, *Populus*, *Turanga*, and then the remaining three (CERVERA *et al.* 2005). Thus, section *Populus* is less related to *Turanga* than *Aigeiros*, for example, due to its earlier split. Accordingly, natural hybridization among *Tacamahaca*, *Aigeiros* and *Leucoides* is common, and likewise among section *Populus* members, but intersectional crosses between section *Populus* and *Aigeiros* (for example) is extremely rare. *P. alba* and *P. tremula* hybridise commonly in sympatry (producing *P.* × *canescens* or grey poplar), but *P. alba* and *P. nigra*, despite common sympatric occurrence, hardly cross in nature. However, the latter two species surely hybridised in the past, when reproductive isolation was less severe, as suggested by the study of chloroplast DNA, captured by *P. nigra* from section *Populus* (SMITH and SYTSMA 1990). Hybridization in the genus may have been facilitated by the high stability of its genome structure, as revealed by the elevated synteny observed between its species (CERVERA *et al.* 2001).

Ecological factors

Apart from the large variety of climatic regions where *Populus* species can live (from subpolar to subtropical latitudes), they all share some ecological constraints, as they all are intolerant to drought and to shade (SLAVOV and ZHELEV 2010). In consequence, the species that live in areas with seasonal drought must grow close to water bodies, such as river courses or wetlands and indeed, some species of *Populus* can be ecologically classified as riverine, with a direct dependence on water level dynamics. Upland or mountain species, which usually grow on hill slopes, can also occupy lowland plains in boreal regions (SLAVOV and ZHELEV 2010), where they are also characterised by a high tolerance to cold temperatures. Interestingly, this classification is not completely coincident with the phylogeny of the genus (e.g., in section *Populus*, *P. alba* is a riverine tree from warm to

temperate areas, whereas aspens such as *P. tremula* or *P. tremuloides* are typical montane boreal trees).

Populus species occupy an initial stage in the ecological succession, constituting pioneer species. This feature determines some reproductive system traits, especially the sexual system, whose reproductive success depends largely on the occurrence of catastrophic events, such as floods, fires, or ice scorns (SLAVOV and ZHELEV 2010). Their pioneer strategy is also responsible for the species' high growth rate and their intolerance to shade. Finally, some species (such as *P. alba*) present an elevated tolerance to salinity, a condition that permits them to proliferate in these unfavourable habitats.

All these factors make poplars, cottonwoods and aspens highly competitive in areas under recurrent disturbances with elevated soil moisture, but they could also impede the development of stable, large, and continuous populations. However, stable, large populations occur in scarcely disturbed habitats. In these environments with almost absent disturbances, it may seem that the species must rely on clonality to assure persistence. However, neither the details on ecological succession, nor the way in which disturbances influence poplars' ability to outcompete other trees in riverine or hillside sites, are well known.

Phylogeographical patterns

Like in other plant groups, *Populus* phylogeographical patterns have been more thoroughly studied in cold-temperate and boreal regions, where glaciations depleted almost all populations, than in warm-temperate areas, where glacial refugia were abundant. Phylogeographical signatures in European poplars from boreal or oceanic ranges indicate rapid expansions with a strong genetic drift, resulting in low genetic diversity in the colonised range except for contact areas between lineages, which normally show increased genetic diversity and admixed genotypes. Black poplar persisted in Europe in three glacial refugia, located in the Iberian, Italic and Balkan Peninsulas (COTTRELL *et al.* 2005). Higher diversity has been detected in the Iberian Peninsula – the precise location of the refugia is unknown –, although most of these haplotypes could not have recolonised Europe because of the Pyrenees, an effective geographical barrier. A more local study, that did not include the Iberian Peninsula, revealed two white poplar glacial refugia in Italy and Romania and inferred a secondary contact zone for this species in Central Europe (FUSSI *et al.* 2010). With respect to European aspen, COTTRELL *et al.* (2005) did not find any conspicuous

phylogeographic structure, as expected for a boreal species. Other local studies have established the autochthonous origin of white poplar populations in Malta (FUSSI *et al.* 2012) and Sardinia (BRUNDU *et al.* 2008). Likewise, no structure was found in the sister genus *Salix*, specifically in the boreal species sallow (*Salix caprea*), which was attributed to different factors: rapid expansion, the presence of cryptic northern refugia, an elevated mutation rate, or possible hybridization with other willows (PALME *et al.* 2003).

A considerable number of studies in North America have focused on the continental-range genetic structure of boreal *Populus* species (BREEN *et al.* 2012; CALLAHAN *et al.* 2013; KELLER *et al.* 2010). However, the only work on the genetic structure of temperate poplars at continental-scale is the above mentioned study on European black poplar (COTTRELL *et al.* 2005). Another few have been performed at regional-scale in continental areas (Central Europe, FUSSI *et al.* 2010; Tibetan Plateau, PENG and CHEN 2011), and at local-scale in some Mediterranean islands (BRUNDU *et al.* 2008; (FUSSI *et al.* 2012), but no work to date has analysed genetic structure within a main glacial southern refugium.

Two features distinguish European poplars from most temperate forest trees: they grow along river courses – being highly azonal species –, and they can produce clonal propagates. Both qualities can affect how poplar genetic diversity has responded to both glaciations and the subsequent recolonization of formerly glaciated areas. Riverine habitats buffer climatic change but can also difficult post-glacial migration. Likewise, clonal propagation may increase population endurance to adverse climatic conditions where sexual reproduction is precluded, but may come at the cost of a reduced genotypic diversity. For example, a study on a clonal epiphytic bryophyte correlated intensity of glaciations with decreased genetic diversity and higher use of clonal reproduction (CRONBERG 2000).

Population genetic structure, gene flow and adaptation

In general, most *Populus* species can be considered as widespread forest trees (SLAVOV and ZHELEV 2010). Although often scattered over vast territories, *Populus* groves and stands seem to be effectively connected by high rates of gene flow (e.g., SLAVOV *et al.* 2009), probably due to wind dispersal of pollen and seeds, thus blurring the species' genetic structure. However, direct evaluation of seed dispersal in *Populus* is still incomplete (SLAVOV and ZHELEV 2010 and references therein). For example, a continent-wide study on the spatial genetic structure of black poplar found that historical gene flow must have

been considerable across large distances, including between river catchments (SMULDERS *et al.* 2008). In apparent contrast, another two studies on the same species revealed some spatial genetic structure within (RATHMACHER *et al.* 2010) and among populations (IMBERT and LEFEVRE 2003), something that had been attributed to short-distance gene flow in the former study.

The reason may be that gene flow is influenced by population structure and that it is difficult to estimate by indirect methods. For example, a study on *P. trichocarpa* (SLAVOV *et al.* 2009) revealed that observed pollen immigration seems to be negatively affected by the number of mature males in the neighbourhood of mothers, and that structure can also affect male genetic neighbourhood size. A later study on the same populations showed that natural regeneration involving only a few mothers severely affected population genetic structure (SLAVOV *et al.* 2010). A regional analysis on *P. tremuloides* (in Wisconsin, US) found that differentiation among populations was low, hinting at the occurrence of gene flow (COLE 2005). Genetic differentiation was present, albeit also low, between populations of Italian white poplar, two of them separated by more than 400 km (CASTIGLIONE *et al.* 2010). The number of effective migrants between these two populations was above three, which can be interpreted as absence of isolation. Between two populations of Central and Eastern Europe separated by more than 700 km, the number of migrants per generation was similar ($N_e m > 3$), and between two populations of European aspen (separated by 240 km) this value was close to 15 in one direction (LEXER *et al.* 2005).

Despite high gene flow between distant populations, local adaptation has also been detected in *Populus* by some studies. For instance, phenotypic differentiation related to adaptive traits was recently found in a study on *P. balsamifera* (KELLER *et al.* 2011b). Adaptation is also manifested in the clinal variation of genes with adaptive relevance (INGVARSSON *et al.* 2006; MA *et al.* 2010), for example in some flowering-time genes in *P tremula* (HALL *et al.* 2011; INGVARSSON 2010). Signatures of local adaptation have also been found in genes involved in phenology in *P. balsamifera* (KELLER *et al.* 2011a). It is intuitive to consider that clonality may affect adaptive patterns, but to my knowledge, no study has tackled this topic. Given their reduced generation turnover, clonal populations could, for example, slow down micro-evolutionary processes due to longer generation times. Some *Populus* clones are able to persist for several centuries and thus buffer the loss of adaptive or detrimental polymorphisms. Nonetheless, even if sexual reproduction was completely

precluded, micro-evolution would occur anyway, albeit at slow pace, since competition among clones would always be subjected to selection pressure.

Conservation genetics and restoration issues

Most *Populus* species are so widespread that they do not present any risk of whole-species extinction. However, genetic diversity within the species is so ample that some local varieties and populations are surely at higher risk of disappearance. Also, given the genus' economic importance, the numerous artificial plantations growing nearby natural populations pose a potential danger.

Namely, three genetic issues could pose proximal risks in many *Populus* populations: (i) introgression from artificial plantations for timber or pulp yield; (ii) genetic isolation of populations after fragmentation by human-induced changes in land use; and (iii) increase of clonality due to impairment of effective sexual recruitment produced by the alteration of ecological conditions.

The first risk derives from the extensive plantations of hybrid poplars for production of timber and paper pulp. Such plantations are popular due the facility of clonal replication, the high capacity for hybridization, and the elevated vigour of first generation hybrids. Some concerns have been raised as to the effect of these commercial plantations on nearby natural groves of *Populus*, with which they may hybridise and thus suffer genetic introgression (as in FOSSATI *et al.* 2003; HEINZE 2008). Such alloctonous gene flow could damage native populations by the introduction of maladapted genes or by the reduction of genetic diversity, as plantations are genetically poor (HIDALGO *et al.* 2010) – normally, they are composed of one clone possessing specific qualities, and this clone is almost always a hybrid between two species, usually not native. In theory, artificial populations can also naturalise and thus directly displace natural individuals, although this seems to happen very rarely, as the many studies carried out on this topic conclude that there are many pre- and postzygotic reproductive barriers that reduce potential introgression (HIDALGO *et al.* 2010; and references therein). However, more thorough studies on the genomics of introgression and reproductive isolation are still necessary to fully evaluate the extent of this threat.

The second risk, fragmentation, is due to the fact that the habitat of many species of *Populus*, and especially that of riverine poplars, is usually land claimed for human uses, due to its high productivity. Therefore, *Populus* populations are usually constrained to discontinuous stands, thus reducing the effective population size. Although some studies in

wind-pollinated trees have revealed that gene flow can be maintained effectively even between distant populations (e.g., ROBLEDO-ARNUNCIO 2011), it is not clear how poplars can deal with this issue (but see SLAVOV *et al.* 2009).

Finally, the third risk, increased clonality, can result from persistence assurance when sexual reproduction is lacking (VALLEJO-MARIN *et al.* 2010). So ecological conditions that impede sexual reproduction should also promote, or at least permit, clonal reproduction. In poplars, many anthropic factors can foster increased clonal rates and impede sexual reproduction: water regime control, reduction of underground water table, droughts associated to Climate Change, etc. It is known that, at least in first stages, increased clonal rates do not affect allelic diversity (more details in Chapter 6), but rather genotypic diversity, which can weaken a population's evolutionary potential. Besides, the effects on allelic diversity of a long-term reduction in genotypic diversity are unknown, and some authors have already sounded the alarm about possible populations dead-ends (HONNAY and BOSSUYT 2005).

Any of the three genetic threats described above may affect genetic diversity and push valuable populations of poplars into extinction vortexes. Determining their degree of relevance for European poplars would allow to establish the thresholds beyond which populations would be vulnerable to extinction.

Section II. Reproductive systems

Introductory concepts

Reproductive systems can be classified as sexual and asexual. Broadly, sexual reproduction comprises modes of reproduction in which the offspring is product of the genetic mixing of *two parents*, while asexual reproduction implies that the offspring's genes are originated by, and *identical* to, one parent only. These definitions are not completely accurate, though. On the one hand, as selfing is also a form of sexual reproduction, a more accurate definition of sexual reproduction would therefore be the production of offspring with a significantly different genetic composition to that of its parents, which can be one or two. On the other hand, defining asexual offspring as an identical genetic copy of its parent is not accurate either, due to the almost unavoidable occurrence of somatic mutations.

Recombination is the main biological process by which mixing of genetic composition takes place, via meiosis, allowing for sexual reproduction.

Plants, like animals, usually reproduce sexually, with some groups (birds, mammals) only being able to produce sexual offspring. Actually, this mode of reproduction is present in almost all eukaryote groups – with the conspicuous exception of bdelloid rotifers (WELCH and MESELSON 2000). The ubiquity of sexual reproduction contrasts with the additional costs it implies when compared to asexual reproduction. Contacting and selecting partners, the reduction of the overall number of offspring (i.e. the twofold cost of sexual reproduction), or the production of a share of aberrant or maladapted offspring are some examples of these extra expenses (OTTO and LENORMAND 2002). Indeed, the ecological and evolutionary reasons that support sexual reproduction remain undetermined, although many hypotheses (like the existence of a trade-off with offspring fitness) are in the focus of ongoing research (OTTO 2009).

Sexual and asexual reproduction in plants

Plant sexual reproduction is essentially linked to 'hermaphroditism', the biological property of producing female and male fertile gametes by an organism. When dealing with plants, the term usually restricts to the ones whose every flower can produce both gametes. In case of a plant producing both gametes, but in different flowers, it is termed 'monoecy'. Both kinds of plants are known as 'co-sexual'. Plant species that have individuals which produce either female or male gametes are called 'dioecious'. These species are sexually dimorphic. Dioecious species are scarce (*ca.* 4% of higher plants; AINSWORTH 2000), but widespread in the plant phylogenetic tree, and certainly paraphyletic.

Although sexual reproduction is present in all the families of the kingdom Plantae, asexual reproduction is also a frequent mode of reproduction, in combination with sex. The capability of reproducing asexually was not present in the onset of angiosperms (TIFFNEY and NIKLAS 1985) but was a paraphyletic advance that evolved independently several times (KLIMES *et al.* 1997; SACHS 2001).

Many plants (and also some animals) are able to reproduce both sexually and asexually, depending on the circumstances (FRYXELL 1957). Within such plants, a means for asexual (clone) propagation is vegetative growth, which consists in replicating modules from the sourcing plants, through more or less specialised organs. Some plants like willows (genus

Salix) just drop branches, which afterwards root, thus generating new offspring, whereas others have more specialised organs (like stolons) to asexually colonise new areas.

Upon which circumstances asexual reproduction is favoured continues under study. It seems that asexual reproduction prevails when sexual reproduction is difficult to achieve (AARSSEN 2008). During the first life stages, when offspring are more vulnerable, clonal offspring maintain connection to parents, which confers them abundant resources, in contrast to sexual offspring. Therefore asexual reproduction is favoured as a safer propagation mode under harsh environmental conditions. It is also argued that clonal assemblies, when integration among ramets is maintained, permit them to forage different resources in different areas (ALPERT and STUEFER 1997). Accordingly, asexual propagation will be also favoured in limiting soils.

Evolutionary trade-offs between sexual and asexual reproduction

Asexual reproduction requires investing resources drawn from sexual reproduction. Trade-offs on fitness would depend on several factors, which have been moderately studied at ramet-level, but hardly addressed at genet-level (VALLEJO-MARIN *et al.* 2010), a level with a higher evolutionary importance. Fitness returns due to clonality are surely higher when conditions for sexual recruitment are impaired. Moreover, fitness gain curves of investment in sexual reproduction at ramet-level are not linear, but decreasing in slope, so diversion of resources towards clonality would increase total fitness, due to genets taking advantage of the steepest part of the gain curve (VALLEJO-MARIN *et al.* 2010).

Conversely, clonal reproduction can increase the risk of self-pollination, which could decrease fitness by several means: causing inbreeding depression in self-compatible species due to geitonogamy – i.e. the pollination of a flower with the pollen from another flower on the same flowering plant, or pollen discounting in self-incompatible species (CHARPENTIER 2001). The latter may nevertheless use clonality as a means of persistence when mates are scarce. For their part, selfing species will rarely favour clonality when inbreeding depression is low, but will do when inbreeding indeed impairs fitness (VALLEJO-MARIN *et al.* 2010).

Larger plants have higher probabilities of selfing and inbreeding depression, making their clones also more prone to suffer from these drawbacks. Self-pollination can be reduced by dioecy and, indeed, relations between plant size and dioecy have been found (BARRETT 2002; VAMOSI *et al.* 2003). Similarly, dioecy could be induced by clonality: if the gain curves

for sex function are different depending on the sex, dioecy could be favoured as well in clonal plants to facilitate higher replication of one of the sexes, biasing the sex ratio, to take advantage of the steepest part of the curve (DORKEN and VAN DRUNEN 2010).

Consequences of modes of reproduction on genetic diversity and structure

Not much is known about the ecological and evolutionary factors that lead to different modes of reproduction and their implication in the patterns of genetic diversity and structure has hardly been investigated. The main trouble encountered to disentangle these plausible links is the difficulty to test hypotheses on natural populations. Firstly, because many concurrent factors appear to affect the evolution of reproduction modes, so it is extremely challenging to determine the degree of influence of each factor with manageable sample sizes or to set an experiment on natural populations keeping control, at least, of the main known factors.

Nonetheless, some theoretical approaches have been applied, and in some cases, experimental or observational evidence has supported them. Numerical models suggest that asexual reproduction has no effect on genetic diversity levels. If any, it rather helps to conserve them, since it slows the action of genetic drift along time (BALLOUX *et al.* 2003). Some field studies in *Populus trichocarpa* concurred with this suggestion (SLAVOV *et al.* 2010). In any case, the long-term effects of asexual reproduction are not clear, as analyses have been conducted in contrasting populations at a single time. Another suggestion is that outcrossing and "guerrilla-like" clonal propagation could be linked (CHARPENTIER 2001). Clonal populations can present distinctive spatial genetic structures (SGS) dependant on their clonal architecture. Modes of clonal growth deliver different degrees of clonal aggregation ("guerrilla" or "phalanx"). In general, clonality increases SGS, rising kinship coefficients at shorter distance classes in correlograms (as in VAN LOO *et al.* 2008). Estimation of the contribution of clonal growth to the total SGS needs to rely on both the ramet and genet levels to be informative (VALLEJO-MARIN *et al.* 2010).

Reproduction in Populus

Sexual reproductive modes of species from the genus *Populus* are quite uniform (WYCKOFF and ZASADA 2008). All species are primarily pollinated by means of wind (anemophilous),and disperse seeds alike (anemochorous). Almost all are dioecious (except monoecious *P. lasiocarpa*).

Sex determination of flowers in the genus *Populus* takes place before organ initiation (BRUNNER 2010), producing the abortion of inappropriate gender flowers, something that is rare in plants (AINSWORTH 2000). However, sex lability has been reported repeatedly in many *Populus* species.

Most *Populus* can reproduce asexually, by clonality, but this capacity differs among sections (SLAVOV and ZHELEV 2010), being much more common in section *Populus*, for whom huge clones have been found. In many of these species, sexual reproduction is relatively rare because of limited seed longevity – which prevents the formation of seed banks –, and the requisition of concurrent circumstances to accomplish recruitment, in particular related to bareness of sites and moisture levels, which is uncommon (GONZALEZ *et al.* 2010a; and references therein). River regulation has been shown to be an additional impairing factor for sexual reproduction (GONZALEZ *et al.* 2010b). Indeed, under certain circumstances, sexual reproduction is so impeded that clonality is the only way of persistence. For example, white poplar (*P. alba*) resorts to clonal propagation at different rates apparently depending on ecological conditions: the dryer the conditions, the higher the presence of clonal assemblies (BRUNDU *et al.* 2008; CASTIGLIONE *et al.* 2010; FUSSI *et al.* 2012). The existence of ancient large clones in this group of species is an intriguing fact and discerning whether it is a healthy surviving strategy or, on the contrary, a morbid population stage is relevant.

As mentioned above, members of genus *Populus* hybridise easily, especially within sections or between related sections. Hybrid vigour is usually observed in these interspecific crosses, and has been used in agronomy to improve wood yield.

Hybrid zones in Populus

Hybrid zones are contact areas between two populations of hybridising species, where natural hybrids occur. Such extraordinary zones hold valuable characteristics for the study of various genetic and evolutionary processes (described as "natural labs"; HEWITT 1988; LEXER and VAN LOO 2006). Hybrid zones are especially interesting for studying forest trees (LEXER *et al.* 2004), because they can replace alternative extremely time-consuming approaches – due to long generation times in trees – like controlled crosses or long-term common garden experiments. In particular, hybrid zones are ideal to study the mechanics underlying speciation and hybridization, the patterns of adaptive genetic variation and the evolution of sexual systems. For example, through admixture mapping in hybrid zones it is

feasible to link genomic markers with adaptive traits and reproductive isolation factors (BUERKLE and LEXER 2008; LEXER *et al.* 2007). Finally, hybrid zones also facilitate the generation of genetic (linkage) maps in natural populations.

The study of several contact zones between *P. alba* and *P. tremula* across Europe has supplied relevant information on the genomic architecture of this genus, and the extent and mechanisms of speciation (LEXER *et al.* 2005; LEXER *et al.* 2010; STOELTING *et al.* 2013). They have shown that the genomes of these two species are not evenly permeable, and that reproductive barriers prevent species merging. These studies have also unveiled some of the genetic architecture underlying phenotypic traits in poplars (LINDTKE *et al.* 2013), revealed transgressive traits in hybrids (LEXER *et al.* 2009) and provided evidence of hybrid advantage in clonal propagation (VAN LOO *et al.* 2008). Additionally, research in hybrid zones has hinted at the likely importance of introgression in refurbishing the genetic pool of some species, since recombinant hybrids maintain a high level of heterozygozity at many loci (LINDTKE *et al.* 2012).

Sex chromosomes

In sexually dimorphic species, sex determination is extremely varied. In plants, the mechanisms range from exclusively genetic to environment-dependant ones (MING *et al.* 2007). Within partially or completely genetic mechanisms, the variety is also high. Among them, sex chromosomes are an extreme case of genetic sex determination. In species with such chromosomes, one of the sexes is then homogametic (i.e. has the same version of sex chromosome), and the other heterogametic (i.e. has one version of each sex chromosome). In mammals, males are the heterogametic sex (XY chromosomes) and females the homogametic (XX). Conversely, in birds, males are the homogametic (ZZ chromosomes). In homogametic individuals, sex chromosomes recombine similarly to autosomes.

While sex chromosomes can only (but not always) exist in sexual dimorphic species, and are intimately linked to differential sexual development, sexual determination loci are not always placed in the sex chromosomes (i.e. *Rumex acetosa* and *Humulus japonicus*; WESTERGAARD 1958). Not all pairs of sex chromosomes must be heteromorphic (visually different), since their differences could be only in arrangement or composition rather than in size (MING *et al.* 2007). Indeed, the diversity of sexual chromosomes is elevated, from full development of heteromorphic sex chromosomes to the loss of one of them (i.e. the

X0 system). Such diversity appears to be a consequence of the development of sex chromosomes.

Sex chromosomes evolve from autosomes. The initiation of sex chromosomes is linked to the suppression of recombination in normal autosomes (Figure 1.2). When important genes for differential sexual development coincide in the same genomic region, recombination is reduced by selection to avoid recombinants with antagonistic alleles. Depending on which function (male or female) these genes control, recombination will be supressed in one gender, while maintained in the other. Since recombination is suppressed, degeneration to a variable degree in the non-recombining region will take place (BACHTROG 2013). This degeneration takes place by gene dysfunction or complete loss, genomic rearrangements (as repetitions, transposons or inversions), background selection, or Hill-Robertson effects (MING *et al.* 2007). The lack of recombination leads also to an increased divergence in this area, between the degenerating chromosome and its non-degenerated pseudo-homologous. As degeneration goes on, some essential genes are lost, sometimes producing a sex bias towards the homogametic gender, following HALDANE's rule (1922). Further degeneration produces an enlargement of physical size of the non-recombining region, and also the extension of recombination suppression to adjacent areas (new strata), reducing the pseudo-autosomal region (which still recombines with its non-degenerating pair; BERGERO and CHARLESWORTH 2009). As this process goes on and more parts of the heteromorphic chromosome degrade, genomic areas can be lost, leading to a reduction of sex-chromosome size (the current situation in Eutherian mammals). Finally, if the pseudo-autosomal region disappears, the degenerated chromosome will be lost, resulting in an X0 system (as in *Rumex* spp.; MING *et al.* 2011). The rate of evolution of sex chromosomes seems to vary among species (MING and MOORE 2007).

Incipient Populus sex chromosomes

Chromosome base number in genus *Populus* is consistently 19, with high synteny among species (CERVERA *et al.* 2001). Hence, individuals, commonly diploid, have a load of 2n = 38, although triploid and tetraploid genets have been reported in nature, especially in the section *Populus* and its hybrids. Genome sequencing also revealed a whole genome duplication pre-dating the separation of *Populus* and *Salix* lineages, known as 'salicoid' duplication (see TUSKAN *et al.* 2012 and references therein). Their common ancestor possessed 12 chromosomes.

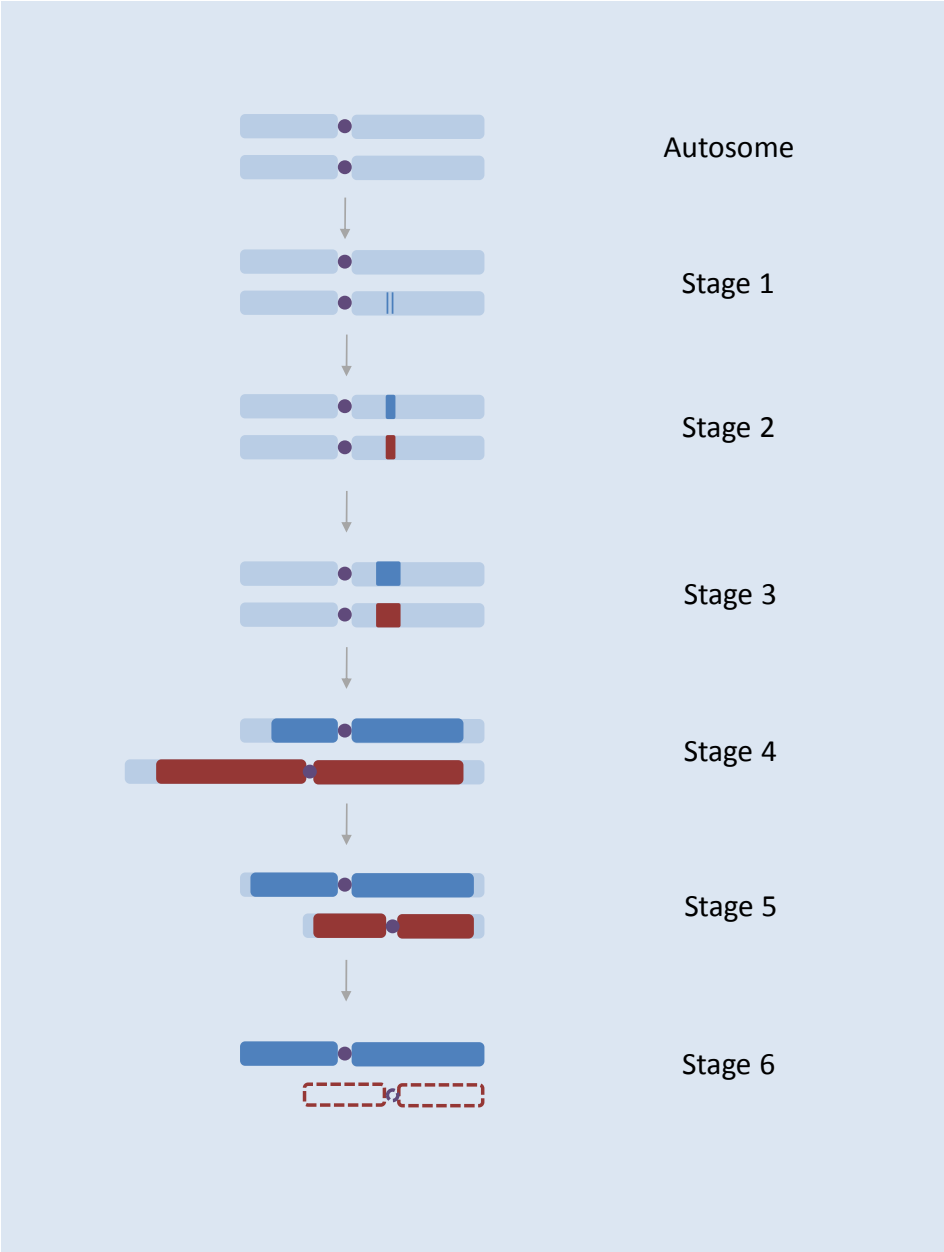


Figure 1.2 Steps in sex chromosome evolution. Stages numbered as in MING *et al.* (2011). Light blue represents autosomal or pseudo-autosomal regions. Intersecting blue lines represent male sterile and female sterile mutations. Maroon represents degenerated male specific region (MSR). Dark blue represents MSR corresponding region on X. Solid dot is the centromere. Adapted from MING *et al.* (2011).

Poplar chromosomes are metacentric, small, and uniform in size (see YIN *et al.* 2008 and references therein). None of these pairs display size heteromorphy (PETO 1938), indicative of advanced sex chromosome development. Nevertheless, the oncoming of *Populus trichocarpa* (section *Tacamahaca*) genome sequence (clone Nisqually-1, female; TUSKAN *et al.* 2006), unveiled that a peritelomeric region of 706 kb of chromosome XIX possessed a sex determining locus and suppressed recombination, only present in one haplotype of the female genotypes (YIN *et al.* 2008), pointing to an incipient ZW system (female is the heterogametic gender). With the aid of fine-scale genetic mapping, and several interspecific crosses, YIN *et al.* (2008) found, across this region, suppressed recombination, distorted segregation and haplotype divergence between the alternate haplotypes, solely in the maternal parent. Moreover, a male-bias sex ratio was detected in the interspecific cross, as Haldane's rule predicts.

Not much before, sex had been mapped in the peritelomeric region of same chromosome (GAUDET *et al.* 2008) in a cross of *P. nigra* (section *Aigeiros*). Afterwards, an intraspecific cross of the closely related aspen *P. tremula* × *P. tremuloides* (both in section *Populus*) placed the sex determining region in the central area of chromosome XIX, in the *P. tremuloides* male parent (PAKULL *et al.* 2009). Soon after, an intraspecific cross in *P. alba* (section *Populus*), sited the sex determination locus also in the pericentromeric region, but in the female parent (PAOLUCCI *et al.* 2010). Recently, KERSTEN *et al.* (2013) validated the result of the aspen cross (*P. tremula* × *P. tremuloides*), and estimated indirectly that the size of the non-recombining region was about two million base pairs.

In spite of the elevated synteny within genus *Populus*, sex determining loci appear to be in different positions in chromosome XIX. Likewise, the heterogametic sex is not constant. Apparently, members of the related sections *Aigeiros* and *Tacamahaca* would have gender-determining loci close to the telomere, while in species of section *Populus* it would be by the centromere (TUSKAN *et al.* 2012). Sections *Aigeiros* and *Populus* would have a male heterogametic (XY system), in contrast to section *Tacahamaca*. Interestingly, TUSKAN *et al.* (2012) cited unpublished data on a *P. alba* cross in which the hetereogametic sex was the male parent, consistent with the findings in aspens.

References

AARSSEN, L. W., 2008 Death without sex - the 'problem of the small' and selection for reproductive economy in flowering plants. Evolutionary Ecology **22:** 279-298.

AINSWORTH, C., 2000 Boys and girls come out to play: The molecular biology dioecious plants. Annals of Botany **86:** 211-221.

ALPERT, P., and J. F. STUEFER, 1997 Division of labour in clonal plants, pp. 137-154 in *Ecology and Evolution of Clonal Plants*, edited by H. DE KROON and J. VAN GROENENDAEL. Backhuys Publishers, Leiden, The Netherlands.

BAACK, E. J., and L. H. RIESEBERG, 2007 A genomic view of introgression and hybrid speciation. Current Opinion in Genetics & Development **17:** 513-518.

BACHTROG, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nature Reviews Genetics **14:** 113-124.

BALLOUX, F., L. LEHMANN and T. DE MEEUS, 2003 The population genetics of clonal and partially clonal diploids. Genetics **164:** 1635-1644.

BARRETT, S. C. H., 2002 The evolution of plant sexual diversity. Nature Reviews Genetics **3:** 274-284.

BELL, C. D., D. E. SOLTIS and P. S. SOLTIS, 2010 The age and diversification of the angiosperms re-revisited. American Journal of Botany **97:** 1296-1303.

BERGERO, R., and D. CHARLESWORTH, 2009 The evolution of restricted recombination in sex chromosomes. Trends in Ecology & Evolution **24:** 94-102.

BREEN, A. L., D. F. MURRAY and M. S. OLSON, 2012 Genetic consequences of glacial survival: the late Quaternary history of balsam poplar (*Populus balsamifera* L.) in North America. Journal of Biogeography **39:** 918-928.

BRUNDU, G., R. LUPI, I. ZAPELLI, T. FOSSATI, G. PATRIGNANI *et al.*, 2008 The origin of clonal diversity and structure of *Populus alba* in Sardinia: evidence from nuclear and plastid microsatellite markers. Annals of Botany **102:** 997-1006.

BRUNNER, A., 2010 Reproductive development in *Populus*, pp. 155-170 in *Genetics and Genomics of Populus*, edited by S. JANSSON, R. P. BHALERAO and A. T. GROOVER. Springer Science+Business Media, LLC, New York.

BUERKLE, C. A., and C. LEXER, 2008 Admixture as the basis for genetic mapping. Trends in ecology & evolution (Personal edition) **23:** 686-694.

BURKE, J. M., and M. L. ARNOLD, 2001 Genetics and the fitness of hybrids. Annual Review of Genetics **35:** 31-52.

CALLAHAN, C. M., C. A. ROWE, R. J. RYEL, J. D. SHAW, M. D. MADRITCH *et al.*, 2013 Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*). Journal of Biogeography **40:** 1780-1791.

CASTIGLIONE, S., A. CICATELLI, R. LUPI, G. PATRIGNANI, T. FOSSATI *et al.*, 2010 Genetic structure and introgression in riparian populations of *Populus alba* L. Plant Biosystems **144:** 656-668.

CERVERA, M. T., V. STORME, B. IVENS, J. GUSMAO, B. H. LIU *et al.*, 2001 Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. Genetics **158:** 787-809.

CERVERA, M. T., V. STORME, A. SOTO, B. IVENS, M. VAN MONTAGU *et al.*, 2005 Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. Theoretical and Applied Genetics **111:** 1440-1456.

CHARPENTIER, A., 2001 Consequences of clonal growth for plant mating. Evolutionary Ecology **15:** 521-530.

COLE, C. T., 2005 Allelic and population variation of microsatellite loci in aspen (*Populus tremuloides*). New Phytologist **167:** 155-164.

COTTRELL, J. E., V. KRYSTUFEK, H. E. TABBENER, A. D. MILNER, T. CONNOLLY *et al.*, 2005 Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA. Forest Ecology and Management **206:** 71-90.

COYNE, J. A., and H. A. ORR, 2004 *Speciation*. Sinauer Associates, Sunderland, MA.

CRONBERG, N., 2000 Genetic diversity of the epiphytic bryophyte *Leucodon sciuroides* in formerly glaciated versus nonglaciated parts of Europe. Heredity **84:** 710-720.

DORKEN, M. E., and W. E. VAN DRUNEN, 2010 Sex allocation in clonal plants: might clonal expansion enhance fitness gains through male function? Evolutionary Ecology **24:** 1463-1474.

ECKENWALDER, J. E., 1996 Systematics and evolution of *Populus*, pp. 7-32 in *Biology of Populus and its implications for management and conservation*, edited by R. F. STETTLER, H. D. BRADSHAW, P. E. HEILMAN and T. M. HINCKLEY. NRC Research Press, Ottawa, Ontario, Canada.

FOSSATI, T., F. GRASSI, F. SALA and S. CASTIGLIONE, 2003 Molecular analysis of natural populations of *Populus nigra* L. intermingled with cultivated hybrids. Molecular Ecology **12:** 2033-2043.

FRYXELL, P. A., 1957 Mode of reproduction of higher plants. Botanical Review **23:** 135-233.

FUSSI, B., J. BONELLO, E. CALLEJA and B. HEINZE, 2012 Combining the use of molecular techniques and archival documentary evidence to trace the origin of *Populus alba* in a Central Mediterranean archipelago. European Journal of Forest Research **131:** 347-354.

FUSSI, B., C. LEXER and B. HEINZE, 2010 Phylogeography of *Populus alba* (L.) and *Populus tremula* (L.) in Central Europe: secondary contact and hybridisation during recolonisation from disconnected refugia. Tree Genetics & Genomes **6:** 439-450.

GAUDET, M., V. JORGE, I. PAOLUCCI, I. BERITOGNOLO, G. S. MUGNOZZA *et al.*, 2008 Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genetics & Genomes **4:** 25-36.

GONZALEZ, E., F. ANTONIO COMIN and E. MULLER, 2010a Seed dispersal, germination and early seedling establishment of *Populus alba* L. under simulated water table declines in different substrates. Trees-Structure and Function **24:** 151-163.

GONZALEZ, E., M. GONZALEZ-SANCHIS, A. CABEZAS, F. A. COMIN and E. MULLER, 2010b Recent changes in the riparian forest of a large regulated Mediterranean river: Implications for management. Environmental Management **45:** 669-681.

HALDANE, J. B. S., 1922 Sex ratio and unisexual sterility in hybrid animals. Journal of Genetics **12:** 101-109.

HALL, D., X.-F. MA and P. K. INGVARSSON, 2011 Adaptive evolution of the *Populus tremula* photoperiod pathway. Molecular Ecology **20:** 1463-1474.

HEINZE, B., 2008 Genetic traces of cultivated hybrid poplars in the offspring of native *Populus nigra* in Austria. Preslia **80:** 365-374.

HEWITT, G. M., 1988 Hybrid zones - natural laboratories for evolutionary studies. Trends in Ecology & Evolution **3:** 158-167.

HIDALGO, E., S. C. GONZÁLEZ-MARTÍNEZ, C. LEXER and B. HEINZE, 2010 Conservation genomics, pp. 15-38 in *Genetics and Genomics of Populus*, edited by S. JANSSON, R. P. BHALERAO and A. T. GROOVER. Springer Science+Business Media, LLC, New York.

HONNAY, O., and B. BOSSUYT, 2005 Prolonged clonal growth: escape route or route to extinction? Oikos **108:** 427-432.

IMBERT, E., and F. LEFEVRE, 2003 Dispersal and gene flow of *Populus nigra* (*Salicaceae*) along a dynamic river system. Journal of Ecology **91:** 447-456.

INGVARSSON, P. K., 2010 Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. Molecular Biology and Evolution **27:** 650-660.

INGVARSSON, P. K., M. V. GARCIA, D. HALL, V. LUQUEZ and S. JANSSON, 2006 Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). Genetics **172:** 1845-1853.

KELLER, S. R., N. LEVSEN, P. K. INGVARSSON, M. S. OLSON and P. TIFFIN, 2011a Local selection across a latitudinal gradient shapes nucleotide diversity in Balsam Poplar, *Populus balsamifera* L. Genetics **188:** 941-U318.

KELLER, S. R., M. S. OLSON, S. SILIM, W. SCHROEDER and P. TIFFIN, 2010 Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, *Populus balsamifera*. Molecular Ecology **19:** 1212-1226.

KELLER, S. R., R. Y. SOOLANAYAKANAHALLY, R. D. GUY, S. N. SILIM, M. S. OLSON *et al.*, 2011b Climate-driven local adaptation of ecophysiology and phenology in Balsam Poplar, *Populus balsamifera* L. (*Salicaceae*). American Journal of Botany **98:** 99-108.

KERSTEN, B., B. PAKULL, K. GROPPE, J. LUENEBURG and M. FLADUNG, 2013 The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. Plant Biology**:** n/a-n/a.

KLIMES, L., J. KLIMESOVA, R. HENDRIKS and J. VAN GROENENDAEL, 1997 Clonal plant architecture: A comparative analysis of form and function. Ecology and Evolution of Clonal Plants**:** 1-29.

LEXER, C., C. A. BUERKLE, J. A. JOSEPH, B. HEINZE and M. F. FAY, 2007 Admixture in European Populus hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences (September 20, 2006). Heredity **98:** 183-184.

LEXER, C., M. F. FAY, J. A. JOSEPH, M. S. NICA and B. HEINZE, 2005 Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Molecular Ecology **14:** 1045-1057.

LEXER, C., B. HEINZE, R. ALIA and L. H. RIESEBERG, 2004 Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (Helianthus spp.). For Ecol Manage **197:** 49-64.

LEXER, C., J. JOSEPH, M. VAN LOO, G. PRENNER, B. HEINZE *et al.*, 2009 The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. Taxon **58:** 349-364.

LEXER, C., J. A. JOSEPH, M. VAN LOO, T. BARBARA, B. HEINZE *et al.*, 2010 Genomic Admixture Analysis in European Populus spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. Genetics **186:** 699-U391.

LEXER, C., and M. VAN LOO, 2006 Contact Zones: Natural Labs for Studying Evolutionary Transitions. Current Biology **16:** R407-R409.

LEXER, C., and A. WIDMER, 2008 The genic view of plant speciation: recent progress and emerging questions. Philos Trans R Soc Lond B Biol Sci **363:** 3023-3036.

LINDTKE, D., C. A. BUERKLE, T. BARBARA, B. HEINZE, S. CASTIGLIONE *et al.*, 2012 Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in Populus. Molecular Ecology **21:** 5042-5058.

LINDTKE, D., S. C. GONZALEZ-MARTINEZ, D. MACAYA-SANZ and C. LEXER, 2013 Admixture mapping of quantitative traits in *Populus* hybrid zones: power and limitations. Heredity **111:** 474-485.

MA, X.-F., D. HALL, K. R. ST ONGE, S. JANSSON and P. K. INGVARSSON, 2010 Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. Genetics **186:** 1033-1044.

MAYR, E., 1963 *Animal Species and Evolution*. The Belknap press, Cambridge, MA.

MING, R., A. BENDAHMANE and S. S. RENNER, 2011 Sex Chromosomes in Land Plants, pp. 485-514 in *Annual Review of Plant Biology, Vol 62*, edited by S. S. MERCHANT, W. R. BRIGGS and D. ORT.

MING, R., and P. H. MOORE, 2007 Genomics of sex chromosomes. Current Opinion in Plant Biology **10:** 123-130.

MING, R., J. WANG, P. H. MOORE and A. H. PATERSON, 2007 Sex chromosomes in flowering plants. American Journal of Botany **94:** 141-150.

NOLTE, A. W., and D. TAUTZ, 2010 Understanding the onset of hybrid speciation. Trends in Genetics **26:** 54-58.

OTTO, S. P., 2009 The Evolutionary Enigma of Sex. American Naturalist **174:** S1-S14.

OTTO, S. P., and T. LENORMAND, 2002 Resolving the paradox of sex and recombination. Nature Reviews Genetics **3:** 252-261.

PAKULL, B., K. GROPPE, M. MEYER, T. MARKUSSEN and M. FLADUNG, 2009 Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). Tree Genetics & Genomes **5:** 505-515.

PALME, A. E., V. SEMERIKOV and M. LASCOUX, 2003 Absence of geographical structure of chloroplast DNA variation in sallow, *Salix caprea* L. Heredity **91:** 465-474.

PAOLUCCI, I., M. GAUDET, V. JORGE, I. BERITOGNOLO, S. TERZOLI *et al.*, 2010 Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across Populus species. Tree Genetics & Genomes **6:** 863-875.

PENG, Y. H., and K. CHEN, 2011 Phylogeographic pattern of *Populus cathayana* in the Southeast of Qinghai-Tibetan Plateau of China revealed by cpSSR markers. Silva Fennica **45:** 583-594.

PETO, F. H., 1938 Cytology of poplar species and natural hybrids. Canadian Journal of Forest Research **16:** 445-455.

QVARNSTROM, A., and R. I. BAILEY, 2009 Speciation through evolution of sex-linked genes. Heredity **102:** 4-15.

RATHMACHER, G., M. NIGGEMANN, M. KOHNEN, B. ZIEGENHAGEN and R. BIALOZYT, 2010 Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures: implications for in situ conservation measures. Conservation Genetics **11:** 1327-1338.

RIESEBERG, L. H., and J. H. WILLIS, 2007 Plant speciation. Science **317:** 910-914.

ROBLEDO-ARNUNCIO, J. J., 2011 Wind pollination over mesoscale distances: an investigation with Scots pine. New Phytologist **190:** 222-233.

ROIRON, P., A. A. ALI, J. L. GUENDON, C. CARCAILLET and J. F. TERRAL, 2004 Evidence of the indigenous status of *Populus alba* L. in the western Mediterranean Basin. Comptes Rendus Biologies **327:** 125-132.

SACHS, T., 2001 Developmental processes and the evolution of plant clonality. Evolutionary Ecology **15:** 485-500.

SLAVOV, G. T., S. LEONARDI, W. T. ADAMS, S. H. STRAUSS and S. P. DIFAZIO, 2010 Population substructure in continuous and fragmented stands of *Populus trichocarpa*. Heredity **105:** 348-357.

SLAVOV, G. T., S. LEONARDI, J. BURCZYK, W. T. ADAMS, S. H. STRAUSS *et al.*, 2009 Extensive pollen flow in two ecologically contrasting populations of *Populus trichocarpa*. Molecular Ecology **18:** 357-373.

SLAVOV, G. T., and P. ZHELEV, 2010 Salient biological features, systematics, and genetic variation of *Populus*, pp. 15-38 in *Genetics and Genomics of Populus*, edited by S.

JANSSON, R. P. BHALERAO and A. T. GROOVER. Springer Science+Business Media, LLC, New York.

SMITH, R. L., and K. J. SYTSMA, 1990 Evolution of *Populus nigra* (sect *Aigeiros*) - introgressive hybridization and the chloroplast contribution of *Populus alba* (sect *Populus*) American Journal of Botany **77:** 1176-1187.

SMULDERS, M. J. M., J. E. COTTRELL, F. LEFEVRE, J. VAN DER SCHOOT, P. ARENS *et al.*, 2008 Structure of the genetic diversity in black poplar (*Populus nigra* L.) populations across European river systems: Consequences for conservation and restoration. Forest Ecology and Management **255:** 1388-1399.

STOELTING, K. N., R. NIPPER, D. LINDTKE, C. CASEYS, S. WAEBER *et al.*, 2013 Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Molecular Ecology **22:** 842-855.

TIFFNEY, B. H., and K. J. NIKLAS, 1985 Clonal growth in land plants: a paleobotanical perspective, pp. 35-66 in *Population biology and evolution of clonal organisms*, edited by J. B. C. JACKSON, L. W. BUSS and R. E. COOK. Yale University Press, New Haven, CT.

TUSKAN, G. A., S. DIFAZIO, P. FAIVRE-RAMPANT, M. GAUDET, A. HARFOUCHE *et al.*, 2012 The obscure events contributing to the evolution of an incipient sex chromosome in *Populus*: a retrospective working hypothesis. Tree Genetics & Genomes **8:** 559-571.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596-1604.

UNGERER, M. C., S. J. E. BAIRD, J. PAN and L. H. RIESEBERG, 1998 Rapid hybrid speciation in wild sunflowers. Proceedings of the National Academy of Sciences of the United States of America **95:** 11757-11762.

VALLEJO-MARIN, M., M. E. DORKEN and S. C. H. BARRETT, 2010 The Ecological and Evolutionary Consequences of Clonality for Plant Mating, pp. 193-213 in *Annual Review of Ecology, Evolution, and Systematics, Vol 41*, edited by D. J. FUTUYMA, H. B. SHAFER and D. SIMBERLOFF.

VAMOSI, J. C., S. P. OTTO and S. C. H. BARRETT, 2003 Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. Journal of Evolutionary Biology **16:** 1006-1018.

VAN LOO, M., J. A. JOSEPH, B. HEINZE, M. F. FAY and C. LEXER, 2008 Clonality and spatial genetic structure in Populus x canescens and its sympatric backcross parent P-alba in a Central European hybrid zone. New Phytologist **177:** 506-516.

WELCH, D. M., and M. MESELSON, 2000 Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science **288:** 1211-1215.

WESTERGAARD, M., 1958 The mechanism of sex determination in dioecious flowering plants. Advances in Genetics Incorporating Molecular Genetic Medicine **9:** 217-281.

WIDMER, A., C. LEXER and S. COZZOLINO, 2009 Evolution of reproductive isolation in plants. Heredity **102:** 31-38.

WU, C. I., 2001 The genic view of the process of speciation. Journal of Evolutionary Biology **14:** 851-865.

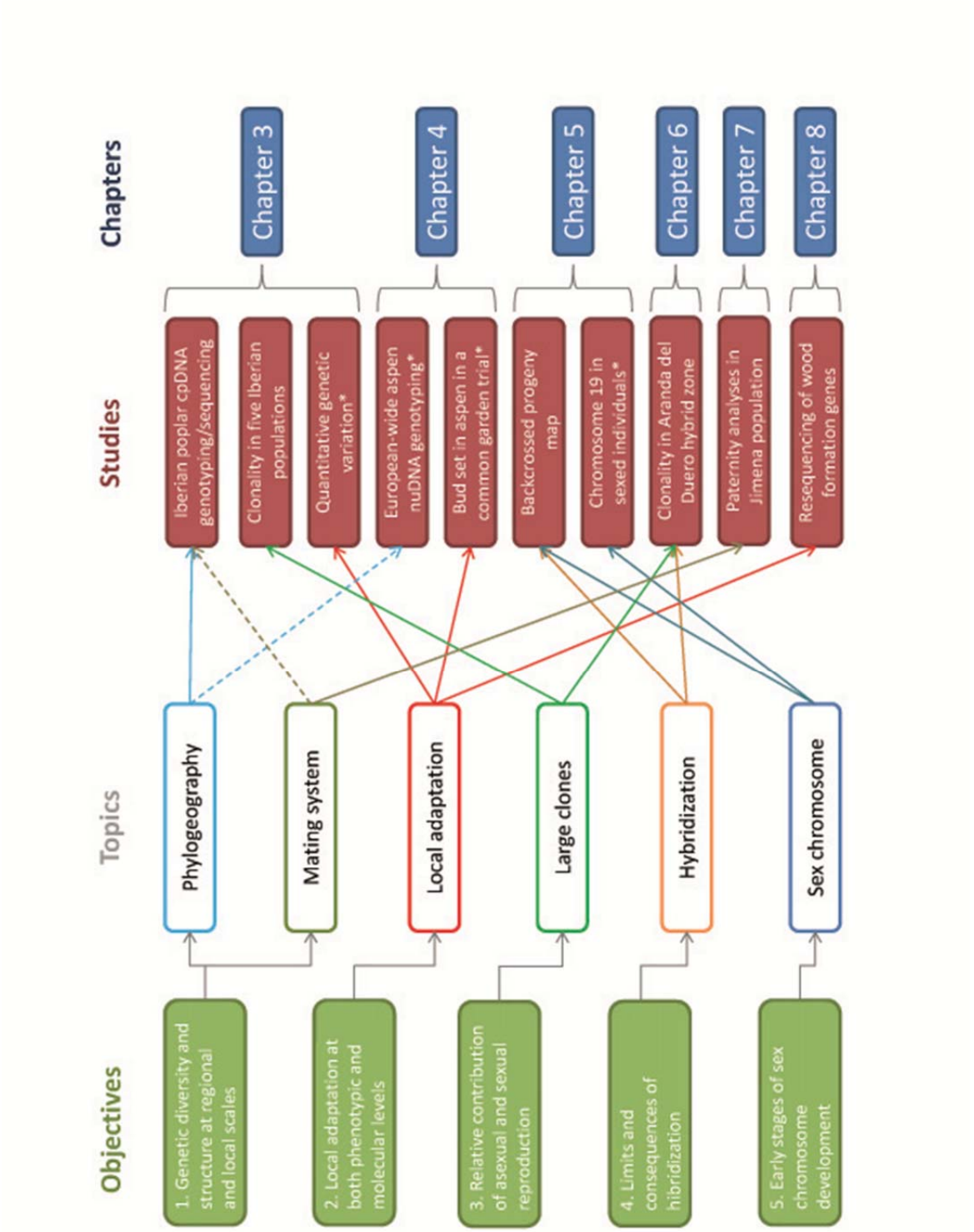
WYCKOFF, G. W., and J. C. ZASADA, 2008 *Populus* L. poplar, cottonwood, aspen, pp. 856-871 in *The Woody Plant Seed Manual*, edited by F. T. BONNER and R. P. KARRFALT. USDA Forest Service.

YIN, T., S. P. DIFAZIO, L. E. GUNTER, X. ZHANG, M. M. SEWELL *et al.*, 2008 Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Research **18:** 422-430.

CHAPTER 2: Objectives

1. To study poplar genetic diversity patterns at regional and local scales in the Iberian Peninsula, with particular focus on understanding the role of mating system (including sex lability) and gene flow in shaping population genetic structure.
2. To evaluate whether and to what extent local adaptation occurs in white poplar, at both phenotypic and molecular levels, and at different spatial scales.
3. To investigate the relative amount of asexual reproduction (clonality) in white poplar, and to tackle the possible causes governing the balance between sexual and asexual reproduction and the ecological and evolutionary consequences of increased asexual reproduction.
4. To assess the limits and consequences of genetic exchange and hybridisation between white poplar and European aspen, as well as the roles of various reproductive barriers in promoting introgression and maintaining the integrity of species identity.
5. To contribute to the knowledge of the early stages of sex chromosome development by studying population genetic parameters of chromosome XIX in white poplar and European aspen.

Overview


Objectives
Topics
Studies
Chapters
1. Genetic diversity and structure at regional and local scales
2. Local adaptation at both phenotypic and molecular levels
3. Relative contribution of asexual and sexual reproduction
4. Limits and consequences of hibridization
5. Early stages of sex chromosome development
Phylogeography
Mating system
Local adaptation
Large clones
Hybridization
Sex chromosome
Iberian poplar cpDNA genotyping/sequencing
Clonality in five Iberian populations
Quantitative genetic variation*
European-wide aspen nuDNA genotyping*
Bud set in aspen in a common garden trial*
Backcrossed progeny map
Chromosome 19 in sexed individuals*
Clonality in Aranda del Duero hybrid zone
Paternity analyses in Jimena population
Resequencing of wood formation genes
Chapter 3
Chapter 4
Chapter 5
Chapter 6
Chapter 7
Chapter 8

CHAPTER 3: The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars

[see publication in Annex A]

Introduction

As the genetic effects of glaciations and post-glacial colonisations at the continental scale become more widely known, researchers have started to focus on the regional scale, in particular on patterns inside glacial refugia. However, few studies to date have focused on the genetic structure of riparian trees, whose dependence on river valleys renders them more resilient to climatic changes, but also more prone to isolation (see Chapter 1). Isolation promotes stochastic processes that constraint both molecular and quantitative standing genetic variation, with a consequent loss of adaptive ability. Gene flow between severely isolated populations contributes to replenish genetic variance (ALLEAUME-BENHARIRA *et al.* 2006), but also to counteract the effects of selection, as it can modify adaptive local allele frequencies (i.e. outbreeding depression). In the case of riparian tree species, which inhabit similar, albeit disconnected and scattered, habitats, gene flow may be critical in keeping adequate levels of genetic variation and assure microevolution.

I will address regional and local genetic structure within the Iberian Peninsula (IP) of three species of poplars: white, grey and black poplars (*Populus alba*, *P.* × *canescens* and *P. nigra*, respectively); with an emphasis on white poplar. The IP, besides having been a glacial refugium, combines a complex orography with climatic heterogeneity, and it is a perfect scenario to test the following questions: (i) which ancient geological events have affected current poplar genetic structure and how?; (ii) can riparian species survive severe climate changes *in situ* or effectively migrate across important barriers?; (iii) how is local genetic structure produced, and in what manner does asexual propagation affect it?; and (iv) does local adaptation take place, imprinting signatures in quantitative traits (as estimated by Q_{ST})?

Materials and Methods

Sampling scheme

Fifty-nine populations of Iberian poplars (n = 628 trees) were sampled, comprising 44 white poplar, 13 black poplar and two grey poplar populations. Sample effort focused on

three major river basins (Duero and Guadalquivir, in the Atlantic Ocean watershed and Ebro, in the Mediterranean Sea watershed), and on several smaller, scattered populations in other basins, like Tajo and Guadiana, also in the Atlantic watershed. An additional Moroccan white poplar population was sampled in order to calibrate coalescence models based on the biogeographic separation time of North Africa and Iberian Peninsula (IP). Leaf tissue was collected of approximately 10 trees per population, separated at least 100 m, to minimise sampling same genet.

Five of such populations were further analysed, one of grey poplar and the rest of white poplar. Three to five additional individuals were sampled around each of the ten core trees. Individual geographical references were registered. DNA was isolated following a modified (DOYLE and DOYLE 1990) protocol.

Genetic markers

Two chloroplast microsatellites (*ccmp2* and *ccmp5*) were selected out of a panel of 13 (WEISING and GARDNER 1999), as they successfully amplified in white poplar and were polymorphic. All samples were amplified and resolved via electrophoresis (details of PCR reaction mix, temperature profile and fragment resolution in MACAYA-SANZ *et al.* 2012). Then, chloroplast region *trnC-petN1* was sequenced in at least one individual per population and cpSSR haplotype (n = 133), assuming that individuals with the same cpSSR within a population would share the variant of the supposedly less variable *trnC-petN1* region. For each combination of cpSSRs and *trnC-petN1*, a *rpl16-poprpl* region was sequenced at least once (n = 107; more details for PCR, sequencing and alignment in MACAYA-SANZ *et al.* 2012). Finally, five highly polymorphic nuclear microsatellites (ORPM 127, ORPM 312, PMGC 2852, ORPM 30 AND ORPM 344; LEXER *et al.* 2005) were used to study the local genetic structure in the more intensively sampled populations.

Haplotypes and haplotype networks

Chloroplast haplotypes were defined by a combination of microsatellites and the two sequenced regions. Given the elevated rate of polymorphisms in cpSSRs and the microsatellite inside the *trnC-petN1* region, homoplasy was expected. As homoplasy can induce bias in coalescence analysis (PROVAN *et al.* 2001), three alternative haplotype sets were made: (1) with whole polymorphisms ('full data set', coded with a 'h' prefix); (2) with unique event polymorphisms, to avoid homoplasy-related reticulations ('UEP data set', coded with a 'H' prefix); and (3) with exclusively single nucleotide polymorphisms ('SNP

data set', coded with a 'HR' prefix). Haplotype frequencies within population were plotted in geographical maps. Additionally, haplotype parsimony networks were constructed using TCS vs. 1.21 (CLEMENT *et al.* 2000) for the three data sets. Polymorphisms were equally weighed because of the lack of mutation rate information. For comparative purposes, twelve Iberian and French European aspens were added to the network.

Coalescence haplotype analyses

By means of Bayesian simulations using BATWING software (WILSON *et al.* 2003), UEPs-based haplotype coalescences were inferred. Times to the most recent common ancestor (TMRCA) were calculated for three subsets of white poplar haplotypes: (set1) all Iberian plus Moroccan haplotypes; (set2) all Iberian haplotypes; and (set3) haplotypes present in the Mediterranean drainage basin. *Salix* sp. Genebank accessions were used as outgroup. BATWING simulations were carried out under a constant size model and parameterisation scaled to population size N (more details in MACAYA-SANZ *et al.* 2012). To unscale coalescence times delivered by BATWING, divergence time of North African and Iberian lineages (set 1) were anchored in the flooding of Strait of Gibraltar, at the end of Messinian Salinity Crisis (5.33 Ma). Weather conditions are thought to impede wind dispersed seeds reaching across the Strait of Gibraltar, and phylogeographical signatures in conifers do indicate that the Strait acts as a barrier (see JARAMILLO-CORREA *et al.* 2010). Ratios of different haplotype sets products of scaled TMRCA (T) and mutation rates (θ) are proportional to rates of TMRCA in years, so set2 and set3 coalescence times can be estimated using the 5.33 Ma anchor point (i.e. forced set1 coalescence; (more details in MACAYA-SANZ *et al.* 2012). Effective population sizes N of different sets were calculated assuming a generation time g of 20-60 years, longer than the previously estimated for European aspen, to account for higher clonality.

Genetic diversity and differentiation

Genetic diversity, calculated as Nei's expected heterozygosity (H_E), was computed by Arlequin vs. 3.1 (EXCOFFIER *et al.* 2005) using the full data set: independently for each population but also for pooled populations according to river basin, drainage basin (Atlantic vs. Mediterranean) and latitude (North vs. South). Finally, raw and rarefied haplotype richness, as well as rarefied number of private haplotypes, were calculated (using RAREFAC; PETIT *et al.* 1998).

To estimate genetic differentiation among populations and regions of white and black poplars, both full and UEP data sets were used. The latter was used for group differentiation, and the former for population pairwise differentiation. By means of Arlequin vs. 3.1, SPAGeDi vs. 1.3d (HARDY and VEKEMANS 2002) and R package DEMEtics (GERLACH *et al.* 2010), I computed F_{ST}, N_{ST} (differentiation accounting for haplotype genetic distances) and Jost's *D* (less biased when populations differ in diversity rates). Precise details can be seen in MACAYA-SANZ *et al.* (2012).

Isolation by distance (IBD) was calculated by regression of population pairwise $F_{ST}/(1 - F_{ST})$ and distance. These two parameters are expected to co-vary linearly in one-dimension analyses; and also in two-dimension ones, although these with log-scaled distance. Then, within basin, IBD was computed under the one and two-dimension approaches, and among basins IBD only under the latter. I used the so-called 'resistance' distance (in my case, the itinerary between populations following the river course) as the one-dimension distance (further details in MACAYA-SANZ *et al.* 2012).

Genetic structure at local scale

Nuclear markers were used to genotype the five focal populations in which sampling was more intense. Afterwards, I assembled genets by exact genotype identity among sampled ramets, using Gimlet vs. 1.3.2 (VALIERE 2002). Among-ramet relative kinship coefficient F_{ij} of LOISELLE *et al.* (1995) was computed within population using SPAGeDi, referred to overall allele frequencies (n=201), and regressed on the Euclidean distance. Regression slope significance was tested by permutation. Finally, spatial genetic structure (SGS) patterns were plotted averaging F_{ij} in five distance classes.

Quantitative genetic differentiation

Four quantitative traits [total height right after the first year (HT1); and same measure (HT3), stem diameter at the base (DSB3), and stem form (FOR3) right after the third year] were measured in a common garden experiment that comprised two to four open pollinated families of seven populations (15 families in total: eight from two genotyped populations from the Ebro basin, and seven from three populations from the Guadalquivir basin). Approximately 40 plants per family were randomised in a complete block design with eight blocks. Variance components for basin, population and family were obtained by Restricted Maximum Likelihood (REML; more details in MACAYA-SANZ *et al.* 2012). Overall and among-group quantitative genetic differentiation (Q_{ST}) was estimated from

such variance components. Hypothesis $Q_{ST} > F_{ST}$ was contrasted by bootstrapping (as outlined by WHITLOCK 2008), to disentangle genetic drift effects from among-basin adaptive divergence. F_{ST} values were calculated from the allozyme data set published in ALBA (2000), since allozymes' lower polymorphism makes them more suitable for unbiased F_{ST} estimation (JOST 2008).

Results

Haplotypes and haplotype networks

Chloroplast loci showed a total of 36 polymorphic sites, including three mononucleotide microsatellites (one within the *trnC-petN1* region), 13 SNPs, 17 indels and three short tandem repeats (STRs). Combining all polymorphisms, 54 haplotypes were resolved: 36 for *Populus alba* (one shared with *Populus* × *canescens*), 2 exclusive for *P.* × *canescens*, and 16 for *Populus nigra*. The highly polymorphic *trnC-petN1* region alone resolved 26 haplotypes, while the less variable *rpl16-poprpl* resolved only ten. Considering only UEPs, the number of haplotypes was reduced to fourteen. In terms of the full and UEP data sets, there were no shared haplotypes across species (Figures 3.1 and 3.2). However, cpSSRs alone were unable to distinguish among species, with *P. nigra* and *P. alba* sharing seven haplotypes. This fact highlights the limited value of cpSSRs for phylogenetic inference in poplars. The majority of white and grey poplar haplotypes were confined to one basin (70%).

The full data set delivered a heavily reticulated network. However, UEPs and SNPs data set networks lacked reticulations and were easily interpretable (Figure 3.2). Two salient facts must be highlighted: (i) a divergent *P. nigra* haplotype, and (ii) the short distance between Iberian white and black poplars, shorter than between Iberian and Moroccan white poplar haplotypes or between Iberian white poplar and European aspen (related species, phylogenetically closer to each other than to black poplar).

Time to the most recent common ancestor

Distributions of the product of scaled TMRCA and mutation rate ($T \times \theta$) were unimodal and asymmetric, as predicted by theory. Comparison among them dated the spread of Iberian haplotypes to *ca.* 2.76 Ma, and the divergence across Mediterranean lineages to *ca.* 1.23 Ma. Effective population sizes (N) ranged 21 000 – 63 000 for the Iberian white poplar and 10 667 – 32 000 for the Mediterranean group.

CpDNA diversity and differentiation

Based on the full data set, overall haplotypic diversity per population was lower in *P. alba* (average of 0.317) than in *P. nigra* (average of 0.409). At population level, haplotypic variation was extremely variable for both species, but no consistent differential spatial pattern was detected. By contrast, genetic diversity at the regional level (i.e. pooling populations within regions) was higher in the southern than in the northern white poplar metapopulations (Table 1 in MACAYA-SANZ *et al.* 2012). Private haplotype numbers revealed the same result. Conversely, the Atlantic and Mediterranean ranges did not differ in diversity.

In white poplar, genetic differentiation as estimated by F_{ST} and Jost's D was significant for almost all spatial scales (Table 3.1), with overall values of F_{ST} = 0.670 (0.735 for the UEP data set) and D = 0.929 (0.559 for the UEP data set, Table 3.1). Main factors causing genetic structure (based on the more reliable UEP data set) in this species were river and drainage basins with F_{CT} /N_{CT} /D values of 0.320 / 0.223 / 0.511 and 0.374 / 0.260 / 0.569, respectively, and latitudinal differentiation was low. Moreover, only five haplotypes (out of 36) were shared across drainage basins and numbers of private haplotypes were alike in both.

In black poplar, patterns were less clear, probably due to the reduced sample and higher human mediation. Despite overall genetic differentiation being similar to that of white poplar (F_{ST} = 0.627 and D = 0.600 with the UEP data set), black poplar showed lower genetic differentiation across drainage basins and inconsistent values for differentiation across river basins.

Isolation by distance (i.e. positive slopes) was found in white poplar, but with different strengths depending on scales. Regression slopes were steepest among river basins than within them, reflecting the isolation effect produced by white poplar's dependency on river courses.

Levels of clonality

The four intensively-studied white poplar populations displayed similar levels of clonality, with 6-13 genets per population, with an average of 4.2 ramets per genet (see Figure 3.3 for examples). However, clone size was highly variable within populations, from mono-ramet genets to genets that ranged tens of kilometres (MACAYA-SANZ *et al.* 2012). Clone size in

three out of four populations averaged below 100 m. All populations had lower and non-significant kinship at > 100 m distance classes. The grey poplar populations had an unusual pattern, with only four large-size clones, probably extensively propagated by humans, given the region's scarcity of better soft wood sources.

Genetic differentiation of quantitative traits in white poplar

HT3 and FOR3 showed significant overall differentiation, as both had over three to six-fold higher values of differentiation among than within basins. When comparing with neutral markers, Q_{ST} for FOR3 was significantly higher than F_{ST} among river basins, but not among populations within them (numbers in MACAYA-SANZ *et al.* 2012). For HT3, a similar pattern occurred, but among river basins the test was no significant.

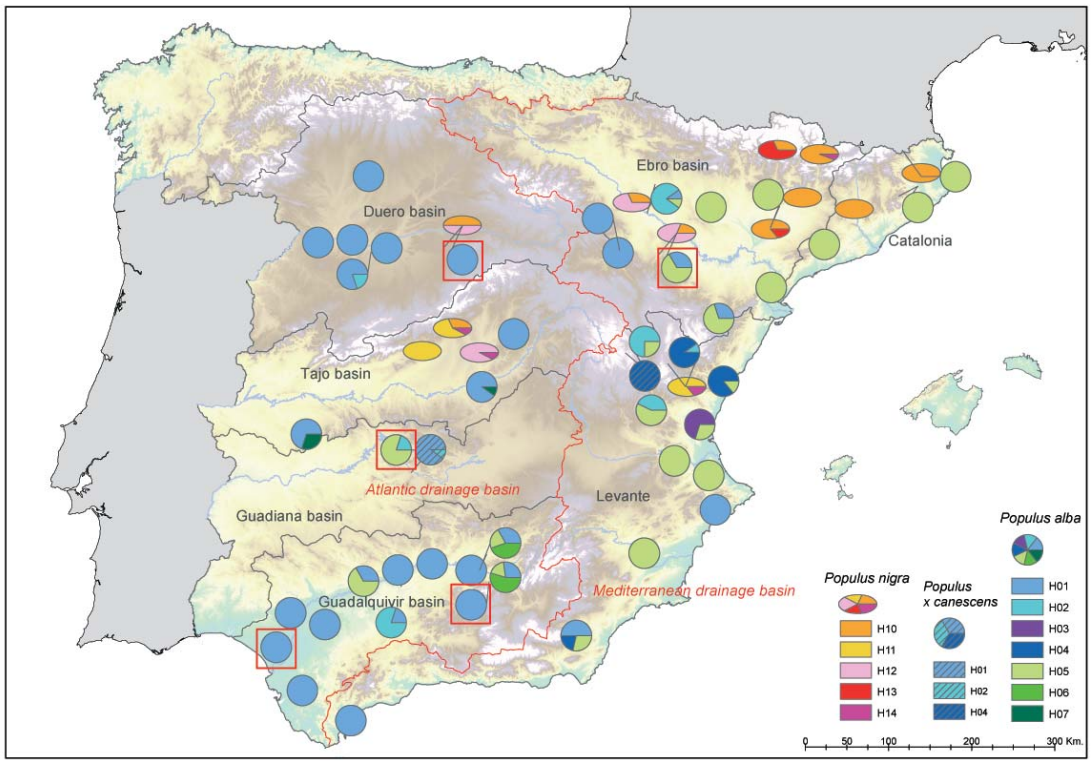


Figure 3.1 Geographic distribution and population frequency of haplotypes based on unique event polymorphisms (UEPs). Red squares indicate populations used to study clonality levels and fine-scale genetic structure. Main hydrographic features and altitudinal pattern (in shadows) are also shown. Figure from MACAYA-SANZ *et al.* (2012; reproduced with permission from John Wiley & Sons, Chichester, UK)

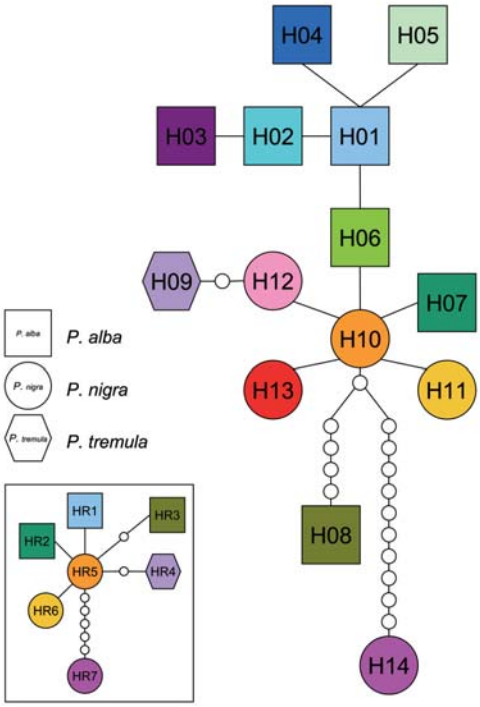


Figure 3.2 A statistical parsimony network representing the minimum number of polymorphic site differences among haplotypes. The network was constructed considering only unique event polymorphisms (UEPs). The inset represents a network using only single nucleotide polymorphisms (SNPs). Notice the unexpected location of H08, H09 and H14. Figure from MACAYA-SANZ *et al.* (2012; reproduced with permission from John Wiley & Sons, Chichester, UK)

Discussion

Haplotype networks and shared polymorphism across species

Haplotype networks display some noteworthy facts. Firstly, that most black poplar haplotypes are located among Iberian white poplar, Moroccan white poplar and European aspen haplotypes. This means that black poplar, which is commonly accepted as a member of section *Aigeiros*, would belong to the section *Populus* by chloroplast origin, like white poplar. This finding, previously reported elsewhere (HAMZEH and DAYANANDAN 2004), is contradicted by phylogenetics based on nuclear DNA, which place black poplar in the same section as its morphology does (CERVERA *et al.* 2005). This odd position shown by plasmid DNA, points to ancient hybridisation events (SMITH and SYTSMA 1990), when

white and black poplar could hybridise naturally. Secondly, that the highly divergent position of the H14 haplotype is explained by hybridisation with commercial Euroamerican clones. This was verified by sequencing an array of 14 Euroamerican clones and Lombardy cultivars. Euroamerican clones had H14 haplotype. Thirdly, that the distant location in the network of the Moroccan haplotype H08 indicates an ancient divergence of North African and Iberian lineages (see below).

Whereas DNA sequences segregated among species, microsatellites (*ccmp2*, *ccmp5* and the one within *trnC-petN1*) shared variants, which may be due to homoplasy. My conclusion is that chloroplast microsatellites are more useful in local and contemporary studies, but less applicable to phylogeographical approaches.

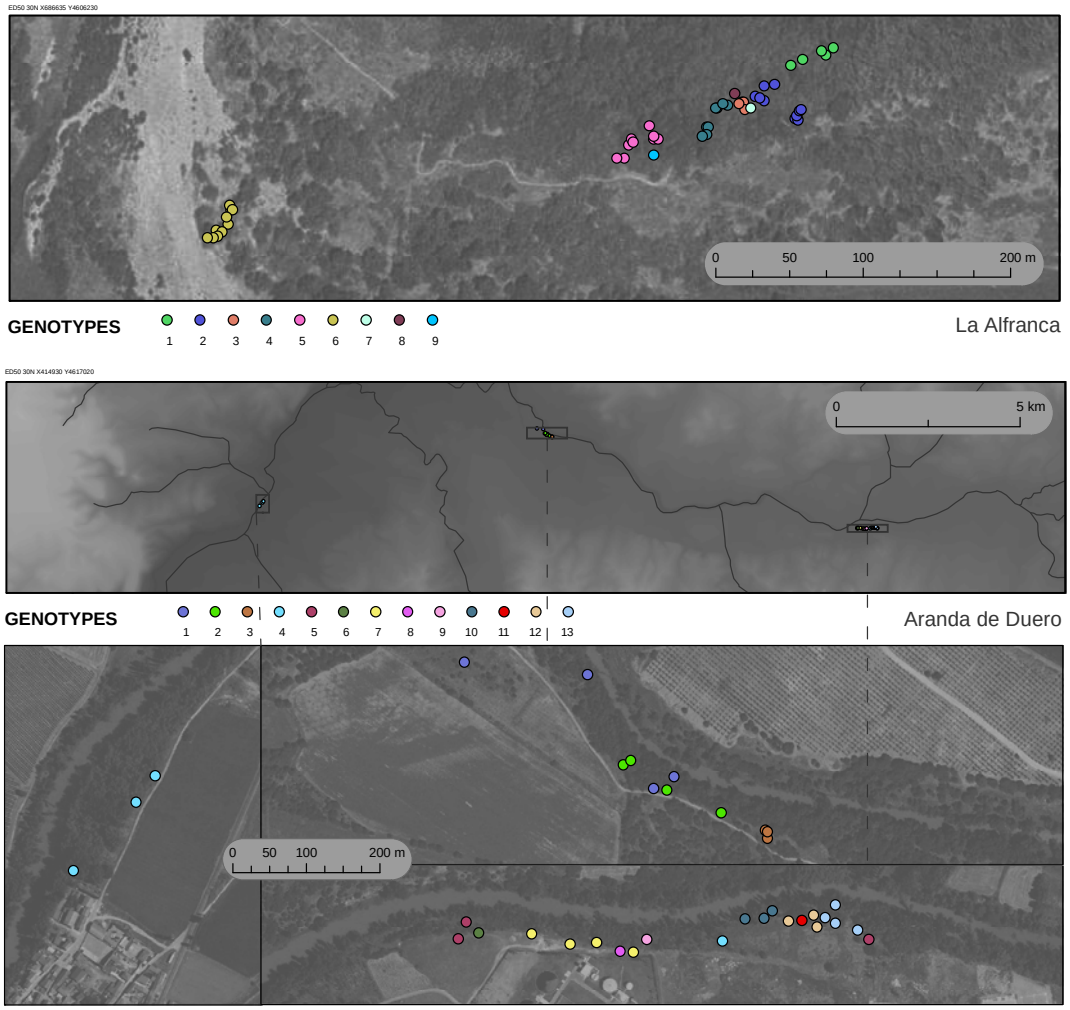


Figure 3.3 Spatial distributions of ramets from two contrasting white poplar populations. The La Alfranca population (top) had smaller and less spread clonal assemblies than the Aranda de Duero population (bottom), which includes clones stretching over c. 15 km. Figure from MACAYA-SANZ *et al.* (2012; reproduced with permission from John Wiley & Sons, Chichester, UK)

Table 3.1 Genetic differentiation among populations/groups at various hierarchical levels in white and black poplars from the Iberian Peninsula. Differentiation was measured considering haplotypic frequencies (*F*-statistics and Jost's *D*-statistics) or, alternatively, taking into account genetic distances among haplotypes (*N*-statistics). Estimates are provided for haplotypes resolved using the complete dataset or, alternatively, using only UEPs (see text for details). All genetic differentiation estimates are significant at α = 0.05 unless stated otherwise (ns). NA: not available or not possible to compute.

		White poplar			Black poplar		
		Full set	UEPs		Full set	UEPs	
Group	Level	*F*-stats	*F*-stats	*N*-stats	*F*-stats	*F*-stats	*N*-stats
River basin	F_{CT}	0.165	0.320	0.223	0.048^{ns}	0.170^{ns}	0.043^{ns}
	F_{SC}	0.616	0.632	0.586	0.524	0.568	0.312
	F_{ST}	0.679	0.750	0.678	0.547	0.642	0.341
Latitude	F_{CT}	0.090	-0.001ns	0.028^{ns}	NA	NA	NA
	F_{SC}	0.653	0.735	0.660	NA	NA	NA
	F_{ST}	0.685	0.734	0.669	NA	NA	NA
Drainage basin	F_{CT}	0.082	0.374	0.260	0.043^{ns}	0.130^{ns}	0.032^{ns}
	F_{SC}	0.655	0.654	0.605	0.532	0.600	0.324
	F_{ST}	0.683	0.783	0.707	0.552	0.652	0.346
Overall	F_{ST}	0.670	0.735	0.665	0.542	0.627	0.331
		Jost's *D*-statistics					
River basin	Among	0.892	0.511		0.773	0.586	
Duero	Within	0.353	0.008ns		NA	NA	
Catalonia	Within	0.076^{ns}	0.000ns		1.000	0.000^{ns}	
Ebro	Within	0.764	0.544		0.794	0.465	
Levante	Within	0.895	0.678		NA	NA	
Tajo	Within	0.985	0.031ns		0.661	0.723	
Guadalquivir	Within	0.807	0.268		NA	NA	
Latitude	Among	0.896	0.052		NA	NA	
North	Within	0.759	0.534		NA	NA	
South	Within	0.935	0.576		NA	NA	
Drainage basin	Among	0.926	0.569		0.744	0.432	
Atlantic	Within	0.888	0.227		0.737	0.695	
Mediterranean	Within	0.846	0.549		0.892	0.501	
Overall	Among	0.929	0.559		0.889	0.600	

*F_{CT} refers to genetic differentiation (GD) among groups (i.e. basins), F_{SC} to GD among populations within groups and F_{ST} to GD among populations without considering groups.

As European aspen and white poplar are known to hybridise in nature, the lack of shared haplotypes found is noteworthy. However, a broader sampling in European aspen should be conducted to conclude that no chloroplast trade has occurred between these species in the IP.

Different species, different histories

Black and white poplars presented a similar degree of overall genetic differentiation, but a different structure in a smaller scale: black poplar showed lower population structure among river and drainage basins, and higher regional haplotypic diversity. This reveals a higher rate of gene flow across these areas, perhaps due to two causes: higher tolerance to cold temperatures, which may facilitate black poplar's crossing dividing mountain chains, and higher seed and cutting transfer by humans due to its higher wood value. However, the high number of private haplotypes found in the IP weakens the latter cause. Similar patterns of high diversity and low differentiation due to tolerance to cold have been observed, for instance, in the distribution of the six native Iberian pine species (SOTO *et al.* 2010) or of European ashes (HEUERTZ *et al.* 2006). Such cold tolerant species may have endured cold stages better and, subsequently, may have colonised defrosted territories faster.

Genetic signatures of ancient events in white poplar

The end of the Messinian Salinity Crisis (ca. 5.33 Ma) coincided with the flooding of the Strait of Gibraltar, creating a barrier that is still reflected today in the genetic imprint. The divergence between North African and Iberian lineages could possibly have started then. Mediterranean and Atlantic white poplar lineages present a marked differentiation, revealing a limited genetic exchange between their drainage basins and resembling that found in other tree species (RODRIGUEZ-SANCHEZ *et al.* 2010). This reflects an ancient event. As a matter of fact, that the signal obtained using only UEPs is stronger than the one obtained using complete polymorphism data (0.374 vs. 0.082) points to such a past event, one whose signal fades when highly mutable molecular markers are used.

The fact that major Iberian mountain systems run west- and eastwards, preventing latitudinal migration, is counterintuitive with the greater differentiation found among drainage basins than the one found among latitudes. Past vegetation altitudinal shifts may provide an explanation. Before the Pleistocene, a benign climate could have favoured migration among basins, even for lowland riparian trees. Cold cycles in the Pleistocene

could have displaced populations towards lower altitudes and along the coast line. Due to the orography of the IP, white poplars would have settled in the western and eastern coastal fringes, where migration along the coast line would have allowed latitudinal exchange, but migration between drainage basins would have been more severely restricted. My results indicating to an increasing isolation between Atlantic and Mediterranean drainage basins ca. 1.12 Ma agree with this scenario.

Regional and population effects of glacial times in white poplar

Current patterns of genetic diversity and structure in white poplar reflect Pleistocene climatic oscillations as shown by several lines of evidence. First, regional genetic diversity was higher (and also number of private haplotypes) in the southern river basins than in the northern ones. Secondly, genetic structure among populations was more pronounced as well in the southern basins. Thirdly, IBD was detected in the Duero basin when considering Euclidean distances, but not under 'resistance' distances, a pattern consistent with a rapid isotropic postglacial spread. Fourthly, clonal assemblies seem to be larger in the Duero basin. Asexual propagation might have favoured persistence during glacial maxima and facilitated subsequent colonisation.

Pleistocene glacial oscillations lowered temperature and humidity globally, although in the North and West of the IP the reduction was higher, hampering tree population survival. Indeed, white poplar presents reduced genetic diversity in this region. Such genetic pattern is consistent with a population bottleneck with survival in 'cryptic refugia', but not with a local extinction followed by colonisation from other basins, due to the presence of private haplotypes.

Evidence for local adaptation in white poplar

The higher quantitative (Q_{ST}) than molecular (F_{ST}) genetic differentiation found across river basins for some traits points to local adaptation in white poplar, albeit at a wider scale than in other temperate trees. White poplar populations have typically low population sizes, due to their dependence on phreatic water, and the places where they grow, which are very valuable for crops. Indeed, white poplar would rather grow in the lower and medium water courses, where human impact is stronger. Such small population sizes are balanced by the detected gene flow over mesoscale distances, revealed by the occurrence of regional but not local adaptation, and the reduced regional genetic structure. Gene flow replenishes genetic variation and counteracts local genetic drift. The relative homogeneity of riparian

habitats would have prevented the arrival of maladapted genotypes, hampering the development of 'outbreeding depression'.

Conclusions

Although overall genetic differentiation in IP's white and black poplar is similar, their genetic structure has marked differences. Black poplar genetic structure seems to have been more affected by gene flow among basins, probably due to its higher resistance to cold. In white poplar, gene flow has been more intense between latitudes than between drainage basins (i.e. east-westwards), especially during the Pliocene, possibly facilitated by the Iberian orography. Glaciations have affected white poplar genetic structure, reducing northwestern diversity, yet not so strongly as to eliminate completely private haplotypes. Gene flow over mesoscale distances allows white poplar to maintain genetic diversity, despite its local small population sizes. It also prevents local adaptation, whereas regional adaptation has been observed.

References

ALBA, N., 2000 Variabilidad genética de *Populus alba* L. mediante caracteres isoenzimáticos y fenotípicos. PhD Dissertation, pp. 55-75. Technical University of Madrid, Madrid.

ALLEAUME-BENHARIRA, M., I. R. PEN and O. RONCE, 2006 Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. Journal of Evolutionary Biology **19:** 203-215.

CERVERA, M. T., V. STORME, A. SOTO, B. IVENS, M. VAN MONTAGU *et al.*, 2005 Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. Theoretical and Applied Genetics **111:** 1440-1456.

CLEMENT, M., D. POSADA and K. A. CRANDALL, 2000 TCS: a computer program to estimate gene genealogies. Molecular Ecology **9:** 1657-1659.

DOYLE, J. J., and J. L. DOYLE, 1990 Isolation of plant DNA from fresh tissue. Focus **12:** 13-15.

EXCOFFIER, L., G. LAVAL and S. SCHNEIDER, 2005 Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics **1:** 47-50.

GERLACH, G., A. JUETERBOCK, P. KRAEMER, J. DEPPERMANN and P. HARMAND, 2010 Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics! Molecular Ecology **19:** 3845-3852.

HAMZEH, M., and S. DAYANANDAN, 2004 Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast TRNT-TRNF region and nuclear rDNA. American Journal of Botany **91:** 1398-1408.

HARDY, O. J., and X. VEKEMANS, 2002 SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes **2:** 618-620.

HEUERTZ, M., S. CARNEVALE, S. FINESCHI, F. SEBASTIANI, J. F. HAUSMAN *et al.*, 2006 Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): roles of hybridization and life history traits. Molecular Ecology **15:** 2131-2140.

JARAMILLO-CORREA, J. P., D. GRIVET, A. TERRAB, Y. KURT, A. I. DE-LUCAS *et al.*, 2010 The Strait of Gibraltar as a major biogeographic barrier in Mediterranean conifers: a comparative phylogeographic survey. Molecular Ecology **19:** 5452-5468.

JOST, L., 2008 G(ST) and its relatives do not measure differentiation. Molecular Ecology **17:** 4015-4026.

LEXER, C., M. F. FAY, J. A. JOSEPH, M. S. NICA and B. HEINZE, 2005 Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Molecular Ecology **14:** 1045-1057.

LOISELLE, B. A., V. L. SORK, J. NASON and C. GRAHAM, 1995 Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany **82:** 1420-1425.

MACAYA-SANZ, D., M. HEUERTZ, U. LOPEZ-DE-HEREDIA, A. I. DE-LUCAS, E. HIDALGO *et al.*, 2012 The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars. Molecular Ecology **21:** 3593-3609.

PETIT, R. J., A. EL MOUSADIK and O. PONS, 1998 Identifying populations for conservation on the basis of genetic markers. Conservation Biology **12:** 844-855.

PROVAN, J., W. POWELL and P. M. HOLLINGSWORTH, 2001 Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology & Evolution **16:** 142-147.

RODRIGUEZ-SANCHEZ, F., A. HAMPE, P. JORDANO and J. ARROYO, 2010 Past tree range dynamics in the Iberian Peninsula inferred through phylogeography and palaeodistribution modelling: A review. Review of Palaeobotany and Palynology **162:** 507-521.

SMITH, R. L., and K. J. SYTSMA, 1990 Evolution of *Populus nigra* (sect *Aigeiros*) - Introgressive hybridization and the chloroplast contribution of *Populus alba* (sect *Populus*). American Journal of Botany **77:** 1176-1187.

SOTO, A., J. J. ROBLEDO-ARNUNCIO, S. C. GONZALEZ-MARTINEZ, P. E. SMOUSE and R. ALIA, 2010 Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. Molecular Ecology **19:** 1396-1409.

VALIERE, N., 2002 GIMLET: a computer program for analysing genetic individual identification data. Molecular Ecology Notes **2:** 377-379.

WEISING, K., and R. C. GARDNER, 1999 A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome **42:** 9-19.

WHITLOCK, M. C., 2008 Evolutionary inference from Q_{ST}. Molecular Ecology **17:** 1885-1896.

WILSON, I. J., M. E. WEALE and D. J. BALDING, 2003 Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. Journal of the Royal Statistical Society Series a-Statistics in Society **166:** 155-188.

CHAPTER 4: Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree

[see publication in Annex B]

Introduction

The role of gene flow in adaptation is an intensely disputed topic. In the case of gene flow within species, some studies argue that it hampers local adaptation (STEARNS and HOEKSTRA 2005), while others state that it favours the maintenance of standing genetic variation and species cohesion (PRZEWORSKI *et al.* 2005). Similarly, in the case of gene flow between species, some defend that it generates 'evolutionary noise', whereas others consider that it represents an important evolutionary force that creates opportunities for speciation (ARNOLD 2006; RIESEBERG *et al.* 2003). However, less attention has been paid to an intermediate scenario: gene flow among conspecific populations that have been previously isolated for a long time.

This is the case of populations in temperate areas in which quaternary climatic shifts have been more intense (HEWITT 2000) and cycles of population contraction and expansion have produced, respectively, cycles of isolation and connection. Theoretical predictions estimate an increase of standing genetic variation (HEWITT 2000), while research done for animals – fish, mice – and some plants have provided evidence in this direction (see DE CARVALHO *et al.* 2010 and references therein). Trees in particular should be prone to follow this pattern as they undergo rapid diversifying selection, thanks to large population numbers and great environmental heterogeneity. However, quite unexpectedly, they present a low nucleotide substitution rate per time unit, suggesting that lack of standing genetic variation could be a constraining factor for local adaptation (PETIT and HAMPE 2006; SAVOLAINEN and PYHAJARVI 2007). Additionally, admixture between lineages could limit adaptation via outbreeding depression (STEARNS and HOEKSTRA 2005).

Effects of gene flow may be described by geographic clines of phenotypic traits or allelic frequencies (BARTON and HEWITT 1985), and by single-locus clines within genome-wide admixture gradients (GOMPERT and BUERKLE 2009; REICH *et al.* 2005). The latter analysis can reveal the exact target loci and the nature of the selective forces acting on them (GOMPERT and BUERKLE 2009; LEXER *et al.* 2007).

European aspen (*P. tremula*) is a poplar species with a large Palearctic distribution and substantial levels of neutral variability (scaled synonymous mutation rate theta = 0.012; one single nucleotide polymorphism for every 100-150 bp; INGVARSSON 2008). It survived during glaciations in separate refugia, but suffered extinction in the Scandinavian region, where populations show strong differentiation and clinal variation.

In this chapter, our objective is to gain insights on how admixture between divergent lineages affects adaptation from standing variation in this tree. We also aim to infer patterns of neutral and non-neutral population divergence at the continental scale, and to test for admixture in Northern Europe. Finally, we will try to answer whether admixture explains cline shape, variances, and selection differentials for an adaptive phenological trait: namely, bud set.

Materials and Methods

Seven European aspen natural populations were sampled, comprising 17 to 40 individuals across several dozens of square kilometres each. This scheme intended to avoid sampling of clonal genotypes and to capture the high dispersal capacity of this species. One of them was a population from the Central Mountain Chain of the Iberian Peninsula (Sierra de Gredos), comprising 32 individuals. Three scales were used in the population pairwise comparison in order to describe spatial structure: regional (among the two populations of Austria, and among the two others of Russia), European (among the remaining pairs) and range-wide (between all sampled populations and a white poplar reference group described in LEXER *et al.* 2005).

In order to assess adaptive trait differentiation, a controlled common garden trial was analysed (described in INGVARSSON *et al.* 2006). Bud setting was measured in two Swedish common gardens during four years. Each trial comprised a total of eight replicas of 116 different trees from 12 Swedish populations selected from a latitudinal gradient.

Individuals from the seven natural populations, as well as from the common garden trial were genotyped using a set of seventy markers (see DE CARVALHO *et al.* 2010 for further details). Linkage relationship among them was referred to physical distance displayed in the *Populus trichocarpa* genome assembly and, to a lesser extent, to the linkage maps described in Chapter 5.

Genetic structure and diversity was estimated for all populations. Thus, allelic richness, expected and observed heterozygosity, and inbreeding coefficient were computed. Variance in allele size and genetic differentiation (F_{ST}, G'_{ST}) were computed for each locus. Population bottlenecks were tested using sign and Wilcoxon sign-rank tests (PIRY *et al.* 1999). Isolation by distance (IBD) was examined by a Mantel test. Finally, to test for linkage disequilibrium (LD), we used exact tests available in GENEPOP 007 (ROUSSET 2008).

Among population genetic structure was evaluated using a UPGMA cluster analysis and a Bayesian method. STRUCTURE software was used for the Bayesian analyses (FALUSH *et al.* 2003), following two different approaches: a standard admixture model, and a linkage model with information from the *P. trichocarpa* genome assembly. Different runs were carried out to estimate optimal prior parameters (see DE CARVALHO *et al.* 2010 for further details).

Imprints of selection were analysed with genetic divergence tests, namely a migration and drift outlier-F_{ST} based approach (BEAUMONT and BALDING 2004) and an alternative drift-based approach that simulates explicitly glacial refugia (VITALIS *et al.* 2001). The first is known to be robust across a wide range of demographic scenarios, while the second is useful with departures from equilibrium conditions (as those related to survival in glacial conditions). Neutral F_{ST} was estimated under a realistic population model (see DE CARVALHO *et al.* 2010 for further details). Selection imprints were surveyed as well using diversity-based tests, in order to detect evidence of selective sweeps. Two ratios specific for microsatellite data were used: gene diversity (lnRH) and variance in repeat number (lnRV), which were calculated for pairs of populations. Three different spatial scales were analysed (see above) to examine the interplay of gene flow and selection in each of them.

Data on phenotypic traits (bud set and growth) were taken from a previous study (LUQUEZ *et al.* 2008). Phenotypic clines were depicted after estimation of clone-specific breeding values by Best Linear Unbiased Predictors (BLUPs) and regression between bud set and latitude. A simple linear and a nonlinear model were essayed, and the best fitting model was selected using the Akaike Information Criterion (AIC) (more details in DE CARVALHO *et al.* 2010). Selection differentials (i.e. covariances between bud set and fitness) were also computed (LANDE and ARNOLD 1983). Growth rate was used as fitness proxy.

Locus-specific ancestries were calculated using the R script INTROGRESS (Z. GOMPERT and C. A. BUERKLE) only for the Swedish common garden individuals. Since previous analyses displayed that the main sources of the Swedish genomic background were the Scottish and the Russian pools, these two were used as referring populations. Each locus in the Swedish individuals was categorised as homo- or heterozygous for Scottish or Russian alleles.

Results

The genomic divergence and significant IBD revealed the importance of gene flow and drift in genetic structuration. The conventional cluster analyses depicted the geographical structure vaguely, placing the Spanish population in a radically different branch (Figure 4.1 A). The Bayesian approach showed the admixed origin of the Swedish and Central European populations. Indeed, the northernmost population was clearly composed of Scottish and Russian gene pools, while the Spanish population was composed of several groups of closely related genotypes and showed imprints of a recent bottleneck (Figure 4.1 B).

Genetic diversity estimated with microsatellites was intermediate (H_E = 0.563 ± 0.030 SE; H_O = 0.468 ± 0.029 SE), albeit low enough to conduct unbiased detection of F_{ST} outliers. Observed values of significant inbreeding (overall F_{IS} ranging from 0.083 to 0.173), pointed to a possible cryptic population subdivision. The Spanish population displayed lower values of genetic diversity (H_E = 0.474; allelic richness = 3.497) than any other population, as well as lower inbreeding (F_{IS} = 0.022).

Selection analysis based on pairwise population divergence yielded from 1 to 16 candidate loci. Diversity-based comparisons delivered 1 to 3 loci potentially under selection at the European scale, 1 to 4 at the regional scale, and no loci at the range-wide scale. Only three loci showed selection evidence for both methods (i.e. divergence and diversity-based). LD among chromosome XIX loci significantly exceeded the general decay observed in genus *Populus* (>400 kb vs 100 bp; INGVARSSON 2008).

The admixed background of Scandinavian populations affected the shape of the phenotypic bud set cline: the nonlinear model fitted better, revealing a conspicuous step in the middle of the curve; also, in the same interval, bud set variance and selection differentials were elevated. Ancestries of individual loci varied greatly along the

chromosomes. For instance, when compared to the normally recombinating chromosome VI (YIN *et al.* 2004), chromosome XIX's ancestry pattern was consistent with the lesser recombination rate expected (YIN *et al.* 2008), since ancestries changed gradually from Scottish to Russian along the chromosome.

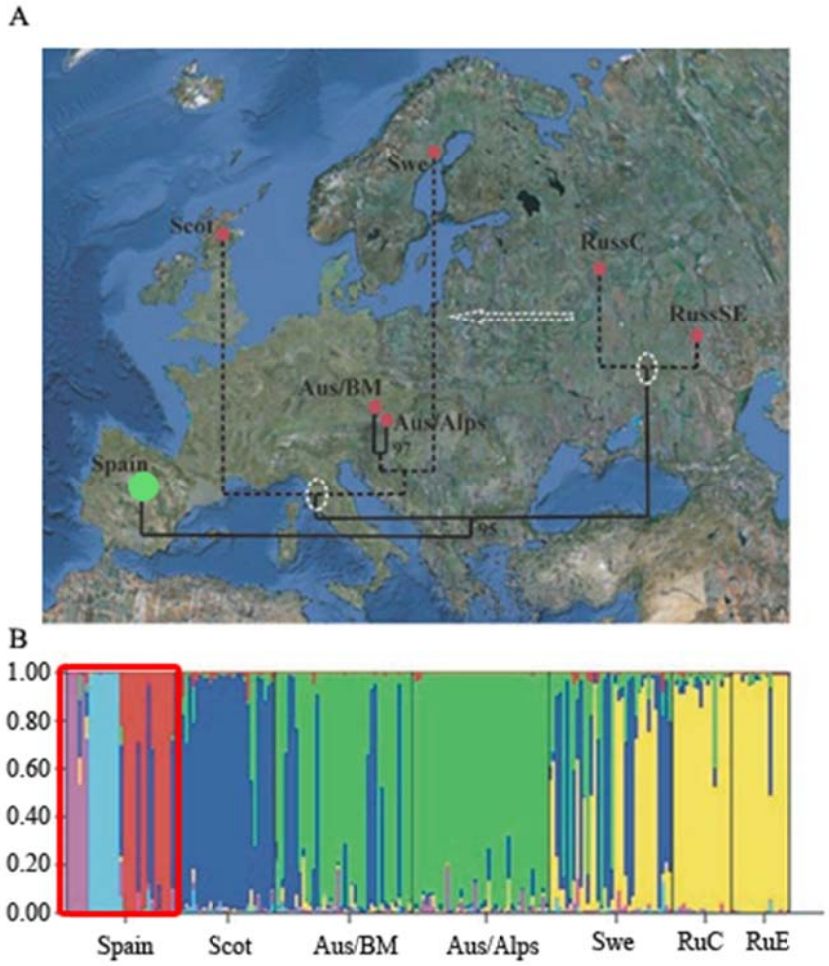


Figure 4.1 Conventional (A) and Bayesian (B) analysis of the population structure in *Populus tremula*, based on 70 microsatellite loci. (A) UPGMA tree based on Nei's standard genetic distance. The tree is informative regarding branching patterns but not branch lengths. The bigger green dot indicates the Spanish population. (B) Bayesian admixture coefficients for individual trees estimated within a linkage model with $K = 6$. More details in DE CARVALHO *et al.* (2010; reproduced with permission from John Wiley & Sons, Chichester, UK).

Discussion

This study revealed that locally varying selection is detectable in European aspen at the continental scale, implying linkage between putatively 'neutral' marker alleles and selected regions. Such finding contrasts with the high level of interpopulation gene flow in this species (N_em = up to 15 effective migrants per generation; LEXER *et al.* 2005), and the rapid decay of LD (within just a few 100 bp; INGVARSSON 2008). Detection of 8% outliers (±2% SE across all population comparisons) in the divergence-based tests indicates that selection on particular genome regions of *P. tremula* is stronger or that recombination in these regions was lower.

Three of the twelve loci with consistent departures from neutrality are located in the proximal end of chromosome XIX. This region is thought to be an incipient sex chromosome; so a low recombination rate could then account for such departures (YIN *et al.* 2008). Its within-population LD reaches 400 kb in some studied populations.

Alternatively, ten of the twelve outlying loci are close (< 3 kb) to genes. Some are even within transcribed regions, and others could be adjacent to *cis*-regulatory elements. A confrontation of the 12 candidate loci with the *Populus* genome assembly revealed that all were located in gene-rich regions.

Analyses for detection of 'selective sweeps' detect limited diversity-based outlier loci, and they also rarely coincide with the ones detected by divergence-based methods. This pattern was not due to the inclusion of the bottlenecked Spanish or the admixed Swedish populations, since many population-pairwise comparisons followed the same trend, nor to errors in the spatial scale selected, since the number of selective sweeps was low at the three studied scales. Quite the contrary, this pattern is expected when sweeps are driven by local adaptation acting on standing variation (soft sweeps), rather than from new mutations (hard sweeps; PENNINGS and HERMISSON 2006). Presence of soft sweeps is theoretically expected when scaled mutation rate ($\theta = 2N_e\mu$) exceeds 0.01 (PENNINGS and HERMISSON 2006). This value is easily reachable for European aspen, whose effective population size (N_e) is in order of 10^5 (INGVARSSON 2008). Moreover, migration among geographically disconnected populations would increase standing variation.

The Bayesian genetic structure analysis exhibits the admixed background of Scandinavian population, mostly from the Western and Eastern lineages. The Scandinavian overlay of ice during glacial maxima suggests that this pattern could not be generated by anything other

than postglacial colonisations, which have created a similar pattern in other species (HEWITT 2000). This admixture holds the simplest explanation for the 'step' in the geographic cline for the bud set. Of course, clinal variation cannot be attributed exclusively to admixture, as relevant environmental factors (especially day length) also vary along sampled populations (HALL *et al.* 2007). However, admixture contributes to this variation, as visible from the step in the cline, but also from the variance and selection differentials for bud set. Both parameters are likewise elevated in the middle of the gradient. Thus, admixture between differentiated postglacial lineages contributes to the standing variation available for natural selection and adaptation. Moreover, it is unlikely that genetic admixture reduced adaptation, as no sign of outbreeding depression was found in the populations in the centre of the gradient and geographic clines for fitness-related growth traits are shallow and linear, as expected from the gradual change in the length of the growth season (HALL *et al.* 2007).

References

ARNOLD, M. L., 2006 *Evolution through Genetic Exchange*. Oxford University Press, Oxford.

BARTON, N. H., and G. M. HEWITT, 1985 Analysis of hybrid zones. Annual Review of Ecology and Systematics **16:** 113-148.

BEAUMONT, M. A., and D. J. BALDING, 2004 Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology **13:** 969-980.

DE CARVALHO, D., P. K. INGVARSSON, J. JOSEPH, L. SUTER, C. SEDIVY *et al.*, 2010 Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. Molecular Ecology **19:** 1638-1650.

FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2003 Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics **164:** 1567-1587.

GOMPERT, Z., and C. A. BUERKLE, 2009 A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Molecular Ecology **18:** 1207-1224.

HALL, D., V. LUQUEZ, V. M. GARCIA, K. R. ST ONGE, S. JANSSON *et al.*, 2007 Adaptive population differentiation in phenology across a latitudinal gradient in European Aspen (*Populus tremula*, L.): A comparison of neutral markers, candidate genes and phenotypic traits. Evolution **61:** 2849-2860.

HEWITT, G., 2000 The genetic legacy of the Quaternary ice ages. Nature **405:** 907-913.

INGVARSSON, P. K., 2008 Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics **180:** 329-340.

INGVARSSON, P. K., M. V. GARCIA, D. HALL, V. LUQUEZ and S. JANSSON, 2006 Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). Genetics **172:** 1845-1853.

LANDE, R., and S. J. ARNOLD, 1983 The measurement of selection on correlated characters. Evolution **37:** 1210-1226.

LEXER, C., C. A. BUERKLE, J. A. JOSEPH, B. HEINZE and M. F. FAY, 2007 Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. Heredity **98:** 74-84.

LEXER, C., M. F. FAY, J. A. JOSEPH, M. S. NICA and B. HEINZE, 2005 Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Molecular Ecology **14:** 1045-1057.

LUQUEZ, V., D. HALL, B. R. ALBRECTSEN, J. KARLSSON, P. INGVARSSON *et al.*, 2008 Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. Tree Genetics & Genomes **4:** 279-292.

PENNINGS, P. S., and J. HERMISSON, 2006 Soft sweeps II-molecular population genetics of adaptation from recurrent mutation or migration. Molecular Biology and Evolution **23:** 1076-1084.

PETIT, R. J., and A. HAMPE, 2006 Some evolutionary consequences of being a tree, pp. 187-214 in *Annual Review of Ecology Evolution and Systematics*.

PIRY, S., G. LUIKART and J. M. CORNUET, 1999 BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. Journal of Heredity **90:** 502-503.

PRZEWORSKI, M., G. COOP and J. D. WALL, 2005 The signature of positive selection on standing genetic variation. Evolution **59:** 2312-2323.

REICH, D., N. PATTERSON, P. L. DE JAGER, G. J. MCDONALD, A. WALISZEWSKA *et al.*, 2005 A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nature Genetics **37:** 1113-1118.

RIESEBERG, L. H., O. RAYMOND, D. M. ROSENTHAL, Z. LAI, K. LIVINGSTONE *et al.*, 2003 Major ecological transitions in wild sunflowers facilitated by hybridization. Science **301:** 1211-1216.

ROUSSET, F., 2008 GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources **8:** 103-106.

SAVOLAINEN, O., and T. PYHAJARVI, 2007 Genomic diversity in forest trees. Current Opinion in Plant Biology **10:** 162-167.

STEARNS, S. C., and R. HOEKSTRA, 2005 *Evolution: An Introduction*. Oxford University Press, Oxford.

VITALIS, R., K. DAWSON and P. BOURSOT, 2001 Interpretation of variation across marker loci as evidence of selection. Genetics **158:** 1811-1823.

YIN, T., S. P. DIFAZIO, L. E. GUNTER, X. ZHANG, M. M. SEWELL *et al.*, 2008 Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Research **18:** 422-430.

YIN, T. M., S. P. DIFAZIO, L. E. GUNTER, D. RIEMENSCHNEIDER and G. A. TUSKAN, 2004 Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. Theoretical and Applied Genetics **109:** 451-463.

CHAPTER 5: Genetic analysis of post-mating reproductive barriers in hybridising European *Populus* species

[see publication in Annex C]

Introduction

Reproductive Isolation (RI) is a key feature of speciation and both Mayr's 'biological species' concept and the more recent 'genic view' theory of species, are linked to it. RI starts when genomic regions ('speciation genes' or other isolation factors) stop being interchangeable, leading to genomic islands of divergence (NOSIL *et al.* 2009; WU 2001) that subsequently spread across the genome. Understanding the genesis and development of RI across the genome could help explain the way speciation occurs.

The genetics of the post-mating components of RI can be investigated using controlled interspecific multi-generation crosses (BOUCK *et al.* 2005; COYNE and ORR 2004; FISHMAN *et al.* 2001). These crosses produce meiosis of first-generation hybrids (F_1), through which parental chromosomes start to break up and create hybrid chromosomes, so that new components of RI are revealed in their structure (LEXER and WIDMER 2008). Segregation distortion – i.e. distortions from the expected Mendelian segregation – in the progeny's genomic regions explain whether such regions resist introgression (for example FISHMAN *et al.* 2001) or, conversely, favour it (for example BOUCK *et al.* 2005).

These experimental crosses – in which F_1 recombination is assayed – are rare in forest trees, since long generation times difficult the creation of recombinant hybrid generations. To circumvent this complication, researchers have used natural hybrids (for example WOOLBRIGHT *et al.* 2008).

In this chapter, I have studied segregation distortion in a backcross progeny BC_1 (i.e. a cross between a F_1 hybrid and a pure individual from one of the parental species) to assess the strength and the genetic architecture of the post-mating reproductive barriers between white poplar and European aspen, which are related species. The potentiality for interspecific introgression in *Populus*, and the role its sex determination region plays in blocking interspecific gene flow were also investigated.

Materials and Methods

Plant material

A controlled backcross of an F_1 natural poplar hybrid (*Populus tremula* × *P. alba*) with a pure *P. alba* was produced. The female, BET3, was a natural hybrid collected from a population in the headwaters of the Tajo, while the white poplar, J1, was a male tree from a natural population in the riverbanks of the Jalón, in the Ebro basin. The F_1 status of the female hybrid was assessed first morphologically and then verified with molecular markers known to be able to segregate both parental species (as in Chapter 6). The plastid and cytoplasmic origin of this individual was determined by sequencing its chloroplast region, *trnC-petN1*, and comparing it with the pure species' haplotypes, resulting to be highly divergent (studied in Chapter 3). This controlled cross yielded 131 individuals under greenhouse conditions, of which just 86 seedlings survived the first year.

To determine the origin of the alleles of the parental species from molecular markers, I used allelic information from one sample of white poplar and one of European aspen, both comprising several European populations. These populations had been previously genotyped by collaborators (see Chapter 4; LEXER *et al.* 2010). Finally, twenty white poplar individuals with known sex (12 females and 8 males) were sampled across Spain.

Molecular genetic markers and genotyping reactions

Genomic DNA was purified from young leaves using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). A genome-wide set of 98 nuclear markers was used for segregation analysis in the interspecific BC_1 (see MACAYA-SANZ *et al.* 2011 for further details). These markers included microsatellite loci available from the *Populus* genome consortium (SMULDERS *et al.* 2001; TUSKAN *et al.* 2006; VAN DER SCHOOT *et al.* 2000; YIN *et al.* 2009), microsatellites isolated *de novo* by my collaborators from expressed sequence tags and from genomic sequence for contig 117 of *Populus trichocarpa* genome assembly v.1, homologous to chromosome XIX of *Populus* (Chapter 4; JOSEPH and LEXER 2008), and single nucleotide polymorphisms. Insertion–deletion (indel) markers were used as well, isolated from expressed sequence tags representing candidate genes for traits involved in ecological divergence between *P. alba* and *P. tremula* (JOSEPH and LEXER 2008). Amplifications, band resolving and sequencing are described elsewhere (Chapter 4; JOSEPH and LEXER 2008; LEXER *et al.* 2005; MACAYA-SANZ *et al.* 2011). Plastid region *trnC-petN1* was amplified and sequenced as described in Chapter 3.

Data analysis

Maker segregation and linkage analyses were carried out in MAPMANAGER QTX. This program tests allelic (= gametic) deviations from Mendelian expectations using Ji-square statistics. Likewise, although using JOINMAP 3.0, I tested deviations in genotype (= zygotic) frequency.

Origin of hybrid parent (BET3) alleles was calculated by comparison to allele frequencies in pure natural populations. This was possible because the majority of allelic variation is located between the two parental species rather than between conspecific populations (LEXER *et al.* 2005). Odds-ratio tests were based on a contingency table constructed using frequency of each allele in the two species, using proc FREQ in SAS version 9 (SAS Institute Inc, Cary, NC, USA). When allele frequency was zero or one in any of the species, odds ratios were undefined. In that case, a Fisher's exact test was performed to test for significant differences in the contingency table. For a small number of loci, for which the odds-ratio test was not significant, species origin was inferred based on linkage with markers with clearly assigned species origin.

Linkage groups obtained after linkage disequilibrium were compared to *P. trichocarpa* physical map to assess synteny and quality of the backcross genotype data. Linkage groups were obtained based in log-of-odds likelihood (threshold set to 3.00). Maps distances in centimorgans were calculated from recombination frequencies using a Kosambi mapping function. Marker physical positions in *P. trichocarpa* were obtained by BLASTn searches in genome assembly v.2.

Loci sited in chromosome XIX were characterised in natural populations using general population genetics parameters, using program FSTAT (GOUDET 1995). LD among markers was assessed as common marker correlation, and P-values computed with GENEPOP (ROUSSET 2008). Between-species divergence of these loci was calculated using F_{ST} and G'_{ST} statistics, and compared to genome-wide F_{ST} values (LEXER *et al.* 2010). The 20 sexed individuals from Spain were characterised for their diversity (H_E) and heterozygosity (H_O).

Results

Marker polymorphism and species origin of donor alleles

Out of the 98 genetic markers analysed, 39 (40%) were polymorphic only in the female F_1 hybrid parent (BET3), 11 (11%) were polymorphic only in the male *P. alba* backcross parent (J1), and 26 (27%) were polymorphic in both parents, whereas 22 (22%) were monomorphic in both parents of the cross (see MACAYA-SANZ *et al.* 2011). The odds-ratio test and Fisher's exact test facilitated the statistical assignment of species origin of alleles segregating from the female F_1 hybrid parent (BET3) for 47 of the markers. For 13 further loci, putative species origin could be assigned based on linkage to markers with clear species assignments (see MACAYA-SANZ *et al.* 2011). Plastid DNA sequencing revealed a *P. tremula* haplotype for BET3, thus indicating species origin of the cytoplasm of the interspecific BC_1 cross.

Segregation distortion

Thirteen markers on six chromosomes (20%) displayed a significant segregation distortion of alleles segregating from the BET3 hybrid parent of the BC_1 (Table 5.1), compared to the three markers expected by chance alone. All loci displayed genotypic segregation distortion as well. For 12 of these loci, the *P. tremula* allele was significantly overrepresented in the backcross progeny.

Synteny with P. trichocarpa

All detected linkages were conserved between *P. trichocarpa* physical map and the interspecific backcross (see MACAYA-SANZ *et al.* 2011). Marker order was completely conserved on chromosome VI, which previously exhibited normal levels of recombination (YIN *et al.* 2004), and on chromosome XIX, where no recombination was detected among four markers (comprising 4560 kb on the *P. trichocarpa* physical genome map), agreeing with previously reported suppressed recombination (Figure 5.1; YIN *et al.* 2008).

Segregation distortion and diversity of chromosome XIX

The proximal end of chromosome XIX held three segregation distorted loci towards *P. tremula* (Table 5.1). Sexed individuals did not present consistent departures from random mating in this region (Table 5.2), thus behaving as autosomal regions. Considering natural populations, LD extended over >560 kb in the white poplar population and in the Swedish

aspen population, whereas no LD was found in the Eastern Alps aspen population (see MACAYA-SANZ *et al.* 2011 for further details). Markers in LD coincide with non-recombining markers of the controlled cross.

Congruence between BC_1 segregation patterns and genomic divergence in natural populations

Reduced interspecific divergence in chromosome XIX in natural populations (measured as F_{ST}) agreed with the increased share of *P. tremula* alleles in this region. No such reduction was observed in chromosome VI markers (Figure 5.2). Similar results were observed using G'_{ST} statistics (see MACAYA-SANZ *et al.* 2011 for further details).

Table 5.1 Genetic markers with segregation distortion in an interspecific BC_1 between *P. tremula* and *P. alba*, including chromosome assignment on *P. trichocarpa* genome assembly v.2, significance levels of segregation distortions in the BC_1, identity of the overrepresented allele for each locus, odds ratios for parental species assignments of alleles in natural populations, and inferred species assignment of the overrepresented allele

Locus	*Populus trichocarpa* Chr.	Distortion *P. alba* × *P. tremula* F_1 parent (♀)	Over-represented allele	Odds ratio *P. alba* / tremula F1 (♀) allele 1	Odds ratio P. alba / tremula F1 (♀) allele 2	Species origin of over-represented allele from F1 (♀)
GCPM 1274	1	****	2	1.55/0.45	0.93/1.32	(P. tremula)
ASP 112376	1	*	1	3.18/0.05	0.00/3.34	P. alba
GCPM 124	1	**	2	42.20/0.07	0.11/2.90	P. tremula
GCPM 1629	3	*****	1	0.62/5.02	1.24/0.00	P. tremula
Thau	9	**	2	2.44/0.00	NA	(P. tremula)
ORPM 23	9	*****	1	0.00/3.49	1.61/0.00	P. tremula
ORNL 149	10	*	1	0.00/3.8	14.00/0.00	P. tremula
ORPM 344	10	****	1	0.00/3.73	2.27/0.25	P. tremula
GCPM 1250	10	****	2	4.03/0.04	0.25/23.05	P. tremula
GCPM 154	12	*****	1	0.00/2.79	NA	(P. tremula)
Yin1	19	**	1	NA	1.28/0.73	P. tremula
Yin2	19	*	2	2.64/0.00	0.00/1.90	P. tremula
ORPM 206	19	***	1	0.00/4.06	10.27/0.00	P. tremula

Abbreviations: BC, backcross; NA, not applicable.
Species assignments supported by genotype data, but not significant in the odds-ratio test, are shown in parentheses.
Significance thresholds from χ^2 tests.
*P<0.05, **P<0.01, ***P<0.005, ****P<0.001, *****P<0.00005.

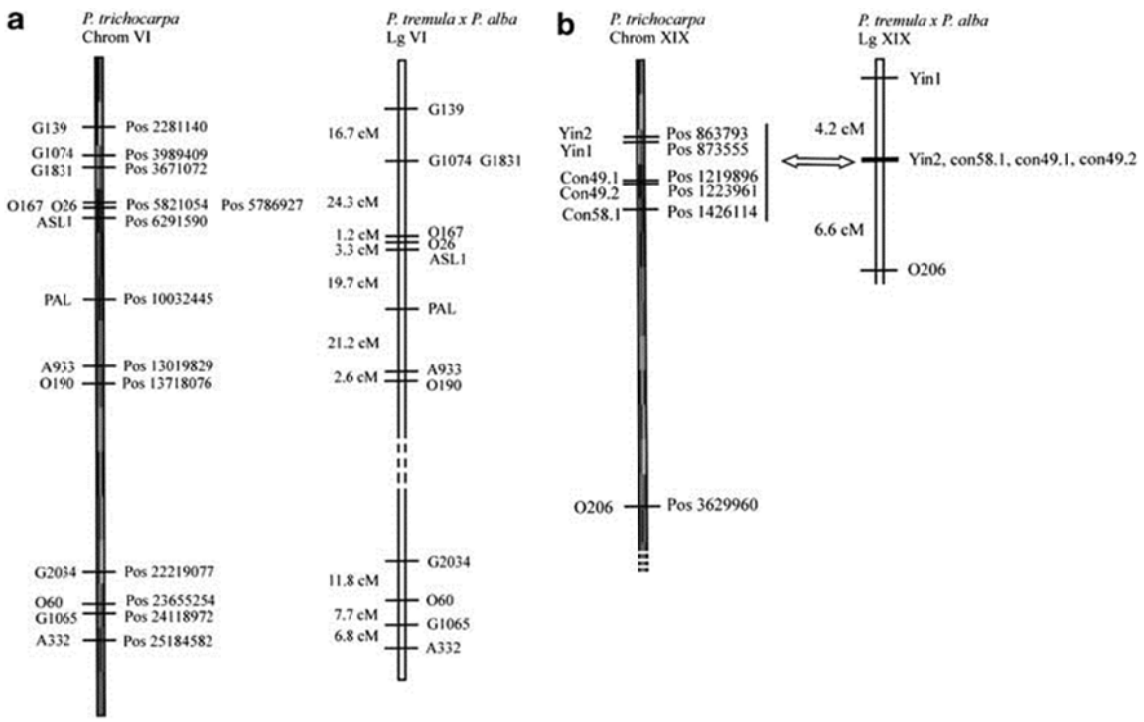


Figure 5.1 Comparison of the *P. tremula* × *P. alba* linkage map to *P. trichocarpa* genome assembly v.2 for chromosome VI (a), known to exhibit normal levels of recombination (YIN *et al.* 2004), and chromosome XIX (b), known to exhibit greatly reduced recombination (YIN *et al.* 2008). Complete synteny between the two maps is indicated by the conserved marker order on chromosome VI. On chromosome XIX, zero recombination was observed between markers Yin2, con58.1, con49.1 and con49.2 on the *P. tremula* × *P. alba* linkage map (indicated by the arrow), which corresponds to >560 kb on the *P. trichocarpa* genome assembly. Figure from MACAYA-SANZ *et al.* (2011; reproduced with permission from Nature Publishing Group)

Discussion

The study of this backcross has revealed extensive segregation distortions (20% of polymorphic loci), the majority showing a prevalence of *P. tremula* alleles (12 out of 13; >90%) amid a *P. alba* genome. These distortions manifest at the zygotic level as well. High levels of mortality (34%) were observed in seedlings, a critical life stage in forest trees. Such patterns could be explained by three non-exclusive hypotheses: (1) epistatic interactions between the hybridising genomes (COYNE and ORR 2004); (2) overdominance or heterozygote advantage in hybrids (HARTL and CLARK 1997); and (3) cyto-nuclear co-adaptation (FUTUYMA 2009).

The role of epistasis in the hybridisation of these two species derives from a recent analysis of genomic admixture in hybrid zones, heterospecific genomic interactions clearly contributing to steep clines in localities where these species co-occur (LEXER *et al.* 2010).

However, epistasis cannot explain by itself the unidirectional bias of the segregation (FISHMAN *et al.* 2001) as, in fact, epistasis would actually favour a bias in the opposite direction (*P. alba* alleles prevalence).

Heterozygote advantage implies that heterozygous genotypes show more fitness. Selection would thus favour *P. tremula* alleles in a *P. alba* background (actual BC_1 pure species). This hypothesis is supported by a high seedling mortality (34%) and the low heterozygosity of the *P. alba* parent (38%), compared to that of the hybrid (67%). Heterozygosity increases can reduce the effects of inbreeding; biparental inbreeding in *P. alba* becomes apparent from the short-range kinship coefficients among individuals (F_{ij}) observed in recent studies (VAN LOO *et al.* 2008), and from the extraordinary dimensions of clones in the southern European populations (Chapter 6).

Table 5.2 Chromosome XIX diversity statistics for *P. alba* individuals with known sex, including expected (H_E) and observed (H_O) heterozygosity and inbreeding coefficients (F_{IS}) in each group, and the female/male ratio of H_O

Locus	Females				Males				H_O ratio
	A	H_E	H_O	F_{IS}	A	H_E	H_O	F_{IS}	♀/♂
Yin2	4	0.645	0.273	0.589	4	0.692	0.625	0.103	0.437
Yin1	3	0.301	0.333	-0.114	3	0.492	0.625	-0.296	0.533
Con03.1	3	0.610	0.909	-0.527	3	0.689	1.000	-0.539	0.909
Con49.1	5	0.667	0.500	0.258	4	0.592	0.625	-0.061	0.800
Con49.2	5	0.825	0.500	0.411	5	0.842	0.375	0.571[a]	1.333
Con58.1	9	0.892	1.000	-0.128	9	0.858	0.750	0.134	1.333
O206	2	0.290	0.333	-0.158	2	0.325	0.375	-0.167	0.888
O276	5	0.656	0.750	-0.151	5	0.700	0.625	0.114	1.200

[a] Significantly different from zero at the 0.05 level.

Increased introgression can be accounted by cyto-nuclear interactions, as long as BC_1 plastids and cytoplasm (inherited from the mother in *Populus*) are from *P. tremula* (note that the BC_1 mother, BET3, had in turn a *P. tremula* mother). Nuclear genes that are co-adapted to specific cytoplasm composition should be selected, so that *P. tremula* alleles should prevail in these genic areas. In the absence of reciprocal crosses, it is impossible to accept or reject either of these hypotheses (GALLOWAY and FENSTER 1999), although congruent results obtained with natural populations favour the generality of these findings. Our results suggest that asymmetries in post-mating barriers in these forest trees may result in

introgression rather than evolution of reinforcement following secondary contact. In fact, a higher than expected frequency for two overrepresented loci in the controlled cross was found in a well-studied hybrid zone (LEXER *et al.* 2010).

Regarding the sex determination region of *Populus*, its reduced level of divergence is not consistent with its expected implication on speciation. Moreover, our study has detected an increased linkage disequilibrium (LD) in the controlled crosses and in the natural populations of white poplar, but almost none in the natural populations of European aspen. The reduced effective population size (N_e) in white poplar and its metapopulation structure could account for this LD pattern.

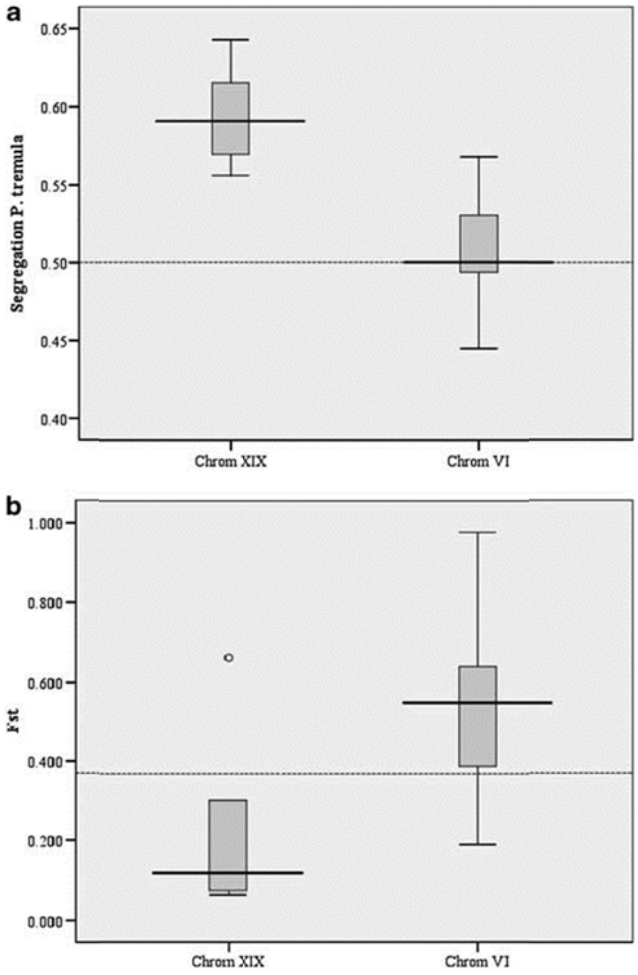


Figure 5.2 Box plots showing introgression and divergence of genetic markers on chromosomes VI and XIX relative to genome-wide expectations. (a) Segregation of *P. tremula* alleles in the interspecific BC_1 (dotted line, Mendelian expectation of 0.5). (b) Interspecific divergence (F_{ST}) in natural populations (dotted line, genome-wide expectation of 0.369; LEXER *et al.* 2010). Increased introgression of *P. tremula* alleles in the controlled interspecific BC_1 (a) and reduced interspecific divergence in natural populations (b) are visible for chromosome XIX. Figure from MACAYA-SANZ *et al.* (2011; reproduced with permission from Nature Publishing Group)

These counter-intuitive results could be due to the special features of *Populus*' sex determination region. First, because this region is rich in R-genes, an important functional class, highly amplified in the *Populus* genome (KOHLER *et al.* 2008). It is assumed that these genes are under balancing selection (FUTUYMA 2009) and, if so, introgression would be favoured. Second, because variability in the sex determination systems of *Populus* (as shown in genetic maps of *P. alba* and *P. tremula* × *P.tremuloides* in contrast to *P. trichocarpa*; PAKULL *et al.* 2009; PAOLUCCI *et al.* 2010; YIN *et al.* 2008) suggests that *Populus*' sex chromosome is in an early stage of evolution. Our results back up this latter statement, since microsatellites in this region behave as codominant, with no consistent pattern of reduced heterozygosity in either sex (Table 5.2).

References

BOUCK, A., R. PEELER, M. L. ARNOLD and S. R. WESSLER, 2005 Genetic mapping of species boundaries in Louisiana irises using IRRE retrotransposon display markers. Genetics **171:** 1289-1303.

COYNE, J. A., and H. A. ORR, 2004 *Speciation*. Sinauer Associates, Incorporated, Sunderland, MA.

FISHMAN, L., A. J. KELLY, E. MORGAN and J. H. WILLIS, 2001 A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. Genetics **159:** 1701-1716.

FUTUYMA, D. J., 2009 *Evolution*. Sinauer Associates, Incorporated, Sunderland, MA.

GALLOWAY, L. F., and C. B. FENSTER, 1999 The effect of nuclear and cytoplasmic genes on fitness and local adaptation in an annual legume, *Chamaecrista fasciculata*. Evolution **53:** 1734-1743.

GOUDET, J., 1995 FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity **86:** 485-486.

HARTL, D. L., and A. G. CLARK, 1997 *Principles of Population Genetics*. Sinauer Associates, Incorporated, Sunderland, MA.

JOSEPH, J. A., and C. LEXER, 2008 A set of novel DNA polymorphisms within candidate genes potentially involved in ecological divergence between *Populus alba* and *P. tremula*, two hybridizing European forest trees. Molecular Ecology Resources **8:** 188-192.

KOHLER, A., C. RINALDI, S. DUPLESSIS, M. BAUCHER, D. GEELEN *et al.*, 2008 Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. Plant Molecular Biology **66:** 619-636.

LEXER, C., M. F. FAY, J. A. JOSEPH, M. S. NICA and B. HEINZE, 2005 Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Molecular Ecology **14:** 1045-1057.

LEXER, C., J. A. JOSEPH, M. VAN LOO, T. BARBARA, B. HEINZE *et al.*, 2010 Genomic Admixture Analysis in European *Populus* spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. Genetics **186:** 699-U391.

LEXER, C., and A. WIDMER, 2008 The genic view of plant speciation: recent progress and emerging questions. Philosophical Transactions of the Royal Society B-Biological Sciences **363:** 3023-3036.

MACAYA-SANZ, D., L. SUTER, J. JOSEPH, T. BARBARA, N. ALBA *et al.*, 2011 Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species. Heredity **107:** 478-486.

NOSIL, P., D. J. FUNK and D. ORTIZ-BARRIENTOS, 2009 Divergent selection and heterogeneous genomic divergence. Molecular Ecology **18:** 375-402.

PAKULL, B., K. GROPPE, M. MEYER, T. MARKUSSEN and M. FLADUNG, 2009 Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). Tree Genetics & Genomes **5:** 505-515.

PAOLUCCI, I., M. GAUDET, V. JORGE, I. BERITOGNOLO, S. TERZOLI *et al.*, 2010 Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree Genetics & Genomes **6:** 863-875.

ROUSSET, F., 2008 GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources **8:** 103-106.

SMULDERS, M. J. M., J. VAN DER SCHOOT, P. ARENS and B. VOSMAN, 2001 Trinucleotide repeat microsatellite markers for black poplar (*Populus nigra* L.). Molecular Ecology Notes **1:** 188-190.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596-1604.

VAN DER SCHOOT, J., M. POSPISKOVA, B. VOSMAN and M. J. M. SMULDERS, 2000 Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theoretical and Applied Genetics **101:** 317-322.

VAN LOO, M., J. A. JOSEPH, B. HEINZE, M. F. FAY and C. LEXER, 2008 Clonality and spatial genetic structure in *Populus* x *canescens* and its sympatric backcross parent *P. alba* in a Central European hybrid zone. New Phytologist **177:** 506-516.

WOOLBRIGHT, S. A., S. P. DIFAZIO, T. YIN, G. D. MARTINSEN, X. ZHANG *et al.*, 2008 A dense linkage map of hybrid cottonwood (*Populus fremontii* x *P. angustifolia*) contributes to long-term ecological research and comparison mapping in a model forest tree. Heredity **100:** 59-70.

WU, C. I., 2001 The genic view of the process of speciation. Journal of Evolutionary Biology **14:** 851-865.

YIN, T., S. P. DIFAZIO, L. E. GUNTER, X. ZHANG, M. M. SEWELL *et al.*, 2008 Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Research **18:** 422-430.

YIN, T. M., S. P. DIFAZIO, L. E. GUNTER, D. RIEMENSCHNEIDER and G. A. TUSKAN, 2004 Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. Theoretical and Applied Genetics **109:** 451-463.

YIN, T. M., X. Y. ZHANG, L. E. GUNTER, S. X. LI, S. D. WULLSCHLEGER *et al.*, 2009 Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome. New Phytologist **181:** 498-503.

CHAPTER 6: Causes and consequences of large clonal assemblies in a poplar hybrid zone

[see manuscript in Annex D]

Introduction

Vegetative asexual reproduction (or clonal reproduction) is an ability easily acquired by angiosperm plants (SACHS 2001), hence it is quite common (KLIMES *et al.* 1997). However, the ecological and genetic factors that govern it are more complex, and poorly known. Plant persistence under adversity (in suboptimal or recently colonised environments), and better foraging or ramet specialisation can independently favour the development of clonality.

In general, asexual reproduction increases when sexual reproduction is impaired. However, the first can also limit the second by competition of propagules, mutation load in sexual traits or geitonogamy. Some authors argue that this detrimental feedback could reach a tipping point, beyond which populations fall into an extinction vortex (HONNAY and BOSSUYT 2005). Although asexual reproduction sometimes favours local persistence, theory predicts that selection promotes sexual reproduction. Common extinction of phylogenetic branches with asexual reproduction suggests the prevalence of sexual reproduction in the medium term.

At ecological scale, clonal levels vary across populations and geographical ranges (see Chapter 1 for examples in *Populus*). Clonal structure can be broadly depicted by three parameters. Since ramet distribution follows a Pareto distribution, the strength of its slope, together with genotypic richness and evenness, serve to describe clonal structure (ARNAUD-HAOND *et al.* 2007). These three parameters depict the frequency of sexual recombination and the intensity of clonality and dominance among genets.

Competition among genets appears to be affected by ecological and genetic factors, but it can just also obey randomness. Flood control, soil moisture, or water table level are implied in the clonal spread of several populations of *Populus* (e.g. SLAVOV *et al.* 2010). Besides, some genets could be genetically better adapted to asexual reproduction, as it has been shown under controlled conditions in *Populus* (e.g. STENVALL *et al.* 2005). This variation could come from admixture or hybridisation, i.e. genetic ancestry could be contributing to determine the ability to spread asexually (e.g. VAN LOO *et al.* 2008).

The effects of prolonged clonal reproduction on genetic and genotypic diversity are not well known. In first term, clonal reproduction reduces the number of genotypes by competition or just by drift, but it also protects allelic richness, slowing the pace of genetic drift. Individual-based simulations reveal that allelic diversity is maintained despite high clonality (BALLOUX *et al.* 2003). Long-term effects of a reduction of genotypic richness in allelic diversity or population fitness have been, however, barely explored.

Natural hybrid zones in poplars allow to investigate how genetic ancestry, and in particular introgression by related species, influences asexual propagation rates, a factor previously overlooked in clonal studies, albeit a known factor in genet vigour. As long as clonal levels differ among poplar hybrid zones across Europe, comparative approaches can also be developed.

My study focused on a white poplar population in the surroundings of Aranda de Duero (province of Burgos, Spain), with little but noticeable levels of introgression by European aspen. Using a panel of 73 microsatellites, I identified clones and sketched their genetic background, in order to: i) examine the spatial structure of large clones and compute population genetic statistics in a largely clonal hybrid zone; ii) discern links among genetic ancestry and clonal success; iii) appraise the impact of large clones on population demography, and on sexual outcome and population allelic and genotypic diversity.

Materials and Methods

The study focused on a poplar population on the riverbanks of the Duero River, around Aranda de Duero (UTM zone 30T, X 438000, Y 4613000), in Central Spain. This population is a discontinuous, mixed riparian forest including poplars (white, grey and black), willows, ashes, elms and alders. European aspen is present in reduced spots. Together with the core hybrid zone, other areas have been prospected, less intensively, to complete the analysis. Spacing of the core population was at least 100 metres as a general rule, except when two ramets belonged obviously (by morphology) to different genets, in which case smaller distances were allowed. Spacing out of the core area was larger of 100 metres. I collected leaf tissue of 533 poplar trees (362 white poplars, 143 grey poplars, and 28 European aspens), recording their UTM coordinates.

After DNA isolation, all the samples were genotyped using a panel of 20 SSR markers. This information was used to resolve different MLGs (Multi-Locus Genotypes) and MLLs (Multi-Locus Lineages; i.e. ignoring somatic mutations) using GIMLET software (VALIERE 2002). Afterwards, at least one sample of each MLL (in total 137 samples) was further genotyped with 53 SSR markers (for details of DNA isolation, PCR, and fragment electrophoresis see Annex).

Some common population genetic parameters were calculated by means of SPAGeDi 1.3 (HARDY and VEKEMANS 2002) and HP-Rare 1.0 (KALINOWSKI 2005), as well as some descriptors for clonal structure [genotypic richness (*R*), Simpson's evenness (*V*), and the additive inverse of the slope of the log-scaled Pareto distribution (β)]. These calculations were also done for a hybrid population from Austria and Slovakia (Danube River; LEXER *et al.* 2010; VAN LOO *et al.* 2008), and another one from North Italy (Ticino River; CASTIGLIONE *et al.* 2010).

The high genotyping effort allowed a rough estimation of the age of some genets, whose sampling had been more intense (5 genets). Two methods were used. The first takes into account the genetic divergence within genets (ALLY *et al.* 2008). In this method, genet size trends (constant vs. growing) are initially inferred by comparison of several estimators of population mutation rate (θ). Once a size model is chosen, TMRCA (Time since Most Recent Common Ancestor) can be calculated. The second is based on the fact that the process of somatic mutation accumulation follows a Poisson distribution (e.g. THOMSON *et al.* 2000). Thus, the number of accumulated mutations should be a function genet age (more details in Annex).

Genomic ancestry was assessed by analysing locus specific ancestries (LSAs), and determining specific homozygosity and interspecific heterozygosity of each locus (as in LINDTKE *et al.* 2012). Individual-level inbreeding was estimated by computing intra-individual kinship (F_i) using SPAGeDi 1.3 (HARDY and VEKEMANS 2002). Allelic frequency reference was constructed to avoid overrepresentation of large clonal assemblies (see Results). Singular patterns in large clones were surveyed.

Kinship among genets was inferred using COLONY 2.0 (JONES and WANG 2010), and FRANz 2.0.0 (RIESTER *et al.* 2009; RIESTER *et al.* 2010). FRANz considers clone size as prior information, in contrast to COLONY. With these results, correlations of sexual

offspring number with ramet number and with genet extension were tested using the Spearman rank method.

Finally, population demographic history was investigated using MSVAR 1.2 (BEAUMONT 1999; STORZ and BEAUMONT 2002). Linear and exponential models were tested, and the most fitting one was selected using Akaike information criterion (details of the runs in Annex). Chain convergence was evaluated using Gelman-Rubin statistic (GELMAN and RUBIN 1992).

Results

SSR genotyping (20 loci) of the whole sample set (533 samples) resolved 132 MLGs (96 white poplar out of 362 samples, 19 grey poplar out of 143, and 17 European aspen out of 28). After discarding somatic mutations, 82 MLLs were recognised in white poplar and 13 MLLs in grey poplar. Two very large genets (in terms of number of ramets; white poplar MLL009, with 189 ramets; and grey poplar MLL006, with 124 ramets) caused this striking reduction of genotype numbers. Also, two rather large genets were found (MLL025 with 26 ramets, and MLL073 with 17 ramets, both white poplars). Geographic extension of clones was not completely correlated with number of ramets (Table 6.1).

Genotypic richness (R) and evenness (V) were comparatively low in the Duero hybrid zone, in comparison with the Danube and Ticino populations, both for the white and grey poplar subpopulations (Table 2 in Annex D). Expected heterozygosity (H_E) and rarefied allele richness (A) were however similar in all the populations, somewhat higher in the Danube. Grey poplar subpopulations were consistently more diverse (Table 2 in Annex D), as expected from their hybrid ancestry.

From the comparison of different estimators of population mutation rate (θ), I concluded that the genets were increasing their size, because the average number of pairwise differences per locus (π_k) was equivalent to $2S_k/n$ (ratio 1:1; more details in Annex). The average genet age was over one thousand years, and apparently smaller genets were older, although their error rates were larger also. The two largest clones (MLL009 & MLL006) were dated to 547 (± 424) and 553 (± 520) years old, respectively. Under the method based on accumulation of somatic mutations since common ancestor, these huge clones were aged 8,411 (± 3,726) and 5,127 (± 3,119) years old, respectively.

Table 6.1 Representative clones (MLLs) found in the Duero hybrid poplar zone, including large clones. *N*: Number of ramets; Extension: longest distance among ramets within clones; *Q*: ancestry value (1 for pure *P. alba*); F_i: inbreeding coefficient; Offspring: number of descendants detected by either COLONY or FRANz software (see main text for details). Star indicates hybrid ancestry (*P.* × *canescens*); the remaining genets were assigned to *P. alba*. NA: not applicable. The last two rows report averaged values for remaining genets. Bold font indicates significantly different from zero ($P < 0.05$). Parentheses indicate standard deviation.

MLL	***N***	**Extension (km)**	***Q***	F_i	**Offspring**	
					COLONY	**FRANz**
MLL009	189	99.5	1.00	-0.013	29	21
MLL006*	124	158.6	0.42	-0.121	7	2
MLL025	26	74.6	1.00	-0.116	23	23
MLL073	17	5.6	1.00	-0.048	1	2
MLL002	8	22.6	1.00	-0.066	0	0
MLL074	7	10.7	1.00	**0.226**	1	1
MLL011	6	4.1	1.00	-0.022	13	8
MLL049	5	17.5	1.00	-0.022	0	0
MLL057*	5	0.6	0.80	0.002	0	2
MLL083	5	2.6	1.00	-0.059	0	2
MLL126	3	4.9	0.99	0.215	5	2
MLL086	2	1.5	1.00	0.162	5	4
MLL053	2	5.2	1.00	0.080	0	4
MLL111	2	17.1	0.99	0.166	0	2
MLL058	1	NA	1.00	0.170	3	2
MLL030*	1	NA	0.90	0.002	0	1
MLL120	1	NA	1.00	**0.328**	8	4
Rest of *P. alba*	1.3 (0.5)	NA	1.00 (0.00)	0.064 (0.090)	0.08 (0.32)	0.24 (0.43)
Rest of hybrids	1.3 (0.5)	NA	0.74 (0.10)	-0.006 (0.121)	0.00 (0.00)	0.40 (0.52)

The genetic background of grey poplar was consistent with first generation hybrids (F_1) or backcrosses of different white poplar generations. Whereas numerous pure white poplars showed slight signatures of introgression, few pure European aspens did so. Some genomic regions of hybrids (including backcrosses) showed an aspen background dominance (chromosomes II, IV, VI, VIII, and XII). Within these regions, two makers (GCPM1809 and GCPM1065) were homozygous for *P. tremula* background in Duero, but also in three other European hybrid zones (LINDTKE *et al.* 2012). Another locus (GCPM154) displayed a similar bias in the Duero and Danube populations (LEXER *et al.* 2010), and was also distorted towards aspen in the backcrossed progeny (Chapter 5). This marker is placed near two genes in *P. trichocarpa* genome assembly v. 3. Some large clones were highly interspecies heterozygous across this region. Moreover, many of the larger clones were severely introgressed by aspen around the GCPM1274 marker, with a significantly abnormal abundance of the allele 207. This marker also segregated in distortion in the backcrossed progeny (Chapter 5) and laid beside three genes. Inbreeding coefficients at the individual level (F_i) followed a normal distribution, but the values of the largest four genets were consistently below the mode (i.e. more outbreeding).

In general, large clones had more sexual descendants, as observed in parentage inferences of COLONY and FRANz (Table 6.1). Spearman's rank tests displayed significant correlations between ramet and offspring numbers, either considering all genets or just the ten largest ones. These correlations turned non-significant when spatial expansion was contrasted with offspring number. The most salient exception was MLL006, a huge hybrid clone with only seven offspring, all probably backcrosses to white poplar (averaged STRUCTURE's Q-value = 0.75).

Finally, demographical analyses by MSVAR showed an old but soft decline in population effective size. AIC slightly favoured an exponential model over a linear one. Both models indicated a persistent reduction in effective population size of one tenth during the last hundreds of thousands years. Notwithstanding the decline, current effective population size was still considerable ($N \approx 2{,}240$). These calculations were done considering a generation time of 40 years, but given the high clonality and probably low genet turnover of the population, a longer generation time may have been more realistic, and would have implied that population decline would have begun earlier.

Discussion

The Duero hybrid zone is characterised by a few large and widespread clones and many small ones (SANTOS-DEL-BLANCO *et al.* 2012). Such extensive genets could have resulted from propagation by humans, but this explanation is unlikely because the quality of white and grey poplar wood is low and the age of the large clones appears to exceed the time of complex human settlements in the region (Romanisation). Additionally, large clones do not possess outstanding phenotypes that could have favoured their cultivation, nor root-suckering expansion is a sufficient explanation due to the large extension of some clones (> 150 km). More probably, twig translocation by water (BARSOUM *et al.* 2004) or by other means (e.g. birds) might have contributed to long-distance spread.

Compared to the two other European white poplar hybrid zones, the Duero population displayed lower values of genotypic richness and evenness. Other studies in poplar have revealed higher genotypic richness in less stressful environments, as in areas with more accessibility to ground water (VONLANTHEN *et al.* 2010) or more humid climate (SLAVOV *et al.* 2010), while an increase in clonal propagation rate (in opposition to sexual reproduction) is expected to reduce genotype number (BALLOUX *et al.* 2003) , given the increasingly higher probability of losing genotypes by drift. In conclusion, sexual reproduction difficulties may underlie genotypic richness differences. However, besides the above mentioned water-related handicaps, other factors could impair sexual reproduction, which in poplars is extremely delicate due to the fragile constitution of their seeds. River course regime may surely be implicated in recruitment success, since flooding bares spaces for seedling establishment, but late events can also ruin it. A lack of snowmelt floods in the Duero population, besides the much drier summers of the region could also contribute to explain reduced genotype richness. The Ticino and Danube rivers receive waters from the heights of the Alps (in particular in the part of the courses where the hybrid zones are located), whereas the Duero carries water from more distant and less snowed-in mountain ranges.

The lower genotypic richness was not coupled with a lower genetic (allelic) diversity at the genet level, in line with numerical modelling (BALLOUX *et al.* 2003) and other studies in clonal species (HONNAY and BOSSUYT 2005). However, the dynamics of genetic variation in clonal populations are not sufficiently understood yet. Considering allelic richness, the more clonal population of *P. trichocarpa* studied by SLAVOV *et al.* (2010) displayed reduced values. Moreover, numerical modelling did not include selection as a driving force

(BALLOUX *et al.* 2003), although selection was the factor suggested to produce the right-tailed distribution of genet sizes commonly found in clonal organisms.

Although heterosis in poplar hybrids causes extreme phenotypes, including more vigorous clonal propagation (SCHWEITZER *et al.* 2002; VAN LOO *et al.* 2008), in the Duero population hybrids did not show wider clonal spread (from the ten largest clones, just two were hybrid, matching to the sampling share). However, many of such large clones possessed a *P. tremula* allele in a region with unusual interspecific heterozygosity, by the end of chromosome I (GCPM 1274). This telomeric region is conspicuous for displaying an exceptional density of NBS-class resistance genes (KOHLER *et al.* 2008) and expressed small RNAs (KLEVEBRING *et al.* 2009). Other studies hint at local interspecific heterozygosity as a cause of increased fitness (LEXER *et al.* 2004; LINDTKE *et al.* 2012; SCHWEITZER *et al.* 2002).

Values for individual inbreeding coefficient (F_i) were atypically low for the largest four clones, indicating that, besides having local genomic spots of interspecific heterozygosity, they possess higher overall heterozygosity too. Few studies have analysed the correlation between heterozygosity and clonal success, although some have found aligned results. For example, genotypes of mono-clonal populations of *Decodon verticillatus* are usually highly heterozygous (ECKERT 2001).

Larger clones, which are also probably older, were found to be parents of more genets than smaller ones. This sexual dominance was surely caused by the higher number of ramets and the longer persistence of the large genets. Furthermore, this finding implies that ageing sterility has not affected these clones yet, and that effective population size is being constrained because of this assortative mating (not all the genets reproduce in the same manner). Ageing sterility originates when prolonged clonal reproduction impairs sexual function by genetic erosion, since not using sex impedes selection to purge sex-related deleterious mutations. ALLY *et al.* (2010) estimated in 500 to 20000 years the time needed to reach male sterility in *P. tremuloides*. Regarding effective population size, although an unbalance in sexual reproduction among genets appears to increase effects of genetic drift, it has not prejudiced the population to the present, as its levels of inbreeding (F_{IS} = 0.067) are still similar to other poplar hybrid zones with lesser levels of clonality (F_{IS} = 0.09; COLE 2005). Interestingly, MLL006, the large hybrid female has fewer descendants than expected from its size. Although hybrid sterility and hybrid vigour could be factors involved, barriers to backcrossing with white poplar is probably the most determining factor. This genet produces viable seeds, with healthy seedlings, thus hybrid sterility cannot be a major factor.

Besides, hybrid vigour could have produced a sudden expansion of this clone in a brief period of time, not allowing time for extensive sexual reproduction, but accumulation of somatic mutations refutes this scenario (see above). So probably a conjunction of pre- and post-zygotic (e.g. phenological asynchronies) ecological barriers underlie the reduced sexual success of this genet.

Bayesian MSVAR analyses to ascertaining population demography indicate a long-term, but soft, decline. This contrasts with the continuous growth found in large clones. High levels of genetic diversity added to the moderate levels of inbreeding suggest that sexual dominance of large clones is not responsible for such decline, but rather intraspecific competition during clone expansion. Several processes may underlie this clonal expansion. Desiccation and cooling owed to Quaternary climatic changes may have hampered white poplar sexual reproduction in the Mediterranean basin, thus promoting clonal propagation. Besides, human impact on the natural ecology and dynamics of rivers, namely, river regulation, surely impedes normal sexual recruitment. Lastly, habitat fragmentation could also have reduced sexual function due to pollen limitation and reduced habitat for establishment.

The balance between sexual and asexual contributions to reproduction depends on ecological conditions, so it changes when ecology changes. Populations are resilient to these oscillations, but some authors argue that crossing a tipping point could lead to an irreversible point. Our Duero hybrid population appears to be far from this point, according to its population genetic characteristics and demography, but new, stressful incidents could drive it beyond a non-return point. Two mechanisms can drive this situation: the loss of less competitive genotypes by intraspecific exclusion, and the ageing sterilisation of the large clones.

References

ALLY, D., K. RITLAND and S. P. OTTO, 2008 Can clone size serve as a proxy for clone age?An exploration using microsatellite divergence in *Populus tremuloides*. Molecular Ecology **17:** 4897-4911.

ALLY, D., K. RITLAND and S. P. OTTO, 2010 Aging in a Long-Lived Clonal Tree. Plos Biology **8**.

ARNAUD-HAOND, S., C. M. DUARTE, F. ALBERTO and E. A. SERRAO, 2007 Standardizing methods to address clonality in population studies. Molecular Ecology **16:** 5115-5139.

BALLOUX, F., L. LEHMANN and T. DE MEEUS, 2003 The population genetics of clonal and partially clonal diploids. Genetics **164:** 1635-1644.

BARSOUM, N., E. MULLER and L. SKOT, 2004 Variations in levels of clonality among *Populus nigra* L. stands of different ages. Evolutionary Ecology **18:** 601-624.

BEAUMONT, M. A., 1999 Detecting population expansion and decline using microsatellites. Genetics **153:** 2013-2029.

CASTIGLIONE, S., A. CICATELLI, R. LUPI, G. PATRIGNANI, T. FOSSATI *et al.*, 2010 Genetic structure and introgression in riparian populations of *Populus alba* L. Plant Biosystems **144:** 656-668.

COLE, C. T., 2005 Allelic and population variation of microsatellite loci in aspen (*Populus tremuloides*). New Phytologist **167:** 155-164.

ECKERT, C. G., 2001 The loss of sex in clonal plants. Evolutionary Ecology **15:** 501-520.

GELMAN, A., and D. B. RUBIN, 1992 Inference from iterative simulation using multiple sequences. Statist. Sci. **7:** 457-472.

HARDY, O. J., and X. VEKEMANS, 2002 SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes **2:** 618-620.

HONNAY, O., and B. BOSSUYT, 2005 Prolonged clonal growth: escape route or route to extinction? Oikos **108:** 427-432.

JONES, O. R., and J. L. WANG, 2010 COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources **10:** 551-555.

KALINOWSKI, S. T., 2005 HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes **5:** 187-189.

KLEVEBRING, D., N. R. STREET, N. FAHLGREN, K. D. KASSCHAU, J. C. CARRINGTON *et al.*, 2009 Genome-wide profiling of *Populus* small RNAs. Bmc Genomics **10**.

KLIMES, L., J. KLIMESOVA, R. HENDRIKS and J. VAN GROENENDAEL, 1997 Clonal plant architecture: a comparative analysis of form and function, pp. 1-29 in *The Ecology and Evolution of Clonal Plants*, edited by H. DE KROON and J. VAN GROENENDAEL. Backhuys, Leiden.

KOHLER, A., C. RINALDI, S. DUPLESSIS, M. BAUCHER, D. GEELEN *et al.*, 2008 Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. Plant Molecular Biology **66:** 619-636.

LEXER, C., J. A. JOSEPH, M. VAN LOO, T. BARBARA, B. HEINZE *et al.*, 2010 Genomic Admixture Analysis in European *Populus* spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. Genetics **186:** 699-U391.

LEXER, C., Z. LAI and L. H. RIESEBERG, 2004 Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytologist **161:** 225-233.

LINDTKE, D., C. A. BUERKLE, T. BARBARA, B. HEINZE, S. CASTIGLIONE *et al.*, 2012 Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. Molecular Ecology **21:** 5042-5058.

RIESTER, M., P. F. STADLER and K. KLEMM, 2009 FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics **25:** 2134-2139.

RIESTER, M., P. F. STADLER and K. KLEMM, 2010 Reconstruction of pedigrees in clonal plant populations. Theoretical Population Biology **78:** 109-117.

SACHS, T., 2001 Developmental processes and the evolution of plant clonality. Evolutionary Ecology **15:** 485-500.

SANTOS-DEL-BLANCO, L., A. DE-LUCAS, S. GONZÁLEZ-MARTÍNEZ, R. SIERRA-DE-GRADO and E. HIDALGO, 2012 Extensive Clonal Assemblies in *Populus alba* and *Populus* x *canescens* from the Iberian Peninsula. Tree Genetics & Genomes**:** 1-12.

SCHWEITZER, J. A., G. D. MARTINSEN and T. G. WHITHAM, 2002 Cottonwood hybrids gain fitness traits of both parents: A mechanism for their long-term persistence? American Journal of Botany **89:** 981-990.

SLAVOV, G. T., S. LEONARDI, W. T. ADAMS, S. H. STRAUSS and S. P. DIFAZIO, 2010 Population substructure in continuous and fragmented stands of *Populus trichocarpa*. Heredity **105:** 348-357.

STENVALL, N., T. HAAPALA, S. AARLAHTI and P. PULKKINEN, 2005 The effect of soil temperature and light on sprouting and rooting of root cuttings of hybrid aspen clones. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere **35:** 2671-2678.

STORZ, J. F., and M. A. BEAUMONT, 2002 Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. Evolution **56:** 154-166.

THOMSON, R., J. K. PRITCHARD, P. D. SHEN, P. J. OEFNER and M. W. FELDMAN, 2000 Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. Proceedings of the National Academy of Sciences of the United States of America **97:** 7360-7365.

VALIERE, N., 2002 GIMLET: a computer program for analysing genetic individual identification data. Molecular Ecology Notes **2:** 377-379.

VAN LOO, M., J. A. JOSEPH, B. HEINZE, M. F. FAY and C. LEXER, 2008 Clonality and spatial genetic structure in *Populus* x *canescens* and its sympatric backcross parent *P. alba* in a Central European hybrid zone. New Phytologist **177:** 506-516.

VONLANTHEN, B., X. M. ZHANG and H. BRUELHEIDE, 2010 Clonal structure and genetic diversity of three desert phreatophytes. American Journal of Botany **97:** 234-242.

CHAPTER 7: Hermaphroditism and pollen gene flow in white poplar

[unpublished]

Introduction

Gene flow and mating systems are key features affecting the genetic structure of plant species. They also play a role in facilitating adaptation, maintaining standing genetic variation, and protecting from inbreeding and outbreeding depression or genetic drift; hence their evolutionary importance. Major parameters in understanding gene flow and mating systems are: correlated paternity, pollen dispersal kernels and sexual reproduction modes.

Correlated paternity, i.e. proportion of full sibs within a maternal progeny, is important because it may promote biparental inbreeding, furthering the risk of inbreeding depression (DE-LUCAS *et al.* 2008). Such predicament is especially dangerous in outbreeding dioecious forest trees such as poplars, and made even worse by their intense clonality, which can effectively reduce the number of potential male parents in some localities (Chapter 6).

Pollen dispersal can be studied by analysing the paternity of natural offspring of known mothers in natural conditions, as determining actual fathers serves to describe the shape of the dispersal curve and to evaluate the immigrated pollen share. These methods need a complete sampling of potential males within the sampling area, and a high marker resolution (DE-LUCAS *et al.* 2008). Accordingly, they are only feasible in some small isolated populations from some well-studied species (i.e. those for which numerous genetic markers have been developed). Alternatively, mating models can be used to describe dispersal kernels as well as other features of the mating system (GAUZERE *et al.* 2013).

Sexual reproductive modes (from cosexuality to dioecy) are intrinsically related to a species' ecology and life strategies, and play a role in its evolutionary fate. Most species in the genus *Populus*, and its sister genus *Salix*, are dioecious, although the ancient species *P. lasiocarpa* is monoecious (WYCKOFF and ZASADA 2008) and *S. martiana* has been described as a 'regularly hermaphrodite willow' (ROHWER and KUBITZKI 1984). While sporadic cases of monoecy has been found in many putative dioecious species (WYCKOFF and ZASADA 2008; and references therein) and recent reports of monoecy or hermaphroditism in poplar have been provided (*P. euphratica*; ROTTENBERG 2000; *P. deltoides*; ROWLAND *et al.* 2002), studies

have usually considered these findings as anecdotal or aberrant. However, such findings could be the fruit of adaptation: poplars could have been forced to maintain a loose sexual determination, probably through balancing selection, in contrast to the long-time strictly fixed sex determination observed in other groups, like mammals or birds. Deviations from dioecy would then demonstrate that the pathway between reproductive modes is not completely closed, and the fact that almost all the current species of these two tree groups are dioecious would suggest that such reproductive mode could be adaptively advantageous.

A previous study found several hermaphrodite individuals across a small region in southern Spain. These individuals, besides having hermaphrodite flowers in the catkins, possess different degrees of femininity, which are correlated to seed germination rate. However, male function success has not been assessed in these individuals yet. The aim of this chapter is to describe the spatial pattern of pollen dispersal in a white poplar population with presence of natural hermaphrodite individuals and to estimate their male reproductive fitness.

Materials and Methods

Studied population

To study pollen dispersal and hermaphroditism in white poplars, I chose Jimena, an isolated medium-sized population in the riverbanks of the Bédmar River within the Guadalquivir basin. This population was chosen because it is: (1) situated in an accessible area, (2) surrounded by an extensive olive grove, which helped in recognising all the individuals, (3) quite isolated, not having another white poplar population in the neighbourhood, and because (4) it presents hermaphrodite individuals (i.e. a subdioecious population). Additionally, its genotypic structure had been previously studied (see Chapter 3 and Figure 7.1).

Hermaphrodites and female individuals from two close populations were also sampled (located close to the town Puente del Obispo, < 20 km away from the Jimena population). These two populations presented hermaphrodites with a higher level of femininity (see below).

Figure 7.1 Map of the Jimena population (see Chapter 3). Dots indicate sampled ramets and filling colour shows genet (=genotype). Mothers (B44 and B8) are represented by red squares.

Plant material

In 2006 and 2007, I collected open-pollinated, fruited catkins from two female individuals (B44 and B8) and extracted their seeds. Hermaphrodite individuals (B1 and B6) did not produce seeds any of those years, since their female flowers aborted before seed maturation. Afterwards, seeds were sown and grown. In total (and after initial stage mortality), 200 offspring were obtained (136 from B44: 2006-2007 seed crops, and 64 from B8: 2007 seed crop). When seedlings had fully developed at least two true leaves, plantlets were harvested and dried in silica gel. DNA was then isolated from this material (as in Chapter 6). To study feminitity levels (FL hereafter), flowered and fruited catkins were collected in 2007 from three females (B42, PO1-89 & PO2-1) and five hermaphrodites (B1, B6, PO1-2, PO1-95 & PO1-99) from the Jimena (coded 'B') and Puente del Obispo (coded 'PO') populations.

Genetic markers

Eight microsatellites markers (ORPM 29, ORPM 30, ORPM 60, ORPM 220, ORPM 312, PMGC 2852, WPMS 05 and WPMS 15; more details in Chapter 6) were used in order to infer offspring paternal identity. These markers were chosen because of this population's known variability (see Chapter 3), or because they had been found highly polymorphic in Aranda de Duero, our main study region (see Chapter 6). PCR amplification and microsatellite band resolving were done following the same procedures described above for these markers (see Chapter 6).

Paternity inference

After microsatellite scoring, father-offspring relationships were identified using standard likelihood methods. For that, CERVUS software vs. 3.0.3 (KALINOWSKI *et al.* 2007) was used. Allele frequencies were calculated considering only natural population individuals at the genet-level (i.e. only including one individual per genet and excluding the offspring). To reduce uncertainty, maternal identity was forced, since this information was known. Genotyping error was set to zero, as genotyping a progeny with known maternal alleles greatly decreases the possibilities of error. Moreover, CERVUS runs with genotyping error of 0.01 produced similar results (data not shown). Correlated paternity was afterwards inferred with a kinship method based on co-ancestry among inferred pollen gametes within families (HARDY *et al.* 2004). For this purpose, the allele frequencies of the Aranda de Duero population were used as reference. Note that the inverse of correlated paternity is

the effective number of fathers (N_{ep}). Correlated paternity was also calculated with the maximum-likelihood method implemented by the program MLTR (RITLAND 2002). Finally, a third estimate of effective number of fathers was computed, by means of a Bayesian method that makes direct inferences from the calculated parentage probabilities (NIELSEN *et al.* 2001), using the software PATRI.

Level of femininity and germination rate

Femininity level (FL) was evaluated counting the number and sex of floral pieces of three flowers (located in the proximal, medium and distal end of the catkin) per catkin, ten catkins per individual. For the three strict females, only 3-6 catkins were assessed as long as no variance in number or sex of flowers was detected.

Seed germination rate was estimated by placing 100 seeds (when possible) on soaked paper and maintaining them protected by Petri dishes inside a germination chamber. Individuals PO1-95 and PO1-99 did not produce enough seeds, despite the large number of fruited catkins collected, so only 35 were assayed. B1 and B6 did not produce seeds at all: the scarce female flowers aborted during the first stages of maturation. Germinated seeds were counted after 12 days of incubation. Seedling vigour was not registered, since it was not the aim of the experiment, although clear differences among mothers were observed.

Results

The selected panel of microsatellite markers permitted a satisfactory parental assignment, despite two markers being uninformative (ORPM 220 was almost monomorphic at allele 181; and WPMS 05 was homozygous in mothers, and heterozygous for the same two alleles in all the fathers), and another marker not segregating codominantly in this population (ORPM 29), so they were discarded for further analyses. With the information from the remaining five markers, only 16% of B44-offspring and 57% of B8-offspring lacked any assigned father. Most father-offspring relationships were assigned unambiguously. This suggests that the fathers of unassigned offspring were unsampled individuals rather than that we were dealing with a low power marker set. Thus, the high rate of unassigned B8-offspring was probably due to the presence of one or a few unsampled close males or to a higher incoming pollen flow to the mother (see below).

Overall values of correlated paternity (r_p) ranged from 0.12 (by Ritland's method) to 0.63 (by Hardy's method). This broad range was expected, given that Ritland's method tends to underestimate r_p for low values of the statistics ($r_p < 0.15$); while the use of Aranda de Duero's allele frequencies as reference for Hardy's method should result in an overestimation, as the Aranda de Duero and Jimena populations are genetically differenced, although only slightly ($F_{ST} = 0.080$). However, these values can serve as approximate interval limits. Nielsen's method (PATRI) delivered an intermediate effective number of parents (N_{ep}) with respect to the values calculated by Ritland's and Hardy's methods [N_{ep} (Nielsen's) = 6.55; N_{ep} (Ritland's) = 8.33; N_{ep} (Hardy's) = 1.58; see Table 7.1].

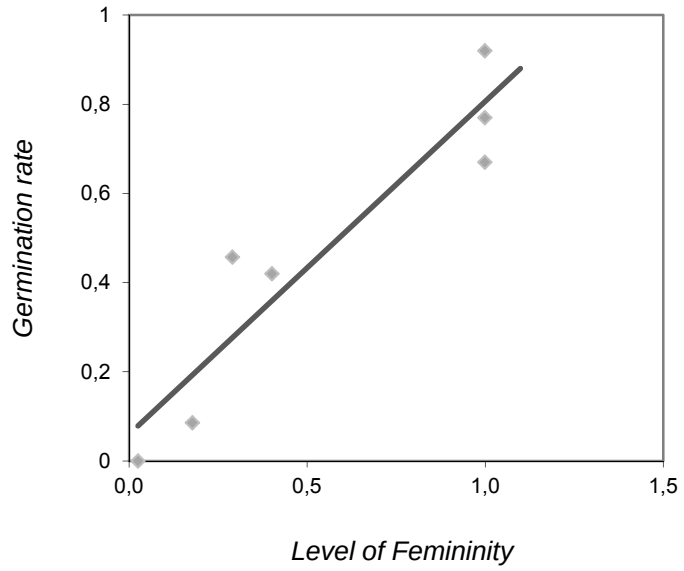


Figure 7.2 Correlation between the level of femininity (FL) of female (FL = 1.0) and hermaphrodite (0.0 < FL < 1.0) individuals and the germination rate of their seeds. The leftmost bottom point represents B1 hermaphrodite.

Regarding father contributions to each mother, B6 did not sire with B8, but did, in contrast, with B44 (31 offspring out of 133). B1 had a noteworthy higher number of offspring with B8 than with B7, despite being located farther away (Figure 7.3).

Femininity level was highly correlated with seed germination rate (Figure 7.2; Spearman rank test $R = 0.927$; $P = 0.003$). Strictly female individuals displayed germination levels of 0.67-0.92, whereas moderately female ones varied from 0.09 to 0.46. B1 and B6 individuals produced few female flowers, and not every year. In 2007, B6 did not produce any female

flowers in any of the collected catkins, although female flowers had been observed the previous year. Moreover, no female flowers ending maturation to produce seeds were observed in any of these two hermaphrodite individuals. In conclusion, these two individuals behaved like effective males, as they could not produce viable seeds, but could, in contrast, produce viable pollen.

TABLE 7.1 Male (B7) and hermaphrodite (B1 and B6) share on female (B44 and B8) offspring. F_s: relatedness between the paternal genes of sibs; r_p: correlated paternity; N_{ep}: effective number of sires.

Year	2006	2007		Overall
Mother	B44	B44	B8	
B1-offspring	0.54	0.57	0.27	0.47
B6-offspring	0.21	0.25	0.00	0.16
B7-offspring	0.07	0.03	0.17	0.09
Not assigned	0.17	0.14	0.57	0.28
F_s	0.35	0.35	0.31	0.32
r_p	0.70	0.70	0.62	0.63
N_{ep}	1.42	1.43	1.60	1.58
r_p [a]	0.04	0.05	-0.05	0.12
N_{ep} [a]	27.03	19.23	NC	8.33
N_{ep} [b]	5.74	4.16	NC	6.55

[a] Calculated with the program MLTR

[b] Calculated with the program PATRI

NC not computable

Discussion

To my knowledge, this is the first study that reports effective siring by a hermaphrodite poplar, indicating that hermaphroditism does not significantly reduce pollen fecundity. Not only its pollen yields seeds that can germinate, but they can also successfully grow to produce seedlings and, eventually, mature trees. However, this result must be qualified with two caveats: B1 and B6 genets were some of the less 'feminine' hermaphrodites sampled in the region (FL = 0.03 and 0.00 in 2007, respectively, vs FL = 0.29 ± 0.11, in the remaining hermaphrodites), and no fully-developed seeds from their hermaphrodite flowers were found. Other, more 'feminine', hermaphrodite clones produced seeds, but neither seed viability nor pollen success have been tested for those. Thus, despite observations of hermaphrodites, no single poplar individual has been yet observed capable of effectively reproducing by pollen and seeds, or to self-fertilising.

The two main pathways to dioecy revealed by phylogenetic analyses and field observation of mixed populations from transitioning species are gynodioecy and monoecy. Dioecy can arise either from gynodioecy through selection acting on new mutants, favouring the male function in hermaphrodite morphs; or from monoecy due to disruptive selection on quantitative genetic variation in floral sex ratios (BARRETT 2002). Thus, sex lability in dioecious species could be indicative of a monoecius origin (AINSWORTH 2000), though empirical studies oppose this expectation (EHLERS and BATAILLON 2007).

In *Populus*, although the pathway to dioecy has not been determined, the only cosexual species of the genus is monoecious (*P. lasiocarpa*; FAO 1980), suggesting a monoecious origin. Additionally, monoecy is often related to abiotic pollination (CHARLESWORTH 1993), and poplars are anemophilous. Abiotic pollination could be an indicator of the monoecious pathway to dioecy, while biotic pollination could be more linked to the pathway to gynodioecy (as inferred from literature survey summary in EHLERS and BATAILLON 2007). However, in our case, the hermaphrodite catkins of the Jimena population were found in male individuals that also developed pistils in some flowers. The number of floral pieces of hermaphrodite flowers coincides with that of male flowers [8 ± 2 (SD)], yet not with that of female flowers (always one). Also, hermaphrodite flowers usually share catkin with pure male flowers, but not with female flowers (data not shown). This phenomenon is known as 'male sex inconstancy' and it is relatively frequent in dioecious species (BARRETT 2013). Besides, these hermaphrodite catkins never present pure female flowers. Both the presence of male sex inconstancy and perfect flowers and

the absence of alternate gender flowers support a hypothetic gynodioecious origin (EHLERS and BATAILLON 2007), but it does not discard a monoecious one.

Surprisingly, previous studies of hermaphroditism in *Populus* (namely *P. tremuloides* and *P. trichocarpa*) have reported inconstancy primarily in females (WYCKOFF and ZASADA 2008). A great variety of inconstant morphs has been found in the related, mostly dioecius, genus *Salix* (willows) (ZASADA *et al.* 2008). However, willows sometimes use animal mediated pollination, in contrast with poplars, which are strictly anemophilous. In the unrelated genus *Fraxinus* (ashes), sex inconstancy is common, although this group presents a wide variety of floral systems, from dioecious to perfect hermaphrodites, and also biotic and abiotic pollination (BONNER 2008), so probably molecular devices for sex determination are not as firmly established as in poplars. Finding common patterns in flower systems is a complex and laborious task, due to easier than expected transitions among mating systems (BARRETT 2013) and among pollination modes – thus obscuring the evidence provided by phylogenetic relations –, and their dependence on multiple concurring ecological factors.

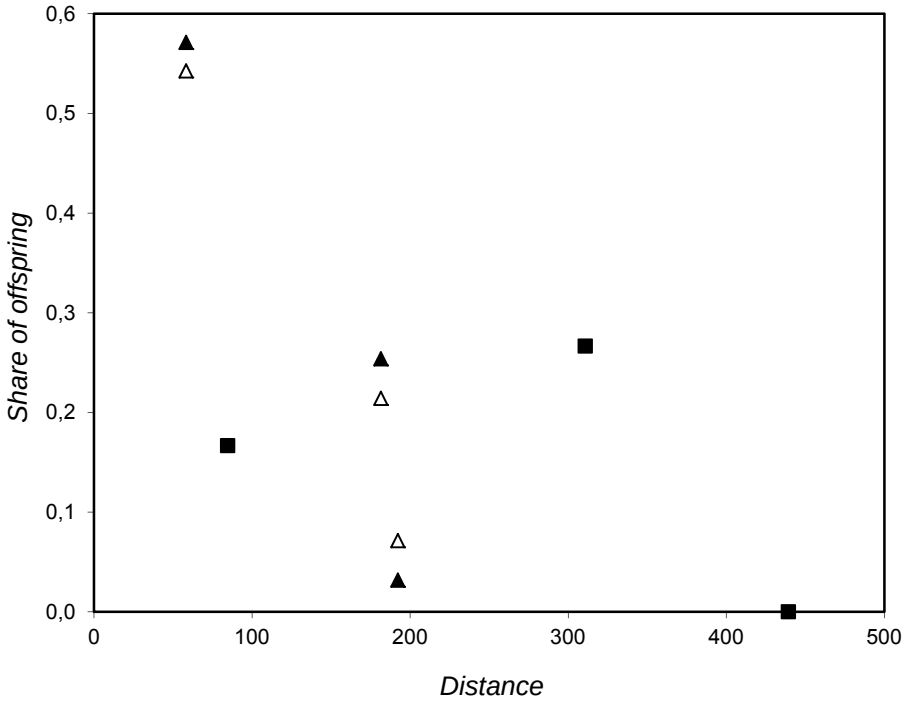


Figure 7.3 Distance between mothers and the geometrical centre of each paternal clonal assembly. Triangles represent B44-offspring and squares B8-offspring; blank for 2006 and solid for 2007.

Remarkably, father identity remained unknown for a substantial amount of B8-offspring (57%), while analyses for B44 recovered a much lower number of unassigned offspring (16%). This higher value for B8 was probably due to the presence of one unsampled male in close proximity, as pollen gametes from unassigned offspring were compatible with the existence of a single sire. An alternative explanation could be a higher rate of long-distance incoming pollen for B8, but this hypothesis is not supported: the large difference in unknown fathers found between B44 and B8's offspring, and the values of correlated paternity calculated for B8-offspring, which were similar to B44's (see Table 7.1). Considering this, a more intense sampling should be executed to determine unknown male parents and to be able to fit dispersal kernels.

In black poplar (*P. nigra*), a sister sympatric poplar species with an almost identical mating system, most gene flow (about 75%) occurred in a distance under 1 km in a population besides the Eder River, in Germany (RATHMACHER *et al.* 2010). Given that the sampled Jimena population spreads along a similar distance, we could expect similarly low levels of incoming gene flow. However, population structure must be acknowledged, as it can heavily influence dispersal distances, and the Jimena and Eder populations although they have similar linear arrangements, the latter has a denser population structure. Density can impair long-distance pollen flow and the low density of the Jimena population may imply a higher than expected incoming gene flow. In *P. trichocarpa*, in the west coast of North America, observed pollen immigration lowered with the increase of reproductively mature male trees located close to the mothers (SLAVOV *et al.* 2009).

Finally, paternity analyses for B8-offspring revealed that distance is not the only factor affecting siring, as evidenced by the larger progeny of the more distant B1 than the closer B7 (Figure 7.2; Table 7.1). These differences could have resulted from phenological asynchronies (GERARD *et al.* 2006). Alternatively, B7's lower pollen production, probably related to its smaller clonal size, could explain our results. The lack of offspring sired by B6 is also remarkable. Non-exclusive factors could be: (1) distance; (2) uncoupled phenology; and (3) genetic incompatibilities or inbreeding depression. With the current analyses, none of these hypotheses can be discarded or advanced, but the likely implication of clone size in male reproductive success must be investigated. Indeed, some studies have been carried out in fruit trees from the Rosaceae family (i.e. apple or cherry trees), given its economic importance. These trees are zoophilous and self-incompatible, so pollination issues could seriously impair fruit yield. For example, it has been shown that male reproductive success

by ramet in *Prunus ssiori* was favoured in larger trees and ramets belonging to mono-ramet genets (in contrast to genets with multiple ramets) (MORI *et al.* 2009). However, absolute male reproductive success by genet was not analysed in this paper, although it seemed to be higher in genets with many ramets due to multiplicative effects. Another study on apple orchards detected no effect of genet (distributed by blocks) size on male reproductive success (ROUTLEY *et al.* 2004). The authors of this study considered that pollen discounting due to pollinators' tendency to remain in the same genotype (floral constancy) explained the lack of significance of genet size. Salicaceae trees, however, are wind pollinated, making pollinator activity irrelevant, while the multiplicative effect of mass pollination by genets with several ramets is probably relevant to explain pollination patterns.

Acknowledgements

This work was made in collaboration with Santiago C. González-Martínez, José M. Climent Maldonado and Nuria Alba Monfort.

References

AINSWORTH, C., 2000 Boys and girls come out to play: The molecular biology dioecious plants. Annals of Botany **86:** 211-221.

BARRETT, S. C. H., 2002 The evolution of plant sexual diversity. Nature Reviews Genetics **3:** 274-284.

BARRETT, S. C. H., 2013 The evolution of plant reproductive systems: how often are transitions irreversible? Proceedings of the Royal Society B-Biological Sciences **280**.

BONNER, F. T., 2008 *Fraxinus* L. ash, pp. 537-543 in *The Woody Plant Seed Manual*, edited by F. T. BONNER and R. P. KARRFALT. USDA Forest Services.

CHARLESWORTH, D., 1993 Why are unisexual flowers associated with pollination and unspecialized pollinators? American Naturalist **141:** 481-490.

DE-LUCAS, A. I., J. J. ROBLEDO-ARNUNCIO, E. HIDALGO and S. C. GONZALEZ-MARTINEZ, 2008 Mating system and pollen gene flow in Mediterranean maritime pine. Heredity **100:** 390-399.

EHLERS, B. K., and T. BATAILLON, 2007 'Inconstant males' and the maintenance of labile sex expression in subdioecious plants. New Phytologist **174:** 194-211.

FAO, 1980 *Poplars and willows in wood production and land use.* Food and Agriculture Organization of the United Nations (FAO), Rome.

GAUZERE, J., E. K. KLEIN and S. ODDOU-MURATORIO, 2013 Ecological determinants of mating system within and between three *Fagus sylvatica* populations along an elevational gradient. Molecular Ecology **22:** 5001-5015.

GERARD, P. R., E. K. KLEIN, F. AUSTERLITZ, J. F. FERNANDEZ-MANJARRES and N. FRASCARIA-LACOSTE, 2006 Assortative mating and differential male mating success in an ash hybrid zone population. Bmc Evolutionary Biology **6**.

HARDY, O. J., S. C. GONZALEZ-MARTINEZ, B. COLAS, H. FREVILLE, A. MIGNOT *et al.*, 2004 Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. Genetics **168:** 1601-1614.

KALINOWSKI, S. T., M. L. TAPER and T. C. MARSHALL, 2007 Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology **16:** 1099-1106.

MORI, Y., T. NAGAMITSU and T. KUBO, 2009 Clonal growth and its effects on male and female reproductive success in *Prunus ssiori* (Rosaceae). Population Ecology **51:** 175-186.

NIELSEN, R., D. K. MATTILA, P. J. CLAPHAM and P. J. PALSBOLL, 2001 Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. Genetics **157:** 1673-1682.

RATHMACHER, G., M. NIGGEMANN, M. KOHNEN, B. ZIEGENHAGEN and R. BIALOZYT, 2010 Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures: implications for in situ conservation measures. Conservation Genetics **11:** 1327-1338.

RITLAND, K., 2002 Extensions of models for the estimation of mating systems using *n* independent loci. Heredity **88:** 221-228.

ROHWER, J., and K. KUBITZKI, 1984 *Salix martiana*, a regularly hermaphrodite willow. Plant Systematics and Evolution **144:** 99-101.

ROTTENBERG, A., 2000 Fertility of exceptional bisexual individuals in four dioecious plant species. Sexual Plant Reproduction **12:** 219-221.

ROUTLEY, M. B., P. KRON and B. C. HUSBAND, 2004 The consequences of clone size for paternal and maternal success in domestic apple (*Malus* x *domestica*). American Journal of Botany **91:** 1326-1332.

ROWLAND, D. L., R. GARNER and M. JESPERSEN, 2002 A rare occurrence of seed formation on male branches of the dioecious tree, *Populus deltoides*. American Midland Naturalist **147:** 185-187.

SLAVOV, G. T., S. LEONARDI, J. BURCZYK, W. T. ADAMS, S. H. STRAUSS *et al.*, 2009 Extensive pollen flow in two ecologically contrasting populations of *Populus trichocarpa*. Molecular Ecology **18:** 357-373.

WYCKOFF, G. W., and J. C. ZASADA, 2008 *Populus* L. poplar, cottonwood, aspen, pp. 856-871 in *The Woody Plant Seed Manual*, edited by F. T. BONNER and R. P. KARRFALT. USDA Forest Service.

ZASADA, J. C., D. A. DOUGLAS and W. BUECHLER, 2008 *Salix* L. willow, pp. 1000-1009 in *The Woody Plant Seed Manual*, edited by F. T. BONNER and R. P. KARRFALT. USDA Forest Service.

CHAPTER 8: Local adaptation imprints on wood formation genes

[unpublished]

Introduction

Local adaptation is a crucial process for evolution and speciation (HOFFMANN and WILLI 2008) that involves genetic divergence of the specific loci drawn by selection. The identification of such loci could facilitate tasks related to conservational or industrial purposes and could also cast some light on how adaptation occurs in natural populations (e.g. GONZALEZ-MARTINEZ *et al.* 2006a).

Massive genotyping methods and the release of *P. trichocarpa* genome assemblies (TUSKAN *et al.* 2006) have permitted the analysis of the variability among individuals of specific metabolic pathways. This variability may be concretised in nucleotide variants, namely in single-nucleotide polymorphisms (SNPs). Theoretically, differentiation among populations is caused either by genetic drift or by selection. In particular, diversifying selection would result in adaptive loci having higher differentiation than neutral loci. Methods to detect selection that use this approach are known as divergence-based tests, while other methods either correlate environmental variables to genetic clines to infer selection, or focus in the detection of selective sweep signals (as described in Chapter 4).

Trees present good features to study loci genetic divergence due to natural selection. They are scarcely domesticated, they usually have large distributions that span different ecological environments with high levels of genetic and phenotypic variation, and they also have large population numbers, normally displaying low genetic differentiation between populations (GONZALEZ-MARTINEZ *et al.* 2006b; TSUMURA *et al.* 2012). Poplars, in particular, are expected to share these properties, as both their reproductive system (wind dispersal of pollen and seeds) and population distribution (widespread presence at the continental scale) promote them.

Wood composition and structure play a fundamental role in the survival and adaptation capacity of trees. Cellulose and lignin, the two main components of wood, confer it essential qualities that allow two main functions: mechanical support, and water transport. Specifically, the hardiness and hydrophobicity of lignin influence mechanical support and resistance to vessel cavitation, as essays in transgenic poplars modified for lignin metabolism have revealed (AWAD *et al.* 2012). This, in turn, implies resistance to drought

and freezing, giving lignin a fundamental role in the evolution of land plants (PETER and NEALE 2004). Therefore, genes involved in the formation of wood, and especially in the metabolism of lignin, are likely to contribute to local adaptation, and thus are expected to be under selection.

My aim in this chapter is to uncover selection signatures on the wood formation pathway genes, with emphasis on the lignin route, among three contrasting populations of Iberian white poplars, *Populus alba*, and a population of European aspen. The geographical situation of the Iberian Peninsula makes it one of transition from a humid temperate climate to a dry tropical one. This puts populations under strikingly different conditions in a short spatial range, while at the same time they remain connected through gene flow by pollen exchange, reducing genetic differentiation. In the case of the Iberian white poplar, numerous populations survived in river basins during glacial times, making regional extinctions rare (see Chapter 3). After glaciations, pollen was the main gene exchange conductor, thus maintaining high levels of standing genetic variation. Adaptation has been observed in some phenotypic traits, in particular across river basins (see Chapter 3), and we expect at the same time natural selection to have imprinted a differential signal on adaptive polymorphisms.

Material and methods

Plant material

Leaf tissue from 15 poplars was collected and sampled: 11 white poplars from different populations (four from La Alfranca in the Ebro riverbank; four from Jimena, in the Guadalquivir basin; and three from Aranda de Duero, in the Duero riverbank); one grey poplar from the Aranda de Duero population; and three European aspens (one from Aranda de Duero, and two from Gredos) (see Chapters 3 and 4). Sampled trees included female and male individuals, as well as two rare hermaphrodite specimens from the Jimena population (see Chapter 7). DNA was isolated following the procedures indicated in Chapter 3.

Amplicon resequencing

After DNA isolation and quality checking using a Nanodrop 1000 Spectrophotometer, 199 amplicons were resequenced, comprising regions of 40 genes related to wood formation,

the majority from the lignin synthesis pathway (Table 8.1). Primers for amplicon sequencing were designed using the JGI Poplar Genome Assembly v. 2.0 (www.phytozome.org/poplar).

Data analyses

A custom software pipeline, developed by collaborators, aligned amplicon sequences and detected SNPs among the sequences. Quality of SNPs was checked manually, and questionable SNPs and tree samples with a poor amount of sequences removed. A final quality trimming was performed by removing from the remaining trees those with SNPs with only one variant, and those with triallelic variants or with missing data, as our reduced sample size could lead to spurious results if SNPs with missing information were tested.

Correlation among SNPs was evaluated using R package LDHEATMAP (SHIN *et al.* 2006). SNPs pairs with correlation higher than $r^2 > 0.25$ were detected, and one of them was discarded. Finally, genetic differentiation (calculated as F_{ST}) among various contrasting subsets of samples was computed for each SNPs, using BAYESCAN software (FOLL and GAGGIOTTI 2008), and outlier values were identified. Contrasting subsets were: (i) between-populations (i.e. Aranda de Duero vs. Jimena), (ii) between-sexes (i.e. males vs. females), and (iii) hermaphrodite vs. remaining trees within Jimena. We fixed false positive discovery rate (FDR) at 5% to perform calculations and to identify reliable candidate loci under selection, contrasting allele frequencies between populations. This was done because BAYESCAN uses a multinomial-Dirichlet model that incorporates some degree of uncertainty on allele frequencies due to small sample sizes. A previous work that tested different computing methodologies to test for selection revealed that the underlying BAYESCAN method delivered a good power of detection with very low rates of false positives (DE MITA *et al.* 2013).

In addition, we used FDIST, another F_{ST} outlier detection method (BEAUMONT and NICHOLS 1996), as implemented in the software LOSITAN (ANTAO *et al.* 2008). To run this program, I used the complete set of SNPs after the initial trimming by missing data. In other words, no locus was discarded because of Linkage disequilibrium (LD) (see Results). Two runs of 1,000,000 simulations each were executed: the first to determine outlier loci and discard them for the second run, where neutral envelopes for confidence limits were calculated. Confidence intervals were set at $\alpha = 0.95$, and false discovery rate at 0.1.

Results

Number of sequences and SNPs

Amplicon resequencing successfully yielded sequences for 137 amplicons from 38 genes, of which only 55 produced sequences for 7 or more samples. In particular, European aspen, grey poplar and the samples from the Ebro basin produced sequences of very bad quality. After a manual quality check, 674 SNPs were kept. Then, the eight samples with high missing data were discarded. Afterwards, we removed SNPs with only one variant in the remaining trees (42), with triallelic variants (2), and with missing data (219). Finally, 411 SNPs were selected for further analyses.

Table 8.1 Summary of candidate gene resequencing for 40 wood-related genes. N_A: number of succesfully resequenced amplicons; N_S: number of SNPs detected in such amplicons.

Gene family	**Gene code**	***P. trichocarpa* locus**	N_A	N_S
4-Coumarate:CoA ligase	**4CL1**	POPTR_0006s18490	8	52
	4CL3	POPTR_0001s07400	1	7
	4CL5	POPTR_0003s18710	3	20
Coumarate 3-hydroxylase	**C3H3**	fgenesh4_pg.C_LG_VI000268	1	8
Cinnamate 4-hydroxylase	**C4H1**	POPTR_0019s15110	1	3
	C4H2	POPTR_0013s15380	3	23
Cinnamyl alcohol dehydrogenase	**CAD**	POPTR_0009s09870	4	8
Cinnamoyl-CoA reductase	**CCR**	POPTR_0003s17980	4	13
Cellulose synthase	**CesA1A**	POPTR_0011s07040	9	4
	CesA1B	POPTR_0004s05830	10	61
	CesA2A	POPTR_0018s11290	6	35
	CesA2B	POPTR_0006s19580	6	18
	CesA3A	POPTR_0002s25970	6	26
Caffeoyl CoA O-methyltransferase	**CoAOMT1**	POPTR_0009s10270	1	6
	CoAOMT2	POPTR_0001s31220	1	5
Caffeate O-methyltransferase	**COMT1**	POPTR_0015s00550	3	8
	COMT2	POPTR_0012s00670	3	13
Ferulate 5-hydroxylase	**F5H1**	POPTR_0005s11950	3	5
	F5H2	POPTR_0007s13720	2	7
Hydroxcinnamoyl-CoA quinate/shikimate hydroxycinnamolytransferase	**HCT1**	POPTR_0003s18210	5	24
	HCT6	POPTR_0001s03440	5	20
	gdcH1	**POPTR_0012s14960**	3	7

Linkage disequilibrium

Correlation evaluation with R package LDHEATMAP (Figure 8.1) depleted the number of unlinked SNPs to 18. The low number of haplotypes (14), compared to the high amount of SNPs (411) surely caused spurious detection of linkage among probable unlinked loci. Thus, divergence tests for selection were executed both for the total amount of SNPs (411) and for the unlinked subset (18).

Selection testing

A Bayesian test using BAYESCAN delivered no significant outlier loci in any of the different sets of contrast performed (Figure 8.2). However, the FDIST detection method revealed a fairly high amount of outlier loci, both for diversifying (67) and for balancing selection (141) (Figure 8.3).

Table 8.1 Continued.

Gene family	**Gene code**	***P. trichocarpa* locus**	N_A	N_S
Glycine decarboxylase complex, T subunit	**gdcT2**	POPTR_0004s01030	0	0
Cellulase	**KOR1**	POPTR_0001s11870	2	5
Laccase	**LAC1a**	POPTR_0016s11950	3	23
	LAC2	POPTR_0008s06430	3	8
	LAC90a	POPTR_0008s07370	4	21
Phenylalanine ammonia-lyase	**PAL2**	POPTR_0008s03810	4	25
	PAL4	gw1.X.2713.1	3	12
	PAL5	POPTR_0010s23100	0	0
S-Adenosylmethionine synthetase	**SAM1**	POPTR_0008s09870	2	4
Serine hydroxymethyltransferase	**SHMT1**	POPTR_0001s32770	3	8
	SHMT3	POPTR_0002s10990	5	47
	SHMT6	POPTR_0017s08600	2	12
Sucrose synthase	**SUSY1**	POPTR_0018s07380	6	37
α-Tubulin	**TUA1**	POPTR_0002s11250	3	15
	TUA5	POPTR_0009s08850	2	28
β-Tubulin	**TUB9**	POPTR_0001s09330	2	30
	TUB15	POPTR_0001s27960	3	17
	TUB16	POPTR_1455s00200	2	9

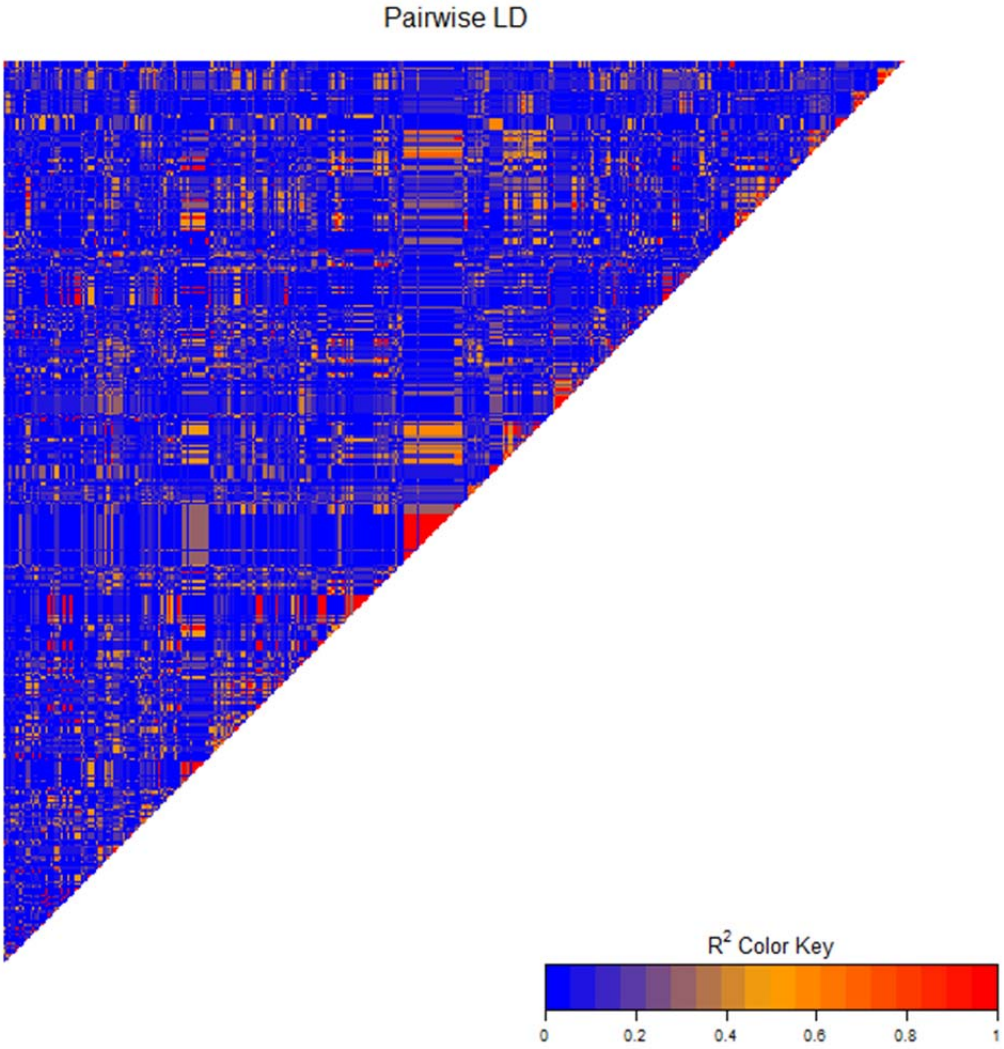


Figure 8.1 Pairwise SNP correlation plot produced by LDheatmap package. Rows and columns contain the same dataset of 411 SNPs. Colour indicates the intensity of correlation (i.e. LD). Notice that warm coloured pixels (coding for high LD) are copious in areas distant from the diagonal. Pixels near to the diagonal represent usually SNP pairs of the same gene.

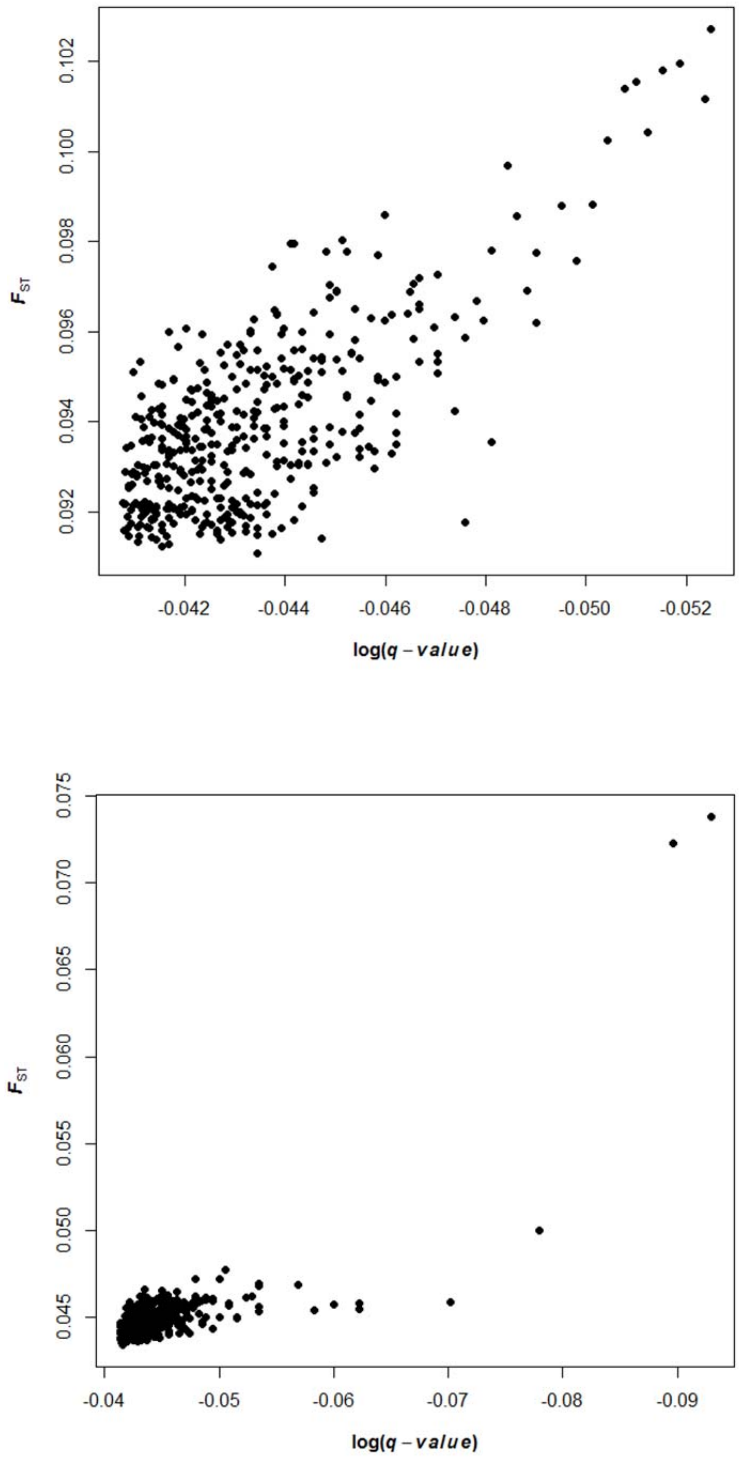


Figure 8.2 Values of F_{ST} of SNPs and correspondent *Q*-values calculated by BAYESCAN. Notice that none of the SNPs reaches the *Q*-value threshold to consider them as an outlier. Top plot: contrast across natural populations from different river basins (Jimena vs. Aranda de Duero); bottom plot: contrast for sexes (male vs. females)

Discussion

BAYESCAN delivered no significant outlier loci in any of the different sets of contrast performed. Directional selection testing using Bayesian methods including false positive control, like the one implemented by BAYESCAN, requires a high statistical power (DE MITA *et al.* 2013). Furthermore, theoretical studies have showed that, for this test, the number of populations is more important than the sample size per population. A method comparison study has highlighted that even eight populations were too little for this approach to be successful (DE MITA *et al.* 2013). In contrast, the FDIST method revealed an elevated number of loci under selection. However, this completely different result could be due to spurious detection given the small amount of samples and populations, as well as the unrealistic demographic model this method uses. Indeed, BAYESCAN is less prone to false positive error (type I) than FDIST-based implementations such as LOSITAN (NARUM and HESS 2011). Although the insufficient dataset we have employed could have produced these extremely opposing results, there could be other reasons that also explain the lack of significant results inferred by BAYESCAN.

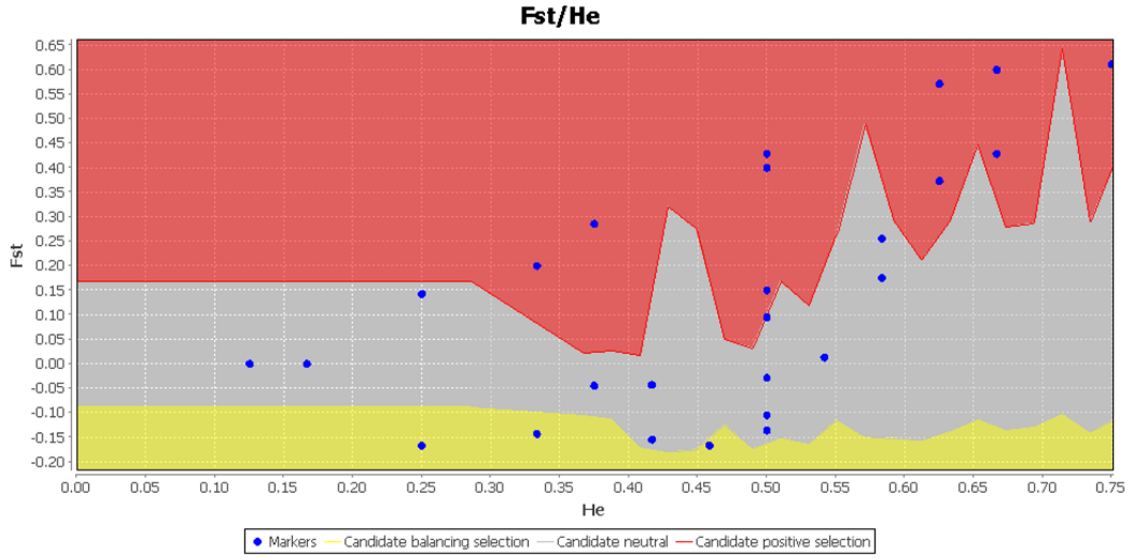


Figure 8.3 Values of F_{ST} and H_e for each SNP calculated by LOSITAN using FDIST method, as well as confidence intervals for loci under positive (directional) and balancing selection. Dots beyond the confidence lines denote outlier loci. Note that many SNPs overlap on the same dot, hence the small number of dots.

For instance, directional selection could be missing the required intensity to imprint detectable signatures in this group of genes. This could be due to a lack of adaptive pressure on this genic pathway, but also to the fact that a large group of genes could be controlling these traits under selection. In the latter case, adaptation is expected to be produced by changes in the among-gene covariance, rather than in the variance of single genes, causing only minor allele frequency shifts (LE CORRE and KREMER 2012). In fact, BAYESCAN has been found less efficient in cases where selection was weak (DE MITA *et al.* 2013).

A second cause could be that neutral differentiation among populations is high, making it difficult to detect divergence outliers, whether it is due to poor gene flow among populations and/or to strong genetic drift associated with demographic contingences (LE CORRE and KREMER 2012). However, poplars' reproductive system favours gene flow, so populations are expected to be little differentiated in their nuclear genome. Indeed, the two populations contrasted have a moderately low genetic differentiation (F_{ST} = 0.080; computed from nuclear microsatellites data of Chapter 3).

In summary, the concurring factors present in this essay result in such a reduced sample size that it is not possible to detect outlier loci associated with diversifying selection when using a conservative model as BAYESCAN. With the same flawed dataset, the more prone to false positives software LOSITAN, based in FDIST, identified a suspiciously high number of outliers.

References

ANTAO, T., A. LOPES, R. J. LOPES, A. BEJA-PEREIRA and G. LUIKART, 2008 LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. Bmc Bioinformatics **9**.

AWAD, H., S. HERBETTE, N. BRUNEL, A. TIXIER, G. PILATE *et al.*, 2012 No trade-off between hydraulic and mechanical properties in several transgenic poplars modified for lignins metabolism. Environmental and Experimental Botany **77:** 185-195.

BEAUMONT, M. A., and R. A. NICHOLS, 1996 Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society B-Biological Sciences **263:** 1619-1626.

DE MITA, S., A. C. THUILLET, L. GAY, N. AHMADI, S. MANEL *et al.*, 2013 Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Molecular Ecology **22:** 1383-1399.

FOLL, M., and O. GAGGIOTTI, 2008 A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics **180:** 977-993.

GONZALEZ-MARTINEZ, S. C., E. ERSOZ, G. R. BROWN, N. C. WHEELER and D. B. NEALE, 2006a DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. Genetics **172:** 1915-1926.

GONZALEZ-MARTINEZ, S. C., K. V. KRUTOVSKY and D. B. NEALE, 2006b Forest-tree population genomics and adaptive evolution. New Phytologist **170:** 227-238.

HOFFMANN, A. A., and Y. WILLI, 2008 Detecting genetic responses to environmental change. Nature Reviews Genetics **9:** 421-432.

LE CORRE, V., and A. KREMER, 2012 The genetic differentiation at quantitative trait loci under local adaptation. Molecular Ecology **21:** 1548-1566.

NARUM, S. R., and J. E. HESS, 2011 Comparison of F_{ST} outlier tests for SNP loci under selection. Molecular Ecology Resources **11:** 184-194.

PETER, G., and D. NEALE, 2004 Molecular basis for the evolution of xylem lignification. Current Opinion in Plant Biology **7:** 737-742.

SHIN, J.-H., S. BLAY, B. MCNENEY and J. GRAHAM, 2006 LDheatmap: An R function for graphical display of pairwise linkage disequilibria between Single Nucleotide Polymorphisms. Journal of Statistical Software **16**.

TSUMURA, Y., K. UCHIYAMA, Y. MORIGUCHI, S. UENO and T. IHARA-UJINO, 2012 Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. Heredity **109:** 349-360.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596-1604.

CHAPTER 9: Synthetic discussion

Main results

The phylogeographic analysis revealed 54 haplotypes (fourteen considering just unique event polymorphisms) were found across 59 populations of poplars (mostly Iberian; n = 628). Overall haplotype diversity was lower in *Populus alba* (average of 0.317) than in *Populus nigra* (average of 0.409). In both species, genetic diversity values varied largely among populations with no clear pattern, but at meta-population level, southern areas displayed more genetic diversity and private allele richness than northern ones. However, no difference was detected between the Mediterranean and the Atlantic drainage basins. Overall genetic differentiation values were significant and similar for both species, although in white poplar river and drainage basins were the most important factors causing genetic divergence, while in black poplar the role of drainage basin was lower and river basin was not significant. Levels of clonality averaged 4.2 ramets per genet, with 6-13 genets per population (32-49 ramets per population). Average clone size was below 100 m in three out of four populations. Regarding quantitative traits, total height at age 3 (HT3) and stem form at age 3 (FOR3) showed significant overall differentiation. Q_{ST} for FOR3 was significantly higher than F_{ST} among river basins, but not among populations within them (Chapter 3).

In the case of *Populus tremula*, European-wide spatial genetic structure analyses based on clustering separated the Spanish population (Sierra de Gredos) from all others. Aspen from Spain displayed lower values of genetic diversity (H_E = 0.474; allelic richness = 3.497), but also lower inbreeding (F_{IS} = 0.022). Furthermore, it showed imprints of a recent bottleneck and was composed of several groups of closely related genotypes. An admixed origin of the Swedish and Central European populations was detected. In particular, the Scandinavian population was integrated in the Scottish and Russian gene pools. Its cline for bud set revealed a conspicuous step, and elevated variance and selection differentials in the middle of the gradient. Finally, LD along chromosome XIX loci significantly exceeded the overall level observed in genus *Populus* (Chapter 4).

In a hybrid (*P* × *canescens*) backcross to *P. alba* (BC_1), thirteen (20%) loci showed segregation distortion in the hybrid parent, of which twelve displayed an overrepresentation of *P. tremula* alleles. Marker order was syntenic to the *P.trichocarpa* genome assembly and no

recombination was detected among four markers from chromosome XIX. In this region, genetic diversity of individuals with different sex was congruent with random mating, and moreover, one population showed no LD therein. However, this genomic region also showed reduced interspecific divergence in natural populations, in contrast to the autosomal chromosome VI (Chapter 5).

The clonal, hybrid population of Aranda de Duero was pervaded by two very large clones, but it was also composed of a few additional large ones and many small ones. Genotypic richness and evenness was lower than in other European hybrid populations, yet genetic diversity was alike. With respect to historical demography, large genets were increasing in size, but the population as a whole was suffering a long-term decline in effective population size (N_e). Furthermore, larger clones were the progenitors of more offspring than smaller clones, also contributing to reducing N_e. Finally, large clones showed introgression from *P. tremula* in a genome region that could be involved in accelerated clonal spread and had lower levels of the inbreeding coefficient (Chapter 6).

In the paternity analyses in Jimena's population, effective number of fathers ranged from 1.58 to 8.33 (depending on the method for calculating). Clone size and distance among parents were factors that apparently influenced pollination success. Besides, an inconstant male (i.e. a hermaphrodite) effectively sired with a female. Lastly, in hermaphrodites, level of feminity was correlated with germination rate (Chapter 7).

Regarding molecular imprints of adaptation in wood formation genes, one method detected no selection while another detected selection in tens of loci. Low sample sizes probably underlie these contrasting results (Chapter 8).

Phylogeographic patterns

Contrasting genetic structure and amounts and distribution of haplotypic genetic diversity hint at different histories for white and black poplar populations in the Iberian Peninsula (Chapter 3). The latter species presents higher values of regional chloroplast genetic diversity and shrunk levels of genetic structure among basins. The most likely reason for this pattern is its higher tolerance to cold, which may have conferred it higher endurance to glaciations and capacity to leap over mountain ranges, promoting gene flow. Concordant observations have been made in the distributions of Iberian pine species (SOTO *et al.* 2010)

and European ashes (HEUERTZ *et al.* 2006). This fact reveals that riparian trees, despite being azonal, are also heavily affected by regional climate. Moreover, this differential response by species to climate fluctuations highlights the likely threat to cold tolerant species in a context of competitive exclusion under climate warming.

Chloroplast haplotypic distribution points to ancient gene flow among latitudes, higher than among longitudinal ranges, which results in differentiated haplotype pools between Atlantic and Mediterranean watersheds (Chapter 3). This pattern has also been observed in other Iberian tree species (RODRIGUEZ-SANCHEZ *et al.* 2010). A probable factor related to these differences could be the altitudinal range shift towards sea shores during glaciations. In the Iberian Peninsula, this shift would have displaced population westwards in the Atlantic drainage and eastwards in the Mediterranean area, due to orography. It would have improved connections north- and southwards – given that Iberian coastal mountain ranges are lower than inland ones –, but limiting exchanges between these two sea drainages.

Genetic adaptation

Quantitative genetic differentiation (Q_{ST}) in some growth traits was higher than molecular genetic differentiation (F_{ST}) among river basins, but not within them, revealing disruptive selection causing local adaptation (Chapter 3). However, the spatial scale at which adaptation takes places is wider than in other temperate trees (LATTA 2004; SAVOLAINEN *et al.* 2007). Gene flow among populations of the same basin apparently impeded local adaptation at such fine scale, but also permitted the replenishment of standing genetic variation, without signatures of outbreeding depression. This might be explained by the relative homogeneity of riparian habitats. Outbreeding depression was absent as well in the admixed populations of European aspen of Northern Europe (Chapter 4). Therein, local selection on bud set was detected by the presence of a phenotypic cline. But also, admixture contributed to increase standing genetic variation, as was shown by the step in the phenotype cline, and the elevated bud set variance and selection differentials in the middle of such cline.

As for genes involved in wood formation, no evidence for adaptive differentiation was detected (Chapter 8). A likely cause for this result could be that studying local adaptation for polygenic phenotypic traits like wood composition needs a larger number of samples, as phenotypic changes might be caused by slight shifts in the allelic frequencies of many genes

(LE CORRE and KREMER 2012). Thus, differences in allelic frequencies are often so small that a high statistical power is required. Additionally, changes in phenotypes are often produced by epistatic interactions among genes, making even more difficult to detect significant signatures of local adaptation at the molecular level (LE CORRE and KREMER 2003).

Asexual reproduction

Clonality was present in all the five populations in which fine-scale spatial genetic structure was investigated (Chapter 3). Average sizes of clonal assemblies varied from 26 m to 164 m, but much larger, conspicuous clones were reported in Aranda de Duero (Northwestern Iberian Peninsula) and Montes de Toledo (Central Iberian Peninsula). In Aranda de Duero in particular, two extraordinarily large clones were widely spread (Chapter 6). White poplar apparently survived during glaciations in this region, as revealed by the occurrence of private chloroplast haplotypes (Chapter 3). Although I did not test for causality, these two facts could be connected, and it may be interesting to gain insights in the likely relation between clonality and endurance to face harsh glacial conditions (by reproduction assurance). Specifically, it could be interesting to test: (i) for differences in cold resistance among the genets with a large number of ramets and the rest, to verify if clone size is linked to endurance to past climate; (ii) whether large clones are effectively more capable in root-sucker or twig shedding propagation; and (iii) whether clonal Mediterranean species could have survived better than non-clonal species in glaciated regions.

In the thoroughly studied population of Aranda de Duero, two other remarkable facts were observed. First, that it has been softly contracting (in effective population size) during several hundred generations. Second, that its large clones have been overall the most successful in sexual breeding (Chapter 6), which could have produced the population decline mentioned. However, the observed elevated genetic diversity and moderate inbreeding suggest that population genetic contraction has been more influenced by expansion of some genets, which have outcompeted others. The presence of clones with numerous ramets also appear to have influenced the capacity to sire offspring in the population of Jimena (Chapter 7), probably by the mere multiplicative effect of having many ramets.

In summary, clones with many ramets seem to reproduce better sexually. This fact does not seem to affect population evolutionary potential, as no increased inbreeding has been detected. The likely trade-offs between sexual and asexual propagation due to resource allocation (VALLEJO-MARIN *et al.* 2010) is therefore not appreciable in our studies. However, no deeper conclusions can be drawn, due to the fact that the high number of ramets in large clones could have been produced by age, and not by higher resource allocation. Specific essays are necessary to test to what extent clones propagating asexually are also more efficient in sexual breeding despite expected reduced resource allocation to this function. For instance, it would be interesting to assess the relative amounts of resources that a genet must invest to effectively propagate asexually and sexually. For this, it is important to consider that (i) propagation by clonality permits to increase the amount of resources accessed, and that (ii) the necessary investment for effective reproduction varies with the ecological circumstances, especially for sexual reproduction, whose optimal conditions are intuitively narrower than for clonal reproduction – at least in highly clonal species.

An obvious, direct consequence of clonality is a reduction of genotypic richness. Additionally, it also reduces genotypic evenness, given the consistent right-tailed distributions of ramets by genet (ARNAUD-HAOND *et al.* 2007). Such skewness was also observed in the population of Aranda de Duero (Chapter 6). However, a significant reduction of genetic diversity at genet-level is neither expected by numerical simulations (BALLOUX *et al.* 2003) nor observed in Aranda de Duero in comparison with other, less clonal, populations (Chapter 6). Genetic diversity is lower at ramet-level, though. Long-term consequences of clonal reproduction for genetic diversity are not clear, especially considering that sexual dominance also reduces the effective population size, increasing genetic drift. Moreover, despite genetic diversity being maintained, the reduction of genotypic diversity could represent a serious menace, as long as endurance to severe drawbacks could be more affected by the overall number of genotypes (as combinations of different alleles in different genes) than by the overall number of alleles (shuffled in less genotypic combinations), as it has been reported for example for *Zostera marina* (REUSCH *et al.* 2005). Hence, linkage disequilibrium (LD) at ramet-level should also be taken into account as an element for proper diagnosis of the population evolutionary potential.

Another plausible long-term consequence of an imbalance between the sexual and asexual reproductive functions could be a complete loss of sexual function (Chapter 6). This may

happen by depletion of smaller genets, reducing the population to very few incompatible genotypes, or by erosion of sex reproductive functions, by the "use it or lose it" evolutionary principle. As a matter of fact, fully clonal and almost fully clonal populations have been found in white poplar (BRUNDU *et al.* 2008; FUSSI *et al.* 2012). Thus, this is a realistic scenario for this species.

Admixture and hybridisation

Admixture between lineages appears to increase standing genetic variation that facilitates adaptation, as observed in *P. tremula* (Chapter 4). In contrast, despite the collision of two different genomic backgrounds, an outbreeding depression was not observed. In white poplar, no signal of outbreeding depression was detected among populations that exchange genes within the river basins of the Iberian Peninsula (Chapter 3). Furthermore, the study of the backcrossed progeny of *P.* × *canescens* × *P. alba* (BC_1) revealed that asymmetries in post-mating barriers resulted in introgression rather than an evolution of reinforcement (Chapter 5). Pollen and seed gene flow among populations, or even assisted migration (MCLACHLAN *et al.* 2007), could benefit from the absence of outbreeding depression and genetic reinforcement, allowing the creation of new gene pools with enriched standing genetic variation, in the context of promoting adaptive responses to impending climate change.

I also found extensive segregation distortion in our BC_1, towards the *P. tremula* background, probably due to a combination of heterozygote advantage and cyto-nuclear interactions (Chapter 5). Besides, local interspecific heterozygosity and genetic background may also underlie clonal success (Chapter 6). Heterozygous genomes should be favoured in outcrossing species. On the one hand, these results highlight the importance of heterozygosity to promote fitness, which makes this issue worthy of further research in poplars. On the other hand, cyto-nuclear interactions could have limited the generation of natural F_2 and BC_1 progenies. Indeed, lower sexual reproductive success of the large hybrid clone found in the Duero basin (Chapter 6) could have been related to cyto-nuclear interactions, although more probable factors are ecological pre- and post-zygotic barriers (Chapter 6). Such barriers could have resulted in the absence of shared haplotypes between *P. alba* and *P. tremula* in the Iberian Peninsula, where even *P. nigra* haplotypes are closer to *P. alba* than *P. tremula* ones (Chapter 3), partially discarding the possibility of intense

hybridisation and introgression between these two species. However, since both species grow in ecologically (and often geographically) very distant habitats (low flatlands vs. high hillsides), gene flow between populations must have been driven primarily by pollen, However, chloroplasts are not transported via pollen in broadleaves, keeping chloroplast integrity of populations intact, and concealing a possible intense hybridisation or introgression between both species in our cpDNA studies. Under this framework, nuclear gene flow could have occurred, but chloroplasts would hardly been exchanged.

Chromosome XIX

From the twelve candidate loci (out of 62) for locally varying selection in five populations of European aspen, three were situated at the proximal end of chromosome XIX (Chapter 4). LD in this region was elevated within the studied populations (extended up to 400 kb, assuming *P. trichocarpa* physical map distances). In two white poplar populations from Central Europe, LD was also high among telomeric markers of chromosome XIX (560 kb), but not in a European aspen population (Chapter 5). Both results are consistent with the lack of recombination in this genomic region, a key feature of sex chromosomes (CHARLESWORTH *et al.* 2005; MING and MOORE 2007). On the other hand, this region also showed distorted segregation in the backcrossed progeny (towards *P. tremula* background) and presented reduced divergence between *P. tremula* and *P. alba*. Therefore, and against expectations for sex chromosomes (QVARNSTROM and BAILEY 2009), this genomic region is not protected against gene flow between species. Although a lack of interspecific gene flow is not an essential characteristic of sex chromosomes, theory and observation support some association (QVARNSTROM and BAILEY 2009). Moreover, population genetic analysis in either sex displays that microsatellites in this region behave like codominant, autosomal markers and show no consistent pattern of reduced heterozygosity (Chapter 5).

Regarding all the features of chromosome XIX, it appears that it is in an initial stage of development, but the reason why sex chromosomes have not evolved more in poplars, where dioecy seems to be an ancient trait (TUSKAN *et al.* 2012), is not clear, especially considering that in some more modern plant groups it has evolved further. For instance, the heteromorphic sex chromosome of a group within genus *Silene* evolved in less than 20 million years (CHARLESWORTH 2002). Three possible hypotheses could be argued. Firstly, dioecy in poplars is not such an ancient trait, so sex chromosomes have not had time to

progress. Secondly, selection forces that favour the development of sex chromosomes are not strong enough and/or evolution in poplars is so slow, due to long generation times (fostered also by clonality; Chapter 6), that the process is still ongoing, but at a slower pace than in other plant groups. And lastly and alternatively, sex lability could have been favoured in poplars, because it can confer an adaptive advantage (CHARNOV and BULL 1977). If the strength of selection for maintaining sex lability is higher than the selection for developing sex chromosomes, this development should be arrested, since sex chromosomes usually impede sex lability (EHLERS and BATAILLON 2007).

The discovery of several hermaphrodites in the white poplar population of Jimena, of which some were effectively pollinators (Chapter 7), opens the possibility that hermaphroditism may be favoured in some conditions, upon which sex lability could be maintained. Interestingly, in subdioecious species (i.e. with sex lability), abundance of hermaphroditic individuals (in contrast to monoecious ones) is normally not associated to clonality (EHLERS and BATAILLON 2007), perhaps because clonality seems to solve the same predicament as hermaphroditism assuring reproduction. Although monoecy also enhances reproduction under partner scarcity, it does it to a lesser extent. The reasons why in poplars and willows sex lability is often reported, while both genera are clonal, remain unknown. Both strategies resolve reproduction adversity, although hermaphroditism permits reproduction under pollen or ovule limitations while clonality assures reproduction even when effective sexual reproduction is impaired. Nevertheless, clonality and hermaphroditism entail serious drawbacks: genotypic impoverishment and inbreeding depression, respectively. Thus, each strategy should be adaptive depending on the ecological circumstances, without a complete overlap of their functionalities, as for example during colonisation or persistence under specific unfavourable conditions. It is worthwhile to remember that standing genetic variation in sex expression and hybridisation could give the basis to sexual lability (BARRETT 2013). For instance, returns from dioecy to monoecy could occur through an increase in the frequency of sex-inconstant males, extant due to an elevated standing variation in sex expression, but also, they can happen after hybridisation (or introgression) with a monoecious relative.

References

ARNAUD-HAOND, S., C. M. DUARTE, F. ALBERTO and E. A. SERRAO, 2007 Standardizing methods to address clonality in population studies. Molecular Ecology **16:** 5115-5139.

BALLOUX, F., L. LEHMANN and T. DE MEEUS, 2003 The population genetics of clonal and partially clonal diploids. Genetics **164:** 1635-1644.

BARRETT, S. C. H., 2013 The evolution of plant reproductive systems: how often are transitions irreversible? Proceedings of the Royal Society B-Biological Sciences **280**.

BRUNDU, G., R. LUPI, I. ZAPELLI, T. FOSSATI, G. PATRIGNANI *et al.*, 2008 The origin of clonal diversity and structure of *Populus alba* in Sardinia: evidence from nuclear and plastid microsatellite markers. Annals of Botany **102:** 997-1006.

CHARLESWORTH, D., 2002 Plant sex determination and sex chromosomes. Heredity **88:** 94-101.

CHARLESWORTH, D., B. CHARLESWORTH and G. MARAIS, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity **95:** 118-128.

CHARNOV, E. L., and J. J. BULL, 1977 When is sex environmentally determined? Nature **266:** 828-830.

EHLERS, B. K., and T. BATAILLON, 2007 'Inconstant males' and the maintenance of labile sex expression in subdioecious plants. New Phytologist **174:** 194-211.

FUSSI, B., J. BONELLO, E. CALLEJA and B. HEINZE, 2012 Combining the use of molecular techniques and archival documentary evidence to trace the origin of *Populus alba* in a Central Mediterranean archipelago. European Journal of Forest Research **131:** 347-354.

HEUERTZ, M., S. CARNEVALE, S. FINESCHI, F. SEBASTIANI, J. F. HAUSMAN *et al.*, 2006 Chloroplast DNA phylogeography of European ashes, Fraxinus sp (Oleaceae): roles of hybridization and life history traits. Molecular Ecology **15:** 2131-2140.

LATTA, R. G., 2004 Gene flow, adaptive population divergence and comparative population structure across loci. New Phytologist **161:** 51-58.

LE CORRE, V., and A. KREMER, 2003 Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics **164:** 1205-1219.

LE CORRE, V., and A. KREMER, 2012 The genetic differentiation at quantitative trait loci under local adaptation. Molecular Ecology **21:** 1548-1566.

MCLACHLAN, J. S., J. J. HELLMANN and M. W. SCHWARTZ, 2007 A framework for debate of assisted migration in an era of climate change. Conservation Biology **21:** 297-302.

MING, R., and P. H. MOORE, 2007 Genomics of sex chromosomes. Current Opinion in Plant Biology **10:** 123-130.

QVARNSTROM, A., and R. I. BAILEY, 2009 Speciation through evolution of sex-linked genes. Heredity **102:** 4-15.

REUSCH, T. B. H., A. EHLERS, A. HAMMERLI and B. WORM, 2005 Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proceedings of the National Academy of Sciences of the United States of America **102:** 2826-2831.

RODRIGUEZ-SANCHEZ, F., A. HAMPE, P. JORDANO and J. ARROYO, 2010 Past tree range dynamics in the Iberian Peninsula inferred through phylogeography and palaeodistribution modelling: A review. Review of Palaeobotany and Palynology **162:** 507-521.

SAVOLAINEN, O., T. PYHAJARVI and T. KNURR, 2007 Gene flow and local adaptation in trees. Annual Review of Ecology Evolution and Systematics **38:** 595-619.

SOTO, A., J. J. ROBLEDO-ARNUNCIO, S. C. GONZALEZ-MARTINEZ, P. E. SMOUSE and R. ALIA, 2010 Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. Molecular Ecology **19:** 1396-1409.

TUSKAN, G. A., S. DIFAZIO, P. FAIVRE-RAMPANT, M. GAUDET, A. HARFOUCHE *et al.*, 2012 The obscure events contributing to the evolution of an incipient sex chromosome in *Populus*: a retrospective working hypothesis. Tree Genetics & Genomes **8:** 559-571.

VALLEJO-MARIN, M., M. E. DORKEN and S. C. H. BARRETT, 2010 The ecological and evolutionary consequences of clonality for plant mating, pp. 193-213 in *Annual Review of Ecology, Evolution, and Systematics, Vol 41*, edited by D. J. FUTUYMA, H. B. SHAFER and D. SIMBERLOFF.

CHAPTER 10: Conclusions/*Conclusiones*

1. White poplar (*Populus alba*) and black poplar (*P. nigra*) have experienced different demographic histories in the Iberian Peninsula, probably due to differences in ecological requirements and their distinct response to climatic adversities.

   *El álamo blanco (*Populus alba*) y el álamo negro (*P. nigra*) han experimentado historias demográficas diferentes en la Península Ibérica, probablemente debido a diferencias en sus requerimientos ecológicos y a su diferente capacidad de respuesta ante las adversidades climáticas.*

2. White poplar has persisted through glaciations in all major river basins of the Iberian Peninsula, although glaciations have affected its patterns of genetic diversity. Migration across latitudinal ranges has been easier than across longitudinal ones, resulting in the differentiation of Atlantic and Mediterranean populations.

   El álamo blanco ha persistido durante el paso de las glaciaciones en todas las grandes cuencas de ríos de la Península Ibérica, si bien las glaciaciones han afectado a sus patrones de distribución de diversidad genética. La migración entre franjas latitudinales ha sido más fácil que entre las longitudinales, lo que ha resultado en una diferenciación entre poblaciones atlánticas y mediterráneas.

3. Local adaptation occurs in white poplar, but at a broader scale than in other forest trees due to gene flow among populations within river basins, although there is no outbreeding depression. Local adaptation is also present in European aspen (*P. tremula*). Genetic admixture of different aspen lineages causes increased standing genetic variation, but without producing outbreeding depression.

   *Se aprecia adaptación local en el álamo blanco, si bien a una escala más amplia que en otras especies debido al flujo genético entre poblaciones dentro de cuencas fluviales que, sin embargo, no ha llegado a causar depresión por exogamia. También el álamo temblón (*P. tremula*) presenta adaptación local. Además, la mezcla genética de diferentes linajes ha incrementado la variación genética presente en esta especie, sin producir depresión por exogamia.*

4. Large clonal assemblies (>10 km) and elevated clonal rates characterise some white poplar populations of the Iberian Peninsula. In Aranda de Duero, dominance of both asexual and sexual reproduction by a few large clones reduces genotypic diversity and results in a soft but prolonged decline in effective population size. However, genetic diversity is still large and no detrimental genetic effects were observed in this population.

   La existencia de grandes conjuntos clonales (>10 km) y las altas tasas de clonalidad caracterizan algunas de las poblaciones de álamo blanco de la Península Ibérica. En Aranda de Duero, la dominancia tanto asexual como sexual de unos pocos clones de gran tamaño reduce la diversidad genotípica y tiene como consecuencia un descenso suave pero prolongado del tamaño efectivo poblacional. A pesar de ello, la diversidad genética es todavía elevada y no se observan efectos genéticos negativos en esta población.

5. Cyto-nuclear interactions can result in increased interspecific heterozygosity which, together with reduced inbreeding, may underlie local adaptation patterns and the spread of large clones.

   Las interacciones citonucleares pueden provocar un incremento en la heterozigosidad interespecífica, lo que, junto con un endogamia reducida, puede que produzcan los patrones de adaptación local y la expansión de los clones de gran tamaño.

6. Some findings in this thesis agree with an expected sex chromosome status for chromosome XIX, while others do not. This reinforces the idea that chromosome XIX is, at best, a sex chromosome in an initial stage of development.

   Algunos resultados de esta tesis sobre el cromosoma XIX concuerdan con lo esperable para un cromosoma sexual, pero otros no. Esto reafirma la idea de que el cromosoma XIX es, en el mejor de los casos, un cromosoma sexual en un estado inicial de desarrollo.

Acknowledgements/*Agradecimientos*

First of all, I would like to thank Santiago César (Santi) González-Martínez and Christian Lexer for their magnificent direction of this doctoral thesis, not only because it has enabled me to finish it in a very satisfactory way, but also because they have taught me and trained me to continue a scientific career in a hopefully successful way.

También me gustaría agradecer a mis padres, Pilar y José María, su apoyo incondicional en todos los ámbitos, desde anímicos a económicos, para lograr esta meta. Asimismo a mi hermano José Ignacio, que siempre que me ha hecho falta me ha ayudado.

Then, I would like to thank the generous support and excellent hosting that David B. Neale, in the University of California, at Davis, and Michael Fay from the Royal Botanical Gardens, Kew have given to me. *De igual manera quiero agradecer a Myriam Heuertz, Ricardo Alía y Nuria Alba (in memoriam) el apoyo prestado desde el INIA en Madrid.*

Del INIA también tengo que agradecer a Carmen su inestimable ayuda en el laboratorio, y a Fátima, José Antonio, Santi, y Fernando su compañía y ayuda en los numerosos viajes de campo. Por su puesto, también a mis compañeros del despacho de becarios Javi, José Alberto, Raúl, Katha, Quique, Marina, Natalia, Paloma y todos los demás, que tanto apoyo moral intercambiamos. Otras personas que también me prestaron consejo y ayuda fueron Juanan, Rosana, Patri, Salva, Ana, Marta, Isra, Marga, Bea, Fede, María, Raquel, Juanjo, Ismael y Maite. En particular, me gustaría agradecer a los dos primeros, como viejos amigos, porque ellos, probablemente sin saberlo, me motivaron para que iniciase mi tesis, y me han estado guiando con su ejemplo durante estos años. También quiero dedicarle una mención especial a Katha por sus valiosos consejos y su apoyo.

De la Universidad de Valladolid, en la que si bien poco me he dejado ver, pero que mucho me ha ayudado en la distancia, tengo que recordar a Ana, por su especial atención, a Celia, por ayudarme a solucionar todo con tanta eficiencia y rapidez, y a Elena Hidalgo, Felipe Bravo y Juan Andrés Oria.

I must acknowledge many people from my stays in Jodrell Laboratory at Kew Gardens, but in particular I would like to recall Jeff for his help and company, and Thelma, Clarisse, Fabio, the two Marcelas, Timothy, Dulce, Annelies, Benedetta, Robyn, Dion, and many others, for their support. From Davis, I would like to thank to Jeff, Joost, Andrew, Jill, and the rest of the people of the lab. Finally, I would like to acknowledge the support given by Doro, Léonie and Celine in the University of Fribourg (Switzerland).

From multiple places, I also want to thank to Unai López, Arantxa Prada, Carmen Maestro, Alex Widmer, Joan Cottrell, Pär Ingvarsson, Berthold Heinze, Denes Bartha, Ivan Schanzer, and Claudio Sedivy their interest and help at different stages of my PhD.

En último lugar, pero no menos importante, me gustaría agradecer su ayuda real y completamente desinteresada a varios amigos que no pertenecen al mundo de la ciencia, pero que en ocasiones me acompañaron a hacer muestreos, para que no tuviera que irme solo. Ellos son Javi, Manuel, Ángel, Chus y Jorge.

Also, the friends I met in Davis, whom I could hang out and relax with, that is sometimes so necessary when you are far away from home: George, John, Eva, and Ana, thank you! *Je tiens remercier Sébastian et Marion pour demeurer chez eux à Londres pendant plusieurs semaines.*

Por supuesto, también quiero agradecer al resto de mis amigos y familiares por haber estado ahí, a las duras y a las maduras, y en particular, a mis dos compañeros de piso, Coché y José Luis, sencillamente porque me han hecho la vida mucho más placentera durante estos años de mi vida.

Por último, querría agradecer la ayuda institucional prestada desde el Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), en particular desde el Centro de Investigación Forestal (CIFOR), y también desde la Universidad de Valladolid y su Instituto Universitario de Investigación en Gestión Forestal Sostenible (IUFGS), así como el apoyo de sus trabajadores.

¡Gracias a todos!

Esta tesis doctoral ha sido financiada gracias al programa de becas y contratos de Formación de Personal Investigador (FPI) del antiguo Ministerio de Ciencia e Innovación, y del proyecto del Plan Nacional de I+D+i "Sistemas de reproducción, flujo genético efectivo y efectos parentales: aplicación a la gestión y conservación de recursos genéticos forestales" (AGL2005-07440-C02-01/FOR).

ANNEX

Annex A

The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars.

MACAYA-SANZ, D., HEUERTZ, M., LOPEZ-DE-HEREDIA, U., DE-LUCAS, A. I., HIDALGO, E., MAESTRO, C., PRADA, A., ALÍA, R., GONZÁLEZ-MARTÍNEZ, S. C., 2012. Molecular Ecology **21**: 3593-3609.

Annex B

Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree.

DE CARVALHO, D., INGVARSSON, P. K., JOSEPH, J., SUTER, L., SEDIVY, C., **MACAYA-SANZ, D.**, COTTRELL, J., HEINZE, B., SCHANZER, I, LEXER, C., 2010. Molecular Ecology **19**: 1638-1650.

Annex C

Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species.

MACAYA-SANZ, D., SUTER, L., JOSEPH, J., BARBARÁ, T., ALBA, N., GONZÁLEZ-MARTÍNEZ, S. C., WIDMER, A., LEXER, C., 2011. Heredity **107**: 478-486.

Annex D

Causes and consequences of large clonal assemblies in a poplar hybrid zone.

MACAYA-SANZ, D., HEUERTZ, M., LINDTKE, D., LEXER, C., GONZÁLEZ-MARTÍNEZ, S. C, 2014. Manuscript.

Annex A

The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars.

MACAYA-SANZ, D., HEUERTZ, M., LOPEZ-DE-HEREDIA, U., DE-LUCAS, A. I., HIDALGO, E., MAESTRO, C., PRADA, A., ALÍA, R., GONZÁLEZ-MARTÍNEZ, S. C., 2012. Molecular Ecology **21**: 3593-3609.

MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 3593–3609 doi: 10.1111/j.1365-294X.2012.05619.x

The Atlantic–Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars

D. MACAYA-SANZ,* M. HEUERTZ,* U. LÓPEZ-DE-HEREDIA,† A. I. DE-LUCAS,‡ E. HIDALGO,§ C. MAESTRO,¶ A. PRADA,** R. ALÍA* and S. C. GONZÁLEZ-MARTÍNEZ*

**Department of Forest Ecology and Genetics, Forest Research Centre, INIA. Carretera de A Coruña km 7.5 – 28040 Madrid, Spain, †Forest Genetics and Physiology Research Group, ETS Forestry Engineering, Technical University of Madrid (UPM), Ciudad Universitaria s/n – 28040 Madrid, Spain, ‡Laboratorio de Diagnóstico Genético, Departamento de Biotecnología, ITAGRA.CT, Campus Universitario ''La Yutera'', Avenida de Madrid 44 – 34004 Palencia, Spain, §Departamento de Producción Vegetal y Recursos Forestales, E.T.S. Ingenierías Agrarias, Universidad de Valladolid, Avenida de Madrid 44 – 34004 Palencia, Spain, ¶Unidad de Recursos Forestales, CITA, Gobierno de Aragón, Avenida de Montañana 930 – 50059 Zaragoza, Spain, **Banc de Llavors Forestals, Dirección General de Gestión del Medio Natural, Conselleria de Territorio y Vivienda, Avda. Comarques del País Valencià 144 – 46913 Quart de Poblet, Valencia, Spain*

Abstract

Recent phylogeographic studies have elucidated the effects of Pleistocene glaciations and of Pre-Pleistocene events on populations from glacial refuge areas. This study investigates those effects in riparian trees (*Populus* spp.), whose particular features may convey enhanced resistance to climate fluctuations. We analysed the phylogeographic structure of 44 white (*Populus alba*), 13 black (*Populus nigra*) and two grey (*Populus x canescens*) poplar populations in the Iberian Peninsula using plastid DNA microsatellites and sequences. We also assessed fine-scale spatial genetic structure and the extent of clonality in four white and one grey poplar populations using nuclear microsatellites and we determined quantitative genetic differentiation (Q_{ST}) for growth traits in white poplar. Black poplar displayed higher regional diversity and lower differentiation than white poplar, reflecting its higher cold-tolerance. The dependence of white poplar on phreatic water was evidenced by strong differentiation between the Atlantic and Mediterranean drainage basins and among river basins, and by weaker isolation by distance within than among river basins. Our results suggest confinement to the lower river courses during glacial periods and moderate interglacial gene exchange along coastlines. In northern Iberian river basins, white poplar had lower diversity, fewer private haplotypes and larger clonal assemblies than in southern basins, indicating a stronger effect of glaciations in the north. Despite strong genetic structure and frequent asexual propagation in white poplar, some growth traits displayed adaptive divergence between drainage and river basins ($Q_{ST} > F_{ST}$), highlighting the remarkable capacity of riparian tree populations to adapt to regional environmental conditions.

Keywords: genetic differentiation, glaciations, Iberian Peninsula, Populus, $Q_{ST} > F_{ST}$, spatial genetic structure.

Received 15 December 2011; revision received 16 March 2012; accepted 27 March 2012

Introduction

Historical demographic processes caused by the Pleistocene glaciations have contributed to shape the current patterns of phylogeographic structure in widespread temperate tree species. As the location of glacial refugia and the ways postglacial colonization took place became elucidated (Nichols & Hewitt 1994; Palmé *et al.* 2003), researchers increasingly started focusing on phylogeographic patterns within former glacial refugia such

Correspondence: Santiago C. González-Martínez, Fax: +34 913572293; E-mail: santiago@inia.es

as the Mediterranean peninsulas (Rodríguez-Sánchez *et al.* 2010). The existence of numerous refugia in the Mediterranean Peninsulas is believed to have enabled species and populations to persist in these buffered habitats throughout the Quaternary (Medail & Diadema 2009). Even in the southern regions that experienced the severest conditions (e.g. the north-western Iberian Peninsula), Mediterranean species could have persisted in isolated benign locations, similar to the 'cryptic refugia' described for boreal species (Anderson *et al.* 2006; Cheddadi *et al.* 2006; Petit *et al.* 2008). Typical signatures of former refuge populations are low within-population genetic diversity, accompanied by large amounts of regional diversity, and high levels of genetic differentiation among populations (Hampe & Petit 2005). Besides the effects of the Quaternary glaciations, some recent studies have also highlighted the importance of older geological events, such as Tertiary plate tectonics, in shaping the current phylogeographic structure of forest trees in former refugial areas (e.g. *Quercus suber*, Magri *et al.* 2007; *Taxus baccata*, González-Martínez *et al.* 2010).

Although there is a large body of phylogeographic studies on former glacial refugia regions (see review in Medail & Diadema 2009 for the Mediterranean basin), few have focused on riparian temperate species, whose particular attributes could have enhanced their resilience to past climate changes. First, as riparian temperate species performance depends largely on a single environmental condition, phreatic water availability, that is more related to orography than to climate, climate factors may affect them less than to other plant species. Second, their preferred habitats (valley bottoms, wetlands and deep gorges) are considered ideal to buffer climatic oscillations due to warmer and moister conditions, making them good candidates for 'refugia within the refugia' (Medail & Diadema 2009; Nieto-Feliner 2011). Third, many typical temperate riparian species (e.g. *Populus* spp., *Salix* spp., *Tamarix* spp., *Ulmus* spp.; Stuefer *et al.* 2002; Ruiz de la Torre 2006; Slavov & Zhelev 2010) have high levels of clonality, which could reinforce population survival by securing local persistence through unfavourable conditions and allowing rapid colonization of disturbed areas. As a drawback, the dependence of riparian trees on phreatic water leads to a scattered pattern of suitable habitats, separated by large inhospitable areas (e.g. elevated plateaus between river valleys). As a result, riparian populations are exceptionally prone to isolation, and consequently substantial genetic structure has been reported at regional level in different riparian trees (e.g. Cottrell *et al.* 2005; Fussi *et al.* 2010).

Pervasive population isolation promotes stochastic processes reducing both molecular and quantitative genetic variation (at a rate that depends on the reciprocal of effective population size). Depletion of genetic variation reduces the ability of populations to adapt. Theoretical models have shown an ambiguous role of gene exchange in local adaptation. Gene flow counteracts the effects of selection, as it dilutes local changes in allele frequencies (Lenormand 2002). However, Alleaume-Benharira *et al.* (2006) showed, using individual-based simulations, that gene flow can also mitigate the effect of drift by replenishing genetic variance in small marginal populations. In species with more specialized ecological requirements, the gene flow from core populations necessary to ensure adaptability of isolated marginal populations is expected to increase (Alleaume-Benharira *et al.* 2006). As specialized species are prone to geographic isolation but, at the same time, inhabit similar environments across large ranges, gene flow may be critical to keep levels of genetic variance high enough to maintain their evolutionary potential. To our knowledge, no study has hitherto reported on the adaptive consequences derived from the ecological specificity of riparian tree species.

In this study, we assessed the genetic diversity and structure of wind-dispersed Iberian poplar species (especially white poplar, *Populus alba* L.) at local, regional and wide spatial scales using chloroplast and nuclear DNA markers. Additionally, common garden data in white poplar provided insights into the adaptive significance of river-basin isolation in this species. The Iberian Peninsula (IP hereafter) represents an ideal setting for this study, as it harbours numerous refuge areas with distinct environmental features (Gómez & Lunt 2007 and references therein). With regard to riparian habitats, high climatic heterogeneity is accompanied by a complex river system consisting of two drainage basins (watershed of the Atlantic Ocean and the Mediterranean Sea), several main river basins and numerous smaller watercourses originating in coastal mountain ranges. Different parts of the IP currently inhabited by poplar species (notably the Duero basin, in the north-western range) were severely affected by Quaternary glaciations and present particularly amenable conditions for testing the persistence of riparian species in harsh environments. Within this range, we focused on the white poplar, a widespread species with high colonization capability and marked tolerance to temperature changes, atmospheric dryness and salt stress, if groundwater is available (Ruiz de la Torre 2006). Despite the scarcity of palynological records resulting from poor pollen preservation (Huntley & Birks 1983), leaf fossils of white poplar found in travertine formations have shown its undoubted native presence in the IP (Roiron *et al.* 2004). Furthermore, in contrast with other European poplar species, white poplar has limited utility to

humans and negligible commercial value. Human mediated movement of reproductive material is therefore unlikely to have modified the pattern of natural genetic structure in this species.

Analysing different aspects of genetic diversity and spatial genetic structure (SGS) in Iberian poplars, we aimed at clarifying the role of climatic fluctuations and orographic barriers on population dynamics in riparian species. The use of different types of molecular markers [plastid DNA (cpDNA) sequences and microsatellites (cpSSRs) and nuclear microsatellites (nSSRs)] allowed us to discriminate among distinct overlapping patterns of SGS and to control for allele homoplasy. The comparison of neutral genetic differentiation patterns to quantitative traits facilitated understanding the role of isolation in promoting local adaptation. More specifically, our goals were to (i) examine the genetic signature of ancient geological divides (the flooding of the Strait of Gibraltar and the rise of the Mediterranean/Atlantic watershed), setting a temporal framework for main phylogeographical events; (ii) assess regional patterns of diversity and differentiation, informing on the capability of riparian species to survive severe climatic changes *in situ* and to migrate across extensive inhospitable areas; (iii) evaluate the role of asexual reproduction and fine-scale genetic structure for maintaining population persistence and connectivity within river basins in a water-dependent species; and (iv) test for signs of local adaptation based on genetic differentiation for quantitative traits (as estimated by Q_{ST}). The comparison of levels of genetic differentiation for molecular markers and quantitative traits addressed the specific question of adaptive divergence vs. genetic drift in a narrow-niche but widespread species.

Materials and methods

Plant material, sampling and DNA extraction

Fifty-nine Iberian poplar populations (n = 628 trees) were sampled (see Fig. 1, details in Table S1, Supporting information), with a focus on Iberian white poplar (*Populus alba* L.; 44 populations), and representative samples of black (*Populus nigra* L.; 13 populations) and grey (*Populus x canescens* (Aiton) Sm.; two populations) poplars. Black poplar was not sampled in the south of the IP as it is relatively scarce in that region. Sampling included three major river basins, two draining to the Atlantic Ocean (Duero and Guadalquivir rivers, in northwestern and southern Spain, respectively) and one to the Mediterranean (Ebro river, northeastern Spain). Several smaller Mediterranean river basins and scattered populations in two additional major Atlantic river basins (Tajo and Guadiana) were also sampled. In addition, one white poplar population from the High Atlas in Morocco was sampled to obtain time calibrations based on major biogeographic events separating the Iberian Peninsula from northern Africa (see below).

For each population, six leaves from each of n = 10 trees, separated by at least 100 m (to reduce the chance of sampling clonal replicates or related trees), were collected and dried in silica gel prior to cpDNA analysis. To study genetic structure at the local scale using nuclear microsatellites (nSSRs), one population of grey poplar (from the Guadiana river) and four of white poplar (from the Duero, Guadalquivir and Ebro basins; Fig. 1) were more intensively sampled, collecting material from three to five additional trees around each of the ten core individuals (n = 200 trees). All trees were geographically referenced.

Total DNA was purified from dried leaves following a slightly modified protocol from Doyle & Doyle (1990).

Molecular markers (cpSSRs, nSSRs and cpDNA sequences)

Thirteen chloroplast microsatellites (cpSSRs) from Weising & Gardner (1999) and Bryan *et al*. (1999) were tested in a panel consisting of 17 individuals sampled across the white poplar range in the Iberian Peninsula. Out of 11 cpSSRs that produced a PCR product, only two (*ccmp2* and *ccmp5*) were polymorphic. All samples were amplified at *ccmp2* and *ccmp5* (missing data of ca. 6%) in 10 μL of final volume, including 5 ng of DNA template, 0.4 units of Taq (Bioline, London, UK), 0.15 μM of each primer (the forward primers labelled with IRD800; Li-Cor Biosciences, Lincoln, NE, USA), 0.1 mM of each dNTP and 2 mM of $MgCl_2$. The PCR profile consisted of 5 min at 94 °C, 12–24 cycles (samples with weak amplification at 12 cycles were repeated using 24 cycles) of 1 min at 94 °C, 30 s at 49 °C (*ccmp2*) or 50 °C (*ccmp5*), and 1 min at 72 °C and a final extension of 10 min at 72 °C. PCR fragments were resolved on a Li-Cor 4300 DNA analyser (Li-Cor Biosciences). To reduce the probability of scoring errors, a selection of samples that covered the fragment size range was included as internal standard in each gel. SAGA^{GT} vs. 3.3. was used for gel calibration and scoring (Li-Cor Biosciences). The chloroplast DNA region *trnC-petN1* was sequenced in at least one individual per population and cpSSR haplotype (n = 133 individuals), assuming that individuals with the same cpSSR haplotype within populations would also share their *trnC-petN1* haplotype, because of lower expected mutation rates in the latter. For each combination of cpSSRs and *trnC-petN1*, the *rpl16-poprpl* cpDNA region was sequenced in at least one individual (n = 107 individuals). To sequence the samples in both directions, 30 μL of PCR product

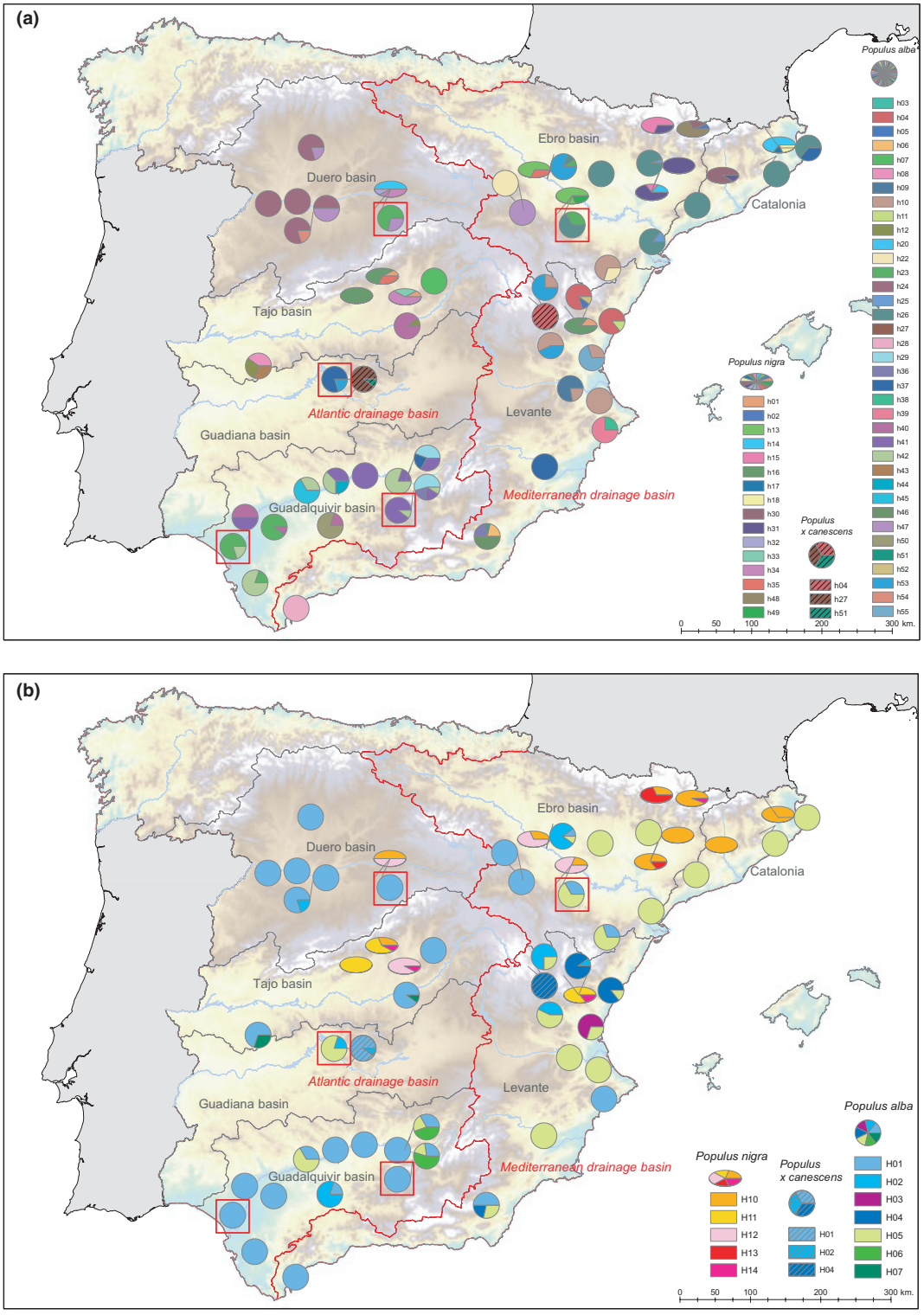


Fig. 1 Geographic distribution and population frequency of haplotypes based on (a) the full data set (cpSSRs and cpDNA sequences) and (b) unique event polymorphisms (UEPs). Red squares indicate populations used to study clonality levels and fine-scale population genetic structure. Main hydrographic features and the altitudinal pattern (in shadows) are also shown.

was yielded for each sample and cpDNA region. The PCR mix included 7.5 ng of DNA template, 0.75 units of Taq (Bioline), 0.3 μM of each primer (unlabelled), 0.15 mM of each dNTP and 3 mM of $MgCl_2$. The PCR profile was 3 min at 94 °C, 39 cycles of 30 s at 94 °C, 30 s at 50 °C and 80 s at 72 °C, and a final extension of 10 min at 72 °C. Fragments were purified with Exo-SAP (Affymetrix, Santa Clara, CA, USA), and sequenced (standard Sanger sequencing) in external facilities (Secugen, Madrid). SEQMAN software (DNASTAR Inc., Madison, Wisconsin, USA) was used to edit and align cpDNA sequences.

Five highly-polymorphic nSSRs (ORPM127, ORPM312, PMGC2852, ORPM30 and ORPM344; Lexer *et al.* 2005) were used to study genetic structure at the local scale. Fourteen additional nSSRs (see Table S2, Supporting information) were used to confirm clonal identity of large clonal assemblies. Protocols for amplification and fragment resolution were the same as for cpSSRs with the following PCR modifications: 2 min initial denaturation, 30 s denaturation during cycles, 5 min final elongation, and annealing temperatures and number of cycles as given in Table S2 (Supporting information).

Haplotypes and haplotype networks

CpSSR scores and cpDNA sequences were combined into haplotypes (see Table S3, Supporting information). Owing to SSR mutation mechanisms and high polymorphism in *trnC-petN1*, homoplasy events (and, therefore, network reticulations) were considered likely in our data. Homoplasy events can introduce severe biases in some analyses, such as those based on coalescence (Provan *et al.* 2001). Therefore, three haplotype sets were defined: (i) based on the whole set of polymorphisms ('full data set', haplotypes coded with a '*h*' prefix; see Table S3, Supporting information); (ii) based on unique event polymorphisms, that is excluding any polymorphisms that produced reticulations in the haplotype network ('UEP data set', haplotypes coded with a '*H*' prefix; Table S3, Supporting information); and (iii) based on single nucleotide polymorphisms ('SNP data set', haplotypes coded with a 'HR' prefix; Table S3, Supporting information). Haplotype frequencies per population (see Table S4, Supporting information) were plotted using pie charts on a GIS (ARCMAP version 9.2; ESRI, Redlands, CA, USA) that also included main physiographical features.

Statistical parsimony networks were constructed with TCS vs. 1.21 (Clement *et al.* 2000) based on genetic distance matrices among haplotypes for the three haplotype sets. Because locus-specific mutation rates were not available, distances among different variants at each polymorphic site were considered equal. Likewise, all polymorphism types [insertions/deletions (indels), SSRs, SNPs or short tandem repeats (STRs)] were weighted identically. For comparative purposes, haplotypes from 12 Iberian and French aspens (*Populus tremula*) were added to the network.

Time to the most recent common ancestor in white poplar

To obtain insights into the temporal scale of relevant phylogenetic events in white poplar, we obtained estimates of the times to the most recent common ancestor (TMRCAs) for different data sets, based on the standard coalescence (constant size) model and Bayesian analysis using BATWING (Wilson *et al.* 2003). *Salix* sp. was used as outgroup (GenBank accessions AJ849602.1, DQ875043, DQ875044.1, DQ875047.1, DQ875048.1 and DQ875049.1). Analyses considered UEP sites to define tree topology and included (set1) all white poplar haplotypes (including Morocco), (set2) all Iberian white poplar haplotypes (excluding Morocco) or (set3) the white poplar haplotypes present in the Mediterranean drainage basin (upper part of the haplotype network, Fig. 2). The analyses were performed based on parameters

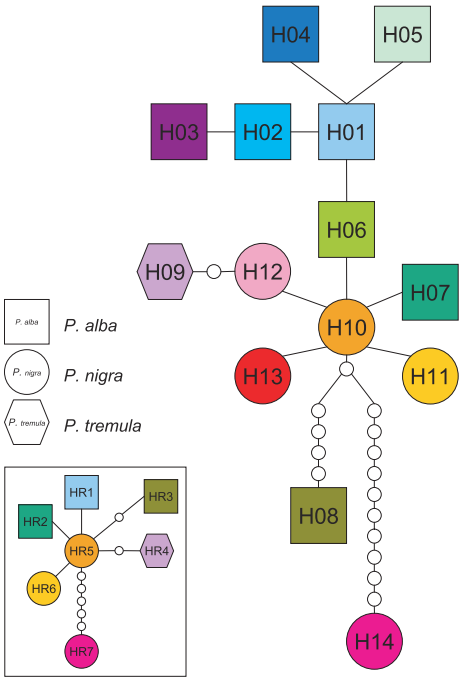


Fig. 2 Statistical parsimony network representing the minimum number of polymorphic site differences among haplotypes. The network was constructed considering only unique event polymorphisms (UEPs). The inset represents a network using only single nucleotide polymorphisms (SNPs). Notice the unexpected location of H08, H09 and H14.

scaled by population size N (which have higher precision; Wilson *et al.* 2003). A uniform prior distribution was assumed for θ ($\theta = 2N\mu$, μ being the mutation rate), and an estimation of TMRCA, T, was obtained in coalescence units, that is, scaled by N. For each data set, three independent MCMC were run with a burn-in period of 10 000 iterations and a main run of 20 000 to obtain a total of 60 000 iterations for the estimation of T.

To obtain unscaled coalescence times in years, we assumed that Iberian and African white poplar lineages diverged after the flooding of the Strait of Gibraltar (some 5.33 Ma; Duggen *et al.* 2003). This assumption seems reasonable as post flooding seed dispersal across the Strait of Gibraltar has been suggested to be very rare (if any) for other forest trees with large dispersal capabilities (see Jaramillo-Correa *et al.* 2010). The barrier to seed dispersal results from distance across the strait (14.3 km at its narrowest point) and from accelerated wind speed through the strait due to funnelling by the steep-sided land masses on both sides (wind normally blows from the East, that is, 'Levanter', during poplar dispersal season; Dorman *et al.* 1995). Considering that the product $T \times \theta$ is proportional to the unscaled TMRCA (t) in years (i.e. $T \times \theta = 2\mu \times t \times g$, where g is the generation time), rough estimates of t were obtained for set2 (the spread of Iberian haplotypes) and set3 (the divergence across Mediterranean lineages) by computing $T \times \theta$ proportions with respect to set1 [i.e. $(T \times \theta)_{set2}/(T \times \theta)_{set1} = t_{set2}/t_{set1}$ where t for set1, t_{set1}, is fixed at 5.33 Ma] under the assumption of constant mutation rates across runs (a reasonable hypothesis within species). Finally, effective population sizes for each set run were computed as $N = t/(T \times g)$ considering a generation time, g, of 20–60 years. These values are slightly higher than those found in the literature (e.g. 15 years for *P. tremula* in Ingvarsson 2008) because in white poplar generation times are probably extended due to extensive clonal propagation.

Genetic diversity and differentiation across drainage and river basins

Nei's (1978) unbiased haplotypic within-population genetic diversity (expected heterozygosity, H_E) was calculated based on the full data set (cpSSRs and cpDNA sequences) for all poplar species and populations using Arlequin vs. 3.1 (Excoffier *et al.* 2005). H_E was also computed by pooling individuals according to drainage basin (Atlantic vs. Mediterranean), river basin or latitude (North vs. South). Non-parametric Kolmogorov-Smirnov tests were used to compare levels of genetic diversity at each spatial scale. Computation of haplotypic richness, before and after rarefaction, and the number of private haplotypes after rarefaction was carried out using RAREFAC (Petit *et al.* 1998).

Estimates of genetic differentiation among populations and regions for white and black poplars were obtained based on the full and the UEP data sets. Hierarchical Analyses of Molecular Variance (AMOVAS) were performed with Arlequin vs. 3.1 grouping populations by drainage basins, river basins or latitude. AMOVAS for the UEP data set considered haplotype frequencies (F_{ST}) or distances among haplotypes (N_{ST}). Global F-statistics were computed using SPAGEDI vs. 1.3d software (Hardy & Vekemans 2002). Because F-statistics have recently been shown to perform badly when levels of diversity are dissimilar among populations, we also computed the D estimator by Jost (2008) using DEMEtics vs. 0.8.0 in R environment (Gerlach *et al.* 2010). Significance of D estimates was evaluated using bootstrap resampling.

In the extensively sampled white poplar, isolation by distance (IBD) analysis was used to test the effect of water-dependency on genetic structure. In the absence of barriers to gene flow, the ratio $F_{ST}/(1 - F_{ST})$ is expected to increase linearly as a function of the distance between pairs of populations (or its logarithm in two-dimensional analyses). Hence, a single linear regression slope is expected for all population pairs. If water-dependency influences the genetic structure, smaller differentiation is expected for pairs of populations belonging to the same river basin than for pairs belonging to different basins. Consequently, a flatter regression slope is expected within and a steeper slope between basins (see box 2 of Guillot *et al.* 2009). IBD was assessed using only the full data set in one (within main river basins) or two (within and between basins) dimensions. In one dimension, both 'resistance' distances (the distance measured following the river course, see Guillot *et al.* 2009) or Euclidean distances were used; otherwise Euclidean distances were applied. The significance of IBD (regression slope greater than zero) was tested with permutations. Statistical analyses were carried out with SPAGeDi vs. 1.3d (F_{ST} estimates) and Statistica vs. 10 (significance of regression slopes).

Genetic structure at the local scale

Five highly-polymorphic nSSRs were used to identify clonal assemblies and to evaluate fine-scale SGS in one grey and four white poplar populations (see Fig. 1). Clone identity in the Aranda de Duero population (where unusually large clonal assemblies were found) was confirmed using 14 additional nSSRs. Individuals with the same genotype (ramets) were identified using Gimlet vs. 1.3.2 (Valière 2002). All individuals (i.e. genets and ramets) were used together to produce an

overall estimate of SGS. The relative kinship coefficient F_{ij} of Loiselle *et al.* (1995) was computed for all pairs of individuals within populations using SPAGeDi, and considering the whole sample (n = 201) allele frequencies as reference. F_{ij} was regressed on the Euclidean distance between individuals (linear environment), and deviation from zero (presence of SGS) of the regression slope b was tested with permutations. Values of b were used to compare SGS strength across populations. The SGS patterns were plotted averaging F_{ij} in five distance classes (0–25, 25–50, 50–100, 100–200 and >200 m) including a similar number of sample pairs. The average kinship in distance classes over 100 m (the distance among trees chosen for wide-range population sampling) is relevant to evaluate the probability of including clonal replicates or related individuals in data analyses performed at larger spatial scales.

Common gardens and genetic differentiation for quantitative traits

Two of the genotyped white poplar populations from the Ebro and three from the Guadalquivir basin were included in a quantitative genetic study to determine trait differentiation among populations and river basins. The Ebro and Guadalquivir were chosen because they represent typical locations of the species and belong to different drainage basins (the Ebro drains to the Mediterranean Sea and the Guadalquivir to the Atlantic Ocean). For each population, two to four open-pollinated families (15 in total: eight from the Ebro and seven from the Guadalquivir) averaging *c.* 40 plants/family were established in a common garden following a complete block design with eight blocks. Total height at age 1 and 3 years (HT1 and HT3), and stem diameter at the base (DSB3) and stem form (FOR3) at age 3 years were measured in all individuals. Stem form was evaluated as a discrete variable, with values from 1 for straight stems, to 2 for arched stems without inflection, and 3 for sinuous stems with at least one inflection.

Variance components for basin, population and family effects were obtained by Restricted Maximum Likelihood (REML) using the following model:

$$y_{ijklm} = \mu + r_m + p(r)_{l(m)} + f(p)_{k(l)} + b_j + (f \times b)_{kj} + \varepsilon_{ijklm}$$

where y_{ijklm} is the phenotypic value of the variable for the ith tree from the kth family in the lth population in the mth river basin located in the jth block, μ is the overall mean, r_m is the effect of the mth river-basin, $p(r)_{l(m)}$ is the effect of the lth population within the mth river-basin, $f(p)_{k(l)}$ is the effect of the kth family within the lth population, b_j is the effect of the jth block, $(f \times b)_{kj}$ is the family per block interaction and ε_{ijklm} the residual. Overall genetic differentiation in quantitative traits (Q_{ST}) and quantitative genetic differentiation at different hierarchical levels (among populations within river basins and among river basins) were estimated from variance components for the four traits. To disentangle the effects of genetic drift from those of adaptive divergence within and among river basins, Q_{ST} estimates for each trait were compared with F_{ST} computed based on the same populations using the allozyme data set (7 loci, average of 60 diploid individuals per populations) in Alba (2000). Allozymes were preferred to nSSRs as their lower polymorphism makes them more suitable for unbiased F_{ST} estimation (Jost 2008). The bootstrap procedure outlined in Whitlock (2008) was used to test for $Q_{ST} > F_{ST}$, generating 1000 bootstraps for each statistic (over individuals for traits and over loci for allozymes). In addition, 95% confidence intervals of the bootstrap distributions of Q_{ST} and F_{ST} were compared.

Results

Haplotypes and haplotype networks

The chloroplast loci showed a total of 36 polymorphic sites, including three mononucleotide microsatellites (one within the *trnC-petN1* region), 13 SNPs, 17 indels and three short tandem repeats (STRs). Combining all polymorphisms, 54 haplotypes were resolved (Table S3, Supporting information): 36 for *Populus alba* (one shared with *Populus x canescens*), 2 exclusive for *P. x canescens* and 16 for *Populus nigra*. The highly polymorphic *trnC-petN1* region alone resolved 26 haplotypes, while the less variable *rpl16-poprpl* resolved only ten. Considering only UEPs, the number of haplotypes was reduced to 14 (Table S3, Supporting information). In terms of the full and UEP data sets, there were no shared haplotypes across species (Figs 1 and 2). However, cpSSRs alone were unable to distinguish among species, with *P. nigra* and *P. alba* sharing seven haplotypes. This fact highlights the limited value of cpSSRs for phylogenetic inference in poplars. Only three haplotypes had a wide distribution, h23, h37, h53, at frequencies of 0.054, 0.046 and 0.032, respectively. Interestingly, the most abundant haplotypes had a restricted distribution (h24, h25 and h41, at frequencies of 0.069, 0.107 and 0.058). Several haplotypes were confined to a single population (over 40% of private haplotypes for white and black poplars) or river basin (over 70% of private haplotypes for white and grey poplars).

Haplotype networks based on the full data set showed a great number of reticulations and could not be resolved. However, the reduced data sets based on

UEPs or SNPs yielded interpretable networks (Fig. 2). Many frequent white poplar haplotypes (H02-H05) were closely related to the widespread H01 that is found in both Atlantic and Mediterranean drainage basins. Some less abundant *P. alba* haplotypes (H06 and H07) lay close to *P. nigra* haplotypes and were exclusive of the southern part of the Atlantic drainage basin, probably indicating an older presence of white poplar in this range. Finally, we have to note one very divergent *P. nigra* haplotype (H14) and the wide separation (with intermingling *P. nigra* haplotypes) of (i) Iberian and North African (H08) *P. alba*; and (ii) *P. alba* and the closely related *P. tremula* (both in section *Populus*).

Time to the most recent common ancestor

The scaled TMRCA estimates from Bayesian inference (T, in coalescent units) in white poplar followed unimodal, asymmetric (gamma-like) distributions with averages of 3.31 for set1 (see Material and methods), of 2.19 for set2 and 1.91 for set3. $T \times \theta$ distributions (Fig. 3) were significantly different for the three data sets (Kolmogorov-Smirnov tests, $P < 0.01$) indicating divergence of Moroccan and Iberian white poplar lineages 1.93 times (CI: 0.74–4.13 at 95%) before the spread of Iberian lineages, and 4.32 times (CI: 1.71–8.81 at 95%) before the lineage radiation of the Mediterranean drainage basin in the IP. Considering the time of the last known direct communication between the IP and northern Africa (the Messinian Salinity Crisis that finished about 5.33 Ma; Duggen *et al.* 2003), the spread of Iberian haplotypes was dated to *c.* 2.76 Ma, and the divergence across Mediterranean lineages to 1.23 Ma. If we assume a generation time, g, for *P. alba* of 20–60 years, effective population sizes (N) are estimated to 21 000–63 000 for Iberian white poplar (10 667–32 000 for the Mediterranean range).

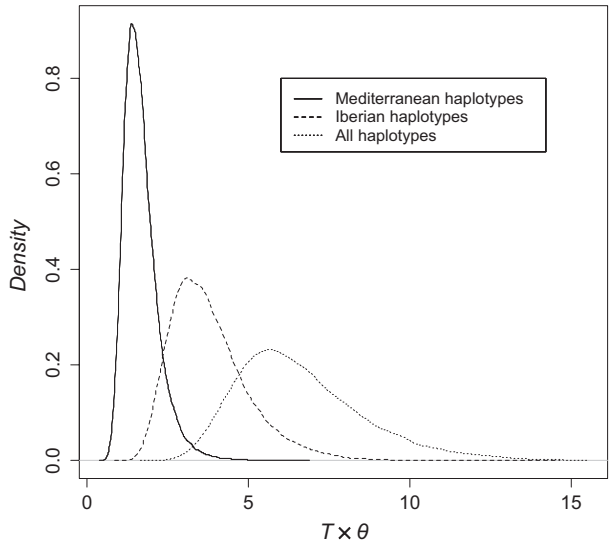


Fig. 3 Density plots of unscaled Time to the Most Recent Common Ancestor, TMRCA, obtained by coalescence simulations using BATWING (see text for details). Three different sets of runs are shown including (from right to left): (i) all white poplar haplotypes; (ii) only Iberian haplotypes; and (iii) haplotypes that are present in Mediterranean populations.

CpDNA diversity at different geographic scales

Overall, based on the full data set, haplotypic diversity per population was lower in *P. alba* (average of 0.317) than in *P. nigra* (average of 0.409) (Table S1, Supporting information). Sympatric populations often showed contrasting diversity levels for each species (e.g. Henares population, *P. alba*: 0.000, *P. nigra*: 0.600; see also Fig. S1, Supporting information) reflecting their different demographic history. Haplotypic diversity was extremely variable at the population level for both white and black poplars, and there were no significant differences between populations belonging to different drainage basins or latitudinal ranges (Kolmogorov-Smirnov tests, $P > 0.1$).

By contrast, genetic diversity at the regional level (*sensu* Hampe & Petit 2005; see Material and methods) was much greater for southern Iberian river basins than for northern basins in white poplar (Table 1). Similarly, after adjustment for uneven sample size via rarefaction (Petit *et al.* 1998), 26 haplotypes (21 private) were found in southern Iberian river basins compared to ten (five private) in the north. Finally, regional genetic diversity was similar for Mediterranean and Atlantic drainage basins (see Table 1), thus indicating a more important role of latitude than drainage basin for explaining the current standing genetic variation of Iberian white poplar. As black poplar is scarce in southern Iberia, comparative data are not available for this species.

CpDNA differentiation across drainage and river basins

In white poplar, genetic differentiation as estimated by F_{ST} and Jost's D was significant for almost all spatial scales (populations, river basins, latitudinal groups and drainage basins; see Table 2 and below for exceptions), with overall values of $F_{ST} = 0.670$ (0.735 for the UEP data set) and $D = 0.929$ (0.559 for the UEP data set, Table 2). The main factors causing genetic structure (based on the more reliable UEP data set) in this species were river and drainage basins with $F_{CT}/N_{CT}/D$ values of 0.320/0.223/0.511 and 0.374/0.260/0.569, respectively. In addition, judging by Jost's D, the river basins with the lowest (and non-significant) levels of internal differentiation were the northern Duero and Catalonia.

Table 1 Haplotypic genetic diversity and allelic richness in white and black poplars from the Iberian Peninsula; number of sampled individuals (n), number of haplotypes (A), number of haplotypes after rarefaction (A'), number of private haplotypes after rarefaction (A_P') and Nei's expected heterozygosity H_E (standard deviation). The minimum sample size in each category was used as reference for rarefaction

Species/group	n	A	A'	A_P'	H_E (SD)
White poplar					
River basin					
Duero	64	4	3.00	1.28	0.543 (0.059)
Catalonia	28	2	1.91	0.00	0.304 (0.094)
Ebro	86	7	4.40	1.12	0.733 (0.037)
Levante	91	15	6.63	3.89	0.899 (0.012)
Tajo	30	5	4.35	3.35	0.779 (0.040)
Guadiana	10	2	2.00	0.00	0.356 (0.159)
Guadalquivir	127	11	5.66	4.00	0.847 (0.016)
Latitude					
Northern	178	10	10.00	5.00	0.797 (0.018)
Southern	258	28	26.49	21.80	0.940 (0.005)
Drainage basin					
Atlantic	231	19	18.95	13.97	0.904 (0.008)
Mediterranean	205	19	19.00	14.00	0.863 (0.016)
Overall	436	33	26.64	11.65	0.939 (0.004)
*Black poplar**					
River basin					
Duero	10	3	2.60	0.60	0.644 (0.101)
Catalonia	24	5	3.21	1.04	0.721 (0.058)
Ebro	54	9	4.03	2.46	0.831 (0.024)
Levante	6	2	2.00	0.00	0.333 (0.215)
Tajo	30	5	3.14	0.50	0.674 (0.076)
Drainage basin					
Atlantic	40	7	7.00	3.00	0.767 (0.046)
Mediterranean	84	13	11.09	7.79	0.892 (0.012)
Overall	124	16	16.00	7.00	0.908 (0.009)

*Black poplar was not sampled in the southern Iberian Peninsula as it is relatively scarce in that region.

Latitudinal differentiation was not significant for F_{ST} or N_{ST} and very low for D. Finally, looking at the full data set, only five haplotypes (out of a total of 36) were shared across drainage basins and numbers of private haplotypes were very similar in each region (16 in the Atlantic vs. 14 in the Mediterranean).

In black poplar, patterns of genetic differentiation were less clear, probably due to reduced sampling (only northern Iberian populations) and higher human-mediated transfer of seeds and cuttings among populations (Galán-Cela *et al.* 2003). Despite overall genetic differentiation similar to white poplar (F_{ST} = 0.627 and D = 0.600 with the UEP data set), black poplar showed lower (and non-significant for F_{ST} or N_{ST}) genetic differentiation across drainage basins (0.130/0.032/0.432 for $F_{CT}/N_{CT}/D$) and inconsistent values for differentiation across river basins (low and non-significant for F- and N-statistics but relatively high for Jost's D; Table 2).

Isolation by distance (i.e. positive slopes) was found in white poplar, albeit with different strengths at different spatial scales. Regression slopes showed stronger (and significant) IBD among river basins than within them, highlighting the isolation effect produced by the dependence of white poplar on water courses (Table 3; see also Fig. S2, Supporting information). Within-basin IBD was found in the Duero when regressing on the logarithm of Euclidean distance and (marginally, $0.05 < P < 0.10$) in the Guadalquivir regressing on resistance distance. No IBD was found in the Ebro basin.

Levels of clonality and genetic structure at the local scale

In white poplar, the five highly polymorphic nSSRs identified 6–13 genets per population, with an average 4.2 ramets per genet (n/N_G, Table 4). Clone size in this species was highly variable (from a few metres to several kilometres). Three out of four white poplar populations had average clone sizes below 100 m. Larger clonal assemblies, with one of them extending over 15 km (Fig. 4), were identified and confirmed with 14 additional SSRs in Aranda de Duero from the northern Duero basin (a tundra-like area during glacial times). This population also contained a higher number of genets (13) than other populations (6–9), resulting in similar levels of (significant) overall fine-scale genetic structure (Table 4). All populations (including Aranda de Duero) had lower and non-significant kinship at >100 m distance classes (Table 4), suggesting that samples of individuals separated by >100 m for cpDNA analysis consisted largely of unrelated individuals.

The grey poplar population was characterized by just four genets with wide-ranging distances (up to 24 km) among ramets. This surprising clonal structure could be a consequence of propagation by farmers, as grey poplar is the only source of softwood in the region, and occurs mostly in managed environments (e.g. abandoned watermills, farms, etc.).

Genetic differentiation for quantitative traits in white poplar

Two of the four quantitative traits showed significant overall genetic differentiation: HT3 with Q_{ST} = 0.569 ± 0.149 (SD) and FOR3 with Q_{ST} = 0.696 ± 0.114 (SD). Both traits had over three to sixfold higher differentiation among basins (0.435 ± 0.195 and 0.592 ± 0.120, respectively) than within them (0.135 ± 0.046 and 0.104 ± 0.102, respectively). Populations from the Ebro basin had generally taller and straighter individuals. HT1 and DSB3 were not significantly different among river basins. Given that secondary growth in trees is

Table 2 Genetic differentiation among populations/groups at various hierarchical levels in white and black poplars from the Iberian Peninsula. Differentiation was measured considering haplotypic frequencies (*F*-statistics and Jost's *D*-statistics) or taking into account genetic distances among haplotypes (*N*-statistics). Estimates are provided for haplotypes resolved using the complete data set or, alternatively, using only UEPs (see text for details). All genetic differentiation estimates are significant at $\alpha = 0.05$ unless stated otherwise (ns). NA: not available or not possible to compute

		White poplar			Black poplar		
		Full set	UEPs		Full set	UEPs	
Group	Level*	*F*-statistics	*F*-statistics	*N*-statistics	*F*-statistics	*F*-statistics	*N*-statistics
River basin	F_{CT}	0.165	0.320	0.223	0.048^{ns}	0.170^{ns}	0.043^{ns}
	F_{SC}	0.616	0.632	0.586	0.524	0.568	0.312
	F_{ST}	0.679	0.750	0.678	0.547	0.642	0.341
Latitude	F_{CT}	0.090	-0.001^{ns}	0.028^{ns}	NA	NA	NA
	F_{SC}	0.653	0.735	0.660	NA	NA	NA
	F_{ST}	0.685	0.734	0.669	NA	NA	NA
Drainage basin	F_{CT}	0.082	0.374	0.260	0.043^{ns}	0.130^{ns}	0.032^{ns}
	F_{SC}	0.655	0.654	0.605	0.532	0.600	0.324
	F_{ST}	0.683	0.783	0.707	0.552	0.652	0.346
Overall		0.670	0.735	0.665	0.542	0.627	0.331
		Jost's *D*-statistics					
River basin	Among	0.892	0.511		0.773	0.586	
Duero	Within	0.353	0.008^{ns}		NA	NA	
Catalonia	Within	0.076^{ns}	0.000^{ns}		1.000	0.000^{ns}	
Ebro	Within	0.764	0.544		0.794	0.465	
Levante	Within	0.895	0.678		NA	NA	
Tajo	Within	0.985	0.031^{ns}		0.661	0.723	
Guadalquivir	Within	0.807	0.268		NA	NA	
Latitude	Among	0.896	0.052		NA	NA	
North	Within	0.759	0.534		NA	NA	
South	Within	0.935	0.576		NA	NA	
Drainage basin	Among	0.926	0.569		0.744	0.432	
Atlantic	Within	0.888	0.227		0.737	0.695	
Mediterranean	Within	0.846	0.549		0.892	0.501	
Overall		0.929	0.559		0.889	0.600	

*F_{CT} refers to genetic differentiation among groups (i.e. river basins, latitudes or drainage basins), F_{SC} to genetic differentiation among populations within groups and F_{ST} to genetic differentiation among populations without considering groups.

less important for early establishment than height differences, significant genetic differentiation for stem diameter may become apparent in later common garden assessments as trees mature. Interestingly, when compared to neutral markers for the same populations using Whitlock's (2008) approach (Table S5, Supporting information), Q_{ST} for FOR3 was significantly higher than F_{ST} among river basins (*P*-value: 0.033, with F_{ST} and Q_{ST} 95% confidence intervals of 0.011–0.238 and 0.337–0.774, respectively), but not among populations within them (*P*-value: 0.781, with F_{ST} and Q_{ST} 95% confidence intervals of 0.086–0.247 and 0.000–0.303, respectively). For HT3, a similar trend was observed (*P*-value for $Q_{ST} > F_{ST}$ among river basins: 0.151, *P*-value for $Q_{ST} > F_{ST}$ within river basins: 0.715), but a high Q_{ST} variance among river basins for this trait (Q_{ST} 95% confidence intervals of 0.065–0.733) rendered the comparison not significant.

Discussion

Haplotype networks and shared polymorphism across species

The paradoxical position of most black poplar haplotypes within the white poplar network and closeness to aspen is consistent with previous hypotheses of ancient hybridization followed by capture of *Populus alba*'s chloroplast by *Populus nigra* (Smith & Sytsma 1990). Hamzeh & Dayanandan (2004) observed a cyto-nuclear incongruence for the phylogenetic position of black poplar. They placed this species in the section *Populus*

Table 3 Isolation by distance (IBD) within and among river basins of white poplar from the Iberian Peninsula (see also Fig. S2, Supporting information). Pairwise genetic distances expressed as $F_{ST}/(1 - F_{ST})$ were regressed on the logarithm of the Euclidean distance. For main river basins, regression slopes with 'resistance' distances (i.e. geographic distances following the river course) are also shown; *$0.05 < P < 0.10$, **$0.01 < P < 0.05$, ***$P < 0.01$. NA: not available or not possible to compute

	Regression slopes	
	log (Euclidean distance)	Resistance distance
Different basins	0.86**	NA
Same basin	0.19	NA
Duero	0.93**	0.3E-05
Ebro	−1.24	1.8E-05
Guadalquivir	0.26	0.3E-05*
Overall	0.79***	NA

close to *P. alba* and *P. tremula* on the basis of cpDNA, but in the section *Aigeiros* on nuclear DNA evidence. Its inclusion in *Aigeiros* is in agreement with classical studies based on morphology (Eckenwalder 1996) and was also supported by nuclear AFLP markers (Cervera *et al.* 2005). Haplotype sharing has been widely described in sympatric, related tree species [for instance in European ashes (Heuertz *et al.* 2006) or in white oaks (Petit *et al.* 2002)]. In species that hybridize readily, such as ashes and oaks, haplotype sharing commonly occurred during hybridization events in shared glacial refugia and post-glacial recolonization (e.g. Petit *et al.* 2002) and is maintained by recurrent interspecific gene flow (Lexer *et al.* 2006). In contrast, in species that do not currently hybridize but that may have hybridized in the past, relatively recent reproductive isolation would result in a progressive loss of shared haplotypes while retaining close phylogenetic relationships. Our results in *P. alba* and *P. nigra* are in agreement with the second explanation. The highly divergent *P. nigra* haplotype H14 (differing by 12 mutations from the closest haplotype in the network) could then be more closely related to the genuine, pre-introgressed *P. nigra* plastid genome. Alternatively, haplotype introgression from commercial Euroamerican clones (Vanden Broeck *et al.* 2006) or from the ornamental Lombardy cultivar (Chenault *et al.* 2011) has been shown for *P. nigra*. Sequencing of a diverse array of commercial clones ($n = 14$) found H14 among them (not shown), pointing to introgression, despite being generally rare (<5%; Heinze & Lickl 2002 and references therein; but see Ziegenhagen *et al.* 2008 and Smulders *et al.* 2008a that reported *c.* 20–50% introgressed offspring in Elbe and Rhine rivers, respectively), as the most-likely explanation.

As DNA sequences revealed a complete segregation among species, shared variants at *ccmp2*, *ccmp5* and the *trnC-petN1* microsatellite are probably due to homoplasy rather than ancient hybridization. Microsatellites usually have higher mutation rates than other regions of the genome (reflected by a higher number of variants

Table 4 Number and size of clonal assemblies, and fine-scale spatial genetic structure for four white poplar and one grey poplar populations. All clone sizes are given as maximum among-ramet distance in metres; max (*L*): size of the largest clone, *L:* mean clone size, min (*L*): size of the smallest clone, *n:* number of samples and N_G: number of genets; standard errors of the regression slope (SE) are computed by a jackknife resampling procedure. NA: not available or not possible to compute

							Average kinship (F_{ij}) by distance class					Slope	
Population	River basin	Max (*L*)	*L*	Min (*L*)	*n*	N_G	0–25	25–50	50–100	100–200	>200	*b**	SE
White poplar													
Aranda de Duero	Duero	558.37†	163.58†	11.40†	36	13	0.324	0.272	0.209	0.074	−0.007	−1.2E-05	3.1E-06
La Alfranca	Ebro	32.32	25.69	5.77	49	9	0.386	0.164	0.038	−0.028	0.082	−3.6E-04	1.2E-04
Jimena	Guadalquivir	62.61	30.30	9.90	33	6	0.421	0.340	0.209	−0.078	0.114	−1.1E-03	2.8E-04
Villamanrique de la Condesa‡	Guadalquivir	NA	NA	NA	32	8	NA	NA	NA	NA	NA	NA	NA
Overall		558.37	86.74	5.77	150	36	0.386	0.245	0.115	0.038	0.092	−8.2E-06	2.9E-06
Grey poplar													
Montes de Toledo	Tajo	23749.85	13641.84	211.24	50	4	0.267	0.286	0.514	0.516	0.198	−1.6E-05	2.1E-06

*All slopes are significantly different from zero with $P < 0.001$, as tested by permutation.
†Excluding one very large and widespread clone (four ramets stretched over 15 km; see Results).
‡Spatial coordinates were not available for this population.

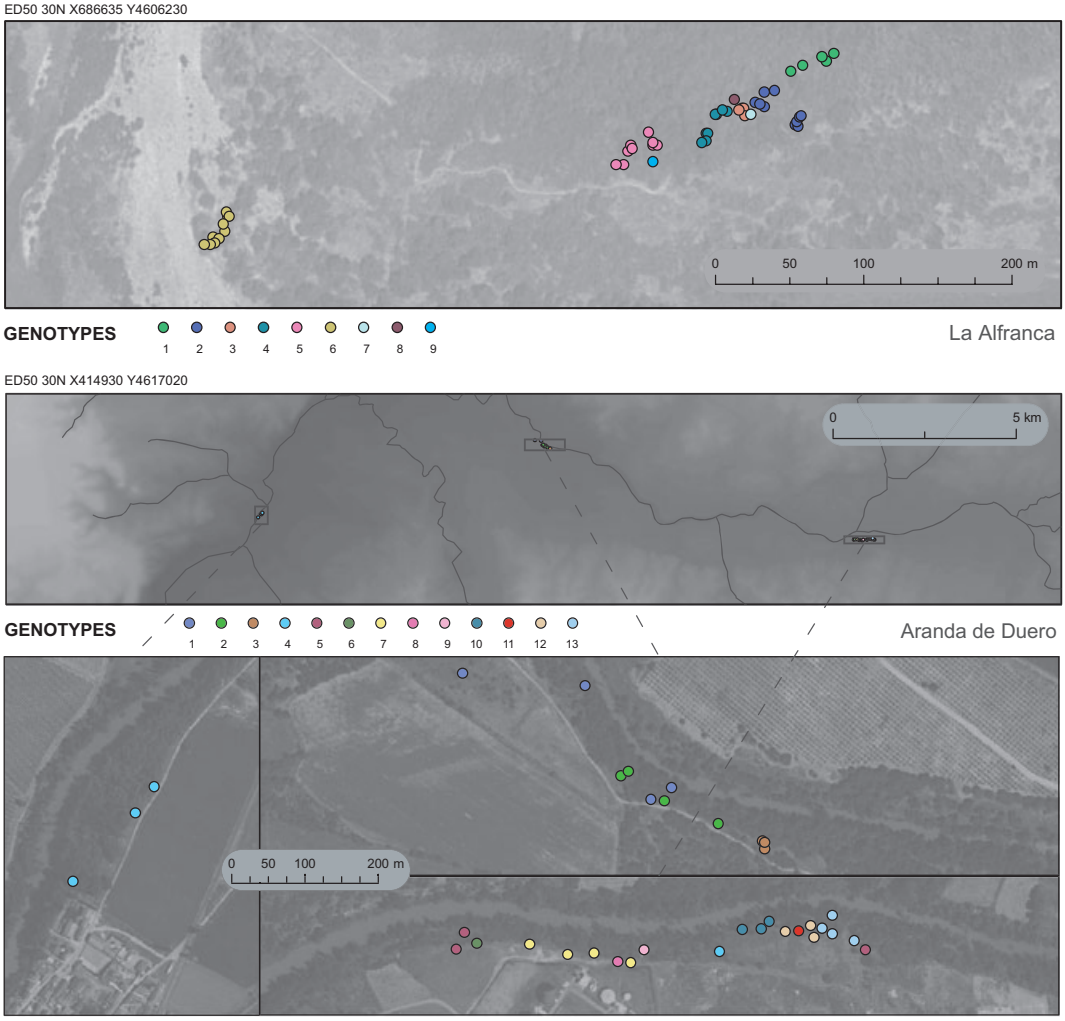


Fig. 4 Spatial distribution of clonal replicates from two contrasting white poplar populations. The La Alfranca population (top) had smaller and less spread clonal assemblies than the Aranda de Duero population (bottom) that includes a clone stretching over *c.* 15 km.

in this study), making them more prone to homoplasy (Provan *et al.* 2001). This condition makes them useful for local and contemporary studies, especially those where high levels of variability are desirable (e.g. parentage analysis), but discourages their use for phylogeographical inference in poplars.

The noteworthy lack of shared, or even closely-related, haplotypes between *P. alba* and *P. tremula* in the IP contrasts with expectation, as these species often hybridize and they are largely sympatric in this range. However, a larger sampling of *P. tremula* should be carried out to confirm this observation. Finally, the pronounced divergence of the Moroccan endemic H08 from Iberian haplotypes indicates an ancient divergence of North African and Iberian lineages (see below), as previously noticed by Fussi *et al.* (2010) based on a limited sample of PCR-RFLP haplotypes.

Different species, different histories

Black poplar, while showing a similar degree of overall genetic differentiation as white poplar, displayed higher regional haplotypic genetic diversity and lower levels of population genetic structure among river and drainage basins. These results indicate that the two species have different demographic histories. In particular, they point to more frequent gene exchange across river and drainage basins, and/or generally higher effective populations sizes in black poplar, which is to be expected in view of its higher tolerance to cold temperatures (Galán-Cela *et al.* 2003; Ruiz de la Torre 2006) and previous literature (e.g. Smulders *et al.* 2008b). Alternatively, the lower genetic structure in black poplar could reflect a higher seed and cutting transfer by humans across regions. However, the high number of private

haplotypes found in the IP in this (*c.* 40%) and in other studies (e.g. Cottrell *et al.* 2005) does not support this alternative hypothesis.

Similar patterns of high diversity and low differentiation have been observed in the more cold-tolerant species of other European tree genera, such as the six native Iberian pine species (Soto *et al.* 2010), or in the thermophilous *Fraxinus angustifolia* and the more cold-tolerant *F. excelsior* (Heuertz *et al.* 2006). The bases of these patterns are likely to be better survival of the cold-tolerant species during the cold stages of past glaciations and early colonization of new territory, compared to thermophilous tree species. Our findings are relevant because they extend these observations to riparian trees that are normally not considered to be dependent on regional climatic patterns and, thus, are excluded from models of future species distributions based on climate predictions (e.g. Benito-Garzón *et al.* 2008 for Iberian trees). Moreover, our findings suggest that near-future predicted climatic change may affect Iberian poplar species differentially, giving a competitive advantage to the more drought- and salt-tolerant white poplar compared to black poplar. Competitive exclusion from the already scarce riparian habitat would possibly drive this already threatened species (Lefèvre *et al.* 1998 and references therein) to lower effective population numbers and, eventually, to local extinctions.

Genetic signatures of ancient events in white poplar

The genus *Populus* appeared during the transition from the Secondary to the Tertiary era and diversified into different sections and species during the warm Paleogene period (Eckenwalder 1996). Modern species are believed to have evolved during the global cooling in the beginning of the Neogene (c. 23 Ma). During this period, still warmer and wetter than today, and before the beginning of the Quaternary, modern poplars would have spread across the IP and northern Africa. The North African and Iberian lineages of white poplar would have diverged after their last possible contact at the end of the Messinian Salinity Crisis *c.* 5.33 Ma, when the Mediterranean Sea was desiccated (Krijgsman *et al.* 1999; Duggen *et al.* 2003). The subsequent flooding of the Mediterranean basin has been associated to a genetic discontinuity at the level of the Strait of Gibraltar in various organisms (Rodríguez-Sánchez *et al.* 2008; see Jaramillo-Correa *et al.* 2010 for some forest trees).

Within the IP, a marked differentiation between the Atlantic and Mediterranean drainage basins was found in white poplar. This pattern has also been found in other Iberian tree species (Rodríguez-Sánchez *et al.* 2010). This disjunction probably reflects a genetic signature of ancient geological events, considering that the main Iberian mountain systems attained their current configuration during the late Miocene. The fact that *F*-statistics using the UEP data indicated stronger differentiation (0.374 vs. 0.082) than using the complete polymorphism set (assumed to be affected by recent mutation) also pointed at ancient phylogeographic processes. Consistent with its lower sensitivity to mutation, Jost's *D* statistic did not reflect these differences.

The reasons for the significant differentiation between drainage basins (Atlantic vs. Mediterranean) are not obvious, considering that major Iberian mountain systems run from west to east, thus mostly preventing latitudinal migration (i.e. among river basins but not between drainage areas). One explanation that can apply to plants, and more specifically to riparian trees, is related to the vegetation altitudinal shifts produced by glacial climatic oscillations (Hewitt 1996; Rodríguez-Sánchez *et al.* 2010). The relatively benign climate before the Pleistocene should have favoured extensive gene exchange across drainage basins, even for lowland species. Then, during the Pleistocene cold periods, altitudinal limits for plant species lowered and riparian trees probably became confined to the lower river courses. This process resulted in distributions close to the western and eastern coastal fringes of the IP where the main Iberian rivers flow into the sea. In this way, the Atlantic and Mediterranean drainage basins would have become effectively separated while migration along the coastlines (where mountain ranges are lower) would have connected river basins. Our results in white poplar suggest that increasing isolation between Atlantic and Mediterranean drainage basins occurred *c.* 1.12 Ma (lower bound of 3.11 Ma), in agreement with the proposed scenario related to Pleistocene cooling.

Regional and population effects of glacial times in white poplar

The patterns of genetic diversity and structure in white poplar reflect the effects of Pleistocene climatic oscillations in several ways. First, regional genetic diversity was higher and private haplotypes were four times more abundant in the southern Iberian river basins, which were warmer than the northern basins. Secondly, genetic structure among populations was much more pronounced in the southern Guadalquivir and Levante basins than in the northern Duero and Catalonia basins. Thirdly, in the formerly tundra-like Duero basin, significant IBD was found only when considering Euclidean geographical distances, but not 'resistance' distances. This IBD pattern is more consistent with a rapid isotropic postglacial spread than with a long-term build-up of SGS along linear favourable environments. Fourthly,

clonal assemblies were apparently larger (with some clones extending up to *c.* 15 km) in the colder Duero basin than in the southern basins. Asexual propagation could have helped Iberian poplars to survive in harsh glacial environments and to colonize new territory rapidly once ecological conditions improved (Silvertown 2008).

Pleistocene glacial oscillations lowered temperature and humidity globally. Palaeoecological inferences indicate that during the glacial maxima, areas in the western- and northernmost IP (like the Duero basin) were barely habitable by arboreal vegetation (González-Sampériz *et al.* 2010). Northern populations of the thermophilous white poplar show a genetic depauperation that seems to reflect these past events. The hostile climatic conditions suffered during Pleistocene glacial periods in these areas could have pushed white poplar populations towards one of two fates: (i) an important population size reduction, but persistence in sheltered 'cryptic refugia' (sensu Stewart & Lister 2001); or (ii) local extinction followed by postglacial recolonization. The first scenario would have resulted in reduced genetic diversity but would have maintained common local haplotypes in surviving populations (Provan & Bennett 2008). In the second situation, diversity would have been reduced due to founder effects, and the region would have been replenished with (non-local) haplotypes from the colonizing populations. Our data support the first scenario, showing significant genetic differentiation among northern and southern river basins and presence of private haplotypes in both latitudes. The existence of cryptic local refugia and recent spread of surviving genotypes is also a plausible explanation for the high haplotypic diversity observed in white poplar populations from Austria (Fussi *et al.* 2010) and the discovery of huge clonal assemblies of the species in Sardinia and Malta (Brundu *et al.* 2008; Fussi *et al.* 2012). Signals of glacial survivorship in scattered populations situated beyond the estimated persistence limit have been widely observed in boreal and alpine latitudes (Hewitt 2004; Opgenoorth *et al.* 2010), including for some Salicaceae (e.g. Palmé *et al.* 2003 for *Salix* sp.).

Evidence for local adaptation in white poplar

Several decades of common garden experiments have revealed the widespread occurrence of locally adapted populations in forest trees (see reviews in McKay & Latta 2002; Latta 2004; Savolainen *et al.* 2007), including some Populus species (see Fig. 4 in Savolainen *et al.* 2007 for *Populus balsamifera* and *Populus tremuloides*). The higher quantitative (Q_{ST}) than molecular (F_{ST}) genetic differentiation found across river basins, albeit not within river basins, for some growth traits in white poplar suggests that this species is also locally adapted, but at wider spatial scales (i.e. river basins that can span hundreds of kilometres) than in other temperate trees. White poplar populations have typically low population sizes, given their dependence on phreatic water and the high anthropization of Iberian riparian habitats, which breaks the continuity of riparian forests. Human impact is most noteworthy in the low- and medium water courses where white poplar is more abundant (Ruiz de la Torre 2006). Our findings suggest the role of gene flow over mesoscale distances, replenishing genetic variation and counteracting local genetic drift. Lack of IBD patterns also suggests frequent gene exchange along river courses within basins, at least for those Iberian rivers where glacial impact was low (see above). In this scenario, the relative homogeneity of riparian habitats would have counteracted the arrival of maladapted genotypes, preventing the development of 'migration meltdown' processes (i.e. self-reinforced processes in which immigration of maladapted genotypes decreases local density, which in turn increases immigration rates bringing in more maladapted genotypes; Lenormand 2002), which can eventually result in population extinction. Theoretical models have shown the potential beneficial effects of gene flow for small peripheral populations (Alleaume-Benharira *et al.* 2006), and experimental evidence is accumulating (e.g. Sexton *et al.* 2011).

The existence of local adaptation and specialized phenotypes has direct consequences for the adaptive response of white poplar to future environments, such as those predicted by the Intergovernmental Panel on Climatic Change (IPCC, http://www.ipcc.ch, accessed on May, 2012). Indeed, past adaptation processes have most likely generated a wide array of standing genetic variation that may prove of utility beyond the current range of the species. This expectation highlights the need for a wider exploration of genetic resources in this species as well as for the establishment of large multisite common gardens.

Conclusions

Past climate conditions have left genetic signatures in riparian tree species, such as the Iberian poplars. Some of these signatures reflect early Pleistocene events that led to differentiation of gene pools in the Mediterranean and Atlantic drainage basins. Genetic diversity is higher and genetic differentiation lower in cold-tolerant black poplar than in thermophilous white poplar, and we speculate that cold-tolerance resulted in better survival and higher gene exchange across geographical barriers of this species during past glaciations, as shown elsewhere for other cold-tolerant trees. Patterns of IBD in white poplar reflect its dependence on phreatic water, resulting in higher IBD among river basins than within.

At the local scale, SGS is greatly influenced by the widespread existence of clonal assemblies extending, in a few cases, up to several kilometres. Nevertheless, the presence of numerous genets of small clone size points out to asexual propagation as a means for maintaining genetic diversity under harsh environments (rather than reducing the effective population size) and for colonizing new territory rapidly. Asexual propagation did not seem to prevent local adaptation in white poplar. Gene flow at mesoscale distances seems sufficient to counteract genetic drift and to promote local adaptation at the river-basin scale. Riparian trees occupy very specialized niches surrounded by large inhospitable areas. The existence of locally adapted phenotypes, even in highly structured species such as the Iberian poplars, is remarkable and suggests some resilience of poplar populations to environmental change and a capacity to adapt when confronted with new environments.

Acknowledgements

We are indebted to José Antonio Mancha, Santiago de Blas and Fernando del Caño for help in sample collection, and Carmen García Barriga and María Teresa Cervera for lab assistance and advice. We are also grateful to P. C. Grant for reviewing the English language. This work was supported by project REPROFOR (AGL2005-07440-C02-01, Ministry of Education and Science). The contribution of the EU EVOLTREE Network of Excellence and the Collaborative Project on 'Conservation of Forest Genetic Resources' between the MARM and the National Institute for Agriculture and Food Research and Technology, INIA (AEG06-054) is also acknowledged. We want to specially acknowledge the contribution of Nuria Alba to this study. She contributed to the collection of the samples of *Populus alba*, the isozyme analysis and the quantitative evaluation. She passed away on January 2nd, 2012.

References

Alba N (2000) *Variabilidad genética de Populus alba* L. mediante caracteres isoenzimáticos y fenotípicos. PhD Dissertation, Polytechnic University of Madrid, Madrid.

Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, **19**, 203–215.

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12447–12450.

Benito-Garzón M, Sánchez de Dios R, Sáinz Ollero H (2008) Effects of climate change on the distribution of Iberian tree species. *Applied Vegetation Science*, **11**, 169–178.

Brundu G, Lupi R, Zapelli I *et al.* (2008) The origin of clonal diversity and structure of *Populus alba* in Sardinia: evidence from nuclear and plastid microsatellite markers. *Annals of Botany*, **102**, 997–1006.

Bryan GJ, McNicoll J, Ramsay G, Meyer RC, De Jong WS (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theoretical and Applied Genetics*, **99**, 859–867.

Cervera MT, Storme V, Soto A *et al.* (2005) Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. *Theoretical and Applied Genetics*, **111**, 1440–1456.

Cheddadi R, Vendramin GG, Litt T *et al.* (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, **15**, 271–282.

Chenault N, Arnaud-Haond S, Juteau M *et al.* (2011) SSR-based analysis of clonality, spatial genetic structure and introgression from the Lombardy poplar into a natural population of *Populus nigra* L. along the Loire River. *Tree Genetics & Genomes*, **7**, 1249–1262.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cottrell JE, Krystufek V, Tabbener HE *et al.* (2005) Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA. *Forest Ecology and Management*, **206**, 71–90.

Dorman CE, Beardsley RC, Limeburner R (1995) Winds in the Strait of Gibraltar. *Quarterly Journal of the Royal Meteorological Society*, **121**, 1903–1921.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Duggen S, Hoernle K, van den Bogaard P, Rüpke L, Morgan JP (2003) Deep roots of the Messinian salinity crisis. *Nature*, **422**, 602–606.

Eckenwalder JE (1996) Systematics and evolution of *Populus*. In: *Biology of Populus and Its Implications for Management and Conservation* (eds Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM), pp. 7–32. NRC Research Press, Ottawa, Ontario.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Fussi B, Lexer C, Heinze B (2010) Phylogeography of *Populus alba* L. and *Populus tremula* L. in Central Europe: secondary contact and hybridisation during recolonisation from disconnected refugia. *Tree Genetics & Genomes*, **6**, 439–450.

Fussi B, Bonello J, Calleja E, Heinze B (2012) Combining the use of molecular techniques and archival documentary evidence to trace the origin of *Populus alba* in a Central Mediterranean archipelago. *European Journal of Forest Research*, **131**, 347–354.

Galán-Cela P, Gamarra-Gamarra R, García-Viñas JI (2003) *Árboles y arbustos de la Península Ibérica e Islas Baleares*, 4th edn. Eds Jaguar, Madrid.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on GST and D: forget GST but not all of statistics!. *Molecular Ecology*, **19**, 3845–3852.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity* (eds Weiss S, Ferrand N), pp. 155–188. Springer, Dordrecht, The Netherlands.

González-Martínez SC, Dubreuil M, Riba M *et al.* (2010) Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: past and present limits to gene movement over a broad geographic scale. *Molecular Phylogenetics and Evolution*, **55**, 805–815.

González-Sampériz P, Leroy SAG, Carrión JS *et al.* (2010) Steppes, savannahs, forests and phytodiversity reservoirs during the Pleistocene in the Iberian Peninsula. *Review of Palaeobotany and Palynology*, **162**, 427–457.

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734–4756.

Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461–467.

Hamzeh M, Dayanandan S (2004) Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast *trnT-trnF* region and nuclear rDNA. *American Journal of Botany*, **91**, 1398–1408.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Heinze B, Lickl E (2002) Rare, but steady, introgression in Austrian black poplar as a long-term risk? In: *Genetic Diversity in River Populations of European Black Poplar* (eds Van Dam B, Bordacs S), pp. 169–175. Csiszár Nyomda Ltd, Budapest.

Heuertz M, Carnevale S, Fineschi S *et al.* (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): roles of hybridization and life history traits. *Molecular Ecology*, **15**, 2131–2140.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 183–195.

Huntley B, Birks H (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13000 Years Ago*. Cambridge University Press, Cambridge.

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, **180**, 329–340.

Jaramillo-Correa JP, Grivet D, Terrab A *et al.* (2010) The Strait of Gibraltar as a major biogeographic barrier in Mediterranean conifers: a comparative phylogeographic survey. *Molecular Ecology*, **19**, 5452–5468.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS (1999) Chronology, causes and progression of the Messinian salinity crisis. *Nature*, **400**, 652–655.

Latta RG (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist*, **161**, 51–58.

Lefèvre F, Legionnet A, Vries S, Turok J (1998) Strategies for the conservation of a pioneer tree species, *Populus nigra* L., in Europe. *Genetics Selection Evolution*, **30**, S181–S196.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology*, **14**, 1045–1057.

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, **15**, 2007–2012.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Magri D, Fineschi S, Bellarosa R *et al.* (2007) The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Molecular Ecology*, **16**, 5259–5266.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, **17**, 285–291.

Medail F, Diadema K (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, **36**, 1333–1345.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal colonization. *Heredity*, **72**, 312–317.

Nieto-Feliner G (2011) Southern European glacial refugia: a tale of tales. *Taxon*, **60**, 365–372.

Opgenoorth L, Vendramin GG, Mao K *et al.* (2010) Tree endurance on the Tibetan Plateau marks the world's highest known tree line of the Last Glacial Maximum. *New Phytologist*, **185**, 332–342.

Palmé AE, Semerikov V, Lascoux M (2003) Absence of geographical structure of chloroplast DNA variation in sallow, *Salix caprea* L. *Heredity*, **91**, 465–474.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Petit RJ, Csaikl UM, Bordacs S *et al.* (2002) Chloroplast DNA variation in European white oaks – phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5–26.

Petit RJ, Hu FS, Dick CW (2008) Forests of the past: a window to future changes. *Science*, **320**, 1450–1452.

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**, 564–571.

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*, **16**, 142–147.

Rodríguez-Sánchez F, Perez-Barrales R, Ojeda F, Vargas P, Arroyo J (2008) The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews*, **27**, 2100–2117.

Rodríguez-Sánchez F, Hampe A, Jordano P, Arroyo J (2010) Past tree range dynamics in the Iberian Peninsula inferred through phylogeography and palaeodistribution modelling: a review. *Review of Palaeobotany and Palynology*, **162**, 507–521.

Roiron P, Ali AA, Guendon JL, Carcaillet C, Terral JF (2004) Evidence of the indigenous status of *Populus alba* L. in the western Mediterranean Basin. *Comptes Rendus Biologies*, **327**, 125–132.

Ruiz de la Torre J (2006) *Flora Mayor*. Organismo Autónomo Parques Nacionales, Dirección General para la Biodiversidad, Madrid.

Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics*, **38**, 595–619.

Sexton JP, Strauss SY, Rice KJ (2011) Gene flow increases fitness at the warm edge of a species' range. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11704–11709.

Silvertown J (2008) The evolutionary maintenance of sexual reproduction: evidence from the ecological distribution of asexual reproduction in clonal plants. *International Journal of Plant Sciences*, **169**, 157–168.

Slavov GT, Zhelev P (2010) Salient biological features, systematics and genetic variation of Populus. In: *Genetic and Genomics of Populus* (eds Jansson S, Bhalerao RP, Groover AT), pp. 15–38. Springer, Dordrecht, The Netherlands.

Smith RL, Sytsma KJ (1990) Evolution of *Populus nigra* (sect *Aigeiros*) – introgressive hybridization and the chloroplast contribution of *Populus alba* (sect *Populus*). *American Journal of Botany*, **77**, 1176–1187.

Smulders MJM, Beringen R, Volosyanchuk R *et al.* (2008a) Natural hybridisation between *Populus nigra* L. and *P. x canadensis* Moench: hybrid offspring competes for niches along the Rhine river in the Netherlands. *Tree Genetics & Genomes*, **4**, 663–675.

Smulders MJM, Cottrell J, Lefèvre F *et al.* (2008b) Structure of the genetic diversity in black poplar (*Populus nigra* L.) populations across European river systems: consequences for conservation and nature development. *Forest Ecology and Management*, **255**, 1388–1399.

Soto A, Robledo-Arnuncio JJ, González-Martínez SC, Smouse PE, Alia R (2010) Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. *Molecular Ecology*, **19**, 1396–1409.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Stuefer JF, Erschbamer B, Huber H, Suzuki J-I (2002) *Ecology and Evolutionary Biology of Clonal Plants*. Kluwer Academic Pub. Group, Dordrecht, The Netherlands.

Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, **2**, 377–379.

Vanden Broeck A, Cottrell J, Quataert P *et al.* (2006) Paternity analysis of *Populus nigra* L. offspring in a Belgian plantation of native and exotic poplars. *Annals of Forest Science*, **63**, 783–790.

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42**, 9–19.

Whitlock MC (2008) Evolutionary inference from Q_{ST}. *Molecular Ecology*, **17**, 1885–1896.

Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **166**, 155–188.

Ziegenhagen B, Gneuss S, Rathmacher G *et al.* (2008) A fast and simple genetic survey reveals the spread of poplar hybrids at a natural Elbe river site. *Conservation Genetics*, **9**, 373–379.

D.M.S. develops a PhD on genetic structure, including clonal structure, and hybridization of Iberian poplars. M.H. is interested in population and evolutionary genetics of plant species from biodiversity hotspots at temperate and tropical latitudes. U.L.H. has broad interests in phylogeography and ecological genetics of forest trees. A.I.L. and E.H's research focuses on the study of genetic diversity and population structure, and the molecular characterization of commercial woody plants. C.M. and A.P. are interested in conservation and use of plant genetic resources, in particular those of poplars and other trees. R.A. and S.C.G.M. have broad interests in population genetics and genomics of forest trees, ecological genetics and the evolution of Mediterranean plants, using quantitative and molecular genetics approaches.

Data accessibility

GenBank accessions for cpDNA haplotypes: JQ782847–JQ782883 (see Table S3, Supporting information for correspondence between accession numbers and haplotypes and Table S4, Supporting information for haplotype counts per population).

Chloroplast microsatellite data deposited in the Dryad repository doi:10.5061/dryad.9hd71135.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Population names, location and basic description, including expected heterozygosity estimates (H_E).

Table S2 Annealing temperatures, number of PCR cycles and source of nuclear microsatellites.

Table S3 Haplotype definition.

Table S4 Haplotype counts per population. Full haplotypes are coded with '*h*' and UEP haplotypes with '*H*'.

Table S5 Test for $Q_{ST} > F_{ST}$ following Whitlock (2008) for total height (HT3) and stem form (FOR3) at age 3.

Fig. S1 Distribution of haplotypic diversity in Iberian poplars calculated as Nei's unbiased expected heterozygosity (H_E) for the full data set of cpSSRs and cpDNA sequences.

Fig. S2 Scatter plots of pairwise genetic distances expressed as $F_{ST}/(1 - F_{ST})$ against the logarithm of the Euclidean distance within and among river basins of white poplar from the Iberian Peninsula.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Annex B

Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree.

DE CARVALHO, D., INGVARSSON, P. K., JOSEPH, J., SUTER, L., SEDIVY, C., **MACAYA-SANZ, D.**, COTTRELL, J., HEINZE, B., SCHANZER, I, LEXER, C., 2010. Molecular Ecology **19**: 1638-1650.

MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 1638–1650 doi: 10.1111/j.1365-294X.2010.04595.x

Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree

DULCINEIA DE CARVALHO,*‡ PÄR K. INGVARSSON,† JEFFREY JOSEPH,‡ LEONIE SUTER,‡ CLAUDIO SEDIVY,‡ DAVID MACAYA-SANZ,‡§ JOAN COTTRELL,¶ BERTHOLD HEINZE,** IVAN SCHANZER†† and CHRISTIAN LEXER‡ ‡‡

**Universidade Federal de Lavras, Caixa Postal 3037 DCF/UFLA, 37200-000 Lavras, MG, Brazil, †Umeå Plant Science Centre, Department of Ecology and Environmental Science, Umeå University, Linneaus väg 6, SE-90187 Umeå, Sweden, ‡Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW 3DS, UK, §CIFOR-INIA, Ctra. de La Coruña, Km. 7.5, 28040 Madrid, Spain, ¶Forest Research, Northern Research Station, Roslin, Midlothian EH25 9SY, UK, **Federal Research Centre for Forests, Department of Genetics, Hauptstraße 7, A-1140 Vienna, Austria, ††Herbarium, Main Botanical Garden, Russian Academy of Sciences, Botanicheskaya Str. 4, 127276 Moscow, Russia, ‡‡Unit of Ecology and Evolution, Department of Biology, University of Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland*

Abstract

Adaptation to new environments can start from new mutations or from standing variation already present in natural populations. Whether admixture constrains or facilitates adaptation from standing variation is largely unknown, especially in ecological keystone or foundation species. We examined patterns of neutral and adaptive population divergence in *Populus tremula* L., a widespread forest tree, using mapped molecular genetic markers. We detected the genetic signature of postglacial admixture between a Western and an Eastern lineage of *P. tremula* in Scandinavia, an area suspected to represent a zone of postglacial contact for many species of animals and plants. Stringent divergence-based neutrality tests provided clear indications for locally varying selection at the European scale. Six of 12 polymorphisms under selection were located less than 1 kb away from the nearest gene predicted by the *Populus trichocarpa* genome sequence. Few of these loci exhibited a signature of 'selective sweeps' in diversity-based tests, which is to be expected if adaptation occurs primarily from standing variation. In Scandinavia, admixture explained genomic patterns of ancestry and the nature of clinal variation and strength of selection for bud set, a phenological trait of great adaptive significance in temperate trees, measured in a common garden trial. Our data provide a hitherto missing direct link between past range shifts because of climatic oscillations, and levels of standing variation currently available for selection and adaptation in a terrestrial foundation species.

Keywords: adaptive divergence, admixture, genome scan, photoperiod, selective sweep, standing genetic variation

Received 19 August 2009; revision received 28 January 2010; accepted 5 February 2010

Introduction

The role of gene flow in adaptive evolution is a hotly debated topic in evolutionary biology. At the *within*-species level, gene flow has traditionally been seen as a homogenizing force that impedes local adaptation (Stearns & Hoekstra 2005), whereas others have stressed its role in contributing alleles to the standing variation available for local adaptation (Przeworski *et al.* 2005; Pennings & Hermisson 2006), or in facilitating species cohesion through the spread of beneficial alleles

Correspondence: Christian Lexer, Fax: +41 26 300 9698; E-mail: christian.lexer@unifr.ch

(Rieseberg & Burke 2001). The role of gene flow at the *between*-species level is no less controversial, opinions ranging from 'evolutionary noise' to 'potent evolutionary force that creates opportunities for adaptive evolution and speciation' (Schemske 2000; Wu 2001; Rieseberg *et al.* 2003; Arnold 2006). Considering the intensity of the debate, surprisingly little attention has been paid to the *intermediate* scenario of gene flow between previously isolated conspecific populations, i.e. lineages that were isolated long enough for adaptive differentiation to arise but over too short time for the evolution of substantial reproductive isolation. This is expected to be the case for many species of the temperate zones, i.e. those regions of the world most strongly affected by the climatic shifts of the pleistocene and holocene (Hewitt 2000).

Cycles of genetic divergence and subsequent gene flow between temporarily isolated populations have been predicted to result in increased standing variation available for adaptive evolution (Hewitt 2000; Pennings & Hermisson 2006; Barrett & Schluter 2008), a hypothesis also supported by recent empirical work on fishes (Colosimo *et al.* 2005), mice (Mullen & Hoekstra 2008) and plants (Whibley *et al.* 2006). This scenario is particularly likely for long-lived organisms, such as trees; in temperate forest trees, rapid diversifying selection (facilitated by great environmental heterogeneity and large effective population size, N_e) and low nucleotide substitution rates per unit time suggest that amounts of standing variation available for local adaptation are often limiting (Petit & Hampe 2006; Savolainen & Pyhajarvi 2007). In contrast, admixture may also constrain adaptation via outbreeding depression or tradeoffs with negative effects on fitness (Stearns & Hoekstra 2005). Limited empirical evidence is currently available in terrestrial species for how admixture between divergent postglacial lineages affects adaptation from standing genetic variation.

The effects of gene flow between populations with varying degrees of isolation can be described in terms of geographic clines for phenotypes or allelic frequencies in space (Barton & Hewitt 1985), or in terms of single-locus clines against genome-wide admixture gradients, estimated by comparing the ancestry of individual loci to expectations obtained from many loci in the genome (Briscoe *et al.* 1994; Reich *et al.* 2005; Gompert & Buerkle 2009). More experience exists for geographic clines, but admixture-based analysis can help pinpoint the exact target loci and nature of selection acting in admixed populations (Lexer *et al.* 2007; Gompert & Buerkle 2009), a task that is not easily achieved by studying the shapes of geographic clines alone (Kruuk *et al.* 1999).

Populus tremula L. is a deciduous forest tree whose distribution ranges from the British Isles to the east tip of Asia and from Northern Africa to Northern Scandinavia. Its mating system is dioecious and both pollen and seeds are wind-dispersed, resulting in substantial levels of neutral variability (scaled synonymous mutation rate theta = 0.012; one single nucleotide polymorphism for every 100–150 bp; Ingvarsson 2008). The species can be used to study the link between biogeographic history and current standing variation available for adaptation, because a growing knowledge base exists for both.

With respect to biogeographic history, fossil pollen records suggest the presence of *Populus* in Central Europe from 9000 BC onwards (Huntley & Birks 1983). The debate over the precise locations of glacial refuges is ongoing (Petit *et al.* 2003), but some safe assumptions can be made. Most importantly, Scandinavian populations must have re-colonized after the last ice age, since Scandinavia was covered by ice during the glacial period up until 7000 years ago and large areas have only been colonized within the last 2000 years because of a continual land uplift of 90 cm per century (Ericson & Wallentinus 1979). With regard to standing variation for candidate adaptive traits and genes, strong genetic differentiation and clinal variation were observed for multiple fitness-related traits in populations from across a wide latitudinal gradient in Sweden (Hall *et al.* 2007). Clinal variation was also observed for DNA polymorphisms located within the phytochrome B2 locus, a key gene controlling adaptation to differences in daylength (Ingvarsson *et al.* 2006).

In this study, we asked how admixture between divergent lineages affects adaptation from standing variation in the European aspen (*P. tremula* L.), a widespread forest tree. We studied a set of seven populations of this species, widely spaced across Europe, for 70 mapped microsatellite loci. Our goal was to infer patterns of neutral and non-neutral population divergence at the European scale and to test for admixture in Northern Europe. We then used phenotypic data from a common garden trial involving 12 populations from across Sweden to test whether admixture explains cline shape, variances and selection differentials for timing of bud set, a phenological trait of great adaptive significance. We first show that locally varying selection is detectable for gene-linked markers in this outcrossing forest tree at the subcontinental scale, contrary to expectations from previous studies with smaller geographic coverage (Hall *et al.* 2007; Ingvarsson 2008). We then demonstrate that admixture between divergent lineages has left its traces in phenotypic clines across the Northern European contact zone, leading to increased variance in and selection on adaptive traits. Our results are relevant to the ongoing discussion regarding the role of intraspecific admixture in adaptive evolution.

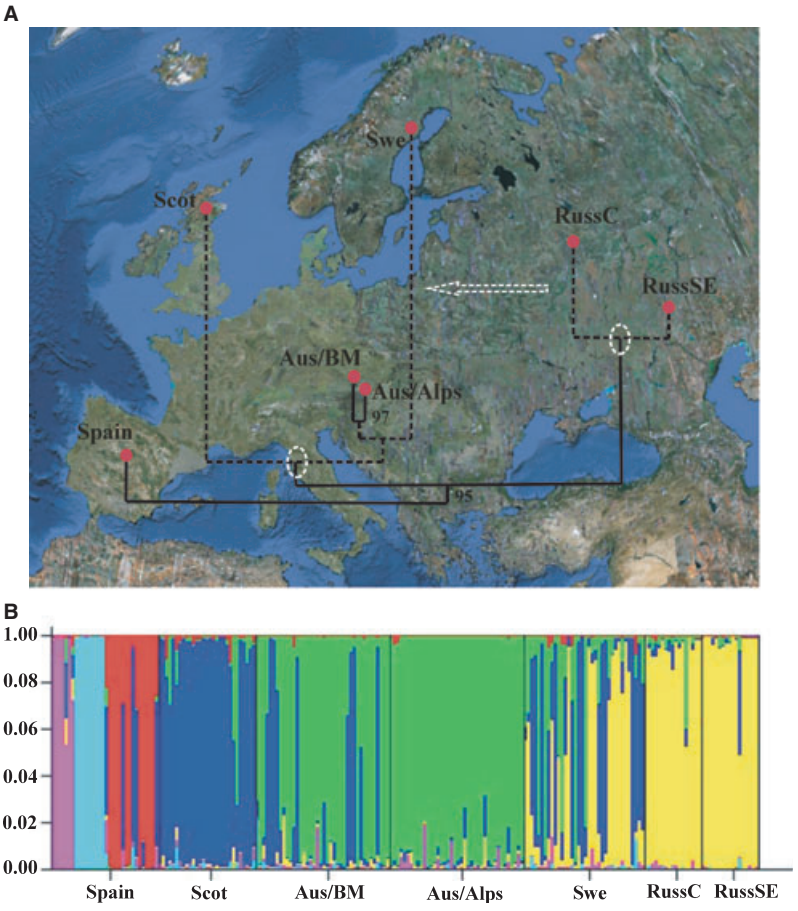


Fig. 1 Conventional (a) and Bayesian (b) analysis of population structure in *Populus tremula*, the European aspen, based on 70 microsatellite loci. (a) UPGMA tree based on Nei's standard genetic distance superimposed on geographic map showing locations of sampling sites. The dendrogram is informative regarding branching patterns but not branch lengths. Bootstrap percentages >60% are indicated beside each node, and nodes with <60% bootstrap support are indicated by dashed branches. Dashed white circles indicate nodes affected by the removal of 12 consistent non-neutral 'outlier' loci (Table 3) from the data set; removal of these loci results in a swap of population RussC to the clade containing populations Swe, Aus/BM, Aus/Alps and Scot, indicated by the white arrow. (b) Bayesian admixture coefficients (*Q*) for individual trees estimated within a linkage model with $K = 6$ gene pools; *Q* estimates for each tree are shown in the form of coloured, vertical bars. For population abbreviations and descriptions, see Table 1.

Materials and methods

Population samples and common garden trial

Population samples for molecular genetic work were collected in seven localities in Europe (Fig. 1a; Table 1) to facilitate analysis of population structure and the detection of non-neutral loci in pairwise comparisons at three different spatial scales: 'regional' (Austria/BM vs. Austria/Alps, Russia/C vs. Russia/SE), 'European' (10 pairwise comparisons) and 'sampling range-wide', i.e. all sampled European *Populus tremula* against Central European material of the closely related species *Populus alba* described in Lexer *et al.* (2005) (Table 2). The sample of *P. alba* was used as a reference to detect locus-specific reductions in genetic diversity because of selective sweeps across all studied populations of *P. tremula*; studying interspecific divergence (F_{ST}) was not a focus of this study. Each population was sampled over several dozens of square kilometres as in previous microsatellite studies of these species (Lexer *et al.* 2005, 2007), to avoid sampling clones and to account for the known great dispersal capacities of these trees. Field collections of this species must weigh the risk of slight population subdivision (Wahlund effects) because of wide local sampling against the risk of sampling multiple ramets of the same clone; cryptic subdivision will convert some of the among-population variation into within-population variation, which renders divergence (F_{ST})-based tests for locally varying selection conservative.

The common garden trial used to study patterns of phenotypic differentiation was previously described by Ingvarsson *et al.* (2006), Hall *et al.* (2007) and Luquez *et al.* (2008). In this study, extensive, previously unpublished multi-year data from this garden trial were analysed to examine adaptive trait differentiation across a Northern European admixture zone identified by molecular markers (below). Data from both the Ekebo garden (55.9°N) and the Sävar garden (63.4°N) described in Luquez *et al.* (2008) were used for this study. Each trial comprises 116 trees from 12 populations across a latitudinal gradient in Sweden (56–66°N), including up to eight replicate clones per tree planted in a randomized block design (Hall *et al.* 2007). Of a

Table 1 Populations of *Populus tremula* sampled for molecular genetic work, including population abbreviations, full names, geographic coordinates, no. of chromosomes sampled (*N*), allelic richness adjusted to the smallest number of individuals sampled by rarefaction, expected (H_E) and observed (H_O) heterozygosities, and within-population inbreeding coefficients (F_{IS})

Population abbreviation	Full name*	Latitude	Longitude	*N*	Allelic richness	H_E	H_O	F_{IS}†
Swe	Sweden/Umeå	63.829N	20.261E	72	4.897	0.549	0.473	0.134
Spain‡	Central Spain	40.017N	3.573W	64	3.497	0.474	0.468	0.022
Scot	Scotland	56.579N	3.606W	58	4.733	0.548	0.456	0.163
Aus/BM	Austria/Bohemian Massif	48.706N	16.097E	80	4.687	0.554	0.482	0.111
Aus/Alps	Austria/Eastern Alps	47.608N	16.099E	80	4.797	0.543	0.468	0.123
Russ/C	European Russia/Central	55.733N	37.674E	34	4.695	0.553	0.481	0.117
Russ/SE	European Russia/Southeast	48.677N	44.452E	34	4.708	0.545	0.457	0.156

*See Fig. 1 for location on a geographic map of Europe.
†Overall F_{IS} was significantly positive, with 95% bootstrap intervals ranging from 0.083 to 0.173.
‡A genetic bottleneck was detected for this population, based on both the sign test and Wilcoxon sign-rank test described by Piry *et al.* (1999).

Table 2 Number of candidate loci for divergent natural selection at different spatial scales out of a total of 70 marker loci analysed, including outlier loci detected by divergence- and diversity-based tests

	Regional scale		European scale*										Sampling range†
Test type	Aus BM-Alps‡	Russ C-SE‡	Swe-Spain	Swe-Scot	Swe-Aus	Swe-Russ‡	Spain-Scot	Spain-Aus	Spain-Russ	Scot-Aus	Scot-Russ	Aus-Russ	Trem-Alba
Divergence§	0	3	7	8	2	2	1	16	9	6	9	5	n/a
Diversity¶	1	4	2	1	3	3	3	3	1	2	1	3	0
Shared	0	2	0	0	1	0	0	0	0	0	0	0	n/a

*The following population pairs were combined at the European scale: Aus/BM and Aus/Alps, and Russia/C and Russia/SE.
†Reduced diversity across the sampled range of *Populus tremula* was tested by using Central European material of *Populus alba* as a reference.
‡Loci under balancing selection within populations were detected in three comparisons: Aus/BM-Aus/Alps, 1 locus; Russ/C-Russ/SE, 4 loci; Swe-Russ, 1 locus. For details about all non-neutral 'outlier' loci see Table S2 (Supporting information).
§'Divergence' outlier loci were significant in both the drift based and the F_{ST} based approach with a FDR of 10%.
¶'Diversity' oulier loci were significant for either one of the two diversity ratios, lnRH or lnRV, at a FDR of 10%.
For expansions of abbreviations in the column head, see Table 1.

number of correlated traits for which measurements were available, timing of bud set was chosen for linear and nonlinear cline fitting in this study. Bud set reveals information about the timing of growth cessation, a trait of known functional significance connected to photoperiod adaptation along latitudinal gradients in temperate trees (Ingvarsson *et al.* 2006; Hall *et al.* 2007). Timing of bud set was measured repeatedly from spring to summer in 2005–2008, as outlined in Luquez *et al.* 2008. Relative growthrate was used as a fitness proxy and was measured as the cumulative growth over the entire period from 2004 to 2008. The cumulative growth includes the effects of frost damage suffered by some of the trees during winters. Frost damage sometimes results in reductions in height (negative growth) as the top (annual) shoots are most likely to be killed by frost.

Microsatellite genotyping

Seventy microsatellite loci were used for population genetic analysis, including 62 markers for which repeat and primer information are available at http://www.ornl.gov/sci/ipgc/ssr_resource.htm and eight additional marker loci. For detailed information about all loci, see Tables S1 and S2, Supporting information and van der Schoot *et al.* (2000), Smulders *et al.* (2001), Tuskan *et al.* (2004) and Yin *et al.* (2009).

All markers were prescreened for single-locus amplification, polymorphism and codominance in the course of a related project on the genetics of species isolation in *P. alba* and *P. tremula*, i.e. the markers were thoroughly tested for robustness prior to this study. Linkage relationships (for graphical representation and genetic

structure analysis; below) were obtained from available *Populus trichocarpa* maps and in a few instances were complemented using recombination data from a controlled cross *P. alba* × *P. tremula*. Genomic DNA was extracted from silica-dried leaves using the Dneasy Plant Mini Kit (QIAGEN), and all markers were polymerase chain reaction (PCR) amplified using the reaction conditions described previously by Lexer *et al.* (2005) and precisely sized using an Applied Biosystems (AB) 3100 Genetic Analyzer, making use of the florescent dyes FAM and JOE as well as size differences among loci for multiplexing. For four loci with particularly low levels of polymorphism in *P. tremula*, the presence of the microsatellite repeat was confirmed by direct sequencing of PCR products, following the protocols of Joseph & Lexer (2008).

Patterns of genomic diversity

All populations of *P. tremula* were characterized for their genetic diversity and structure, including allelic richness (corrected for the smallest sample size by rarefaction), expected (H_E) and observed (H_O) heterozygosity, and the inbreeding coefficient F_{IS} with 95% confidence intervals obtained by bootstrapping over loci (Table S1, Supporting information). Characterization of each locus also included the variance in allele size as a simple diversity parameter under a stepwise mutation model, and single-locus estimates for F_{ST} and Hedrick's (2005) G'_{ST} (Table S1, Supporting information). All *P*-values were adjusted for multiple tests using the Bonferroni method. Population bottlenecks were tested for by the sign test and Wilcoxon sign-rank test described by Piry *et al.* (1999), two standard tests that compare allele numbers with equilibrium gene diversities, and isolation-by-distance (IBD) was tested by comparing the matrices of genetic and geographic distances with a Mantel test. This test allows detecting departures from migration–drift equilibrium which may arise for a variety of reasons, e.g. because of geographic barriers to gene flow, or because of artificial founding of local populations from nonlocal source material. Linkage disequilibrium (LD) among loci was not a major focus of this study and information on this is already available for *P. tremula* (Lexer *et al.* 2007; Ingvarsson 2008). Nevertheless, we were interested in confirming far-ranging LD on the proximal end of chromosome 19 as predicted by Yin *et al.* (2008), which we did using exact tests following Rousset (2008) based on marker order inferred by genetic mapping (Table S1, Supporting information) and assuming *P. trichocarpa* physical map distances.

Population genetic structure was analysed using both conventional and Bayesian methods. Conventional analysis comprised UPGMA cluster analysis in PHYLIP based on Nei's standard genetic distance, using 1000 bootstrap replicates to determine branch support (Fig. 1a). Bayesian analysis was carried out with Structure 2.2 using two different approaches described by Falush *et al.* (2003), namely a standard admixture model assuming correlated allele frequencies and a linkage model which incorporated linkage information from *P. trichocarpa*. The latter assumes good synteny in *Populus*, which we expect at this relatively coarse genomic scale (Cervera *et al.* 2001). Multiple programme runs were conducted for each approach to confirm stabilization of the summary statistics. The two approaches yielded highly similar results, indicating $K = 6$ as the most likely number of gene pools or 'genetic units' based on the model likelihoods and their variances, and the results of the '$k = 6$' linkage model were chosen for presentation and discussion (Fig. 1b), based on a run with 50 000 burn-in followed by 100 000 iterations. Based on their low genetic differentiation in the Bayesian analysis and their close proximity, the two neighbouring populations Aus/BM and Aus/Alps were combined in 'European' scale neutrality tests, as were populations Russia/C and Russia/SE.

Genomic footprint of selection

Neutrality tests were based either on diversity or on divergence. Genetic divergence based tests utilized the migration and drift (F_{ST})-based approach discussed by Beaumont & Balding (2004) and the alternative drift-based approach of Vitalis *et al.* (2001). The former is known to be robust across a wide range of demographic scenarios, and the latter is useful when departures from equilibrium conditions (e.g. bottlenecks) are expected. The migration and drift-based approach was carried out with one initial round of coalescent simulations to estimate 'neutral' F_{ST} and a second round to identify candidate loci subject to selection. We preferred the frequentist approach described by Beaumont & Balding (2004) over the Bayesian method available because more experience exists with it. Parameter settings for the alternative pure drift-based approach were chosen based on available knowledge on the biogeographic history of European forest trees (Huntley & Birks 1983), and the results remained stable across a wide range of parameter settings. The results shown (Table 2; Table S2, Supporting information) are from 100 000 simulations with mutation rate (μ) = 0.01, 0.001, 0.0001; population size before bottleneck (N_e) = 500, 1000, 5000; time since bottleneck (T_0) = 50, 500, 1000 generations; population size before split (N_0) = 50, 500; time (t) since split: 50, 500.

Identifying the most realistic population model for divergence-based outlier detection is subject of intense

ongoing research (Foll & Gaggiotti 2008; Excoffier *et al.* 2009). Here, a combination of the Vitalis *et al.* (2001) and Beaumont & Balding (2004) methods were chosen because extensive experience exists for them. A stringent set of criteria was used to identify loci potentially subject to selection: each candidate locus had to be significant in the purely drift based *and* the F_{ST} based test at a false discovery rate (FDR) of 10% (Benjamini & Hochberg 2000). For completeness, candidate loci for balancing selection were recorded in the form of outliers in the lower tail of the F_{ST} distribution (alpha = 5%) for each pair of populations, but these were not a major focus of this study.

Diversity-based tests for selective sweeps made use of two different diversity ratios, namely the ratio of gene diversity (lnRH) and the ratio of the variance in repeat number (lnRV) for pairs of populations. Significance was assigned using the neutrality tests of Schlötterer (2002) and false positives were accounted for by using a FDR of 10%. We did not use a dynamically adjusted number of linked microsatellites to account for false positives (Wiehe *et al.* 2007) because the number of linked markers available was low. The two ratios, lnRH and lnRV, were interpreted separately because they may reflect different aspects of the genealogical history of a locus.

The divergence- and diversity-based tests were applied to pairwise population comparisons at three different spatial scales as outlined above (Population samples and common garden trial). Our rationale was to examine the interplay of gene flow and selection at these different spatial scales. With respect to timescales under investigation, assuming a generation time of 20 years for poplars yields no more than 600–700 generations since the end of the last ice age 13 000 years ago when most European forest trees were restricted to spatially disconnected refugia. Thus, the relatively long generation time of poplars (compared with annual plants) should render our data ideal for detecting postglacial, locally varying selection with rapidly evolving microsatellites (Schlötterer 2003).

For each candidate locus found to be under locally varying selection at the European scale in at least three pairwise population comparisons, the nearest gene along the chromosome was identified using *P. trichocarpa* genome assembly version 1; version 2 of the genome assembly became available while this study was in review, but its browsing functions at that time were limited. Thus, we preferred the more evolved version 1, with the exception of one locus for which the chromosomal location in version 1 was unclear (footnote to Table 3).

Table 3 Twelve candidate loci for divergent selection identified in three or more pairwise population comparisons in *Populus tremula*, including marker name, linkage group (Lg), closest *Populus trichocarpa* gene model, approximate distance to the nearest gene in kilobases (kb), putative gene function, gene expression information and genomic space of the microsatellite repeat

Marker*	Lg	Closest *Populus* gene model†	Distance to gene (kb)†	Putative gene function	Expression information‡	Genomic space of repeat
G1376	2	fgenesh4_pg.C_LG_II002615	0.4	Transcriptional regulator	No	Intergenic
G1416	3	eugene3.00030612	2.1	26S proteasome regulatory subunit	Yes	Intergenic
ASP322	6	estExt_fgenesh1_pg_v1.C_LG_VI0144	0.0	Adenine phospho- ribosyltransferase	Yes	UTR
ASP933	6	eugene3.00061049	0.0	Amino oxidase	Yes	UTR
G1485	6	grail3.0023026901	1.4	Anion transporter	No	Intergenic
O268	8	fgenesh4_pg.C_LG_VIII001710	0.8	NADH kinase	No	Intergenic
G1353	13	fgenesh1_pg.C_LG_XIII000001	1.0	Pectinesterase inhibitor	No	Intergenic
G1608	15	grail3.0082003901	7.6	NADH2 dehydrogenase	Yes	Intergenic
O14	16	gw1.XVI.1041.1	3.0	Alpha-L-Arabinofuranosidase	Yes	Intergenic
O276	19	gw1.117.48.1	0.1	NBS-LRR–type disease resistance gene	No	Intergenic
Con58.1§	19	grail3.0117003601	3.3	Small auxin-upregulated RNA	Yes	Intergenic
Yin2	19	gw1.XIX.66.1	0.0	Oxysterol-binding protein	Yes	Exon

*Markers with G refer to GCPM, markers with O refer to ORPM microsatellite loci on the *Populus* genome web page (http:// ipgc/ ssr_resource.htm). For complete information about all marker loci including detailed references see Table S1 (Supporting information).

†Gene models and approximate kilobase (kb) distances refer to *P. trichocarpa* genome assembly version 1, except for G1608. This lorus did not map clearly in assembly version 1 but was found on chromosome 15 in *P. trichocarpa* genome assembly version 2. See Materials and methods for the use of the two genome assemblies.

‡Indicates whether or not expressed sequence tag (EST) information is available in JGI *Populus* genome browser version 1.1.

§Locus located on scaffold 117, known to be homologous to *P. trichocarpa* chromosome 19.

Phenotypic clines along a latitudinal gradient

Data on bud set and growth were taken from Luquez *et al.* (2008), where the SwAsp collection and common garden experiments are also described in detail. Clone-specific breeding values for bud set and growth were estimated using linear mixed-model Best Linear Unbiased Predictors (BLUPs) as described by Luquez *et al.* (2008). Within-population variation in bud set was then estimated from variation among clonal BLUPs within the 12 original populations of the SwAsp collection. The relationship between bud set and latitude was studied using a simple linear model ($y = a + b*x$) and a more complex, nonlinear model which allowed for a steeper slope in the centre of the cline ($y = a + b/[1 + \exp[c*x]]$). The two models were fitted to the data using linear and nonlinear regression using the *lm* and *nls* functions in *R*. The two models are not nested and the best-fitting model was therefore selected based on Akaike's Information Criteria (AIC), with $AIC = 2*\log(L) + 2*n$, where L is the likelihood and n is the number of free parameters of a given model (two for the linear model and three for the non-linear model). Selection differentials were calculated within each population in the form of covariances between bud set and fitness (Lande & Arnold 1983), using relative growthrate as a fitness proxy. These analyses were first carried out for the Sävar garden only (the trial with the slightly greater level of replication), then for both gardens combined. Since the result remained the same, the results of the larger, combined data set are presented and discussed below. Note that growth rate is just one of many components of fitness and may thus not fully capture fitness tradeoffs and lifetime reproductive output; the use of partial fitness proxies is common in evolutionary ecology because lifetime fitness is often difficult to measure (Lande & Arnold 1983).

Locus-specific ancestries in the Swedish admixture zone

Genomic admixture and marker locus-specific ancestries were estimated for the Northern European admixture zone based on microsatellite genotype data for trees from Scotland, Sweden and Russia, using the R-script INTROGRESS (Z. Gompert and C. A. Buerkle) which analyses admixture gradients and genomic clines (Gompert & Buerkle 2009). Overall admixture was estimated in the form of a maximum-likelihood (ML) hybrid index and 95% confidence intervals. The ancestry of each locus in each individual was assessed in terms of homo- or heterozygosity for Scottish and Russian alleles, with multi-allelic microsatellites binned into two informative allelic classes per locus as described in Lexer *et al.* (2007). Four loci were removed as they yielded no information about the population origin of alleles in admixed individuals. Scotland and Russia were used to estimate parental allele frequencies for admixture analysis, because the Structure results pointed to these two as predominant source populations contributing to the admixed Swedish sample.

Results

Population differentiation and admixture

Our analysis of molecular genetic variability for European aspen – the most comprehensive molecular data set available for this species to date in terms of geographic and genomic coverage – revealed distinct patterns of diversity and differentiation at the European scale. Whereas overall genomic divergence was low (average F_{ST} over loci = 0.051, 95% CI = 0.040–0.063; average G'_{ST} = 0.181, 95% CI = 0.152–0.211), significant IBD was detected among populations ($r = 0.601$; $P < 0.005$; $R^2 = 36\%$), thus indicating an important role for gene flow and drift in shaping patterns of molecular diversity. Accordingly, geographic structure was detectable by both conventional (Fig.1a; Table S3, Supporting information) and Bayesian (Fig. 1b) methods. Conventional cluster analysis revealed geographic structure but several nodes had <60% bootstrap support, affecting populations from Central, Northern and Eastern Europe (Fig. 1a). Bayesian analysis revealed the likely cause: extensive admixture between Eastern and Western European lineages in Scandinavia and, to a lesser extent, in Central Europe (Fig. 1b, populations Swe and Aus/BM). Note that *all* models from $K = 4$ to $K = 8$ yielded congruent results for Sweden, namely admixture between a Western and an Eastern European lineage (not shown). Bayesian analysis also revealed the presence of several groups of closely related genotypes in Spain (Fig. 1b). This population has undergone a recent genetic bottleneck, detectable by standard tests that compare allele numbers and equilibrium gene diversities (sign test: $P < 0.001$; Wilcoxon test: $P < 0.001$).

In general, microsatellite diversities were intermediate ($H_E = 0.563 \pm 0.030$ SE; $H_O = 0.468 \pm 0.029$ SE), i.e. low enough for meaningful F_{ST}-based analysis of locally varying selection using frequentist approaches. A slight homozygote surplus (positive inbreeding coefficient F_{IS}) was detectable for all populations, the 95% bootstrap interval for overall F_{IS} ranging from 0.083 to 0.173, and this was attributable primarily to cryptic population subdivision (see Materials and methods).

Genomic footprint of locally varying selection

Our stringent divergence-based tests for locally varying selection (a locus had to be significant in both the purely drift-based and the F_{ST}-based tests at a FDR of 10%; see Materials and methods) yielded between 1 and 16 candidate loci for divergent selection at the European scale (6 loci ± 1 SE across all 10 pairwise population comparisons at this geographic scale) and smaller numbers of candidate loci in regional comparisons (Table 2; Table S2, Supporting information).

Diversity-based tests for selective sweeps (at FDR = 10%) yielded between 1 and 3 sweep loci at the European scale, between 1 and 4 at the regional scale and no range-wide sweep was found when *Populus tremula* was compared with Central European populations of its closely related sister species *Populus alba*. Except for three 'sweep loci' with locally reduced diversity, there was no overlap between candidate 'outlier' loci detected by the divergence- and the diversity-based tests (Table 2; Table S2, Supporting information). Six loci were found to be under balancing selection within populations, five of them in regional comparisons.

The top 12 candidate loci for locally varying selection (loci that matched our stringent significance criteria in at least three pairwise population comparisons) comprised three loci located within expressed genes and four other loci that were located 1 kb or closer to the nearest gene model predicted by the *Populus trichocarpa* genome project (average distance to nearest gene: 1.6 ± 0.6 kb SE; Table 3). The gene list includes, among others, a transcription factor and genes involved in plant defence, such as a NBS-LRR-type resistance gene and a pectinesterase inhibitor. Two of the genes are located at the proximal end of chromosome 19, recently put forward as an incipient sex chromosome in *Populus* (Yin *et al.* 2008). Significant within-population LD between markers in this region extended over >400 kb in our study, assuming *P. trichocarpa* physical map distances (markers Con 58.1 and Yin2; details not shown), as opposed to the general prediction of a decay of LD within a few 100 bp in this species (Ingvarsson 2008). Population comparisons including the bottlenecked Spanish and the admixed Swedish population contributed rather moderately to the detection of replicated non-neutral outlier loci: except for loci O268, G1608 and Con58, all loci in Table 3 would have been detected as replicated outliers even without considering Spain, and except for G1608, Con58 and Yin2 all loci would have been detected without considering the admixed Swedish population (Table S2, Supporting information). The contribution of Spain and Sweden to all neutrality tests is visible in Table 2. Likewise, the commonly observed tendency across loci and populations for inbreeding coefficients (F_{IS}) to be positive (Wahlund effects; see Materials and methods; Table 1; Table S1, Supporting information) had little influence on our inference of selection, i.e. there was no significant difference between F_{IS} for the top 12 candidate outlier loci shown in Table 3 and the remaining loci (nonparametric Mann–Whitney *U*-test, $P = 0.35$).

The role of admixture in shaping phenotypic clines in Scandinavia

A nonlinear model provided a better fit for the phenotypic cline for bud set across a latitudinal gradient in Sweden compared with a linear model [Fig. 2a; linear: Akaike's Information Criterion (AIC) = 68.3; nonlinear: AIC = 49.3]. The difference in model fits was caused by a step in the cline between 58° and 63° of latitude (Fig. 2a). This is just south of the admixed Swedish material sampled in the microsatellite-based genome scan (Figs 1a,b). The variance for bud set was greatly increased in this area (Fig. 2b), and selection differentials (covariances between bud set and relative growth-rate as a fitness proxy) were elevated in the contact zone as well (Fig. 2c).

Considerable within-genome variation for marker ancestry was observed when Scotland and Russia were used as proxies for the source gene pools that contributed to the admixed Swedish population (Fig. 3a; choice of Scotland and Russia as source populations motivated by Fig. 1). Whereas a complete admixture gradient was recovered across these populations for the genome-wide panel of 70 microsatellites (Fig. 3b), the ancestries of individual loci varied greatly along the chromosomes (Fig. 3a). This can easily be illustrated by comparing chromosome 6, known to exhibit normal recombination rates (Yin *et al.* 2004), and chromosome 19, known to exhibit greatly reduced recombination for the region studied (Yin *et al.* 2008). Ancestry varied greatly across chromosome 6, its left end exhibiting a noticeable excess of Russian ancestry (light green) and the right half being predominantly Scottish (dark green). In contrast, locus-specific ancestries changed gradually from Scottish to Russian along chromosome 19 (Fig. 3c), consistent with a history of lower recombination.

Discussion

The role of standing variation in adaptive evolution is a topic of fundamental importance for evolutionary biology (Colosimo *et al.* 2005; Stearns & Hoekstra 2005; Pennings & Hermisson 2006; Barrett & Schluter 2008; Mullen & Hoekstra 2008). As beneficial alleles present as standing variation are older than new mutations,

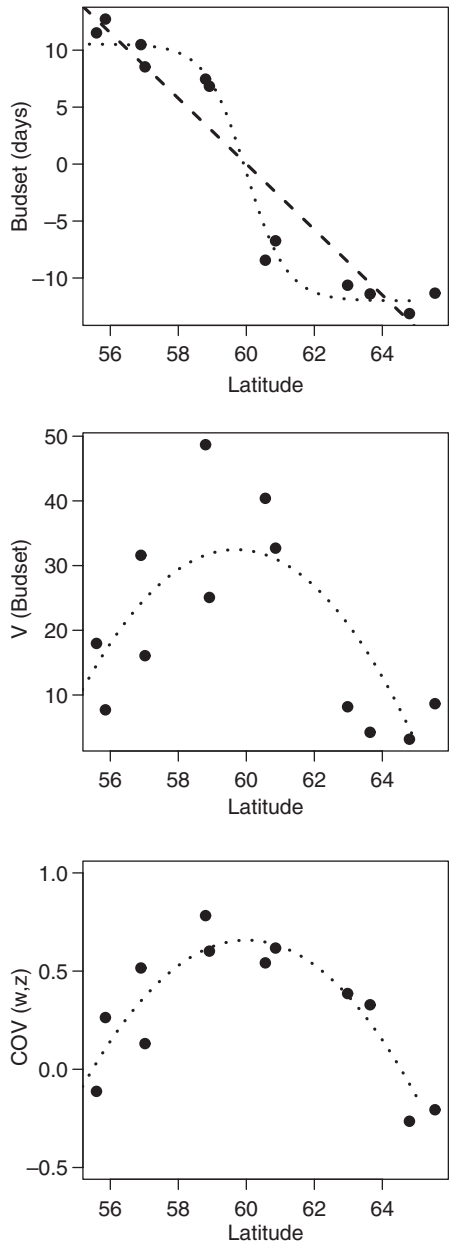


Fig. 2 Cline shapes, trait variances and selection differentials for timing of bud set, a trait of known adaptive significance in deciduous temperate trees, measured in 12 populations of *Populus tremula* from across a latitudinal gradient in Scandinavia grown in two common gardens. (a) Best Linear Unbiased Predictors (BLUPs, expressed in deviations from mean bud set date), with regression curves for linear and non-linear model fits indicated by dashed and dotted lines, respectively. (b) Trait variances, nonlinear fit indicated by dotted line. (c) Selection differentials, estimated as the covariance between bud set and relative growth rate as a fitness proxy, nonlinear fit indicated by dotted line.

their current spatial distributions will reflect key aspects of a species' biogeographic history, particularly in temperate taxa which must undergo regular cycles of range contractions and expansions to track Earth's climatic oscillations (Hewitt 2000). This implies that beneficial alleles have been pre-tested by selection in past environments or in other parts of a species' range (Kremer *et al.* 2002; Barrett & Schluter 2008). In many cases, this past 'selective filter' may even have operated in a different species, followed by interspecific gene flow and recombination in the form of adaptive introgression (Arnold 2006) or recombinational speciation (Rieseberg *et al.* 2003).

Our molecular marker-based scan indicates that locally varying selection is detectable in the wind-pollinated, dioecious *Populus tremula* at the European scale (Table 2; Table S2, Supporting information). This implies linkage of putatively 'neutral' marker alleles with selected coding or regulatory mutations or heritable epialleles, the presize nature of the causal molecular changes under selection being unknown at this stage of the work. At first sight, our detection of locally varying selection is at odds with the known high levels of inter-population gene flow in this species (N_em = up to 15 effective migrants per generation; Lexer *et al.* 2005) and the rapid decay of LD observed at the DNA sequence level (within just a few 100 bp, Ingvarsson 2008). The finding in this study of an average of 8% non-neutral outlier loci (±2% SE across all population comparisons) in our stringent divergence-based tests (Tables 2 and 3) indicates that selection (s) on particular genome regions in *P. tremula* is stronger *or* recombination (r) in these regions is lower than previously thought.

A role for low recombination is indicated by the fact that 3 of the 12 loci with consistent and replicated departures from neutrality in our study (Table 3) are located in the proximal end of chromosome 19, thought to be an incipient sex chromosome conserved across different sections of the genus *Populus* (Yin *et al.* 2008). Within-population LD in this region is greatly elevated in the studied populations (extending up to 400 kb, assuming *Populus trichocarpa* physical map distances). Elevated LD in this region is expected based on genomic sequencing, scaffold assembly and genetic mapping of this region in *P. trichocarpa* (Yin *et al.* 2008). An indication for selection, however, is that 10 of the 12 replicated outlier polymorphisms detected in this study were located less than 3 kb from the nearest gene, and three markers were even located within transcribed sequences (Table 3). This makes it plausible that genes or *cis*-regulatory elements adjacent to the studied polymorphisms were indeed the targets of selection. Of course, the causal molecular changes (mutations or heritable epigenetic modifications) under locally varying selection may also reside in a different gene or DNA region adjacent to each microsatellite. At the very least, our cross-check of the 12 top candidates against the

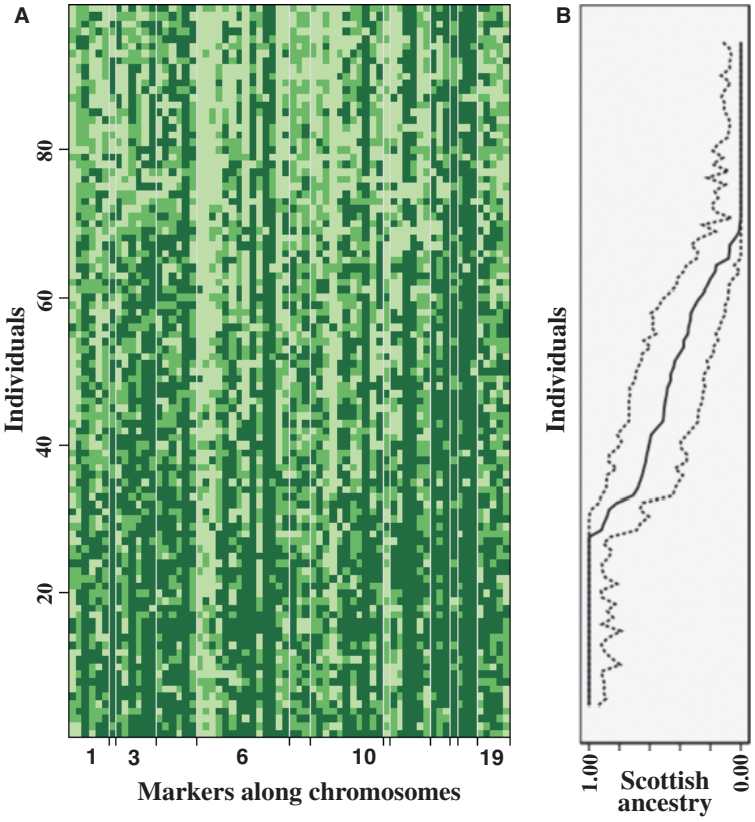


Fig. 3 Admixture and genomic variation for marker ancestry in trees from the Northern European admixture zone of *Populus tremula*. (a) Genomic variation for marker ancestry in terms of homo- or heterozygosity for Scottish and Russian alleles, respectively, with multi-allelic microsatellites binned into two informative allelic classes for each locus. Dark green, homozygotes for Scottish allele; light green, homozygotes for Russian allele; medium green, heterozygotes for Scottish and Russian alleles. Markers along chromosomes are shown in consecutive order on the horizontal axis, chromosomes separated by thin white lines, and individuals are shown on the vertical axis. The five chromosomes with the best genomic coverage are indicated by arabic numbers along the horizontal axis; for all other chromosomal locations, see Table S1 (Supporting information). (b) Overall admixture and 95% confidence intervals for the trees shown, corresponding exactly to the individuals seen along the vertical axis in (a).

Populus genome assembly (Table 3) suggests that all 12 are situated within gene-rich regions.

Two conspicuous findings of our population genomic work were the small number of non-neutral outlier loci detected by diversity based tests for 'selective sweeps' (on average just 3% of the studied loci ±0.4% SE), and the limited overlap between outliers found with divergence vs. diversity based tests (Table 2). This pattern was not caused by the presence of the admixed Swedish and the bottlenecked Spanish populations in the data set, e.g. population comparisons of Scotland–Austria, Scotland–Russia and Austria–Russia followed the same trend (Table 2). The pattern was also not caused by errors in the spatial scale chosen for analysis, as the number of selective sweeps was low at all three geographic scales examined (Table 2). On the contrary, this pattern is expected if local adaptation occurs primarily from standing variation (soft sweeps) rather than from new mutations (hard sweeps; Pennings & Hermisson 2006). Soft sweeps will result in a severe reduction of the chromosomal width of sweep regions (hence a reduction of statistical power to detect selection) compared with hard sweeps, and this is exactly what was observed in this study: locus-specific reductions in diversity were rare (Table 2), and they almost never affected pairs of adjacent loci, whereas the latter was commonly observed for divergence-based tests (Table S2, Supporting information). Note that low levels of neutral divergence as found in *P. tremula* (average F_{ST} = 0.051 ± 0.047 SD) greatly facilitate the detection of locally varying selection in divergence-based tests, because selected loci will emerge from the neutral distribution more clearly.

Although distinguishing between young soft sweeps and old hard sweeps is not trivial, the presence of soft sweeps is expected from theory when the scaled mutation rate, $\Theta = 2N_e\,\mu$, exceeds 0.01 (Pennings & Hermisson 2006). This threshold is easily exceeded in *P. tremula* for a wide range of possible mutation rates (μ), because effective population size (N_e) is in the order of 10^5 (Ingvarsson 2008). Theory also predicts that migration from geographically disconnected source gene pools (e.g. glacial refugia) will contribute to the standing variation available for local adaptation, in a manner similar to recurrent mutation (Pennings & Hermisson 2006). This is what we find here for European populations of the widespread forest tree *P. tremula*.

Our Bayesian structure analysis based on 70 mapped loci clearly indicates genetic admixture between a Western and an Eastern European lineage of *P. tremula* in Sweden, with possible additional contributions from Central European populations (Fig. 1b). The known biogeographic history of Scandinavia (ice cover during last glaciation and recent recolonisation following land

uplift) allows us to exclude that this genetic pattern was produced by anything other than postglacial intraspecific admixture. The Northern European admixture zone of *P. tremula* reported here coincides with a postulated zone of postglacial contact affecting many species of animals and plants (Hewitt 2000; Tollefsrud *et al.* 2008), including bears, rodents, conifers and orchids.

Admixture of postglacial lineages of *P. tremula* in Sweden is the simplest explanation for the significant step in the geographic cline for bud set (Fig. 2a), a heritable trait of clear adaptive significance in temperate trees. Of course, clinal variation in Scandinavia cannot be attributed to admixture alone: the critical daylength for initiating growth cessation varies linearly across the range of the studied populations (Hall *et al.* 2007), and adaptation of forest trees to local photoperiod is known to generate clinal variation in important phenological traits (Howe *et al. 2003*). Accordingly, DNA polymorphisms within the phytochrome B2 locus exhibit clinal variation across the same geographic gradient (Ingvarsson *et al.* 2006), whereas multiple neutral polymorphisms located elsewhere in the genome do not (Hall *et al.* 2007). Admixture contributes to the variation available for adaptation along the gradient, as visible from the step in the cline (Fig. 2a) and from two other observations.

First, the variance of bud set is greatly elevated in the centre of the cline, consistent with increased genetic variation available for selection, stemming from admixture of differentiated gene pools (Fig. 2b). Second, selection differentials (estimated as the covariance between bud set and vegetative fitness) are also elevated in the centre of the cline (Fig. 2c). Thus, admixture between differentiated postglacial lineages contributes to the standing variation available for natural selection and adaptation, as predicted by theory (Arnold 2006; Pennings & Hermisson 2006; Barrett & Schluter 2008). It is unlikely that admixture constrains adaptation in *P. tremula*: no signs of outbreeding depression (=reduced vegetative fitness) were observed in central Sweden, and geographic clines for fitness-related growth traits are shallow and linear as expected from the gradual change in the length of the growing season (Hall *et al.* 2007; Luquez *et al.* 2008). Our data provide the necessary, hitherto missing direct link between range shifts because of past climatic oscillations on the one hand and current levels of standing variation available for adaptive evolution on the other. Our three-pronged approach (genomic scan for selectively differentiated DNA regions, common garden measurements of adaptive trait differentiation, map-based analysis of genomic ancestry) allowed us to demonstrate the effects of admixture on standing variation at a depth rarely (if ever) seen in previous studies of clinal variation in ecological keystone or foundation species.

The most reliable way to clarify the exact source of beneficial alleles in any species is to map and isolate the genes responsible for adaptation, estimate their fitness effects and establish their genealogical histories (Colosimo *et al.* 2005; Barrett & Schluter 2008). In long-lived organisms such as trees, such studies are greatly facilitated by emerging approaches to identify fitness-related genes in natural populations (Savolainen & Pyhajarvi 2007), particularly 'admixture mapping' in intraspecific admixture or contact zones (Buerkle & Lexer 2008). In the case of *P. tremula*, the chromosomal variation in marker ancestry required for admixture mapping appears to be present in the Scandinavian contact zone, as exemplified by the mosaic-like nature of chromosome 6, consisting of blocks of DNA of either 'Scottish' or 'Russian' ancestry (Fig. 3).

Understanding the role of standing variation in facilitating rapid adaptation to new environments is a crucial task for evolutionary biologists, because rapid evolution will be required for species' survival in the face of human-mediated environmental perturbations (Barrett & Schluter 2008). This is particularly relevant for long-lived organisms, such as trees (Davis & Shaw 2001; Petit & Hampe 2006; Savolainen & Pyhajarvi 2007). Detecting the missing links between biogeographic history and current levels of standing variation available for adaptation is a key to both science-based conservation and responsible exploitation of wild species in a changing world.

Acknowledgements

We thank M. Fay and M. van Loo for help in the laboratory, T. Barbará for help with data analysis; C. A. Buerkle, S. P. DiFazio, T.M. Yin, A. Widmer and S. Gonzalez-Martinez for helpful discussions; the *Populus* genome consortium for sharing information and four anonymous referees for reading the manuscript. The research was financially supported in part by grants from the Brazilian CNPq to DdC, from the Russian RFBR to IS, and from the British NERC to CL.

References

Arnold ML (2006) *Evolution Through Genetic Exchange*. Oxford University Press, Oxford.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology Evolution and Systematics*, **16**, 113–148.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Benjamini Y, Hochberg Y (2000) On the adaptive control of the false discovery rate in multiple testing with independent

statistics. *Journal of Educational and Behavioral Statistics*, **25**, 60–83.

Briscoe D, Stephens J, O'Brien S (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. *Journal of Heredity*, **85**, 59–63.

Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution*, **23**, 686–694.

Cervera MT, Storme V, Ivens B *et al.* (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics*, **158**, 787–809.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to quaternary climate change. *Science*, **292**, 673–679.

Ericson L, Wallentinus HG (1979) Sea-shore vegetation around the Gulf of Bothnia. *Wahlenbergia*, **5**, 1–142.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Hall D, Luquez V, Garcia VM *et al.* (2007) Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes and phenotypic traits. *Evolution*, **61**, 2849–2860.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **5**, 1633–1638.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Howe GT, Aitken SN, Neale DB *et al.* (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany*, **81**, 1247–1266.

Huntley B, Birks H (1983) *An Atlas of Past and Present Pollen Maps of Europe: 0–13 000 Years ago*. Cambridge University Press, Cambridge.

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, **180**, 329–340.

Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006) Clinal variation in phyB2, a candidate gene for daylength-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics*, **172**, 1845–1853.

Joseph JA, Lexer C (2008) A set of novel DNA polymorphisms within candidate genes potentially involved in ecological divergence between *Populus alba* and *P. tremula*, two hybridizing European forest trees. *Molecular Ecology Resources*, **8**, 188–192.

Kremer A, Kleinschmit J, Cottrell J *et al.* (2002) Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management*, **156**, 75–87.

Kruuk LEB, Baird SJE, Gale KS, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, **153**, 1959–1971.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

Lexer C, Fay M, Joseph J, Nica M-S, Heinze B (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology*, **14**, 1045–1057.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, **98**, 183–184.

Luquez V, Hall D, Albrectsen BR *et al.* (2008) Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics and Genomes*, **4**, 279–292.

Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, **62**, 1555–1569.

Pennings PS, Hermisson J (2006) Soft sweeps II-molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, **37**, 187–214.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.

Reich D, Patterson N, De Jager PL *et al.* (2005) A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nature Genetics*, **37**, 1113–1118.

Rieseberg LH, Burke JM (2001) The biological reality of species: gene flow, selection, and collective evolution. *Taxon*, **50**, 47–67.

Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.

Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Savolainen O, Pyhajarvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology*, **10**, 162–167.

Schemske DW (2000) Understanding the origin of species. *Evolution*, **54**, 1069–1073.

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753–763.

Schlötterer C (2003) Hitchhiking mapping-functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 32–38.

van der Schoot J, Pospiskova M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). *Theoretical and Applied Genetics*, **101**, 317–322.

Smulders MJM, Van Der Schoot J, Arens P, Vosman B (2001) Trinucleotide repeat microsatellite markers for black poplar (*Populus nigra* L.). *Molecular Ecology Notes*, **1**, 188–190.

Stearns SC, Hoekstra RF (2005) *Evolution: An Introduction*. Oxford University Press, Oxford.

Tollefsrud MM, Kissling R, Gugerli F *et al.* (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular Ecology*, **17**, 4134–4150.

Tuskan GA, Gunter LE, Yang ZK *et al.* (2004) Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. *Canadian Journal of Forest Research*, **34**, 85–93.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.

Whibley AC, Langlade NB, Andalo C *et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science*, **313**, 963–966.

Wiehe T, Nolte V, Zivkovic D, Schlotterer C (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics*, **175**, 207–218.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yin TM, DiFazio SP, Gunter LE *et al.* (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics*, **109**, 451–463.

Yin T, DiFazio SP, Gunter LE *et al.* (2008) Genome structure and emerging evidence of an incipient sex chromosome in Populus. *Genome Research*, **18**, 422–430.

Yin TM, Zhang XY, Gunter LE *et al.* (2009) Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome. *New Phytologist*, **181**, 498–503.

Christian Lexer's team is interested in the evolutionary genomics of adaption and speciation, using Populus and other plants as study systems. Dulcineia de Carvalho is a group leader in Conservation Genetics in Brazil, she carried out this study as a visiting professor in C.L.'s lab. Jeffrey Joseph, Leonie Suter, and Claudia Sedivy were members of C.L.'s group. Pär Ingvarsson, Joan Cottrell, Berthold Heinze, and Ivan Schanzer are group leaders with broad interests in plant evolutionary and conservation genetics; they contributed to this work via sample and data collection, analysis, and interpretation of results.

Supporting Information

Table S1 Microsatellite marker loci used in this study, including locus names, linkage groups (Lg), allele numbers (A), expected (H_E) and observed (H_O) heterozygosities, inbreeding coefficients F_{IS}, and genetic divergence in the form of F_{ST} and Hedrick's (2005) G'_{ST}

Table S2 Results of marker-based genome scans for non-neutral population divergence based on divergence and diversity based tests, pairwise population comparisons arranged side-by-side on separate pages to facilitate inspection of locus-specific outlier patterns across populations and to allow printing of particular population comparisons if desired

Table S3 Matrix of F_{ST} values for pairs of populations used in divergence- and diversity based neutrality tests at the European scale

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Annex C

Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species.

MACAYA-SANZ, D., SUTER, L., JOSEPH, J., BARBARÁ, T., ALBA, N., GONZÁLEZ-MARTÍNEZ, S. C., WIDMER, A., LEXER, C., 2011. Heredity **107**: 478-486.

Heredity (2011), 1–9

www.nature.com/hdy

ORIGINAL ARTICLE

Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species

D Macaya-Sanz[1], L Suter[2,3], J Joseph[2], T Barbará[2,4], N Alba[1], SC González-Martínez[1], A Widmer[3] and C Lexer[2,4]

[1]*Department of Forest Ecology and Genetics, Center of Forest Research, CIFOR-INIA, Carretera de A Coruña, Madrid, Spain;* [2]*Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, UK;* [3]*ETH Zürich, Institute of Integrative Biologie, Universitätstrasse 16, Zürich, Switzerland and* [4]*Department of Biology, University of Fribourg, Unit of Ecology and Evolution, Fribourg, Switzerland*

Molecular genetic analyses of experimental crosses provide important information on the strength and nature of post-mating barriers to gene exchange between divergent populations, which are topics of great interest to evolutionary geneticists and breeders. Although not a trivial task in long-lived organisms such as trees, experimental interspecific recombinants can sometimes be created through controlled crosses involving natural F_1's. Here, we used this approach to understand the genetics of post-mating isolation and barriers to introgression in *Populus alba* and *Populus tremula*, two ecologically divergent, hybridizing forest trees. We studied 86 interspecific backcross (BC_1) progeny and >350 individuals from natural populations of these species for up to 98 nuclear genetic markers, including microsatellites, indels and single nucleotide polymorphisms, and inferred the origin of the cytoplasm of the cross with plastid DNA. Genetic analysis of the BC_1 revealed extensive segregation distortions on six chromosomes, and >90% of these (12 out of 13) favored *P. tremula* donor alleles in the heterospecific genomic background. Since selection was documented during early diploid stages of the progeny, this surprising result was attributed to epistasis, cyto-nuclear coadaptation, heterozygote advantage at nuclear loci experiencing introgression or a combination of these. Our results indicate that gene flow across 'porous' species barriers affects these poplars and aspens beyond neutral, Mendelian expectations and suggests the mechanisms responsible. Contrary to expectations, the *Populus* sex determination region is not protected from introgression. Understanding the population dynamics of the *Populus* sex determination region will require tests based on natural interspecific hybrid zones.

Heredity advance online publication, 18 May 2011; doi:10.1038/hdy.2011.35

Keywords: speciation; hybrid zone; *Populus*; introgression; segregation distortion; sex chromosome

Introduction

The genetics of reproductive barriers is of great current interest in evolutionary genetics, because the extent of gene flow experienced by diverging populations or species depends crucially on it (Coyne and Orr, 2004). Reproductive isolation (RI) is a key feature of Mayr's Biological Species Concept, which includes the notion of whole genome isolation between divergent taxa as a hallmark of 'good' species. More recently, a 'genic view' of species and speciation has found wide acceptance, which recognizes that genomes can be porous and that RI is a property of individual loci or genomic regions, rather than the genome as a whole (Wu, 2001; Lexer and Widmer, 2008).

RI will first arise in genomic regions harboring 'speciation genes' or other isolation factors, leading to genomic islands of divergence (Wu, 2001; Emelianov *et al*., 2004; Nosil *et al*., 2009), and will subsequently spread across the genome. Nevertheless, many groups of taxa do not achieve complete genomic isolation for millions of years, as observed in *Helianthus* (sunflowers), *Populus* (poplars, aspen, cottonwoods), *Quercus* (oaks), *Silene* (campions) or *Iris* (lilies), to name just a few examples among plants (reviewed by Lexer and Widmer, 2008). Despite the shift in perception from whole genome isolation to a genic view of species, the genetics of RI is still at the center of attention in speciation genetics (Widmer *et al*., 2009). This is the case because the genetics of RI will determine how quickly gene flow ceases and which loci are affected first.

Plant speciation geneticists often use controlled crosses to study the genetics of post-mating components of RI (Fishman *et al*., 2001; Coyne and Orr, 2004; Bouck *et al*., 2005; Sweigart *et al*., 2006). Even in species with strong reproductive barriers, interspecific multi-generation crosses can sometimes be obtained in the laboratory. In such crosses, interspecific recombination in the meiosis of the F_1 will effectively start to break up the parental species' genomes, which allows geneticists to isolate and study chromosomal blocks with a role in moderating gene flow (Lexer and Widmer, 2008). One approach to achieve this goal is to search for loci or genomic regions with departures from Mendelian expectations, also known as segregation distortion (Fishman *et al*., 2001; Yin *et al*., 2004; Bouck *et al*., 2005; Sweigart *et al*., 2006;

Correspondence: Professor C Lexer, Department of Biology, Unit of Ecology and Evolution, University of Fribourg, Chemin du Musée 10, Fribourg, CH 1700, Switzerland.
E-mail: christian.lexer@unifr.ch
Received 20 October 2010; revised 16 March 2011; accepted 31 March 2011

Lopez-Fernandez and Bolnick, 2007). As expected from theory (Barton, 2001; Wu, 2001; Turelli and Moyle, 2007), such studies tend to recover the full breadth of departures from Mendelian expectations, including genome regions that cause isolation and resist introgression (Fishman *et al*., 2001; Sweigart *et al*., 2006), and loci that cross the barrier more readily than expected based on Mendel's laws (Tiffin *et al*., 2001; Yin *et al*., 2004; Bouck *et al*., 2005).

Experimental cross-based approaches such as this (beyond the F_1) have rarely been used in forest trees because of the difficulty of creating recombinant hybrid generations in species with long generation times (but see Yin *et al*., 2004; Yin *et al*., 2008). The scarcity of experimental multi-generation crosses in forest trees is unfortunate, as long-lived trees provide the opportunity to dissect post-mating barriers in organisms with juvenile-adult phase change, that is potentially from gametes through early embryonic and juvenile stages to maturity. One way to circumvent the long time needed to generate experimental crosses in forest trees is to take advantage of natural hybrids (for example Woolbright *et al*., 2008), but this approach remains largely unexplored.

Populus alba and *Populus tremula* are two ecologically divergent European members of the 'model tree' genus *Populus*. Reproductive barriers between these species are incomplete, leading to the frequent formation of extensive 'mosaic' hybrid zones (Lexer *et al*., 2005, 2010). Ongoing genomic studies of natural populations indicate that the species boundary is porous, with some loci resisting introgression and others crossing the barrier more readily than predicted by genomic expectations (Lexer *et al*., 2007, 2010). Morphometric data indicate introgression of phenotypic traits from *P. tremula* into *P. alba* (Lexer *et al*., 2009), despite the presence of substantial RI involving both assortative mating and post-zygotic isolation in the form of epistatic interactions (Lexer *et al*., 2010). Studies of controlled interspecific progeny are useful for reducing the complexity of patterns of RI seen *in situ* in natural hybrid zones.

Since *P. alba* and *P. tremula* are dioecious and sex determination in *Populus* appears to be controlled by an incipient sex chromosome (Yin *et al*., 2008; Pakull *et al*., 2009; Paolucci *et al*., 2010), controlled crosses also provide the opportunity to study the role of the sex determination region in blocking interspecific gene flow, which is a topic of great current interest in evolutionary genetics (Qvarnström and Bailey, 2009). Our hypothesis at the outset of this study was that species isolation genes will have accumulated in the sex determination region of poplar, as recently observed for other taxa (Qvarnström and Bailey, 2009).

With this in mind, the questions of this study were as follows: (1) What do marker segregation data from a controlled interspecific backcross (BC_1) tell us about the strength and genetic architecture of post-mating reproductive barriers in these hybridizing forest trees? (2) How great is the potential for interspecific introgression across porous species boundaries in *Populus*? (3) What is the likely role of the *Populus* sex determination region in blocking or moderating interspecific gene flow? To address these questions, we analyzed and interpreted segregation patterns of alleles of known species origin for a genome-wide panel of molecular genetic markers, genotyped in a controlled interspecific BC_1 of *P. alba* and *P. tremula*, and compared the results to patterns of divergence and linkage disequilibrium (LD) in natural populations.

Materials and methods

Plant materials (BC_1 and natural populations)

To study post-mating reproductive barriers creating segregation distortions, we developed a controlled backcross of an F_1 natural poplar hybrid (*Populus tremula* × *alba*) with a pure *P. alba* (BC_1). The male parent of this cross was a known natural clone (J1) of *P. alba* from the Jalón river in the Ebro watershed (Northeast of the Iberian Peninsula), and the female parent an F_1 natural hybrid (BET3) from a hybrid population in the Tajo river headwaters (Central Iberian Peninsula). The F_1 hybrid status of BET3 was assessed through phenotypic features and confirmed prior to this study by genomic admixture analysis in STRUCTURE 2.2 following Lexer *et al*., 2005 (95% credible intervals of admixture coefficient *Q* did *not* include 0.25 or 0.75). The species origin of maternally inherited plastid DNA in BET3 was identified here by sequencing the *trnC-petN1* plastid DNA region in this interspecific F_1 hybrid and comparison with sequences from pure individuals of *P. alba* and *P. tremula*; plastid DNA haplotypes are known to be highly divergent between *P. alba* and *P. tremula* (Lexer *et al*., 2005; Fussi *et al*., 2010).

The controlled backcross was produced at INIA's nursery (Madrid, Spain), yielding 131 seedlings that were further grown under greenhouse conditions. Early mortality in first-year seedlings reduced the number of offspring available for DNA extraction and genotyping to 86 individuals, currently maintained in two clone banks planted in Central Spain.

Apart from the BC_1 used for mapping, up to 201 individuals from European populations of *P. tremula* and up to 167 individuals from European populations of *P. alba* were employed to determine the parental species origin of markers used in segregation analysis, and for additional population genetic analysis of markers located on chromosome XIX (below). Population genetic data for this purpose were taken from de Carvalho *et al*. (2010) and Lexer *et al*. (2010), where detailed documentation of populations and genotypic data can be found. Briefly, populations of *P. tremula* were from Spain, Scotland, Central Sweden and Austria (two populations, one from the Eastern Alps and one from the Bohemian Massif). Populations of *P. alba* were from Spain, the Austrian Danube and the Hungarian Tisza valley. In addition, 20 individuals of *P. alba* whose sex had been determined phenotypically during the flowering season were sampled in Spain (12 females and 8 males). Species assignment for each individual was achieved using STRUCTURE 2.2 as described above.

Molecular genetic markers and genotyping reactions

Genomic DNA was purified from young fresh leaves using the DNeasy Plant Mini kit (QIAGEN, Hilden, Germany). A genome-wide set of 98 nuclear markers was used for segregation analysis in the interspecific BC_1 (Supplementary Tables 1 and 4). These markers included microsatellite loci available from the *Populus* genome

consortium (Van der Schoot *et al*., 2000; Smulders *et al*., 2001; Tuskan *et al*., 2006; Yin *et al*., 2009), microsatellites isolated *de novo* by our group from expressed sequence tags and from genomic sequence for contig 117 of *Populus trichocarpa* genome assembly v.1, homologous to chromosome XIX of *Populus* (Joseph and Lexer, 2008; de Carvalho *et al*., 2010), and single nucleotide polymorphisms as well as insertion–deletion (indel) markers isolated from expressed sequence tags representing candidate genes for traits involved in ecological divergence between *P. alba* and *P. tremula* (Joseph and Lexer, 2008). The *trnC-petN1* plastid DNA region was sequenced in the parents of the mapping cross using essentially the same protocols. All of the marker loci are reported and documented in detail elsewhere (Joseph and Lexer 2008; de Carvalho *et al*., 2010; Lexer *et al*., 2010), with the exception of a small number of new microsatellites from chromosome XIX, the incipient sex chromosome of *Populus*. Because of their special relevance to the objectives of this paper, all chromosome XIX loci are documented once more in detail in the supporting materials of the present paper (Supplementary Table 2).

Forward primers for all nuclear microsatellite markers (86 loci in total; Supplementary Table 1) were M13-tailed, and standard polymerase chain reaction protocols were used for DNA amplification following Lexer *et al*. (2005). Allele sizes were resolved using an Applied Biosystems (ABI) 3100 Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) and FAM and JOE fluorescent dyes. Indel polymorphisms in five expressed sequence tags (Joseph and Lexer, 2008) were genotyped as length polymorphisms, using the same methods. For seven further expressed sequence tags (Supplementary Table 1), single nucleotide polymorphisms were identified by resequencing the parents of the BC_1 cross following Joseph and Lexer (2008), and the progeny was genotyped for each single nucleotide polymorphism using the ABI SNaPshot assay (Applied Biosystems) following the manufacturer's instructions.

Data analysis

Segregation analysis in the interspecific BC_1: Marker alleles segregating from the interspecific F_1 hybrid parent BET3 were analyzed in the interspecific BC_1 progeny. The focus of this study was on patterns of *single marker* segregation–linkage analysis was only used as a quality control (below). Both segregation and linkage analyses were carried out in MAPMANAGER QTX. This program uses χ^2 statistics to test for deviations of marker segregation from Mendelian expectations, equivalent to tests for *gametic* or allelic segregation distortion. In addition, tests for distortions of particular genotypic classes (=*zygotic* distortions) were carried out using χ^2 tests in JOINMAP 3.0.

Species origin of the hybrid parent (BET3) alleles: The parental species frequencies of alleles found in the F_1 hybrid parent (BET3) were estimated from natural populations of *P. alba* and *P. tremula* (see above—Plant materials). Estimating parental frequencies was possible despite the somewhat heterogenous geographic sampling of the parental species (above), because an overwhelming proportion of molecular variance in these European poplars and aspens resides *between* species, rather than between populations of the same species (Lexer *et al*., 2005; de Carvalho *et al*., 2010).

The odds-ratio tests were based on a contingency table constructed using the frequency of each allele in the two species, using proc FREQ in SAS version 9 (SAS Institute Inc, Cary, NC, USA). When the allele frequency was zero or one in any of the species, odds ratios were undefined. In that case, a Fisher's exact test was performed to test for significant differences in the contingency table. For a small number of loci for which the odds-ratio test was not significant, species origin was inferred based on linkage with markers with clearly assigned species origin (Supplementary Table 1; species origin for these markers is given in parentheses). Information about the species origin of each allele in the BET3 hybrid parent was used to determine the direction of segregation distortions in the interspecific BC_1, either toward *P. alba* or toward *P. tremula*.

Synteny with *P. trichocarpa*: Although the main objective of this paper was to study *single marker* segregation distortions, marker coverage on several chromosomes was suitable for defining linkage groups. This was the case because these markers had been picked to tag clusters of quantitative trait loci-controlling species differences in a related project on admixture mapping in *Populus* (Lexer *et al*., 2010). These markers provided a useful quality control for our segregation data, as they allowed us to assess levels of synteny of these linkage groups between our European *Populus* species and the sequenced genome of *P. trichocarpa*.

To examine synteny, linkage groups in the interspecific BC_1 were determined based on log-of-odds likelihood statistics calculated by MAPMANAGER QTX. Marker groupings detected at a log-of-odd threshold of 3.00 were ordered locally by a multi-point analysis using the 'Ripple' command, and map distances in centimorgans were estimated from recombination frequencies using the Kosambi mapping function. Synteny between marker order in the interspecific BC_1 and *P. trichocarpa* was evaluated using *P. trichocarpa* genome assembly v2 (available at http://www.phytozome.net/poplar). Physical positions of all markers in *P. trichocarpa* were determined using blast-n searches of the primer sequences against *P. trichocarpa* genome assembly v2, using an *E*-value threshold of 0.1.

Analysis of chromosome XIX in natural populations: Markers located on the putative sex chromosome XIX were of special interest to this study, because sex chromosomes are often seen as 'hotspots' for species isolation genes. Analyzing natural populations places greater demands on the ease of scoring and robustness of molecular markers than segregation analysis in a simple, controlled pedigree; eight out of nine microsatellites from chromosome XIX (Supplementary Table 2) yielded readily interpretable, fully codominant marker genotypes in the 20 individuals of known sex sampled from natural populations of *P. alba* in Spain (Table 2), and six loci did so across European populations of *both* species, *P. alba* and *P. tremula* (3; Supplementary Table 3).

The loci were characterized via the number of alleles (A), expected (H_E) and observed (H_O) heterozygosity and within-population inbreeding coefficients (F_{IS}) in populations using the computer program FSTAT (Goudet,

1995). LD between pairs of loci was estimated using the common marker correlation and exact *P*-values for LD were computed with GENEPOP (Rousset, 2008). To keep the number of pair wise LD tests manageable, these analyses were restricted to four populations of particular interest to ongoing evolutionary genetics research in these species: Danube (Austria) and Tisza (Hungary) for *P. alba*, and Eastern Alps (Austria) and Central Sweden for *P. tremula*. In addition, the 20 Spanish individuals of known sex were characterized for their diversity and heterozygosity (H_E and H_O).

To yield insights into the role of the putative sex chromosome XIX in blocking gene flow between *P. alba* and *P. tremula*, interspecific genetic divergence for chromosome XIX markers was estimated in the form of F_{ST} and the results confirmed by G'_{ST}, a standardized differentiation measure that takes within-population heterozygosity into account (Hedrick, 2005). To relate interspecific divergence on chromosome XIX to the genome-wide average, F_{ST} values for chromosome XIX were compared with genome-wide expectations for 93 microsatellite- and sequence-based genetic markers reported by Lexer *et al*. (2010).

Results

Marker polymorphism and species origin of donor alleles

Ninety-eight (98) genetic markers (microsatellites, indels and single nucleotide polymorphisms) representing all 19 chromosomes of the *Populus* genome were analyzed in the interspecific BC_1 of *P. alba* and *P. tremula* (Supplementary Tables 1 and 4). Out of these, 39 (40%) were polymorphic in the female F_1 hybrid parent (BET3) only, 11 (11%) were polymorphic in the male *P. alba* backcross parent (J1) only and 26 (27%) were polymorphic in both parents, whereas 22 (22%) were monomorphic in both parents of the cross (Supplementary Table 1). The odds-ratio test and Fisher's exact test facilitated statistical assignment of species origin of alleles segregating from the female F_1 hybrid parent (BET3) for 47 of the markers. For 13 further loci, putative species origin of alleles segregating from the F_1 hybrid parent could be assigned based on linkage to markers with clear species assignments (Supplementary Table 1). With respect to organellar DNA, plastid DNA sequencing revealed a *P. tremula* haplotype for BET3, thus indicating the species origin of the cytoplasm of the interspecific BC_1 cross.

Segregation distortions

Thirteen markers on six chromosomes (20% of polymorphic markers) displayed significant segregation distortion of alleles segregating from the BET3 hybrid parent of the BC_1 (Table 1), compared with three markers expected by chance alone. All of these loci displayed genotypic (= zygotic) segregation distortion as well. For 12 of these loci, the *P. tremula* (= donor) allele was significantly overrepresented in the backcross progeny. Only for 1 of the 13 distorted markers, segregation distortion was against the introgressed *P. tremula* allele (that is the *P. alba* allele segregating from the F_1 hybrid parent was overrepresented instead; Table 1). The results allowed us to discuss the strength of post-mating reproductive barriers between these hybridizing species (see below).

Synteny with *P. trichocarpa*

All detected linkages were conserved between *P. trichocarpa* genome assembly v.2 and the present interspecific BC_1 of *P. alba* and *P. tremula* (Supplementary Table 1). Synteny is best exemplified by chromosome VI, known to exhibit normal levels of recombination (Yin *et al*., 2004), and chromosome XIX, known to exhibit suppressed recombination (Yin *et al*., 2008) (Figure 1). Marker order was completely conserved on chromosome VI, whereas no recombination event was observed between four markers on chromosome XIX in the interspecific BC_1, corresponding to >560 kb on the *P. trichocarpa* physical genome map (Figure 1; see also below).

Segregation distortion and diversity of chromosome XIX

Three loci on the proximal end of chromosome XIX, the incipient sex chromosome of *Populus*, displayed segregation distortion in the female F_1 hybrid parent (BET3) in the form of an overrepresentation of donor alleles from

Table 1 Genetic markers with segregation distortion in an interspecific BC_1 between *P. tremula* and *P. alba*, including chromosome assignment on *P. trichocarpa* genome assembly v.2, significance levels of segregation distortions in the BC_1, identity of the over-represented allele for each locus, odds ratios for parental species assignments of alleles in natural populations and inferred species assignment of the overrepresented allele

Locus	*P. trichocarpa chromosome*	*Distortion P. alba × tremula F1 parent (♀)*	*Overrepresented allele*	*Odds ratio P. alba/ tremula F1 (♀) allele 1*	*Odds ratio P. alba/ tremula F1 (♀) allele 2*	*Species origin of overrepresented allele from F1 (♀)*
GCPM 1274	1	****	2	1.55/0.45	0.93/1.32	(*P. tremula*)
ASP 112376	1	*	1	3.18/0.05	0.00/3.34	*P. alba*
GCPM 124	1	**	2	42.20/0.07	0.11/2.90	*P. tremula*
GCPM 1629	3	*****	1	0.62/5.02	1.24/0.00	*P. tremula*
Thau	9	**	2	2.44/0.00	NA	(*P. tremula*)
ORPM 23	9	*****	1	0.00/3.49	1.61/0.00	*P. tremula*
ORNL 149	10	*	1	0.00/3.8	14.00/0.00	*P. tremula*
ORPM 344	10	****	1	0.00/3.73	2.27/0.25	*P. tremula*
GCPM 1250	10	****	2	4.03/0.04	0.25/23.05	*P. tremula*
GCPM 154	12	*****	1	0.00/2.79	NA	(*P. tremula*)
Yin1	19	**	1	NA	1.28/0.73	*P. tremula*
Yin2	19	*	2	2.64/0.00	0.00/1.90	*P. tremula*
ORPM 206	19	***	1	0.00/4.06	10.27/0.00	*P. tremula*

Abbreviations: BC, backcross; NA, not applicable.
Species assignments supported by the genotype data, but not significant in the odds-ratio test are shown in parentheses.
Significance thresholds from χ^2 tests.
*$P<0.05$, **$P<0.01$, ***$P<0.005$, ****$P<0.001$, *****$P<0.00005$.

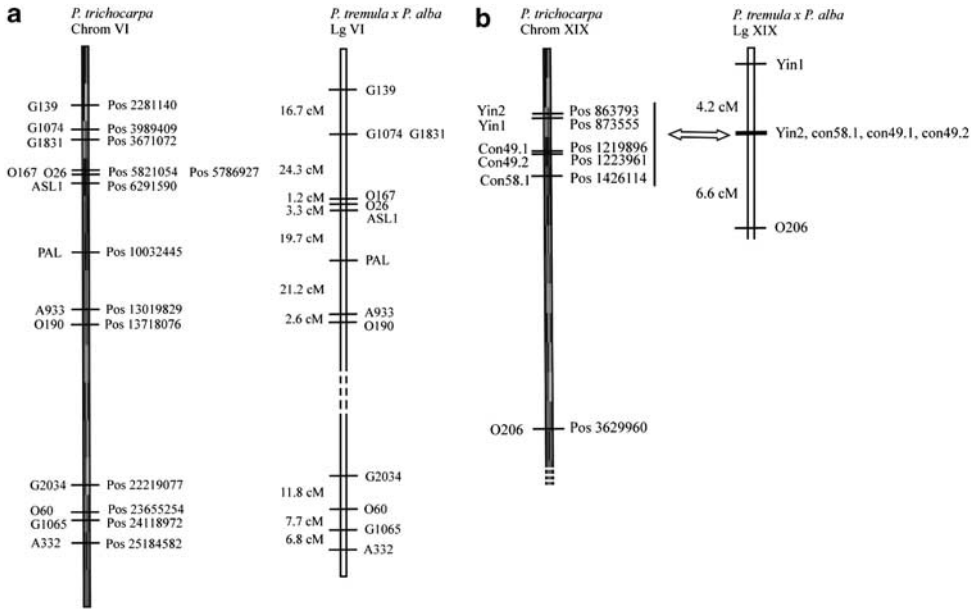


Figure 1 Comparison of the *P. tremula* × *P. alba* linkage map to *P. trichocarpa* genome assembly v.2 for chromosome VI (**a**), known to exhibit normal levels of recombination (Yin *et al*., 2004), and chromosome XIX (**b**), known to exhibit greatly reduced recombination (Yin *et al*., 2008). Complete synteny between the two maps is indicated by the conserved marker order on chromosome VI. On chromosome XIX, zero recombination was observed between markers Yin2, con58.1, con49.1 and con49.2 on the *P. tremula* × *P. alba* linkage map (indicated by the arrow), which corresponds to >560 kb on the *P. trichocarpa* genome assembly.

Table 2 Chromosome XIX diversity statistics for individuals of *P. alba* with known sex, including expected (H_E) and observed (H_O) heterozygosity and inbreeding coefficients (F_{IS}) in each group and the female/male ratio of H_O

Locus	*Females*				*Males*				H_O *ratio* ♀/♂
	A	H_E	H_O	F_{IS}	*A*	H_E	H_O	F_{IS}	
Yin2	4	0.645	0.273	0.589	4	0.692	0.625	0.103	0.437
Yin1	3	0.301	0.333	−0.114	3	0.492	0.625	−0.296	0.533
Con03.1	3	0.610	0.909	−0.527	3	0.689	1.000	−0.539	0.909
Con49.1	5	0.667	0.500	0.258	4	0.592	0.625	−0.061	0.800
Con49.2	5	0.825	0.500	0.411	5	0.842	0.375	0.571[a]	1.333
Con58.1	9	0.892	1.000	−0.128	9	0.858	0.750	0.134	1.333
O206	2	0.290	0.333	−0.158	2	0.325	0.375	−0.167	0.888
O276	5	0.656	0.750	−0.151	5	0.700	0.625	0.114	1.200

[a]Significantly different from zero at the 0.05 level.

P. tremula (Table 1). A survey of trees with known sex indicated autosomal behavior of this chromosome: there were no consistent departures from random mating (measured via F_{IS}) in known females and males and no consistent differences in heterozygosity between known females and males (Table 2). LD on chromosome XIX extended over >560 kb in natural populations of *P. alba* (Table 3), the species with the smaller effective population size, N_e (Lexer *et al*., 2005). The markers in LD corresponded to those loci with zero recombination in the interspecific BC_1 (see Figure 1 and above). LD was also observed in the Swedish population of *P. tremula*, whereas no LD was detectable in *P. tremula* from the Eastern Alps (Table 3).

Congruence between BC_1 segregation patterns and genomic divergence in natural populations

The increased introgression of *P. tremula* (=donor) alleles on chromosome XIX in the interspecific BC_1 (Figure 2a) was mirrored by reduced interspecific divergence in natural populations when measured as F_{ST} (Figure 2b); the median of interspecific F_{ST} was far below the genome-wide expectation of 0.369 reported by Lexer *et al*. (2010). For comparison, no such reduction in interspecific divergence was seen on chromosome VI, consistent with normal segregation on this chromosome (Figure 2). A congruent pattern among chromosomes XIX and VI was recovered when Hedrick's (2005) G'_{ST} was used as a measure of divergence ($G'_{ST} = 0.556 \pm 0.124$ on chromosome XIX vs 0.898 ± 0.049 on chromosome VI, respectively). The results allowed us to compare introgression patterns seen in the BC_1 and patterns of interspecific divergence observed in natural populations.

Discussion

Segregation distortions favor rather than impede introgression in a controlled cross of *P. tremula* and *P. alba*

Segregation distortions of genetic markers in interspecific crosses contain a wealth of information on the strength and genomic architecture of post-mating isolation between species (Fishman *et al*., 2001; Yin *et al*., 2004; Bouck *et al*., 2005; Sweigart *et al*., 2006; Lopez-Fernandez and Bolnick, 2007; Turelli and Moyle, 2007). This can point evolutionary biologists to genetic loci involved in speciation and to possible causes for asymmetries in reproductive barriers (Coyne and Orr, 2004; Turelli and Moyle, 2007) and breeders to genome regions that will resist introgression in marker-assisted breeding programs. Sometimes, however, the direction of segregation distortions in interspecific crosses also suggests the presence of mechanisms that *favor* rather than impede gene exchange (for example Tiffin *et al*., 2001; Yin *et al*.,

Table 3 Linkage disequilibrium (LD) among markers on chromosome XIX in natural populations of *P. alba* and *P. tremula*

Locus[a]	*Yin2*	*Yin1*	*Con49.2*	*Con58.1*	*O206*
(a) P. alba/Austrian Danube valley					
Yin2 (863 793)	—	**0.331***	**0.154***	**0.208***	0.658
Yin1 (873 555)	**0.000***	—	**0.006**	**0.016**	0.532
Con49.2 (1 296 123)	**0.000***	**0.143**	—	0.278	0.587
Con58.1 (1 426 114)	**0.000***	**0.225**	0.148	—	0.363
O206 (3 629 960)	0.107	0.115	0.122	0.478	—
(b) P. alba/Hungarian Tisza valley					
Yin2 (863 793)	—	0.153	**0.168**	0.185	NC
Yin1 (873 555)	0.487	—	0.152	**0.032**	NC
Con49.2 (1 296 123)	**0.008**	0.134	—	**0.008**	NC
Con58.1 (1 426 114)	0.441	**0.160**	**0.165**	—	NC
O206 (3 629 960)	NC	NC	NC	NC	NC
(c) P. tremula/Eastern Alps					
Yin2 (863 793)	—	0.156	0.123	0.118	0.980
Yin1 (873 555)	0.593	—	1.000	0.936	0.510
Con49.2 (1 296 123)	1.000	0.121	—	0.271	0.903
Con58.1 (1 426 114)	0.346	0.118	0.165	—	0.677
O206 (3 629 960)	0.084	0.118	0.113	0.124	—
(d) P. tremula/Sweden					
Yin2 (863 793)	—	0.197	**0.186**	0.192	0.482
Yin1 (873 555)	0.284	—	0.111	0.111	0.784
Con49.2 (1 296 123)	**0.034**	0.166	—	**0.010**	**0.007**
Con58.1 (1 426 114)	0.323	0.188	**0.154**	—	0.723
O206 (3 629 960)	0.101	0.149	**0.142**	0.148	—

Abbreviation: NC, not calculated.
Marker correlations are shown above the diagonal and exact *P*-values are shown below the diagonal. LD extending up to 562 kb (assuming *P. trichocarpa* map distances) is detectable in populations of *P. alba* (a and b), the species with the smaller effective population size (N_e), whereas LD is less readily detectable in populations of *P. tremula* (c and d).
Significant tests at the 0.05 level are indicated by bold type, and significant tests after Bonferroni correction are indicated by an asterisk.
[a]The physical location of each marker in *P. trichocarpa* genome assembly v.2 is shown in parentheses.

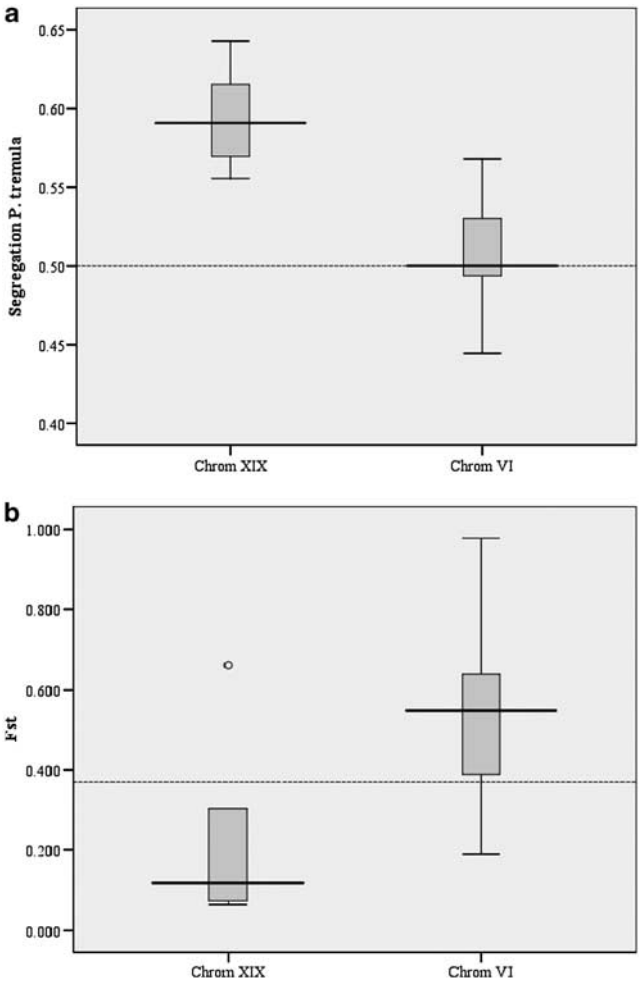


Figure 2 Box plots showing introgression and divergence of genetic markers on chromosomes VI and XIX relative to genome-wide expectations. (**a**) Segregation of *P. tremula* alleles in the interspecific BC_1 (dotted line, Mendelian expectation of 0.5). (**b**) Interspecific divergence (F_{ST}) in natural populations (dotted line, genome-wide expectation of 0.369; Lexer *et al*., 2010). Increased introgression of *P. tremula* alleles in the controlled interspecific BC_1 (**a**) and reduced interspecific divergence in natural populations (**b**) are visible for chromosome XIX.

2004; Bouck *et al*., 2005), a pattern reminiscent of adaptive introgression. These cases are of great interest to our understanding of the genetics of porous species boundaries (Barton, 2001; Tiffin *et al*., 2001; Wu, 2001; Lexer and Widmer, 2008; Widmer *et al*., 2009), but understanding the causes of these patterns is not a trivial task.

Here, we uncovered extensive segregation distortions (20% of polymorphic loci) in a controlled interspecific BC_1 of *P. alba* and *P. tremula*, two naturally hybridizing, ecologically divergent forest trees, and almost all of these distortions (12 out of 13; >90%) *favored* introgression of donor (*P. tremula*) alleles into the heterospecific *P. alba* genomic background (Table 1). These allelic distortions manifested themselves also at the zygotic (=genotypic) level, and high levels of mortality (34%) were observed during early diploid life stages of the progeny (first-year seedlings, a critical life stage in long-lived forest trees). Thus, the observed patterns are most easily explained by three non-exclusive hypotheses: (1) epistatic interactions between the hybridizing genomes (Coyne and Orr, 2004), (2) overdominance or heterozygote advantage in hybrids (Hartl and Clark, 1997), (3) cyto-nuclear coadaptation (Galloway and Fenster, 1999; Futuyma, 2009). The data allow us to address each of these hypotheses in turn.

Hypothesis 1, *epistasis*, refers to a central question in current speciation genetics, namely the relative role of Bateson–Dobzhansky–Muller incompatibilities vs other mechanisms in post-mating RI (Coyne and Orr, 2004; Turelli and Moyle, 2007). Evidence for epistasis comes from a recent analysis of genomic admixture in hybrid zones between these species: heterospecific genomic interactions clearly contribute to steep genomic clines in localities, where these species co-occur (Lexer *et al*., 2010). Nevertheless, epistasis alone is unlikely to generate the observed unidirectional bias in genotype frequencies observed here, consistently favoring alleles of the same species across multiple loci (Fishman *et al*., 2001). In fact, epistasis may be expected to produce the opposite pattern (a bias toward overrepresentation of *P. alba* alleles), so additional mechanisms must be invoked.

Hypothesis 2, *heterozygote advantage* (Hartl and Clark, 1997), is less frequently invoked in speciation genetic studies, but represents a mechanism of great interest in ecological and conservation genetics (Conner and Hartl,

2004). Under this hypothesis, selection favoring interspecific heterozygotes will elevate the frequency of *P. tremula* donor alleles in the backcross. This hypothesis is supported by high mortality (34%) during early life stages of the BC_1 progeny and by the low level of heterozygosity of the *P. alba* backcross parent (only 38%), compared with 67% of heterozygous loci in the F_1 hybrid parent of the backcross. This suggests that increased heterozygosity due to introgression can ameliorate the negative effects of biparental inbreeding (that is of recessive deleterious alleles in homozygous state) in *P. alba*; biparental inbreeding in *P. alba* becomes apparent from the great magnitude of short-range kinship coefficients among individuals (F_{ij}) in recent studies of spatial genetic structure in this species (van Loo *et al*., 2008), and from the extra-ordinary clone sizes of genets of *P. alba* in Southern Europe (Brundu *et al*., 2008; González–Martínez and coworkers, unpublished data).

Hypothesis 3, increased introgression due to *cyto-nuclear interactions* (Galloway and Fenster, 1999; Tiffin *et al*., 2001; Futuyma, 2009), is equally supported by our data: our plastid DNA data indicate a *P. tremula* cytoplasm for the female F_1 hybrid parent of our interspecific BC_1. Maternal inheritance of cytoplasmic genomes implies that all BC_1 progeny will carry cytoplasmic genes from *P. tremula*, so nuclear *P. tremula* alleles segregating in the interspecific BC_1 will be favored by selection in combination with the maternally derived conspecific cytoplasm present in each individual. In effect, cyto-nuclear coadaptation (or genomic conflict between heterospecific cyto-nuclear combinations) will 'pull' *P. tremula* nuclear alleles into a *P. alba* genomic background, resulting in a pattern that resembles, but should not be mistaken with adaptive introgression.

In the absence of reciprocal crosses, it is impossible to reject or accept either of these hypotheses with certainty. The present study was based on a single successful interspecific BC_1 obtained by crossing an individual of *P. alba* with a natural F_1 hybrid. A reciprocal crossing design (each species used as pollen *and* seed donor) would allow us to distinguish between cyto-nuclear and purely nuclear effects (Galloway and Fenster, 1999). While the absence of reciprocal crosses represents a caveat, congruent results (see below) from controlled progeny *and* natural populations indicate the generality of our findings beyond the successful cross used for segregation tests.

The well-documented selection episode in first-year seedlings and the clear consistency with genetic data for natural populations also allow us to interpret our results in the context of those obtained for many other groups of plants (Tiffin *et al*., 2001). Our results suggest that asymmmetries in post-mating barriers in these forest trees may result in introgression rather than the evolution of reinforcement upon secondary contact. In fact, two of the loci with significant overrepresentation of *P. tremula* alleles in the controlled cross (GCPM 1629 on chromosome 3 and ORPM 149 on chromosome 10; Table 1) are already known to exhibit greater than neutral introgression of *P. tremula* alleles in a well-studied natural hybrid zone (Lexer *et al*., 2010). Of course, other mechanisms may be operating in parallel, slowing down introgression and strengthening the 'filter' to interspecific exchange (Coyne and Orr, 2004; Futuyma, 2009; Lexer *et al*., 2010). Thus, our study adds to the ongoing debate regarding the effects of asymmetric barriers on the evolutionary dynamics of species interactions upon secondary contact (Tiffin *et al*., 2001; Coyne and Orr, 2004; Lopez-Fernandez and Bolnick, 2007; Turelli and Moyle, 2007; Veltsos *et al*., 2008; Widmer *et al*., 2009).

The sex determination region of *Populus* is not protected from interspecific gene flow

Sex chromosomes (or sex determination regions more generally) have received considerable attention as 'hotspots' of genes or other genetic factors involved in species isolation (Qvarnström and Bailey, 2009). Particular attention has been paid to the mechanisms underlying 'Haldane's rule,' that is the observation that in hybrid zones between divergent populations or species, the heterogametic sex will often be rare, absent or sterile (reviewed by Coyne and Orr, 2004). More recently, the role of suppressed recombination on sex chromosomes has attracted much attention, triggered by the finding that reduced recombination greatly facilitates the accumulation of speciation genes on sex chromosomes (Qvarnström and Bailey, 2009). Both of these phenomena speak for an important role of sex chromosomes in speciation, and this should manifest itself in reduced introgression and increased divergence of these genome regions in recently diverged species (Nosil *et al*., 2009; Qvarnström and Bailey, 2009).

Our present results for the sex determination region of *Populus* are not consistent with this expectation. The proximal end of chromosome XIX exhibits suppressed recombination in our interspecific BC_1 (Figure 1) and increased LD in natural populations (Table 3); note that LD normally decays within 1 kb in these species (Ingvarsson, 2008; Joseph and Lexer, 2008), and that patterns of LD vary with differences in effective population size (N_e) and metapopulation structure (Lexer *et al*., 2007). Still, interspecific divergence (F_{ST} and G'_{ST}) of this region in natural populations is not greater than elsewhere in the genome (on the contrary, the opposite appears to be the case, see Figure 2 and Results). Hence, this genome region is not protected from interspecific gene flow. Consistent with this observation, genetic markers in this region display significant segregation distortion, consistently favoring *P. tremula* donor alleles in our interspecific BC_1 toward *P. alba* (Table 1). These consistent but counter-intuitive results may be explained by characteristic features that distinguish the *Populus* sex determination region from other, better-studied sex chromosome systems.

First, the *Populus* sex determination region interrogated by our markers on chromosome XIX carries clusters of nucleotide-binding site–leucine rich repeat resistance (R-) genes (Yin *et al*., 2008). The fact that this important functional class of plant R-genes has been amplified in the *Populus* genome to form large clusters (Kohler *et al*., 2008) highlights their functional importance in these trees. If these nucleotide-binding site–leucine rich repeat genes are indeed under balancing selection as widely assumed for plant R-genes (Futuyma, 2009), then introgression may be favored by selection, especially in long-lived forest trees for which levels of standing variation are limited by notoriously low rates of

molecular evolution per unit time (Petit and Hampe, 2006).

A possible alternative explanation for increased introgression on chromosome XIX is based on the known variability of sex determination systems in *Populus*. Whereas genomic data for the North American *P. trichocarpa* are suggestive of a ZW sex chromosome system involving a female-specific chromosomal segment located in the proximal end of chromosome XIX (Yin *et al*., 2008), genetic mapping of the sex locus in *P. alba* and in a cross *P. tremula* × *Populus. tremuloides* indicates the presence of at least two loci-controlling sex in a non-terminal position on this chromosome (Pakull *et al*., 2009; Paolucci *et al*., 2010). Interestingly, the sex locus maps to the female genetic map in *P. alba* (Paolucci *et al*., 2010) and to the male map in *P. tremula* × *P. tremuloides* (Pakull *et al*., 2009), consistent with the presence of two or more sex-controlling loci with different degrees of dominance. This would suggest that the *Populus* sex chromosome is at a very early step of its evolution, in which pairs or groups of sexually antagonistic mutations have accumulated, but full differentiation into heteromorphic sex chromosomes has not yet been achieved (Charlesworth *et al*., 2005). Our data are consistent with this hypothesis, since microsatellites in this region behave like codominant, autosomal markers and show no consistent pattern of reduced heterozygosity in either sex (Table 2). The apparent variation present in the poplar sex determination system (Yin *et al*., 2008; Pakull *et al*., 2009; Paolucci *et al*., 2010) provides a possible alternative explanation for increased introgression of the sex determination region (below).

Conclusions and hypotheses for future work

Our study demonstrates the value of experimental crosses involving natural hybrids of known genomic composition in understanding the genetics of species boundaries and barriers to introgression in *Populus* and other long-lived forest trees. Ongoing studies of natural hybrid zones between *P. alba* and *P. tremula* by our group have started to reveal patterns of genomic admixture and RI in multiple 'replicate' hybrid zone localities of these species across Europe (Lexer *et al*., 2007, 2010), which now provides a basis for picking natural hybrids and parental genotypes for controlled crossing experiments, such as those presented here. Experiments involving reciprocal crosses and both backcrossing directions (toward *P. alba* and *P. tremula*) will provide a more refined picture of post-mating and post-zygotic barriers to gene flow in these ecologically important trees.

With respect to the apparent lack of interspecific isolation of the sex determination region (above), a novel hypothesis has recently been put forward to explain the spread of novel sex-specific genome segments across hybrid zones (Veltsos *et al*., 2008). According to this hypothesis, a new chromosomal sex-determination system can spread across hybrid zones, even though it would normally be selected against within a single, isolated population (Pannell and Pujol, 2009). This apparently is made possible by the interplay between 'identity' disequilibria, commonly observed in hybrid zones, and the sexually antagonistic selection pressures affecting pairs or groups of sex-controlling loci (Veltsos *et al*., 2008; Pannell and Pujol, 2009).

Although Veltsos *et al*. (2008) modeled the specific case of a XY replacing a X0 sex determination system, their model may be more widely applicable to other cases of asymmetric selection pressures in hybrid zones caused by dominance effects (that is 'dominance drive'; Mallet, 1986). The variable genetic architecture of sex determination in poplar (Yin *et al*., 2008; Pakull *et al*., 2009; Paolucci *et al*., 2010) could provide a suitable substrate for this type of process. Tests of the mode of spread and ecological impact of the sex determination region in hybrid zones between these species are currently underway.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Joan Cottrell, Pär Ingvarsson, Berthold Heinze and Denes Bartha for help with obtaining samples from natural populations of poplars and aspens, the International Populus Genome Consortium (IPGC) for making genome assembly and primer sequences publicly available and Robyn Cowan and Dulcineia de Carvalho for help in the laboratory. CL's research on the evolutionary genomics of species barriers in *Populus* is supported by Grant no. NE/E016731/1 of the UK. Natural Environment Research Council (NERC) and Grant no. 31003A_127059 of the Swiss National Science Foundation (SNF).

References

Barton NH (2001). The role of hybridization in evolution. *Mol Ecol* **10**: 551–568.

Bouck A, Peeler R, Arnold ML, Wessler SR (2005). Genetic mapping of species boundaries in Louisiana irises using IRRE retrotransposon display markers. *Genetics* **171**: 1289–1303.

Brundu G, Lupi R, Zapelli I, Fossati T, Patrignani G, Camarda I *et al.* (2008). The origin of clonal diversity and structure of *Populus alba* in Sardinia: evidence from nuclear and plastid microsatellite markers. *Ann Bot* **102**: 997–1006.

Charlesworth D, Charlesworth B, Marais G (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118–128.

Conner JK, Hartl DL (2004). *A Primer of Ecological Genetics*. Sinauer Associates, Sunderland, USA.

Coyne JA, Orr HA (2004). *Speciation*. Sinauer Associates: Sunderland, MA.

De Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D *et al.* (2010). Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Mol Ecol* **19**: 1638–1650.

Emelianov I, Marec F, Mallet J (2004). Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proc Biol Sci* **271**: 97–105.

Fishman L, Kelly AJ, Morgan E, Willis JH (2001). A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**: 1701–1716.

Fussi B, Lexer C, Heinze B (2010). Phylogeography of *Populus alba* (L) and *Populus tremula* (L.) in Central Europe: secondary contact and hybridisation during recolonisation from disconnected refugia. *Tree Genet Genomes* **6**: 439–450.

Futuyma DJ (2009). *Evolution*, 2nd edn. Sinauer Associates: Sunderland, MA.

Galloway LF, Fenster CB (1999). The effect of nuclear and cytoplasmic genes on fitness and local adaptation in an annual legume, *Chamaecrista fasciculata*. *Evolution* **53**: 1734–1743.
Goudet J (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. *J Hered* **86**: 485–486.
Hartl DL, Clark AG (1997). *Principles of Population Genetics*. Sinauer Associates: Sunderland, MA.
Hedrick PW (2005). A standardized genetic differentiation measure. *Evolution* **59**: 1633–1638.
Ingvarsson PK (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula. *Genetics* **180**: 329–340.
Joseph JA, Lexer C (2008). A set of novel DNA polymorphisms within candidate genes potentially involved in ecological divergence between *Populus alba* and *P.* tremula, two hybridizing European forest trees. *Mol Ecol Resources* **8**: 188–192.
Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F *et al.* (2008). Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol Biol* **66**: 619–636.
Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007). Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* **98**: 74–84.
Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B (2005). Barrier to gene flow between two ecologically divergent *Populus* species, *P alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Mol Ecol* **14**: 1045–1057.
Lexer C, Joseph JA, van Loo M, Barbara T, Heinze B, Bartha D *et al.* (2010). Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**: 699–712.
Lexer C, Joseph J, van Loo M, Prenner G, Heinze B, Chase MW *et al.* (2009). The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. *Taxon* **58**: 349–364.
Lexer C, Widmer A (2008). The genic view of plant speciation: recent progress and emerging questions. *Philos Trans R Soc Lond B Biol Sci* **363**: 3023–3036.
Lopez-Fernandez H, Bolnick DI (2007). What causes partial F1 hybrid viability? incomplete penetrance versus genetic variation. *PLoS One* **2**: e1294.
Mallet J (1986). Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning color clines. *Heredity* **56**: 191–202.
Nosil P, Funk DJ, Ortiz-Barrientos D (2009). Divergent selection and heterogeneous genomic divergence. *Mol Ecol* **18**: 375–402.
Pakull B, Groppe K, Meyer M, Markussen T, Fladung M (2009). Genetic linkage mapping in aspen (*Populus tremula* L and Populus tremuloides Michx). *Tree Genet Genomes* **5**: 505–515.
Pannell JR, Pujol B (2009). The paradoxical spread of a new Y chromosome-a novel explanation. *Trends Ecol Evol* **24**: 59–63.
Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E *et al.* (2010). Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genet Genomes* **6**: 863–875.
Petit RJ, Hampe A (2006). Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst* **37**: 187–214.
Qvarnström A, Bailey RI (2009). Speciation through evolution of sex-linked genes. *Heredity* **102**: 4–15.
Rousset F (2008). GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resources* **8**: 103–106.
Smulders MJM, Van Der Schoot J, Arens P, Vosman B (2001). Trinucleotide repeat microsatellite markers for black poplar (*Populus nigra* L). *Mol Ecol Notes* **1**: 188–190.
Sweigart AL, Fishman L, Willis JH (2006). A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics* **172**: 2465–2479.
Tiffin P, Olson MS, Moyle LC (2001). Asymmetrical crossing barriers in angiosperms. *Proc R Soc Lond B* **268**: 861–867.
Turelli M, Moyle LC (2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**: 1059–1088.
Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U *et al.* (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. *and* Gray). *Science* **313**: 1596–1604.
van der Schoot J, Pospiskova M, Vosman N, Smulders MJM (2000). Development and characterization of microsatellite markers in black poplar (*Populus nigra* L). *Theor Appl Genet* **101**: 317–322.
van Loo M, Joseph JA, Heinze B, Fay MF, Lexer C (2008). Clonality and spatial genetic structure in *Populus x canescens* and its sympatric backcross parent *P alba* in a Central European hybrid zone. *New Phytol* **177**: 506–516.
Veltsos P, Keller I, Nichols RA (2008). The inexorable spread of a newly arisen neo-Y chromosome. *PLoS Genet* **4**: e1000082.
Widmer A, Lexer C, Cozzolino S (2009). Evolution of reproductive isolation in plants. *Heredity* **102**: 31–38.
Woolbright SA, DiFazio SP, Yin T, Martinsen GD, Zhang X, Allan GJ *et al.* (2008). A dense linkage map of hybrid cottonwood (*Populus fremontii* × *P. angustifolia*) contributes to long-term ecological research and comparison mapping in a model forest tree. *Heredity* **100**: 59–70.
Wu C-I (2001). The genic view of the process of speciation. *J Evol Biol* **14**: 851–865.
Yin TM, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA (2004). Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theor Appl Genet* **109**: 451–463.
Yin TM, DiFazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA *et al.* (2008). Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res* **18**: 422–430.
Yin TM, Zhang XY, Gunter LE, Li SX, Wullschleger SD, Huang MR *et al.* (2009). Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome. *New Phytol* **181**: 498–503.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)

Annex D

Causes and consequences of large clonal assemblies in a poplar hybrid zone.

MACAYA-SANZ, D., HEUERTZ, M., LINDTKE, D., LEXER, C., GONZÁLEZ-MARTÍNEZ, S. C, 2014. Manuscript.

Causes and consequences of large clonal assemblies in a poplar hybrid zone

David Macaya-Sanz[1], Myriam Heuertz[12], Dorothea Lindtke[2], Christian Lexer[2], Santiago C. González-Martínez[1,*]

[1]Department of Forest Ecology and Genetics, INIA-Forest Research Centre, 28040 Madrid, Spain.
[2]Unit of Ecology and Evolution, Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland.

Reference numbers:

Running head: 'Evolutionary consequences of large clones' or 'Large clonal assemblies in poplar'

Key words:

*Corresponding author:
Santiago C. González-Martínez
INIA, Forest Research Centre,
Carretera de A Coruña km 7.5,
28040 Madrid, Spain
Ph +34 913471499
Fax +34 913572293
santiago@inia.es

Abstract

Clonal propagation is a common and fundamental mode of reproduction in plants. Although persistence in adverse conditions underlies most known cases, proximal ecological and genetic factors of clonal spread remain still unclear, in particular for populations dominated by a few large clonal assemblies. In the same way, the evolutionary consequences of clonal dominance by a few genets are not well known. In this paper, we studied a known clonal population of the riparian tree *Populus alba* in the Douro river basin (northwestern Iberian Peninsula) where low levels of hybridization with *P. tremula* occur, analyzing its genomic structure with a panel of 73 SSR markers. We uncovered that a few ancient (over a few thousand years old) and widespread genets dominate the population, both in terms of clone size and number of sexual offspring. Interestingly, large genet genomes showed local regions of increased interspecific heterozygosity and a reduced overall inbreeding. Heterosis and heterozygote advantage often favor overall fitness and could have also favored clonal spread in this region. At the population level, clonal spread by large genets was accompanied by an overall ancient (>0.1 Myr) but soft decline of effective population size. Despite such a decay, high clonal rates (and low clonal diversity), and sexual reproduction bias, the Douro hybrid zone still displays high genetic diversity at the genet-level, compared to less clonal hybrid zones from Central Europe. In conclusion, interspecific introgression, heterozygote advantage, and sexual prevalence happen to influence clonal population genetics and demography, fostering population changes of great relevance to enhance or reduce resilience in the face of environmental changes.

Introduction

Clonal reproduction is common in plant species, involving ~80% of angiosperms (KLIMES *et al.* 1997). Repeated evolution of clonal or partially clonal plants suggests that asexual reproduction is easily acquired through minor modifications of widespread plant traits (SACHS 2001). Nonetheless, the array of selective pressures, genetic factors and ecological conditions that promote asexual reproduction, in particular the existence of large and ancient clonal assemblies in some taxa, remains poorly known. Although various alternative explanations have been proposed (VALLEJO-MARIN *et al.* 2010), assurance of plant persistence in unpredictable (e.g. desert phreatophytes; VONLANTHEN *et al.* 2010) or recently-colonized (e.g. invasive species; BUDDE *et al.* 2011; LIU *et al.* 2006) environments seems to underlie most cases of clonal plant spread studied to date. Foraging and the division of labor among ramets are other factors that can favor clonality (VALLEJO-MARIN *et al.* 2010; and references therein). Depending on the causes that hamper sexual reproduction (shortage of mates vs. environmental stress), clonal reproduction can be favored by the presence of self-compatibility, as long as clonal self-compatible species (or populations) are more diverse in genotypes (HONNAY and JACQUEMYN 2008). This is expected when pollen limitation costs exceed geitonogamy losses (i.e. losses due to mating between ramets). On the other hand, self-incompatible and dioecious species don't suffer of geitonogamy but are more sensitive to costs of pollen limitation, since potential partners are halved, which is not solved by clonality. Thus, asexual reproduction should only be favored under sexual reproduction difficulties due to environmental stress. Once asexual reproduction has been established, it can further reduce sexual reproduction due to competition of (normally vigorous) sprouts with sexual propagules, accumulation of somatic mutational load on fertility traits (ALLY *et al*. 2010; ECKERT 2001), and increased geitonogamy (BARRETT 2002). The existence of large clonal assemblies exacerbates these processes. Hence, sexual recruitment limitation is both a cause and a consequence of clonal spread and can create a detrimental feedback circle that may end in eventual extinction of sexual function (HONNAY and BOSSUYT 2005).

Asexual reproduction favors local persistence at specific times. Theoretical models, however, predict that selection should favor sex and recombination at evolutionary scales (Otto and Lenormand 2002). In agreement with theoretical predictions, Rice (2002) found medium-term propagation success of clonal species, albeit common extinction of the phylogenetic branches derived from them. He also collected empirical evidence of better performance of sexual lineages in long-term scenarios. An exception to these trends can be found in the Salicaceae family. Within this family, the section Populus (e.g. Populus alba and P. tremuloides; (Brundu et al. 2008; Slavov and Zhelev 2010), includes the largest natural terrestrial clones currently known (Mock et al. 2008; and references therein). In these species, clonal levels are variable depending on the population and geographical range. For example, a large Central European population of white poplar (P. alba) displayed a modest G/N ratio, R, of 0.76 (where G is the number of genets and N the number of ramets) (van Loo et al. 2008) while some populations from Malta and Sardinia islands were found to be

monoclonal (with G/N < 0.05), and in the whole Sardinia island only 26 genets were detected (G/N = 0.16) (Brundu et al. 2008; Fussi et al. 2012). A survey on the phylogenetically related North American aspen (P. tremuloides) showed G/N ratios of 0.11 and 0.23, and clonal assemblies that extended over one kilometer (Mock et al. 2008).

Clonal structure across populations can be typified by its clone-size and spatial distributions. ARNAUD-HAOND *et al.* (2007) described typically exponential distributions of clone size in most species and populations, with few large and many small genets. Only rarely colossal clonal sizes or extreme uneven size distributions have been found. Indeed, ARNAUD-HAOND *et al.* (2007) survey of clonal studies did not report any clone larger than 1 km. Clonal structure is often represented with a Pareto distribution. Consequently, quantitative variation across clonal populations can be depicted by the strength of curve decay, and explained by the slope of the log-scaled Pareto distribution, besides the genotype richness and the evenness. Ecologically, these parameters depend on the intensity of the dominance of the larger clones over the smaller. With respect to the spatial distribution of ramets, two main patterns have been described, dubbed as 'phalanx' and 'guerrilla' growth types, depending on whether different genets intermingle spatially (guerrilla) or not (phalanx). The prevalence of each strategy is determined by the spread mechanisms in each ecological context. For example, branch dropping in a riparian clonal species would produce a 'guerrilla' pattern, whereas sucker sprouts from short roots would create 'phalanx' patches. Intraspecific competition among genets and with sexual propagules is also a main factor that affects clonal structure across populations (ANGELES *et al.* 2011).

The outcome of competition among genets may depend on ecological and adaptation factors, as well as genetic ancestry of competing clones. Firstly, specific ecological conditions appear to affect clonal structure. In the riparian black cottonwood (*Populus trichocarpa*), flood control and soil moisture have crucial roles on sexual recruitment, with a drier population having half the genet richness than a more humid one (SLAVOV *et al.* 2010). In the phreatophyte *Populus euphratica*, clone size increased with distance to ground water (VONLANTHEN *et al.* 2010). The level of moisture seems to be related to the rate of effective sexual recruitment in *Populus alba* too (GONZALEZ *et al.* 2010a). Secondly, there is genetic variation for traits involved in asexual reproduction, which results in better vegetative propagation of some genotypes in control conditions (e.g. greenhouse experiments in *Populus*; STENVALL *et al.* 2005; YU *et al.* 2001). Whether this fact is relevant to influence clonal structure in natural conditions remains unexplored. Thirdly, genetic ancestry could also explain differences in competitive ability across genets. Clones with higher heterozygosity, including those introgressed by closely-related species, may have higher success in natural populations (i.e. heterosis and/or heterozygote advantage). For example, in a Central European hybrid zone of *Populus alba* and *P. tremula*, hybrid genets (named *P. x canescens*) were larger (VAN LOO *et al.* 2008). A classical study on the hybrid sunflower species *Helianthus anomalus* reported 17 transgressive morphological traits, many of them related to growth (SCHWARZBACH *et al.* 2001). Hybridization can indeed initiate new stocks for speciation (NOLTE and TAUTZ 2010). The mechanisms that underlie hybrid

advantage are complex, although it is known that several independent genetic components are involved (BAACK and RIESEBERG 2007; LIPPMAN and ZAMIR 2007). In addition, more heterozygous individuals are also expected to have higher fitness (BRITTEN 1996; CHAPMAN *et al.* 2009). Finally, dominance of some genets may result from opportunity; larger clones may be so just because they were in the right place at the right time.

The effects that prolonged clonal propagation has on genetic and genotypic diversity at the population level are not clear. On the one hand, competition among genets can result in the loss of the less-adapted genotypes, directly reducing the number of genets through time. Still, changes in genotypic diversity may not necessarily affect genetic (allelic) diversity or richness. On the other hand, persistence conferred by clonal propagation extends generation time, slowing the effects of genetic drift. MOCK *et al.* (2008) argued that sexual reproduction, despite its rarity, was an important genetic process in oligoclonal aspen populations, in particular for the maintenance of diversity. However, some theoretical studies state that high levels of clonality has not a negative impact on genetic diversity or sexual performance (BALLOUX *et al.* 2003). Moreover, in some species, low genotypic diversity occurs ($R = 0.31$) without a remarkable loss of diversity, as shown in *Populus trichocarpa* (SLAVOV *et al.* 2010). The long-term consequences of a reduction of genotypic diversity (in cases where genetic diversity is still maintained) are not well-explored. The loss of genotypic diversity could be accompanied by losses in functional diversity and/or co-adapted gene complexes, with unknown consequences for future adaptation to new environments.

Natural hybrid zones of European poplars (*P. alba* and *P. tremula*) are outstanding natural laboratories to study the influence of genomic background in speciation (LEXER *et al.* 2010), and may also be a good system to study whether genetic ancestry underlies differences in clonal spread among genets. While ecological causes for high levels of clonality have been investigated in different poplar species (SLAVOV *et al.* 2010; VONLANTHEN *et al.* 2010), the role of genetic ancestry (in particular introgression and hybridization) remains unexplored despite early evidence of differences between natural hybrids and parental species in European poplars (e.g. VAN LOO *et al.* 2008). Moreover, agronomic studies in these species have shown the impact that hybridization has on many production traits, including growth (e. g. MARRON *et al.* 2010), and naturalists have observed its influence in clonal propagation (SCHWEITZER *et al.* 2002; VAN LOO *et al.* 2008). Furthermore, clonal levels are highly variable across European poplar hybrid zones, from relatively low in Central Europe (largest clone size < 200 m; VAN LOO *et al.* 2008) to large clones spreading over 150 km in the Iberian Peninsula (SANTOS-DEL-BLANCO *et al.* 2012), which allows for comparative approaches. Thus, this species complex represents an interesting case study to provide insights on the causes and consequences of extreme asexual reproduction for long-term population persistence in long-lived species.

In this study, we focus on the white poplar hybrid zone of the Douro river (northwestern Iberian Peninsula), which is characterized by low levels of hybridization with *P. tremula*, with both first-generation hybrids and backcrosses, and the existence of large and geographically extended clonal assemblies (SANTOS-DEL-BLANCO *et al.*

2012). Low-coverage genome-wide genotyping (73 nuclear microsatellites) was used to identify clones and date their age (as in ALLY *et al.* 2008), as well as to discern their genetic background (as in LINDTKE *et al.* 2012). We then (i) analyzed the population genetic structure, and the spatial distribution of large clonal assemblies; (ii) elucidated the effects of genetic background (species ancestry and inbreeding) on clonal success; (iii) assessed the impact of large clones on the overall population demography; and (iv) evaluated the impact of clone size on reproductive fitness and population genetic and genotypic diversity. Finally, we compared the Douro hybrid zone (dominated by large clones) with two other European poplar hybrid zones where sexual reproduction prevails (CASTIGLIONE *et al.* 2010; LEXER *et al.* 2010; VAN LOO *et al.* 2008), using a similar molecular marker panel.

Material and Methods

Study site and sample collection

The sampled Douro poplar hybrid zone comprises a large focal area in the middle course of the Douro river, a major river that flows westwards through the Iberian Northern Plateau to the Atlantic Ocean. The riverbanks in this area are covered by a rather continuous and extensive riparian forest that develops along the main course and its greatest tributaries. This riparian area is well preserved, with zones of low human intervention, and a moderately high floral diversity. It is mainly formed by mixed stands of the genera *Populus*, *Salix*, *Fraxinus*, *Ulmus* and *Alnus*. Because of its ecological value, the area has been proposed as site of Community importance by the European Commission (codes: ES4170083 and ES4120068). White, grey and black poplars (*Populus alba*, *P. x canescens* and *P. nigra*, respectively) naturally grow there, as well as some small stands of European aspen (*P. tremula*). Additionally, the surrounding area was sampled less intensively to fully evaluate the spread of large clones in the region (Figure 1). This area included the higher course of Douro river and some tributaries, riverbanks of other rivers unconnected with Douro river in this area, and mountain ranges.

Leaf tissue was collected from 533 poplar trees, whose position was recorded with a GPS device. The sampling comprised two pure species (*P.alba and P. tremula*) and its natural hybrid species (*P. x canescens*). White (N=362) and grey (N=143) poplar samples came mainly from the focal population extending over 125 km, although some farer away stands were also sampled. European aspen (N=28) samples were collected from stands located mainly in the surrounding mountain systems where this species is more abundant. We spaced the collection at least 100 meters in the core area (Douro middle course and major tributary mouths), and 1,000 meters in the surrounding area. The spacing was set to sample a large geographical area while avoiding overrepresentation of local genets, following previous studies (MACAYA-SANZ *et al.* 2012).

Nuclear microsatellites

DNA was isolated from ground dry tissue using the Invisorb® DNA Plant HTS 96 Kit (STRATEC Molecular, Berlin, Germany), following the producer's protocol. The whole set of samples was initially genotyped with 20 SSR markers normally employed in other populations surveys in the genus *Populus* (see Table S1 in Supplementary Material; LEXER *et al.* 2005). Once Multi-Locus Genotypes (MLGs) and Multi-Locus Lineages (MLLs; i.e. those ignoring somatic mutations) were resolved, at least one sample of each MLL was genotyped (137 samples) with a larger set of 53 SSR markers spaced evenly along the poplar genome (Table S1 in Supporting Material).

The first set of markers (20 SSRs) was amplified following precisely the protocols described by MACAYA-SANZ *et al.* (2012). PCR reactions were run in a 4300 DNA Analyzer (Li-Cor Biosciences, Lincoln, NE, USA), using internal standards to facilitate the scoring. The second set of markers (53 SSRs) was resolved using protocols

described by LEXER *et al.* (2005). Briefly, an M13 tail was attached to forward primers, and fragments were amplified using a touchdown PCR reaction. Allele sizes were resolved using an Applied Biosystems (ABI) 3100 Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) and FAM and JOE fluorescent dyes.

Characterization and spatial distribution of MLGs/MLLs

The first set of 20 loci allowed for initial MLG assignment, as provided by GIMLET software (VALIERE 2002). A further analysis of divergence among MLGs (ARNAUD-HAOND *et al.* 2007) permitted to collapse somatic variants into sexually derived genets (i.e. MLL assignment). The rationale of this assignment method rests in the fact that the divergence among MLGs derived from asexual replication should be much lower (i.e. involving only few somatic mutations) than sexually-produced MLGs. When pairwise genetic distance between MLGs distributes bimodally, distance between peaks is ample, and the smaller peak represent the shorter genetic distances, close MLGs must be grouped under the same MLL, as they are unlikely to be the product of sexual recombination.

Once MLGs were collapsed into MLLs, we plotted and visually analyzed the spatial distribution of MLLs, and MLGs, searching for any conspicuous pattern. We computed some main descriptors of within-species population genetics at MLL level. SPAGeDi 1.3 software (HARDY and VEKEMANS 2002) was used to estimate genetic diversity (measured as expected heterozygosity, H_E), inbreeding coefficient at the population level (F_{IS}), and genetic differentiation between white and grey poplars (F_{ST}). HP-Rare 1.0 (KALINOWSKI 2005) was used to calculate total and rarified allelic richness, and rarified number of private alleles. We also calculated the following descriptors of clonal structure: genotypic richness (R), Simpson's evenness (V), and the additive inverse of the slope of the log-scaled Pareto distribution (β), which correspond to the parameter combination that most completely describes clonal diversity (ARNAUD-HAOND *et al.* 2007). Further, we calculated the same parameters for hybrid zones in the Danube (between Krems, Austria, and Bratislava, Slovakia) and Ticino (Lombardy, Italy) rivers, retrieving raw data from previously published studies (CASTIGLIONE *et al.* 2010; LEXER *et al.* 2010; VAN LOO *et al.* 2008). Geographical representation of the whole population and the most representative genets (e.g. the largest clonal assemblies) was carried out with a GIS (ArcMap version 9.2; ESRI, Redlands, CA, USA).

Dating of genets

The number of somatic mutations within a genet is correlated with its age. Accumulation of somatic mutations within clones is equivalent to mutation accumulation in other systems with suppressed recombination such as sexual chromosomes or some bacterial groups. Although the scope of this paper is not to determine precisely genet age, the high genotyping effort permits its approximation with acceptable confidence level. ALLY *et al.* (2008) developed a method to calculate genet age in *Populus tremuloides*, taking into account genetic divergence within genets.

In a first step, the comparison of the average number of pairwise differences per locus (π_k) with two measures based on the number of segregating sites, S_k, can convey the most adequate genet demographical model. Under a constant-size model, a 1:1 relationship between π_k and θ_s [calculated as $S_k/(\sum_{i=1}^{n-1} 1/i)$] is predicted, whereas under genet growth, π_k correlates 1:1 with $2S_k/n$ (n is the number of sampled ramets). Once a demographical model is chosen, the *TMRCA* (Time since the Most Recent Common Ancestor) can be calculated, which in case of growing-size genets (see Results) follows the equation $\pi_k = 4\ \mu\ TMRCA$, where μ is the mutation rate per generation. This method was applied here to estimate the age of the largest clones in the population (MLL009, MLL006, MLL025, MLL049 and MLL074; see Results) using DnaSP v5 (LIBRADO and ROZAS 2009) for parameter computation. Microsatellites were coded as pseudo-sequences with each locus corresponding to a nucleotide site and each length mutation to a nucleotide point mutation, regardless of the length differences between the original and the mutated alleles. This method was applied to the set of 20 markers that were used to genotype all the sample (i.e. all the ramets), due to it requires the proportion of mutations within all the ramets of eachs MLL. Moreover, only MLLs with more than two ramets and at least one somatic mutation were analysed. Using the remaining MLLs could bias the correlation.

To reduce uncertainty in the estimation of genet age, a second estimation method was employed following THOMSON *et al.* (2000). The accumulation of mutations since the common ancestor in different ramet lineages within a genet follows a Poisson process. Hence, the number of occurred mutations follows a Poisson distribution. The large number of markers genotyped in ramets from the same genet allows then to calculate the within-genet average number of mutations with low error. This value is equal to $\mu_{loc}\ TMRCA$ (THOMSON *et al.* 2000), where μ_{loc} is the mutation rate scaled to twice the number of loci (THOMSON *et al.* 2000). Mutation rate per year was obtained from MSVAR (see below) considering a generation time of 40 years. The ancestral state of somatic mutations was inferred considering allele frequencies. An Infinite Allele Mutation (IAM) model was also assumed, which is reasonable attending the much larger number of loci than mutations. This method was used to estimate the age of the two largest clones, MLL006 and MLL009, based on seven and eight ramets, respectively, in a conservative subset of 66 loci, where less readable markers were discarded.

Genetic background

Genomic ancestry of each genet was evaluated considering both specific homozygosity and interspecific heterozygosity (as defined in LINDTKE *et al*. 2012) based on locus specific ancestries (LSAs) estimated using STRUCTURE vs. 2.1 software (FALUSH *et al.* 2003). Calculations steps followed LEXER *et al*. (2010) and LINDTKE *et al*. (2012). LSAs were plotted using the R package INTROGRESS (GOMPERT and BUERKLE 2010). We only used markers with a known genomic position (listed in Table S1).

Individual-level inbreeding was estimated by computing intra-individual kinship (F_i) using SPAGeDi 1.3 (HARDY and VEKEMANS 2002) and LOISELLE *et al.* (1995)

relative kinship coefficient. To avoid bias due to potential overrepresentation of alleles from large clonal assemblies in the population (which produced more offspring; see Results), we used reference allele frequencies considering only unrelated individuals, as obtained by pedigree reconstruction using FRANz (see below). Besides, intra-individual kinship was calculated for each species (*P. alba*, *P. tremula* and *P. x canescens*) independently, using its own conspecific allele frequency as reference. Singular patterns of specific genetic ancestry and inbreeding in large clones were identified by comparison with the whole population using binomial distributions. In this regard, we considered the number of ramets as a valid proxy for clonal propagation success.

Sexual success of large clones

Family relationships among genets were estimated using COLONY 2.0 (JONES and WANG 2010) and FRANz 2.0.0 (RIESTER *et al.* 2009; RIESTER *et al.* 2010). While the COLONY approach employs a maximum likelihood method to assess relatedness, not considering in any way the actual genet size, FRANz considers the amount of ramets found in each genet as prior information. Three parallel medium-length runs (termed so by the authors) were carried out on COLONY, considering allelic dropout and genotyping error rates of the marker of 0.01, no inbreeding (as the species are dioecious), the full-likelihood method with medium precision and without allele frequency updating. We set the prior probability of finding a father or mother in the population to 0.5, but did not give prior information about sibship size. Besides, all the genets were considered as offspring and candidate mothers and fathers (gender of many genets was unknown). Since true age of genets was neither known, COLONY inferences could have an issue with the directionality of parentage relationship (i.e. which is the parent and which is the offspring in pairwise relationships). Anyway, parents with several offspring, in which we are more interested, should not suffer from this flaw. For FRANz runs, we considered a maximum number of fathers in the population of 1,000,000 (i.e. equivalent to infinite), genotyping error of 0.01 and 20,000 simulation iterations. Number of ramets was included as a prior. Both pedigree reconstructions were computed employing a subset of 56 loci, discarding those that were physically linked or had low polymorphism (Table S1). Finally, we tested for correlations between sexual offspring number and genet size (measured as number of ramets and as spatial extension, i.e. geographical maximum distance among ramets). Significance in different subsets of genets was estimated by Spearman rank tests.

Population size trends

Demographic history was investigated using MSVAR vs. 1.3 software (BEAUMONT 1999; STORZ and BEAUMONT 2002). MSVAR infers the most likely demographic scenario of a population of chromosomes by comparison with iterated Monte Carlo Markov Chains. Both linear and exponential models of population size change were considered. Changes in short periods are prone to be proportional to population size, thus exponential curves are expected to fit better. However, changes along a longer term

are more related to environmental or evolutionary shifts, which often behave linearly (BEAUMONT 1999). Only microsatellites with perfect repeats under a Stepwise Mutation Model (44 loci; listed in Table S1 in Supplementary Material) and one ramet for each genet were used in MSVAR runs. Five MCMC independent simulations were performed for each scenario, with 4.5E9 iterations each, of which 2E9 were treated as burn-in. Priors and hyperpriors were determined by a series of preliminary runs, following the instructions of the software developers: starting current and ancestral population size was 1E4 and starting mutation rate was 1E-4 for all loci; starting time since decline or expansion was 3E6. Prior distributions were all rectangular with the following moments (in logarithmic scale): both population size distributions had a mean 4 and variance of 2.5; mutation rate distributions had a mean of -4 and a variance of 2; and time since decline or expansion distributions had a mean of 6.5 and a variance of 2.5. Generation time was set at 40 years, following MACAYA-SANZ *et al*. (2012). We allowed the program to update the values of the starting parameters. Finally, CODA package (PLUMMER *et al.* 2006) in R environment (R Development Core Team 2009) was used to summarize MCMC outputs and evaluate chain convergence using Gelman-Rubin statistic (GELMAN and RUBIN 1992).

Results

Large and ancient clonal assemblies characterize the Douro poplar hybrid zone

Nuclear microsatellites (20 loci) resolved 132 multilocus genotypes (MLGs) among 533 samples in the Douro poplar hybrid zone, of which 96 were of white poplar (out of 362 samples), 19 were of the hybrid *P. x canescens* (out of 143 samples), and 17 were of European aspen (out of 28 samples). Nonetheless, when putative somatic mutations were taken account, only 82 multilocus lineages (MLLs) were reckoned in white poplar and 13 in the hybrid individuals, while each MLG was assigned to a different MLL in European aspen. The low number of MLLs was owed to the presence of a few clones of colossal sizes (Table 1 and Figure 1): MLL009, a white poplar represented by 189 ramets, spanning 99.5 km (Figure S1a); MLL006, a grey poplar with 124 ramets, spanning 158.6 km (Figure S1b); and two other somewhat smaller white poplar genets (MLL025 with 26 ramets and MLL073 with 17 ramets; Figures S1c and S1d). We have to remark that the number of sampled ramets is correlated with the size of the genet, but it is not the only factor that influences it. A genet with a guerrilla-like growth (e.g. MLL025) can have a similar number of ramets as a genet with a phalanx-like growth (e.g. MLL073), but the former would probably be more spread than the latter (as it occurs in our case, 74.6 km of maximum distance between ramets vs. 5.6 km).

Genotypic (clone) diversity of the Douro population was much lower and more uneven than in the reference populations in the Danube and Ticino river valleys (Table 2). In white poplar, genotypic richness in this Iberian population was between a third and half that found in the Danube and Ticino populations (R = 0.224 vs 0.760 and 0.455, respectively). The distribution was also more uneven (V = 0.539 vs 0.891 in the Danube population), and did not behave similarly to a Pareto distribution, unlike the Danube population (r^2 = 0.721 vs 0.911). Differences in sampling scheme should have favored higher values of clonal diversity in the Douro population than in the others, as spacing between samples was at least 100 meters, approximately tenfold higher than in the reference populations; thus, comparatively, low levels of clonal diversity in the Douro population should be considered even more extreme. In grey poplar, we found lower values of clonal diversity than in white poplar for the Iberian population (R = 0.085 vs 0.224) due to the existence of a large hybrid clone MLL006 (124 ramets out of 143 hybrid samples).

Expected heterozygosity (H_E) and rarified allele richness were almost the same in the Iberian and Italian hybrid zones (e.g. for white poplar: H_E = 0.410 vs 0.428; A' = 2.65 vs 2.80), but somewhat lower than in the Central European population (H_E = 0.508; A' = 3.46). Genetic diversity descriptors in grey poplar revealed higher values of diversity for this species in all populations, an evidence of its hybrid status (Table 2).

The average number of pairwise differences per locus (π_k) was clearly correlated to $2S_k/n$ (r^2 = 0.992; Figure S2), indicating growing genet size. Under this scenario and using formulae from ALLY *et al.* (2008), genet age of the large clonal assemblies varied around one thousand years, from over 500 (MLL006: 553 ± 520; MLL009: 547 ± 424;

MLL025: 629 ± 1117) to over two thousand years old (MLL074: 2,338 ± 3,151; MLL049: 3,359 ± 3,908). The seemingly older genets appeared to be the smaller ones, although the error in the estimates for these clones was also higher. Under the model of accumulation of mutations since the common ancestor (THOMSON *et al.* 2000), MLL006 age was estimated to 5,127 ± 3,119, and MLL009 to 8,411 ± 3,726 years old. A more exact estimation of the age of the large clonal assemblies found in the Douro basin would require a more dedicated investigation, but we can conclude that at least some of the living clones today have ages of over 2,000 years.

Genetic background of genets, in particular of large clones

The genetic background of grey poplars corresponded mainly to first generation hybrids (F1) and backcrosses to white poplar. Many white poplar individuals displayed weak introgression from aspen, represented by a soft cline from first-backcrossed hybrids to pure white poplars (Figure 2). In contrast, only few genets of aspen (3 out of 17) showed white poplar introgression. Concerning these hybrids, some chromosome regions apparently showed high rates of interspecific heterozygosity (i.e. chromosomes VI and XV), whereas other regions displayed more dominance of aspen background (i.e. chromosomes II, IV, VI, VIII and XII), even in the backcrosses to white poplar. Within these regions, two loci (GCPM1809 and GCPM1065) were highly homozygous for aspen in the Douro, and in three other European hybrid populations (LINDTKE *et al.* 2012), and an additional locus (GCPM154) in at least (not tested in the other two: Ticino, and Tisza, Hungary) the Douro and Danube populations (LEXER *et al.* 2010). Furthermore, GCPM154 displayed highly distorted genotype proportions in a controlled backcross of a F1 to white poplar, with dominance of the aspen allele (MACAYA-SANZ *et al.* 2011). Regarding large clones, this marker (GCPM154) presented an increased interspecific heterozygosity with respect to their individual backgrounds in the fourth largest white poplar clone (MLL073) and the two largest hybrid genets (MLL006 and MLL057). This microsatellite lies close (<10kb) to two genes (*P. trichocarpa* assembly v3), one with homology to a calmodulin binding protein-like and the other with unknown function. Large clones were also severely introgressed by *P. tremula* background in the genomic region around locus GCPM1274, even for otherwise pure white poplars. This deviation was not so acute in smaller clones and was also accompanied by segregation distortion in the controlled backcross (MACAYA-SANZ *et al.* 2011). Within this region, allele 207 was more abundant in the large genets than in the rest of the population. This allele was present in 23.21% of genets overall (53.85% of grey poplars, 18.29% of white poplars, and 23.53% of aspens), but it was present in six out of the seven largest clones ($P = 0.0009$; calculated as the tail of a binomial distribution), and in seven out of the ten largest ones ($P = 0.0022$). This locus is located in close proximity (<10 kb) of two annotated genes: a glutaredoxin (GRX), with an antioxidant function, and a myosin-like protein, with an IQ calmodulin binding-motif, and a third one of unknown function.

Inbreeding coefficients at the individual level (F_i) followed an approximate normal distribution (Figure S3). Interestingly, the largest four clones had inbreeding

values (from -0.121 to -0.013) below the mode of the distribution, a result unlikely by chance alone ($P = 0.0625$; calculated as the tail of a binomial distribution). The correlation between number of ramets and individual inbreeding coefficients was significant (Spearman $\rho = -0.21$; $P = 0.0244$), with less inbred clones having a higher number of ramets. However, this correlation was not significant anymore (but still negative) when the largest four clones were excluded (Spearman $\rho = -0.13$; $P = 0.1953$), showing its important effect.

Sexual success of large clones

Large clones had higher numbers of sexual offspring than smaller clones in the population, with the exception of MLL006 (a F1-hybrid; see Figure 2), whose number of offspring was comparatively small (Table 1). All the seven offspring assigned to MLL006 were (as expected) also found to be hybrids by STRUCTURE, with a mean Q-value of 0.75 (0.06), suggesting that they were backcrosses to white poplars. One of the large clones (MLL009) was a male, but the next two largest were female (MLL006 and MLL025). We could only observe the gender of a few genets, thus no general conclusions could be drawn. At the same time, it was conspicuous that some small clones had also produced a fairly large number of descendants. Considering COLONY parental assignments, genets with up to four ramets (91.1% of the MLLs) were the progenitors of 31 offspring (29.5% of the identified parentage relationships) while genets with more than four ramets and less than 10 (5.4%) were responsible of 14 offspring (13.3%). The four largest clones (3.6% of the MLLs), with more than 10 ramets each, were the parents of 60 offspring (57.1%). Spearman rank tests showed significant correlations between clone size (estimated as number of ramets) and offspring number, both considering all genets ($\rho = 0.377$; $P < 0.001$) and for the largest ten ones ($\rho = 0.768$; $P < 0.01$; Figure S4). However, correlations were not significant when clone size was estimated as the maximum distance among ramets both for the whole sample (not including single-ramet genets, given that calculating size needs at least of two ramets; $N = 41$; $\rho = 0.088$; $P = 0.58$) and for the largest clones subset ($\rho = 0.540$; $P = 0.11$; Figure S4). For many individuals (40 clones, i.e. 35.7%), no parent was found within the sampled trees. Hence, considering the high power of analysis conferred by our microsatellite set, these results indicate that many progenitors were not sampled (due to being out of sampling area; within it, but overlooked; or dead) rather than insufficient marker resolution. This fact decrease the dominant role of large clones in sexual reproduction (to ~37% of all offspring vs 57.1% of the identified parentage relationships).

The spatial distribution and clone size of offspring can be used to differentiate between young and old parental genets. For example, 13 offspring were assigned to the relatively small clone MLL011 (6 ramets), but none of them were more than 150 meters apart from a parent MLL011 ramet, whereas the 5 descendants of genet MLL086 (2 ramets) were spread along 12 km. Remarkably, MLL011's offspring were all of small size (1.2 ramets per MLL) while MLL086's offspring had significantly larger genets (2.2 ramets per MLL). Interpreted together, these facts suggest that MLL086 is older

than MLL011. Following this rationale, the largest four clones would be old as well, with the exception of MLL073, which has only one offspring genet (MLL100), with just one ramet sampled, that is separated less than 150 m of a MLL073 ramet. The other large clones have offspring with relatively large clone size (MLL006: 1.7 ramets per offspring; MLL009: 1.9; MLL025: 1.5) that are located rather far from their parents [mean of the minimal distance between offspring ramets and parent ramets: 5.9 km (MLL006); 2.8 km (MLL009); and 4.6 km (MLL025)]. These results should be considered cautiously because of the sampling scheme (larger sampling spacing outside the core population), which probably overestimates these values, but they reveal how massive and dispersed is the offspring of the large clones, reflecting ancient sexual reproduction events together with asexual spread. Moreover, F2 and F3 descendants were detected in genets with numerous offspring [MLL009, MLL025 and MLL011 (only F2s); with the exception of hybrid genet MLL006 (only F1s)]. Interestingly, MLL011 was an offspring of the male huge clone MLL009. Also, this clone paired with two of its "granddaughters". Finally, almost all MLL006's offspring were full-sibs from couplings with the prolific MLL009.

Population size trends

For all the parameters calculated by MSVAR, the upper confidence interval of the Gelman-Rubin diagnostic statistic was equal or lower than 1.02, indicating model convergence. Demographical analyses using MSVAR (44 nuSSRs with perfect repeats) showed a long-term decline in population effective size (Table 3, Figure S5). The exponential decline model fitted better with observed data than the linear model (AIC of 23,776 vs 24,992). Nonetheless, the reduced difference in AIC values indicates that neither of the models is significantly superior. Both models indicated a soft but persistent reduction in effective size to about one tenth of the original one, which appeared to have taken place during the last hundreds of thousands of years. Despite population decline, current effective size was still considerable (~2,240 for the exponential model; Table 3). These calculations considered a generation time of 40 years. Given high levels of asexual reproduction in the population, the existence of ancient clones (see above) and that genet turnover was probably even lower in former times (e.g. during glacial times, see MACAYA-SANZ *et al*. 2012), larger generation times and a more ancient population decline (involving some million years) cannot be discarded.

Discussion

Colossal clones and low genotypic richness, but no loss of genetic diversity

The distribution of genet sizes was leptokurtic (*Kurt* = 60.4 ± 0.5) and positive skewed (right-tailed; *g* = 7,64 ± 0.23; computed by STATISTICA 8.0, StatSoft. Inc., OK, USA), with few large and many small clones, as it has been previously detected in other populations dominated by large clones of white poplars and related species (BRUNDU *et al.* 2008; MOCK *et al.* 2008; VAN LOO *et al.* 2008). Two large and widespread genets were found, one pure male white poplar (MLL009) and a female hybrid (MLL006) (189 and 124 ramets, respectively). Another geographically extended female clone was also found (MLL025), although with fewer ramets (26). Some other genets showed fairly large distributions but were smaller. Large and widespread clones have been repeatedly reported in poplars (e. g. ALLY *et al.* 2008; BRUNDU *et al.* 2008), including for this area (SANTOS-DEL-BLANCO *et al.* 2012). The Douro hybrid zone is located in a region (the Castilian Plateau) where human activity dates back more than two millennia. A study on the European elms revealed that the wide distribution of the so-called English elm, an ancient clone, was due to active human-mediated propagation during Roman times (GIL *et al.* 2004). In our case, spread of poplar clones by humans, though possible, it is not a feasible explanation. Although riverbank reforestation with poplars has surely been conducted in historic times, estimated age of large clones (several thousand years) exceeds the times when social civilization could have fostered this regional expansion, during the Romanization of the Iberian Peninsula. Furthermore, white and grey poplars are not considered valuable timber species in the region, and these large poplar clones (MLL009 and MLL006) do not display any special feature compared to conspecific genets that could have made them especially interesting for reforestation purposes (data not shown). The large extension of some of these clones, the largest over 150 km, is not compatible with an exclusive root-sucker expansion. Twig translocation by water currents has been detected in species of genus *Salix* and *Populus*, and it is known to be a successful long-distance propagation mechanism (BARSOUM *et al.* 2004; and references therein). Large birds, such as the locally abundant storks, could as well have effectively contributed to translocate twigs for nesting, although, to our knowledge, no study has assessed the magnitude of this dispersal mode yet.

Compared to the Danube and Ticino hybrid zones, the Douro population showed a lower level of *genotypic* richness and evenness. Many of the studies reviewed by HONNAY and BOSSUYT (2005) found similar levels of genotypic diversity in clonal and sexual populations, yet they indicated that it may have been overestimated in clonal populations due to DNA profiling misinterpretation or to the oversight of somatic mutations. Simulations (not including selection) showed a decrease of genotype effective number with increasing clonal reproduction rate (BALLOUX *et al.* 2003), which can be interpreted as clonal reproduction increasing the possibility of losing genotypes by drift. Other studies in poplar have revealed that genotypic richness is higher in less stressful environments, as in areas with more water availability due to either more accessible groundwater (VONLANTHEN *et al.* 2010) or more humid climate (SLAVOV *et*

al. 2010). Lower levels of moisture may also underlie low genotypic richness in the Douro hybrid zone, since the summer is more severe in this area of the Iberian Peninsula than in the surroundings of the Alps, where the other two hybrid zones lie. Inadequate groundwater dynamics is known to impair sexual recruitment in white poplar (GONZALEZ *et al.* 2010a). The fragile constitution of poplar seeds delivers weak plantlets that are especially vulnerable to drought during their first years. Thus, it is reasonable to think that water availability crucially influences effective sexual reproduction in this region. Another important ecological factor for sexual reproduction is river water regime. As long as seedling recruitment in riverine poplars occurs in bare mineral substrates after flooding (BRAANTE *et al.* 1996), river regulation or climate change (reduction of snowmelt floods) may decrease effective sexual regeneration through reduction of the intensity of flooding. The three compared rivers (Douro, Danube and Ticino) have been regulated in the last century, although regulation did not avoid occasional floods in any of them. Concerning snowmelt floods, climate and orography make the Iberian Douro river less susceptible to flooding than the alpine Danube and Ticino, so poplars in the Douro basin would not benefit from these occasional opportunities for sexual propagation.

Interestingly, lower levels of *genotypic* richness in the Douro hybrid zone were not accompanied by reduced *genetic* diversity at genet-level (referring to the level where the element is the genet, not the ramet, so calculations include only one ramet per genet). Our findings are in line with theoretical models (BALLOUX *et al.* 2003) and comparative analysis among populations of clonal species (reviewed by HONNAY and BOSSUYT 2005; see for example *P. trichocarpa* in SLAVOV *et al.* 2010) that reported that genetic diversity is not significantly lower in more clonal populations. However, the dynamics of genetic variation in clonal populations is not obvious. In *P. trichocarpa*, allelic richness of the more clonal population was less than two thirds of that of the others, despite the similar levels of genetic diversity (SLAVOV *et al.* 2010). Moreover, BALLOUX *et al.* (2003) did not include selection in their model while the 'universal' leptokurtic distribution of clonal populations suggests that outcompeting by fitter clones (i.e. a form of selection) is a major structuring force. Nevertheless, selection could act simultaneously depleting genotypes but maintaining overall genetic diversity. A study on a partially-asexual Acarus species concluded that negative frequency-dependent selection accounted for the maintenance of genetic variation (WEEKS and HOFFMANN 2008). At ramet-level, genetic diversity was moderately lower than at genet-level in the Douro population (H_E of 0.338 and 0.410, respectively). Thus, at ramet-level, which is more relevant for population dynamics, genetic diversity can be reduced by the existence of large clonal assemblies. A study on the salt marsh perennial *Borrichia frutescens* (RICHARDS *et al.* 2004) showed a negative correlation between clonal and genetic diversities at this level (recomputed from Tables 1 & 2; $R = 0.800$; Spearman rank test, $P = 0.104$), supporting our findings.

In summary, common clonal reproduction results in fewer genotypes. Genetic diversity at genet-level is hardly affected by clonality, unlike at ramet-level, where genetic diversity apparently decays faster, due to the leptokurtic distribution of genet size often found in clonal populations. Thus, caution is needed on the level to evaluate

genetic diversity, with estimates at ramet-level being more adequate to measure the population effects of clonal propagation.

Genetic background and heterozygosity may underlie clonal success

Heterosis in poplar hybrids is known to cause extreme phenotypes, a quality commonly used in agronomy (e.g. the I-214 clone widely-planted in commercial forestry is a cross between *Populus deltoides* and *P. nigra*). However, to what extent heterosis may affect clonal propagation capability in natural populations is unknown. VAN LOO *et al.* (2008) reported that hybrids displayed slightly larger sizes (measured both as number of ramets and as maximum distance among ramets) than pure white poplar genets in a natural hybrid population. Higher clonal propagation was also reported in the hybrids of *Populus angustifolia* and *P. fremontii* (SCHWEITZER *et al.* 2002). Likewise, *Narcissus* hybrids produce in nature more and larger asexual propagules than parental species (MARQUES *et al.* 2011). However, we did not find this pattern: In our population hybrid clones were not larger at a significance level.. Of the ten largest clones, only two were hybrids, a proportion matching the sampling share. Large clones, however, displayed an overrepresented allele (originated from *P. tremula*) in a region of unusual interspecific heterozygosity in a telomere of chromosome I, around locus GCPM1274. The telomeric region of chromosome I is rich in NBS class resistance genes (KOHLER *et al*. 2008) and in expressed small RNAs (KLEVEBRING *et al.* 2009). In the *Helianthus* species-complex emanating from the hybridization of *H. annus* and *H. petiolaris*, the derived *H. paradoxus* tolerates salt environments better than parental species. A candidate gene for salt-tolerance was successfully mapped in a recombinant second-generation backcross (of parental species) on a QTL for survivorship in salt substrates, and was suspected to be more heterozygous in the derived than in the original species (LEXER *et al.* 2004). Moreover, the above mentioned survey on crosses between *P. angustifolia* and *P. fremontii* detected higher clonal propagation not only in F1s but also in backcrossed hybrids (SCHWEITZER *et al.* 2002). Albeit molecular markers were not analyzed, this result reveals that clonal propagation can benefit from local interspecific heterozygosity, and that such heterosis can be maintained for several generations. In line with this, increased interspecific heterozygosity, concentrated in particular areas of the genome, was recently observed in *Populus tremula* × *P. alba* recombinant hybrids (LINDTKE *et al.* 2012). The underlying cause proposed by the authors was persistent selection of coadapted gene complexes, hindering Mendelian segregation across the genome. In that case, clonality was not evaluated, but the study indicated that higher fitness in hybrid populations may come from genome regions with high interspecific heterozygosity.

In the Douro population, the inbreeding coefficient (F_i) of large clones was atypically low (Figure S3), indicating that these individuals, apart from localized "islands" of interspecific heterozygosity, have also higher overall heterozygosity than other genets. Heterozygous genotypes are usually the norm in single-genotype populations of *Decodon verticillatus* (ECKERT 2001). A heterozygote excess due to predominant clonal reproduction was also found in *Prunus avium* populations (STOECKEL *et al.* 2006). However, in this case, heterozygote advantage was discarded as

the underlying mechanism, and no significant differences were found in heterozygosity between single- and multi-ramet individuals. Further research is thus needed to assess whether levels of individual heterozygosity and clonal propagation rates are correlated and, if so, what are the underlying causes. Nonetheless, our findings highlight the role that the overall genomic background, as well as localized regions of interspecific heterozygosity, may play in clonal propagation in natural populations.

Population effects of large, geographically-extended clones

Sexual dominance

Large, geographically-extended clones dominate sexual reproduction in the Douro hybrid zone, as shown by a significant correlation between clone size and number of descendants. Sexual dominance results from both greater opportunities for mating each season (due to the large number of ramets) and the continued contribution to reproduction along several seasons, since clone size is correlated with age. Sexual dominance of large clones has two important implications. (i) Ageing sterility does not seem to have affected these relatively ancient clones yet. Somatic mutation load can reduce sexual performance, but this is a slow process. ALLY *et al.* (2010) calculated that it would take between 500 and 20,000 years to lose male sexual function through pollen quality and quantity depauperation. Furthermore, in case of very large clones, as in this case, it is expected that the time needed for development of sterility will be somewhat longer, as mutations should accumulate throughout the majority of the genet. (ii) At genet level, uneven sexual reproduction reduces the effective population size as it increases the variance of reproductive success (BALLOUX *et al.* 2003), rising genetic drift. However, in the Douro hybrid zone, although significantly higher than zero, population-level inbreeding (F_{IS} = 0.067) was similar to other poplar hybrid zones (see Table 2). Moreover, an extensive study at genet-level in 11 populations of *P. tremuloides* found a similar moderate value of F_{IS} = 0.09 (COLE 2005). Those aspen populations were, however, much less clonal than the Douro's, considering that the sampling scheme was similar to the one followed in our study, but no clonal replicate was detected. In summary, the sexual dominance of Douro's largest clones does not seem to be pervasive enough to increase inbreeding at the population level.

As an exception to the general trend, the large female hybrid genet MLL006 did not have the high number of descendants expected according to its size. Several not exclusive causes can explain this fact: (i) Hybrid sterility produced by a partial genomic incompatibility between parent species; (ii) Backcrossing with white poplar may be prevented by different pre- and post-zygotic barriers; (iii) Hybrid vigor can have conferred an improved capacity for vegetative propagation to this clone compared to other genets, enlarging the size of a relatively young clone. The first explanation is unlikely given the elevated synteny observed in the genus *Populus* (WANG *et al.* 2011), the frequent backcrossing towards white poplar found in hybrid zones (LEXER *et al.* 2010), and that greenhouse sowings of this clone seeds have not detected any decrease in germination (personal observation). In contrast, ecological pre- and post-zygotic barriers may well explain this bias. A delayed flowering phenology of hybrids has been

observed in the field, reducing the co-occurrence with more-abundant white poplar male pollination. Moreover, the consequent delay in seed dispersal could also hamper seedling recruitment due to less favorable conditions for establishment and increased competition from already established seedlings. Finally, the estimated age for this clone is of the same magnitude than for other large clones, over several hundred years. Therefore, a sudden propagation of a young clone seems also implausible.

Long-term population contraction

Bayesian simulations carried out using MSVAR showed a long-term decline of effective population size in the Douro hybrid zone. MSVAR produces more reliable inferences when the population size change is high, fluctuations are ancient ($T_a = 500$) and the population size declines (GIROD *et al.* 2011), as is our case. Long-term contraction of the population as a whole contrasts with the steady growth of large clones, as deduced from the distribution of somatic mutations (see Results). High levels of genetic diversity and moderate inbreeding at the population level (see above) suggest that long-term population decline is due to outcompeting of other genets by larger clones rather than to increased inbreeding due to their sexual dominance.

Several processes may have created opportunity for expansion of current large clones in detriment of sexual reproduction. First, environmental fluctuations, which are common in riparian habitats, would likely hamper sexual reproduction more than asexual propagation, which depends less on ecological factors. It is also admitted that clonal reproduction is more abundant in marginal areas of species' ranges and in more stressful environments (HONNAY and BOSSUYT 2005; KAWECKI 2008; SILVERTOWN 2008), such as those of the Douro basin (as compared to Central European hybrid zones; see above). Second, ancient geological events, such as the transition from the Pleistocene to the colder Pliocene and the Quaternary glaciations, may have promoted clonal spread in the Douro population of white poplar. White poplar evolved during the late Tertiary and early Quaternary, under more humid subtropical conditions. Climate desiccation and cold temperatures are therefore thought to have hindered sexual reproduction in this species, in particular in the Douro basin, which was severely affected by the last glaciation (but not under ice; ALBERTI *et al.* 2004). Third, in more recent times, human activities have certainly affected the ecology of white poplars in this region, so tightly dependent on natural disturbances. Indeed, river flood regulation is directly involved in failure of sexual recruitment in white poplar (GONZALEZ *et al.* 2010b), as well as in other poplar species (e.g. *P. nigra*; BARSOUM 2001). Fire control has direct implications in the clonal structure of poplar species such as *P. tremuloides* (NAMROUD *et al.* 2006), though to date it has not been correlated with white poplar population dynamics, despite the crucial role fire plays in Mediterranean ecosystems. Lastly, habitat fragmentation may as well favor asexual reproduction, as shown in some herbs (HONNAY and BOSSUYT 2005; and references therein). Because of their dependence on phreatic water, white poplars suffer substantial genetic isolation, in particular across different –even adjacent– river basins, which are normally separated by high mountains in the Iberian Peninsula. This is reflected in high genetic differentiation at different spatial scales, from watersheds to populations (MACAYA-SANZ *et al.* 2012).

During colder periods, this genetic isolation may even have been more intense due to downward altitude migrationimposing further limitations on gene flow among riparian tree populations.

Long-term consequences of clonal dominance by few genets

The balance between the sexual and asexual contribution to reproduction is perturbed when ecological conditions change, and a new balance is established. However, it may not always be feasible to recover equilibrium. HONNAY and BOSSUYT (2005) argued that environmental conditions precluding effective sexual reproduction can conduct a population to a tipping point for irreversible extinction of sexual function. Our results suggest that the Douro poplar hybrid zone is contracting, with a few huge clones dominating the scarce sexual reproduction. However, we did not find strong signatures of genetic erosion, and tree health condition and growth is apparently normal (personal observations), suggesting that large-clone dominance is not irreversible. Nevertheless, further environmental changes, new species introductions, or epidemic outbreaks, among other ecological factors, could push this population towards further clonal dominance, to end in a mono- or oligo-clonal population, with virtual sex extinction (as observed in Sardinian white poplar; BRUNDU *et al.* 2008) and low ecological resilience.

Two mechanisms may underpin this process: the loss of less-adapted genotypes by intraspecific competition with larger clones; and the accumulation of somatic mutations that erode the sexual system in successful clones, as predicted by the 'somatic mutation theory' (KLEKOWSKI 1997). Sexual extinction due to intraspecific competition would take place quicker in strongly outcrossing species where 'phalanx' growth is dominant (HONNAY and BOSSUYT 2005). Although the Douro population doesn't seem to display a strict 'phalanx' structure, our widely-spaced sampling scheme cannot fully discard the dominance of this growth type, typically found in other poplar and aspen clonal populations [e.g. *P. nigra*, (CHENAULT *et al.* 2011); *P. tremuloides*, (MOCK *et al.* 2008)]. Accumulation of somatic mutations (not purged by selection in clonal organisms) can produce a decay of genes involved in sexual reproduction (NORMARK *et al.* 2003), and henceforth, also reduce sex efficiency, as has already being reported for male reproductive fitness in *Populus tremuloides* clonal stands (ALLY *et al.* 2010). However, this process may take hundreds of generations and current somatic mutation accumulation is low in the Douro poplars, even for the colossal-size clones.

In conclusion, the Douro poplar hybrid zone provides a case study for a largely clonal population, dominated by ancient and widespread clones, that, nevertheless, maintains substantial levels of genetic diversity and no apparent signs of decline. Common features thought to impoverish clonal populations do not appear to be relevant for the Douro and only drastic demographic events, such as extreme bottlenecks, could eventually compromise the long-term survival of this iconic riparian forest.

Acknowledgements

We would like to acknowledge Mike Fay, Jeffrey Joseph, Carmen García Barriga, Fátima Torrico, Jose Antonio Mancha Núñez, Javier Arenillas Cristóbal, Jesús Pastor Aceves, Jorge Moya García, Manuel Núñez Sanz, Ángel Vicente Blázquez, Luis Santos-del-Blanco, Juan Andrés Oria de Rueda, and Nuria Alba. Also to ReproFor Project from the Spanish Ministry for Economics and Competitiveness (MINECO). MH is a Ramón y Cajal researcher of the MINECO and acknowledges an EU Marie-Curie Intra-European Fellowship.

References

ALBERTI, A. P., M. V. DÍAZ and R. B. CHAO, 2004 Pleistocene glaciation in Spain, pp. 389-394 in *Developments in Quaternary Sciences*, edited by J. EHLERS and P. L. GIBBARD. Elsevier.

ALLY, D., K. RITLAND and S. P. OTTO, 2008 Can clone size serve as a proxy for clone age?An exploration using microsatellite divergence in Populus tremuloides. Molecular Ecology **17:** 4897-4911.

ALLY, D., K. RITLAND and S. P. OTTO, 2010 Aging in a Long-Lived Clonal Tree. Plos Biology **8**.

ANGELES, I. G. C., J. GOLUBOV, B. G. MILLIGAN and M. C. MANDUJANO, 2011 Spatial distribution pattern of a clonal species: effects of differential production of clonal and sexual offspring. Evolutionary Ecology **25:** 1357-1383.

ARNAUD-HAOND, S., C. M. DUARTE, F. ALBERTO and E. A. SERRAO, 2007 Standardizing methods to address clonality in population studies. Molecular Ecology **16:** 5115-5139.

BAACK, E. J., and L. H. RIESEBERG, 2007 A genomic view of introgression and hybrid speciation. Current Opinion in Genetics & Development **17:** 513-518.

BALLOUX, F., L. LEHMANN and T. DE MEEUS, 2003 The population genetics of clonal and partially clonal diploids. Genetics **164:** 1635-1644.

BARRETT, S. C. H., 2002 The evolution of plant sexual diversity. Nature Reviews Genetics **3:** 274-284.

BARSOUM, N., 2001 Relative contributions of sexual and asexual regeneration strategies in Populus nigra and Salix alba during the first years of establishment on a braided gravel bed river. Evolutionary Ecology **15:** 255-279.

BARSOUM, N., E. MULLER and L. SKOT, 2004 Variations in levels of clonality among Populus nigra L. stands of different ages. Evolutionary Ecology **18:** 601-624.

BEAUMONT, M. A., 1999 Detecting population expansion and decline using microsatellites. Genetics **153:** 2013-2029.

BRAANTE, J. H., S. B. ROOD and P. E. HEILMAN, 1996 Life history, ecology, and conservation of riparian cottonwoods in North America., pp. 57-85 in *Biology of Populus and Its Implications for Management and Conservation.*, edited by R. F. STETTLER, J. H. D. BRADSHAW, P. E. HEILMAN and T. M. HINCKLEY. NRC Research Press, Otawa.

BRITTEN, H. B., 1996 Meta-analyses of the association between multilocus heterozygosity and fitness. Evolution **50:** 2158-2164.

BRUNDU, G., R. LUPI, I. ZAPELLI, T. FOSSATI, G. PATRIGNANI *et al.*, 2008 The Origin of Clonal Diversity and Structure of Populus alba in Sardinia: Evidence from Nuclear and Plastid Microsatellite Markers. Annals of Botany **102:** 997-1006.

BUDDE, K. B., L. GALLO, P. MARCHELLI, E. MOSNER, S. LIEPELT *et al.*, 2011 Wide spread invasion without sexual reproduction? A case study on European willows in Patagonia, Argentina. Biological Invasions **13:** 45-54.

CASTIGLIONE, S., A. CICATELLI, R. LUPI, G. PATRIGNANI, T. FOSSATI *et al.*, 2010 Genetic structure and introgression in riparian populations of Populus alba L. Plant Biosystems **144:** 656-668.

CHAPMAN, J. R., S. NAKAGAWA, D. W. COLTMAN, J. SLATE and B. C. SHELDON, 2009 A quantitative review of heterozygosity-fitness correlations in animal populations. Molecular Ecology **18:** 2746-2765.

CHENAULT, N., S. ARNAUD-HAOND, M. JUTEAU, R. VALADE, J. L. ALMEIDA *et al.*, 2011 SSR-based analysis of clonality, spatial genetic structure and introgression from the Lombardy poplar into a natural population of Populus nigra L. along the Loire River. Tree Genetics & Genomes **7:** 1249-1262.

COLE, C. T., 2005 Allelic and population variation of microsatellite loci in aspen (Populus tremuloides). New Phytologist **167:** 155-164.

ECKERT, C. G., 2001 The loss of sex in clonal plants. Evolutionary Ecology **15:** 501-520.

FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2003 Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics **164:** 1567-1587.

FUSSI, B., J. BONELLO, E. CALLEJA and B. HEINZE, 2012 Combining the use of molecular techniques and archival documentary evidence to trace the origin of Populus alba in a Central Mediterranean archipelago. European Journal of Forest Research **131:** 347-354.

GELMAN, A., and D. B. RUBIN, 1992 Inference from iterative simulation using multiple sequences. Statist. Sci. **7:** 457-472.

GIL, L., P. FUENTES-UTRILLA, A. SOTO, M. T. CERVERA and C. COLLADA, 2004 English elm is a 2,000-year-old Roman clone. Nature **431:** 1053-1053.

GIROD, C., R. VITALIS, R. LEBLOIS and H. FREVILLE, 2011 Inferring Population Decline and Expansion From Microsatellite Data: A Simulation-Based Evaluation of the Msvar Method. Genetics **188:** 165-U287.

GONZALEZ, E., F. ANTONIO COMIN and E. MULLER, 2010a Seed dispersal, germination and early seedling establishment of Populus alba L. under simulated water table declines in different substrates. Trees-Structure and Function **24:** 151-163.

GONZALEZ, E., M. GONZALEZ-SANCHIS, A. CABEZAS, F. A. COMIN and E. MULLER, 2010b Recent Changes in the Riparian Forest of a Large Regulated Mediterranean River: Implications for Management. Environmental Management **45:** 669-681.

HARDY, O. J., and X. VEKEMANS, 2002 SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes **2:** 618-620.

HONNAY, O., and B. BOSSUYT, 2005 Prolonged clonal growth: escape route or route to extinction? Oikos **108:** 427-432.

HONNAY, O., and H. JACQUEMYN, 2008 A meta-analysis of the relation between mating system, growth form and genotypic diversity in clonal plant species. Evolutionary Ecology **22:** 299-312.

JONES, O. R., and J. L. WANG, 2010 COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources **10:** 551-555.

KALINOWSKI, S. T., 2005 HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes **5:** 187-189.

KAWECKI, T. J., 2008 Adaptation to Marginal Habitats, pp. 321-342 in *Annual Review of Ecology Evolution and Systematics*.

KLEKOWSKI, E. J., 1997 Somatic mutation theory of clonality, pp. 227-241 in *The Ecology and Evolution of Clonal Plants*, edited by H. DE KROON and J. VAN GROENENDAEL. Backhuys, Leiden.

KLIMES, L., J. KLIMESOVA, R. HENDRIKS and J. VAN GROENENDAEL, 1997 Clonal plant architecture: a comparative analysis of form and function, pp. 1-29 in *The Ecology and Evolution of Clonal Plants*, edited by H. DE KROON and J. VAN GROENENDAEL. Backhuys, Leiden.

LEXER, C., M. F. FAY, J. A. JOSEPH, M. S. NICA and B. HEINZE, 2005 Barrier to gene flow between two ecologically divergent Populus species, P alba (white poplar) and P tremula (European aspen): the role of ecology and life history in gene introgression. Molecular Ecology **14:** 1045-1057.

LEXER, C., J. A. JOSEPH, M. VAN LOO, T. BARBARA, B. HEINZE *et al.*, 2010 Genomic Admixture Analysis in European Populus spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating. Genetics **186:** 699-U391.

LEXER, C., Z. LAI and L. H. RIESEBERG, 2004 Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of Helianthus paradoxus, a diploid hybrid species. New Phytologist **161:** 225-233.

LIBRADO, P., and J. ROZAS, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451-1452.

LINDTKE, D., C. A. BUERKLE, T. BARBARA, B. HEINZE, S. CASTIGLIONE *et al.*, 2012 Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in Populus. Molecular Ecology **21:** 5042-5058.

LIPPMAN, Z. B., and D. ZAMIR, 2007 Heterosis: revisiting the magic. Trends in Genetics **23:** 60-66.

LIU, J., M. DONG, S. L. MIAO, Z. Y. LI, M. H. SONG *et al.*, 2006 Invasive alien plants in China: role of clonality and geographical origin. Biological Invasions **8:** 1461-1470.

LOISELLE, B. A., V. L. SORK, J. NASON and C. GRAHAM, 1995 Spatial Genetic-Structure of a Tropical Understory Shrub, Psychotria Officinalis (Rubiaceae). American Journal of Botany **82:** 1420-1425.

MACAYA-SANZ, D., M. HEUERTZ, U. LOPEZ-DE-HEREDIA, A. I. DE-LUCAS, E. HIDALGO *et al.*, 2012 The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars. Molecular Ecology **21:** 3593-3609.

MACAYA-SANZ, D., L. SUTER, J. JOSEPH, T. BARBARA, N. ALBA *et al.*, 2011 Genetic analysis of post-mating reproductive barriers in hybridizing European Populus species. Heredity **107:** 478-486.

MARQUES, I., G. N. FELINER, M. A. MARTINS-LOUCAO and J. F. AGUILAR, 2011 Fitness in Narcissus hybrids: low fertility is overcome by early hybrid vigour, absence of exogenous selection and high bulb propagation. Journal of Ecology **99:** 1508-1519.

MARRON, N., V. STORME, S. Y. DILLEN, C. BASTIEN, L. RICCIOTTI *et al.*, 2010 Genomic regions involved in productivity of two interspecific poplar families in Europe. 2. Biomass production and its relationships with tree architecture and phenology. Tree Genetics & Genomes **6:** 533-554.

MOCK, K. E., C. A. ROWE, M. B. HOOTEN, J. DEWOODY and V. D. HIPKINS, 2008 Clonal dynamics in western North American aspen (Populus tremuloides). Molecular Ecology **17:** 4827-4844.

NAMROUD, M. C., A. LEDUC, F. TREMBLAY and Y. BERGERON, 2006 Simulations of clonal species genotypic diversity - trembling aspen (Populus tremuloides) as a case study. Conservation Genetics **7:** 415-426.

NOLTE, A. W., and D. TAUTZ, 2010 Understanding the onset of hybrid speciation. Trends in Genetics **26:** 54-58.

NORMARK, B. B., O. P. JUDSON and N. A. MORAN, 2003 Genomic signatures of ancient asexual lineages. Biological Journal of the Linnean Society **79:** 69-84.

OTTO, S. P., and T. LENORMAND, 2002 Resolving the paradox of sex and recombination. Nature Reviews Genetics **3:** 252-261.

PLUMMER, M., N. BEST, K. COWLES and K. VINES, 2006 CODA: Convergence diagnosis and output analysis for MCMC. R News **6:** 7-11.

RICE, W. R., 2002 Experimental tests of the adaptive significance of sexual recombination. Nature Reviews Genetics **3:** 241-251.

RICHARDS, C. L., J. L. HAMRICK, L. A. DONOVAN and R. MAURICIO, 2004 Unexpectedly high clonal diversity of two salt marsh perennials across a severe environmental gradient. Ecology Letters **7:** 1155-1162.

RIESTER, M., P. F. STADLER and K. KLEMM, 2009 FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics **25:** 2134-2139.

RIESTER, M., P. F. STADLER and K. KLEMM, 2010 Reconstruction of pedigrees in clonal plant populations. Theoretical Population Biology **78:** 109-117.

SACHS, T., 2001 Developmental processes and the evolution of plant clonality. Evolutionary Ecology **15:** 485-500.

SANTOS-DEL-BLANCO, L., A. DE-LUCAS, S. GONZÁLEZ-MARTÍNEZ, R. SIERRA-DE-GRADO and E. HIDALGO, 2012 Extensive Clonal Assemblies in Populus alba and Populus x canescens from the Iberian Peninsula. Tree Genetics & Genomes**:** 1-12.

SCHWARZBACH, A. E., L. A. DONOVAN and L. H. RIESEBERG, 2001 Transgressive character expression in a hybrid sunflower species. American Journal of Botany **88:** 270-277.

SCHWEITZER, J. A., G. D. MARTINSEN and T. G. WHITHAM, 2002 Cottonwood hybrids gain fitness traits of both parents: A mechanism for their long-term persistence? American Journal of Botany **89:** 981-990.

SILVERTOWN, J., 2008 The evolutionary maintenance of sexual reproduction: Evidence from the ecological distribution of asexual reproduction in clonal plants. International Journal of Plant Sciences **169:** 157-168.

SLAVOV, G. T., S. LEONARDI, W. T. ADAMS, S. H. STRAUSS and S. P. DIFAZIO, 2010 Population substructure in continuous and fragmented stands of Populus trichocarpa. Heredity **105:** 348-357.

SLAVOV, G. T., and P. ZHELEV, 2010 Salient biological features, systematics, and genetic variation of Populus, pp. 15-38 in *Genetics and Genomics of Populus*, edited by S. JANSSON, R. P. BHALERAO and A. T. GROOVER. Springer Science+Business Media, LLC, New York.

STENVALL, N., T. HAAPALA, S. AARLAHTI and P. PULKKINEN, 2005 The effect of soil temperature and light on sprouting and rooting of root cuttings of hybrid aspen clones. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere **35:** 2671-2678.

STOECKEL, S., J. GRANGE, J. F. FERNANDEZ-MANJARRES, I. BILGER, N. FRASCARIA-LACOSTE *et al.*, 2006 Heterozygote excess in a self-incompatible and partially clonal forest tree species - Prunus avium L. Molecular Ecology **15:** 2109-2118.

STORZ, J. F., and M. A. BEAUMONT, 2002 Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. Evolution **56:** 154-166.

THOMSON, R., J. K. PRITCHARD, P. D. SHEN, P. J. OEFNER and M. W. FELDMAN, 2000 Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. Proceedings of the National Academy of Sciences of the United States of America **97:** 7360-7365.

VALIERE, N., 2002 GIMLET: a computer program for analysing genetic individual identification data. Molecular Ecology Notes **2:** 377-379.

VALLEJO-MARIN, M., M. E. DORKEN and S. C. H. BARRETT, 2010 The Ecological and Evolutionary Consequences of Clonality for Plant Mating, pp. 193-213 in *Annual Review of Ecology, Evolution, and Systematics, Vol 41*, edited by D. J. FUTUYMA, H. B. SHAFER and D. SIMBERLOFF.

VAN LOO, M., J. A. JOSEPH, B. HEINZE, M. F. FAY and C. LEXER, 2008 Clonality and spatial genetic structure in Populus x canescens and its sympatric backcross parent P-alba in a Central European hybrid zone. New Phytologist **177:** 506-516.

VONLANTHEN, B., X. M. ZHANG and H. BRUELHEIDE, 2010 Clonal structure and genetic diversity of three desert phreatophytes. American Journal of Botany **97:** 234-242.

WANG, Y. X., B. ZHANG, X. Y. SUN, B. Y. TAN, L. A. XU *et al.*, 2011 Comparative genome mapping among Populus adenopoda, P. alba, P. deltoides, P. euramericana and P. trichocarpa. Genes & Genetic Systems **86:** 257-268.

WEEKS, A. R., and A. A. HOFFMANN, 2008 Frequency-dependent selection maintains clonal diversity in an asexual organism. Proceedings of the National Academy of Sciences of the United States of America **105:** 17872-17877.

YU, Q. B., N. MANTYLA and M. SALONEN, 2001 Rooting of hybrid clones of Populus tremula L. x P-tremuloides Michx. by stem cuttings derived from micropropagated plants. Scandinavian Journal of Forest Research **16:** 238-245.

Table 1 Representative clones (MLLs) found in Douro poplar hybrid zone, including large clones. *N*: Number of ramets; *Extension*: longest distance among ramets within clones; *Q*: ancestry value (1 for pure *P. alba*); F_i: inbreeding value; *Offspring*: number of descendants detected by either COLONY or FRANz software (see main text for details). Pa code for *P. alba* and Pc code for hybrid *P.* × *canescens*. NA: not applicable. The last two rows report averaged values between remaining genets. Bold font indicates significantly different from zero ($P < 0.05$). Parentheses indicate standard deviation.

MLL	**Species**	***N***	**Extension (km)**	***Q***	F_i	**Offspring**	
						COLONY	**FRANz**
MLL009	Pa	189	99.5	1.00	-0.013	29	21
MLL006	Pc	124	158.6	0.42	-0.121	7	2
MLL025	Pa	26	74.6	1.00	-0.116	nda23	23
MLL073	Pa	17	5.6	1.00	-0.048	1	2
MLL002	Pa	8	22.6	1.00	-0.066	0	0
MLL074	Pa	7	10.7	1.00	**0.226**	1	1
MLL011	Pa	6	4.1	1.00	-0.022	13	8
MLL049	Pa	5	17.5	1.00	-0.022	0	0
MLL057	Pc	5	0.6	0.80	0.002	0	2
MLL083	Pa	5	2.6	1.00	-0.059	0	2
MLL126	Pa	3	4.9	0.99	0.215	5	2
MLL086	Pa	2	1.5	1.00	0.162	5	4
MLL053	Pa	2	5.2	1.00	0.080	0	4
MLL111	Pa	2	17.1	0.99	0.166	0	2
MLL058	Pa	1	NA	1.00	0.170	3	2
MLL030	Pc	1	NA	0.90	0.002	0	1
MLL120	Pa	1	NA	1.00	**0.328**	8	4
Rest of *P. alba*		1.3 (0.5)	NA	1.00 (0.00)	0.064 (0.090)	0.08 (0.32)	0.24 (0.43)
Rest of *P. canescens*		1.3 (0.5)	NA	0.74 (0.10)	-0.006 (0.121)	0.00 (0.00)	0.40 (0.52)

Table 2 Clonal and genetic structure in Douro poplar hybrid zone compared to other European hybrid zones of the species. Genetic structure is based on MLLs only. Minimum sampling distance was 100 meters for the Douro population, whereas 8-10 meters and 50 meters were considered in the Danube and Ticino populations, respectively, to estimate clonal and genetic structure. NA: Not available. Bold font indicates significantly different from zero (P<0.01).

		Douro		**Danube**[a]		**Ticino**[b]	
		P. alba	*P. x canescens*	*P. alba*	*P. x canescens*	*P. alba*	*P. x canescens*
Clonal structure							
Number of ramets	N	362	143	222	185	23	26
Number of MLGs	G'	95	20	NA	NA	NA	NA
Number of MLLs	G	82	13	169	123	11	22
Genotypic richness	R	0.224	0.085	0.760	0.663	0.455	0.840
Simpson evenness	V	0.539	0.110	0.891	0.911	NA	NA
Log-scaled Pareto distribution							
Additive inverse of the slope	β	0.082	0.013	2.203	1.581	NA	NA
Coefficient of determination	r^2	0.721	0.411	0.911	0.935	NA	NA
Genetic structure							
Allelic richness	A	4.99	3.63	7.18	12.91	4.72	8.89
Rarefacted allelic richness[c]	A'	2.65	2.93	3.46	3.99	2.80	3.67
Number of private alleles[c,d]	A_p'	0.71	0.39	1.03	1.07	0.91	0.64
Genetic diversity	H_E	0.410	0.528	0.508	0.596	0.428	0.590
Inbreeding coefficient[e]	F_{IS}	**0.067**	-0.008	**0.143**	**0.180**	**0.173**	**0.124**
Genetic differentiation between species within population	F_{ST}		**0.115**		**0.018**		**0.189**

[a] Clonal and genetic structure statistics from VAN LOO *et al.* (2008) and LEXER *et al.* (2010), respectively.
[b] Clonal and genetic structure statistics from CASTIGLIONE *et al.* (2010) and LEXER *et al.* (2010), respectively.
[c] Rarefied to 10 chromosomes.
[d] Considering *P. tremula* alleles.
[e] Computed using within-species allele frequencies as reference.

Table 3 Posterior densities of main demographic parameters, calculated by MSVAR Bayesian simulations under a linear and an exponential model. L_1 and L_2 indicate the 0.025 and the 0.975 quantiles of the marginal posterior distributions. N_0: current effective population size, N_1: ancestral population size, μ: mutation rate per generation, and t_1: time (in years) since population started to decline or expand. θ is computed indirectly from the posterior distributions of N and μ.

	Linear model			*Exponential model*		
	L_1	*Mean*	L_2	L_1	*Mean*	L_2
N_0	54.20	1,733.80	55,847.02	71.94	2,243.88	69,823.24
θ_0	0.1311	0.2764	0.4782	0.1970	0.3367	0.5232
N_1	1,309.18	42,169.65	1,358,313.45	1,644.37	53,579.67	1,694,337.80
θ_1	2.234	6.449	13.843	2.4464	6.9630	13.8898
μ	2.254E-06	6.998E-05	2.203E-03	1.991E-06	6.109E-05	1.888E-03
t_1	2.911E04	1.019E06	3.648E07	1.148E04	3.908E05	1.352E07

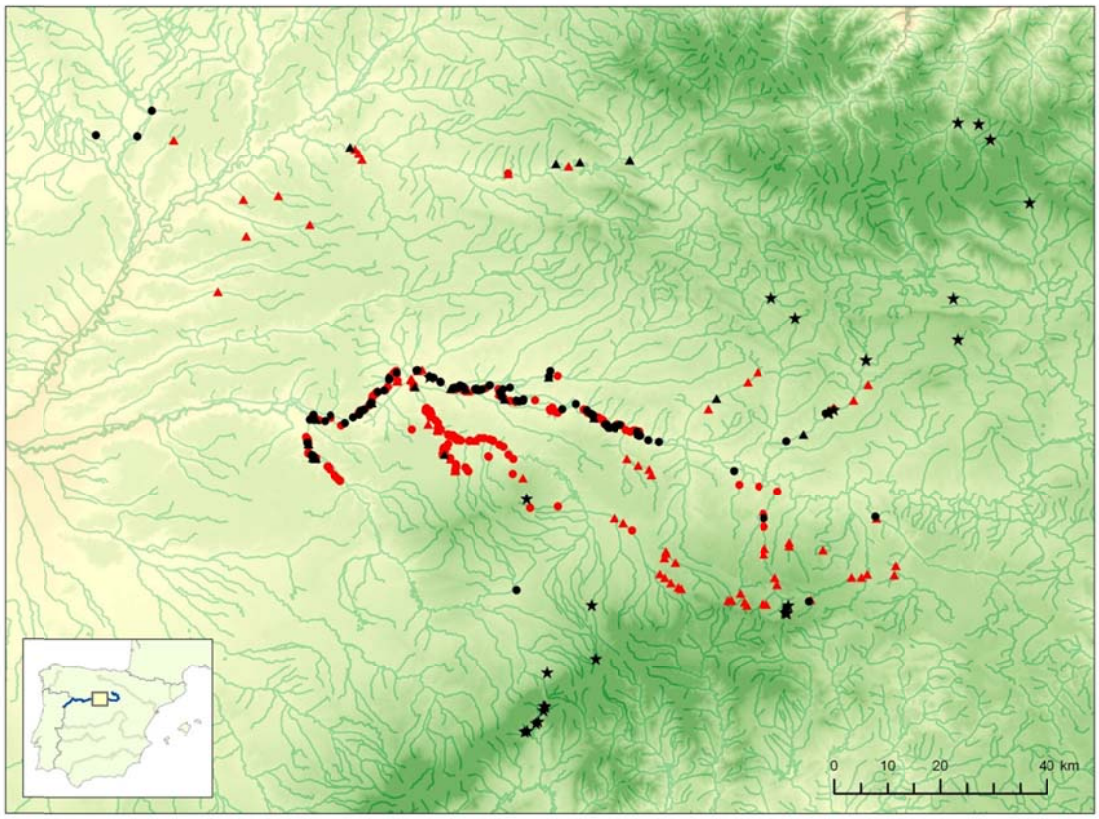


Figure 1 Location of individuals sampled within the Duero poplar hybrid zone. Circles represent *Populus alba*, triangles *P. x canescens*, and stars *P. tremula*. Red color represents large genets (>10 ramets), while black represents small genets (≤10 ramets).

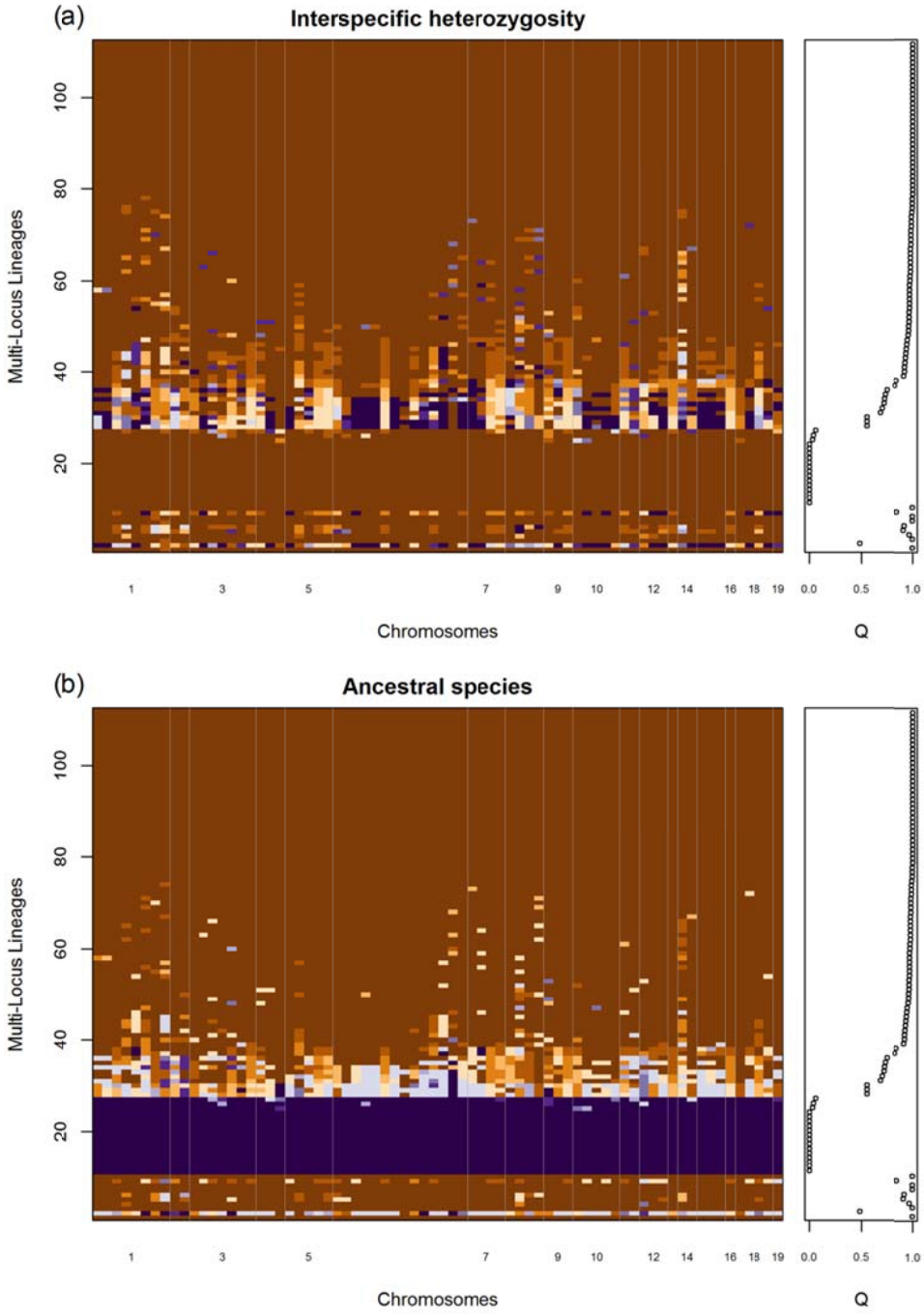
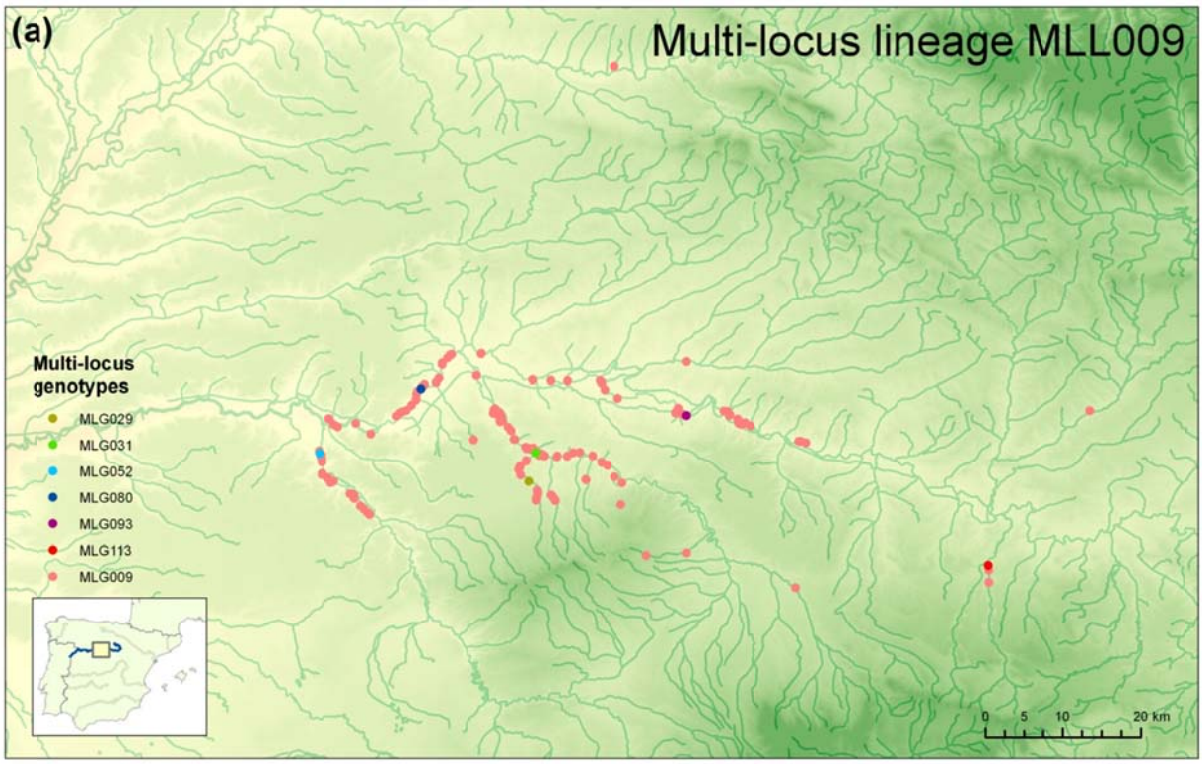


Figure 2 Genomic background of genets (112 multilocus lineages, MLLs) sampled in the Douro poplar hybrid zone. MLLs are represented in the following order from bottom to top: clones, *P. tremula*, *P. x canescens*, *P. alba.* Interspecific heterozygosity is depicted in blue, while homozygosity is in brown (a). In the intraspecific homozygosity plot (b) blue indicates *P. tremula* and brown *P. alba*. The ten bottommost genets represent the largest clones, the lower the larger. Loci are ordered following their genomic position, within chromosomes. Note that Chromosome XVII is not represented. STRUCTURE Q values are also shown (in the side graph).



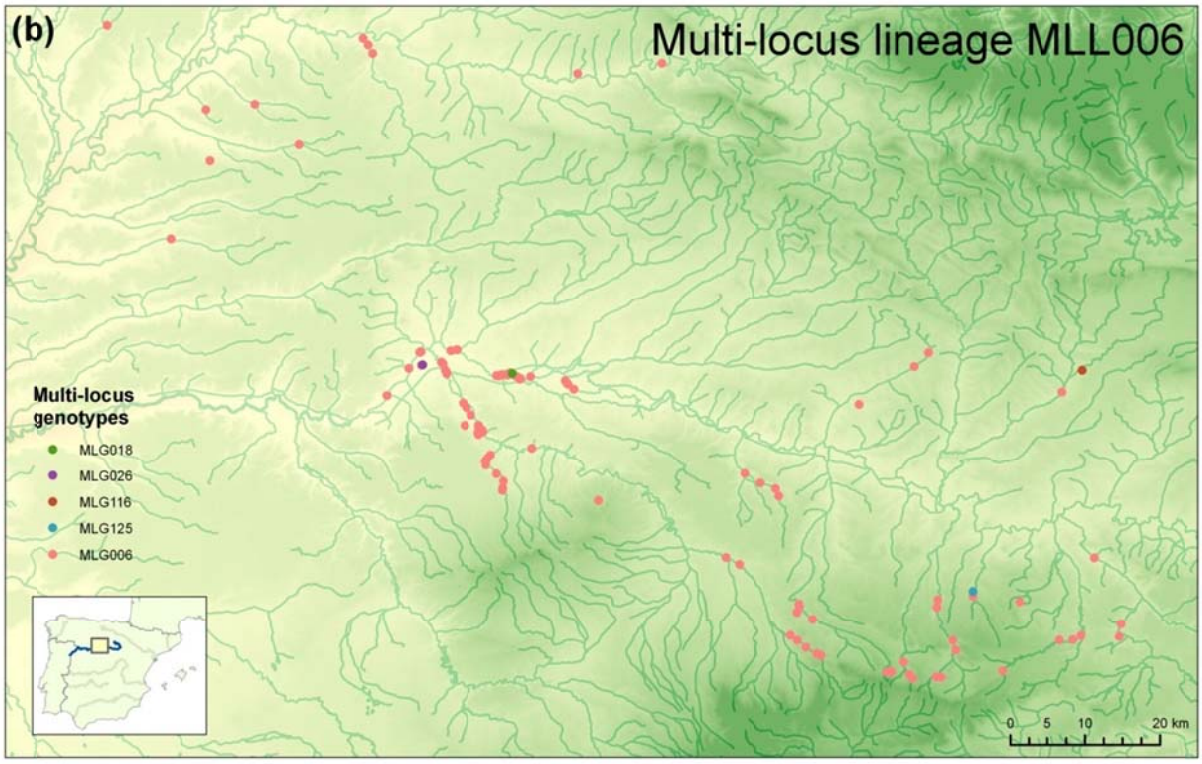
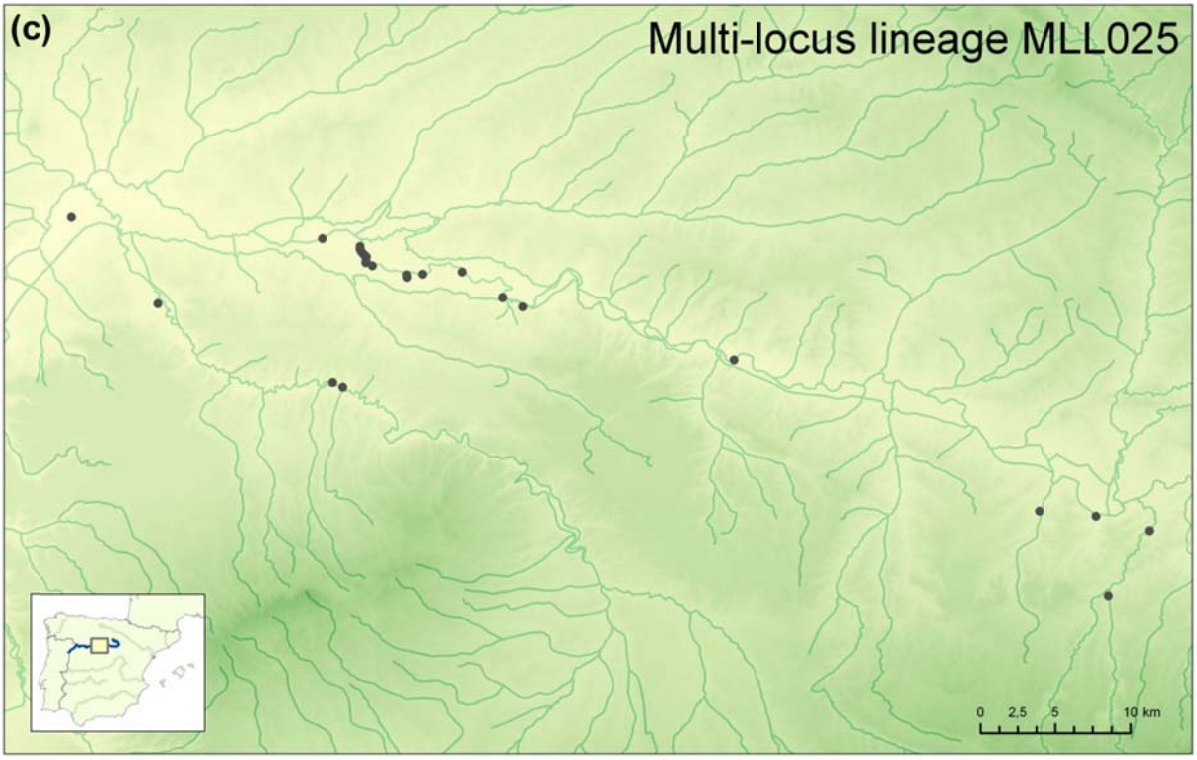


Figure S1 Geographic distribution of ramets for clones MLL009 (a) and MLL006 (b). Color key indicates the genotype (i.e. MLG) of each ramet.



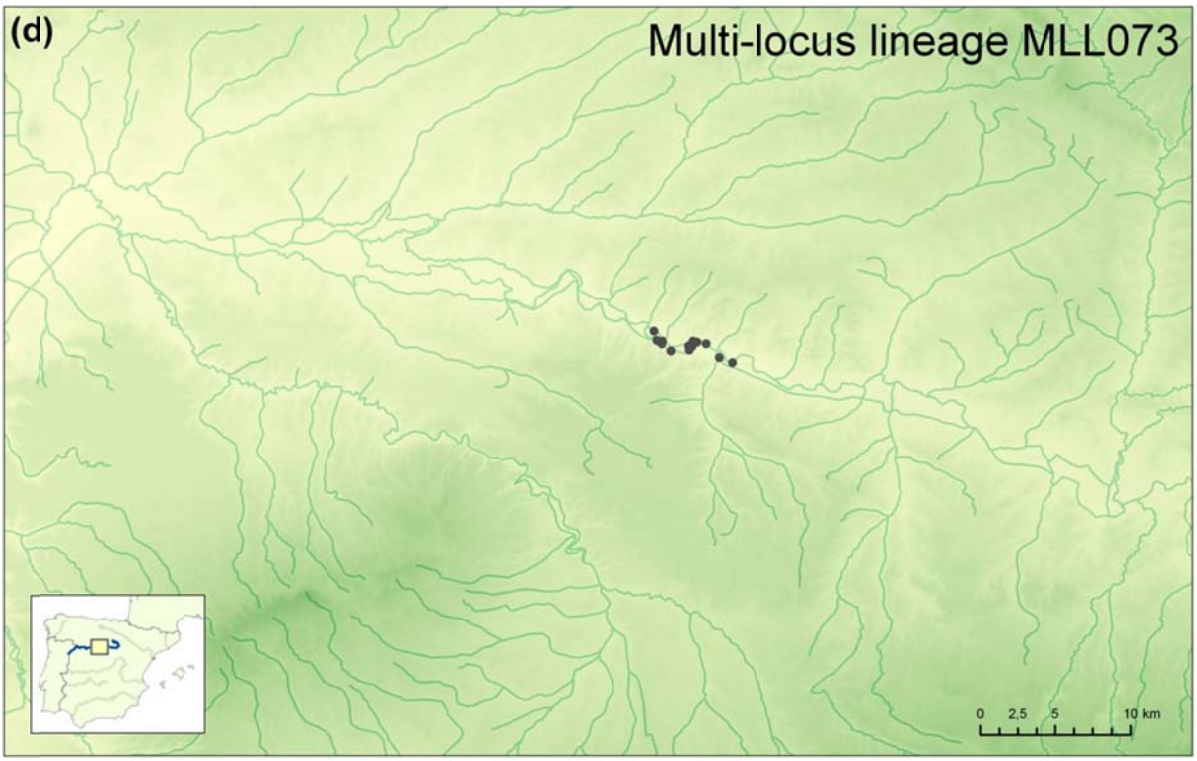


Figure S1 Continued. Clones MLL025 (c) and MLL073 (d).

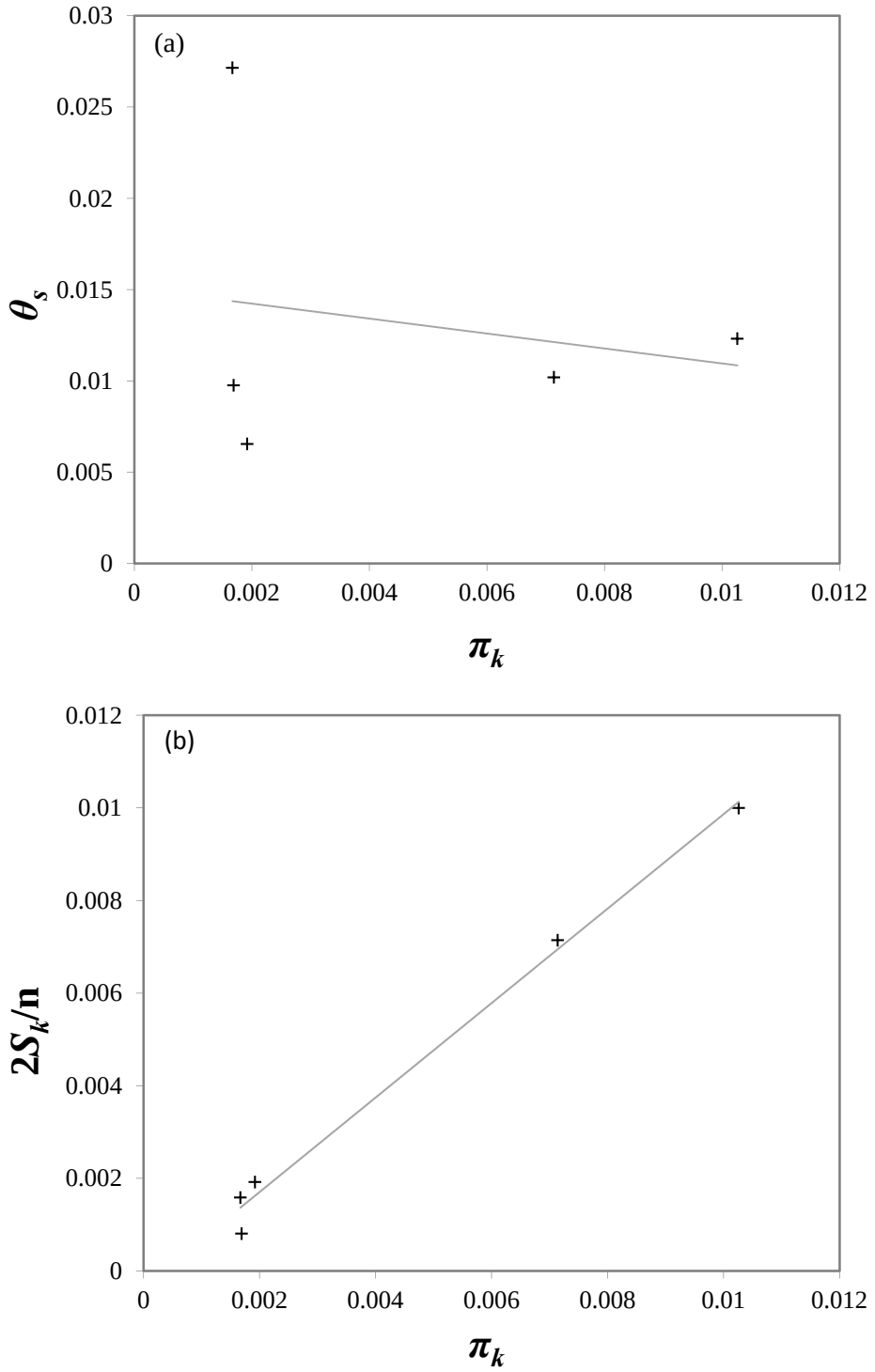


Figure S2 Correlation between number of within-genet pairwise differences per locus (π_k) and two measures based on number of segregating sites (S_k), θ_s (a) and $2S_k$/n (b). Under constant genet size, π_k correlates 1:1 with θ_s whereas for growing genets π_k correlates 1:1 with $2S_k$/n, as observed in this figure. Correlations are based on MLLs (5) for which somatic mutations were observed: MLL006, MLL009, MLL025, MLL049 and MLL074.

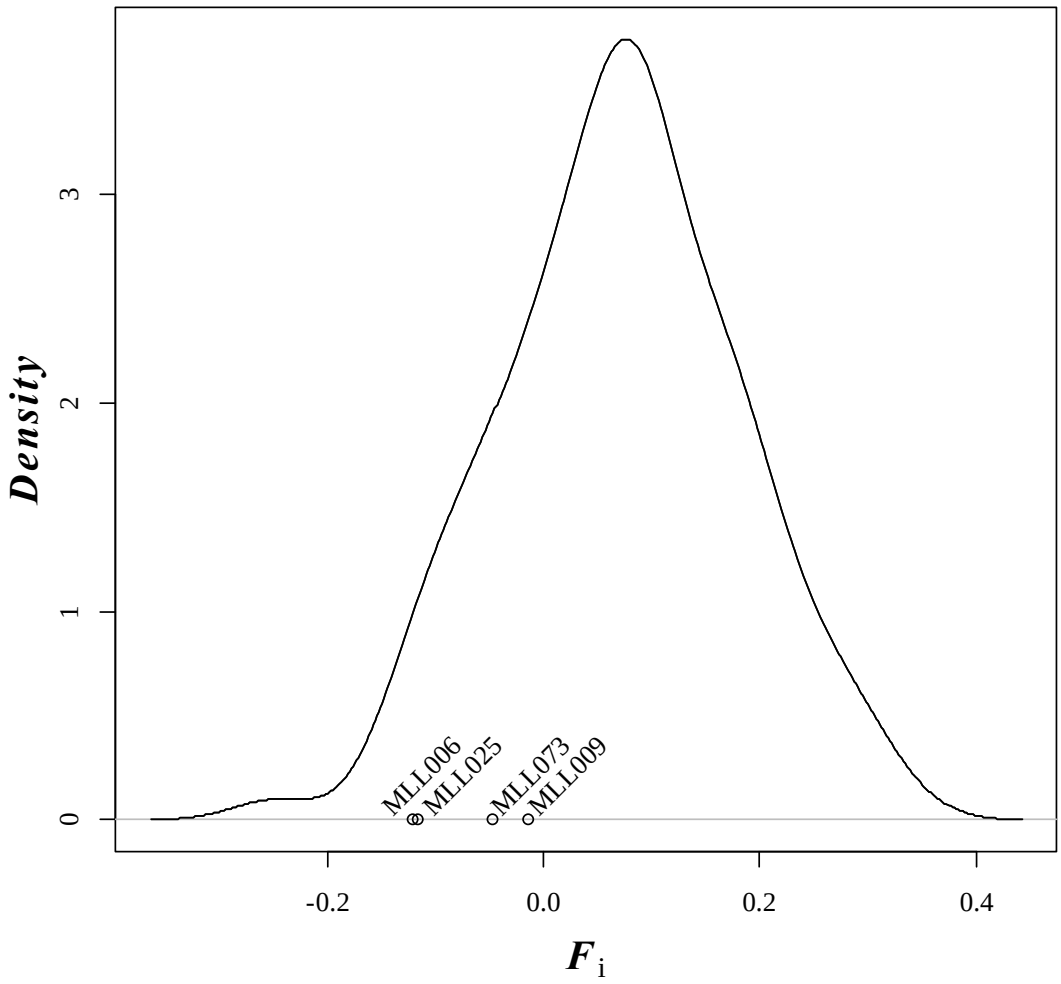


Figure S3 Distribution of individual (genet) inbreeding coefficients for the three species (*P. alba*, *P.tremula* and *P. x canescens*). Three independent computations were carried out for each species, using conspecific allele frequencies as reference for within-individual kinship calculation. Such allele frequencies were calculated only from parental individuals (as estimated by FRANz; see Material and Methods). Four small circles represent the four largest clones.

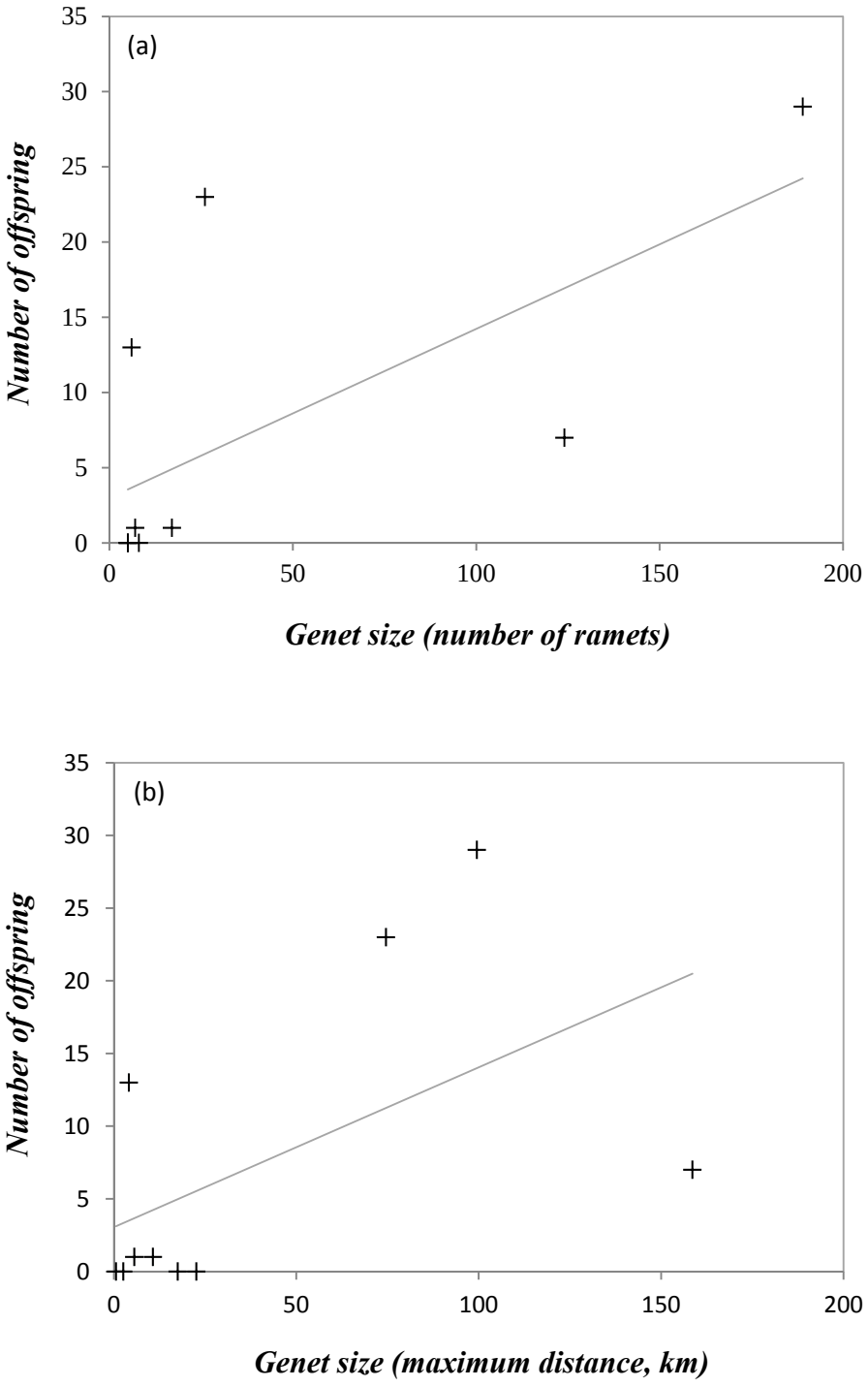


Figure S4 Correlation between genet size and number of offspring, either estimated as number of ramets (a) or as maximum distance between two ramets (b). Only the ten largest clones are represented. A linear trend is also provided.

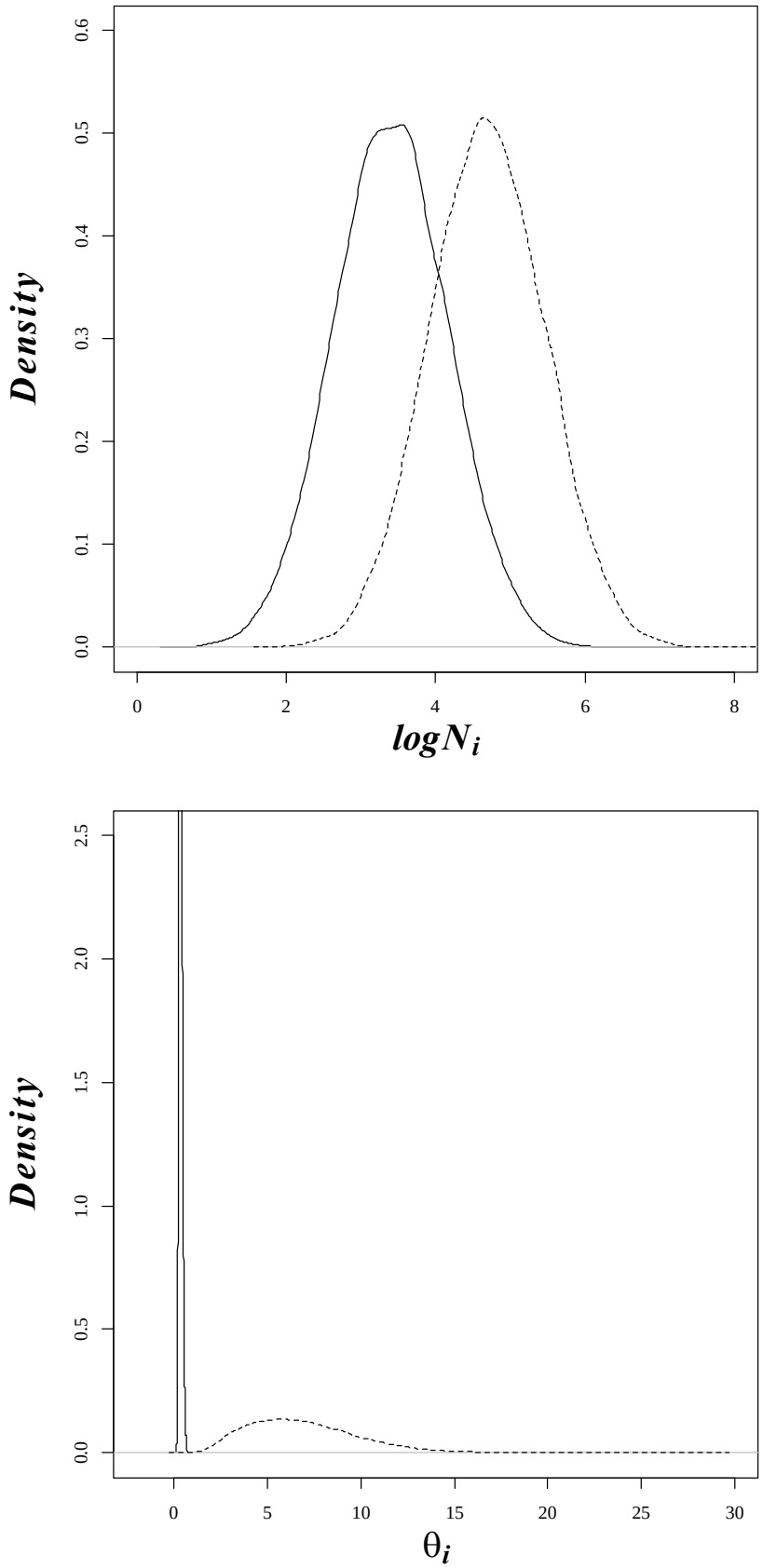


Figure S5 Posterior distributions of main population parameters obtained by MSVAR in an exponential size-change scenario. Solid line corresponds to current values, and dashed line to ancestral values.

Table S1 Nuclear microsatellites, including locus code, chromosome and genetic position, and the marker subsets used in the different data analyses.

Loci (primer code)	Chromosome	Genetic position (cM)	Initial marker subset	STRUCTURE marker subset	MSVAR marker subset	FRANz marker subset	Genet age estimation marker subset
ORPM 137	1	-1.00	X	X	X	X	X
GCPM 124	1	0.22		X	X		X
ORPM 030_1	1	17.34	X	X		X	X
ASP376	1	3.53		X	X	X	X
PMGC 2852	1	6.21	X	X	X	X	X
ASP302	1	79.48		X	X	X	X
GCPM 1719	1	25.22		X	X	X	X
GCPM 1274	1	51.21		X			
GCPM 1158	2	-1.00		X		X	X
GCPM 1376	2	102.18		X			
GCPM 1133	3	-1.00		X	X	X	X
ORPM 30_2	3	30.11	X	X	X	X	X
GCPM 1887	3	1.79		X	X		X
GCPM 1629	3	1.07		X	X	X	X
ORPM 203	3	3.50		X	X	X	X
GCPM 1869	3	23.98		X			
GCPM 1688	3	3.84		X			X
ORPM 127	4	-1.00	X	X		X	X
ORPM 220	4	6.66	X	X		X	X
GCPM 1809	4	6.66		X		X	X
GCPM 1255	5	-1.00		X		X	X
GCPM 1192	5	11.78		X	X	X	X
GCPM 1838	5	23.59		X	X	X	X
GCPM 20	5	71.85		X	X	X	X
WPMS 15	5	11.27	X	X	X	X	X
GCPM 139	6	-1.00		X	X	X	X
GCPM 1831	6	6.95		X	X	X	X
GCPM 1074	6	1.59		X			X
ORPM 26	6	8.99		X			
ORPM 167	6	0.17	X	X		X	X
GCPM 1485	6	6.73		X	X	X	X
ASP933	6	29.26		X		X	X
ORPM 190	6	3.49		X	X	X	X
WPMS 12	6	28.77		X	X	X	X
GCPM 2034	6	13.74		X	X	X	X
ORPM 369	6	2.89		X		X	X
ORPM 60	6	4.29	X	X			X
GCPM 1065	6	2.32		X		X	X
ASP322	6	5.33		X	X	X	X
GCPM 1260	7	-1.00		X	X	X	X
WPMS 17	7	26.29		X	X	X	X
GCPM 1416	7	2.76		X	X		X
ORPM 312_1	7	11.89	X	X	X	X	X
GCPM 2062	8	-1.00		X	X	X	X
ORPM 374	8	7.62		X			
ORPM 202	8	32.88	X	X	X	X	X
ORPM 268	8	1.37		X	X		X
GCPM 1949	9	-1.00		X	X	X	X
ORPM 23	9	13.56		X		X	X
ORPM 21	9	5.11	X	X			X
GCPM 2020	10	-1.00		X	X	X	X
ORPM 344	10	29.98	X	X	X	X	X
GCPM 1574	10	8.95		X	X		X
ORPM 149	10	0.27	X	X	X	X	X
GCPM 114	10	20.28		X	X	X	X
GCPM 1037	11	-1.00		X			
ORPM 29	11	33.77		X		X	X
GCPM 154	12	-1.00		X			
WPMS 05	12	2.06	X	X		X	X
GCPM 1186	12	23.32		X	X	X	X
GCPM 1353	13	-1.00		X	X	X	X
GCPM 1812	14	-1.00		X	X	X	X
GCPM 1306	14	39.86		X		X	X
GCPM 1894	15	-1.00		X	X	X	X
GCPM 1454	15	0.52		X	X		X
GCPM 1608	15	20.33		X	X	X	X
ORPM 14	16	-1.00	X	X	X	X	X
ORPM 214	18	-1.00	X	X		X	X
GCPM 1577	18	7.45		X	X	X	X
ORPM 28	18	29.60	X	X	X	X	X
GCPM 162	18	14.44		X		X	X
ORPM 206	19	-1.00	X	X		X	X
ORPM 312_2	U	-1.00	X			X	X
TOTAL			20	72	44	56	66

Primer sources: loci coded GCPM and PMGC, from http://www.ornl.gov/sci/ipgc/ssr_resource.htm; loci coded ORPM, from Tuskan et al. (2004; Can J Forest Res); loci coded WPMS, from van der Schoot et al. (2000; Theor Appl Genet), and Smulders et al. (2001; Mol Ecol Notes); and loci coded ASP, from de Carvalho et al. (2010, Mol Ecol).

Table S2 General information on MLLs sampled, including the code of an exemplary ramet, the species (Pa for white poplar, Pc for grey poplar, and Pt for European aspen), the number of ramets, the spatial extension of the MLL, the number of offspring assigned by COLONY, the identity of both parents asigned by COLONY, and the number of offspring assigned by FRANz.

MLL code	Code of exemplary ramet	Species	Number of ramets	Extension (km)	# offspring (COLONY)	Parent A identity	Parent B identity	# offspring (FRANz)
9	A012	Pa	189	99.53	29	*4	#3	21
25	A051	Pa	26	74.62	23	*6	#7	23
11	A022	Pa	6	4.14	13	A051	#1	8
120	C064	Pa	1	NA	8	*2	#18	4
6	A008	Pc	124	158.56	7	*13	#17	2
126	C084	Pa	3	4.89	5	*3	#18	2
86	A321	Pa	2	1.55	5	A051	A012	4
58	A217	Pa	1	NA	3	A321	#1	2
59	A220	Pa	2	0.12	2	A321	#1	0
135	T014	Pt	1	NA	2	C056	#19	1
73	A286	Pa	17	5.60	1	C084	A012	2
74	A287	Pa	7	10.68	1	*2	#5	1
42	A141	Pa	2	0.04	1	A022	A012	1
47	A147	Pa	2	0.06	1	A051	#5	1
101	A373	Pa	2	1.16	1	*2	A012	1
129	T003	Pt	2	2.17	1	*11	T014	0
117	C056	Pt	1	NA	1	*11	#16	0
133	T010	Pt	1	NA	1	*11	#19	1
2	AD111	Pa	8	22.60	0	*10	#5	0
49	A150	Pa	5	17.46	0	*6	#11	0
57	A209	Pc	5	0.62	0	A008	A012	2
83	A314	Pa	5	2.59	0	A321	#1	2
128	T001	Pt	4	6.07	0	*11	#19	0
134	T011	Pt	4	28.07	0	*11	#19	0
34	A132	Pa	3	2.38	0	A051	#9	1
108	C002	Pa	3	134.92	0	*5	#6	1
140	T021	Pt	3	1.49	0	*11	#19	0
14	A001	Pa	2	0.05	0	A217	#4	1
12	A003	Pa	2	0.31	0	A022	C064	0
21	A025	Pa	2	0.12	0	A051	#5	0
23	A043	Pa	2	0.06	0	A051	#1	0
24	A049	Pa	2	10.45	0	*5	#6	0
33	A130	Pa	2	3.91	0	A051	#5	1
53	A200	Pa	2	5.24	0	A220	#1	4
55	A203	Pc	2	0.08	0	A008	A012	1
56	A207	Pa	2	0.09	0	A321	#1	1
61	A225	Pa	2	6.80	0	*10	#4	0
82	A312	Pa	2	2.84	0	C084	A012	1
103	A383	Pa	2	0.33	0	*3	A012	0
105	A386	Pa	2	13.62	0	*3	A012	1
4	AD201	Pa	2	0.56	0	A022	#4	0
107	C003	Pa	2	7.71	0	*12	#15	0
111	C045	Pa	2	17.12	0	C084	C064	2
118	C057	Pc	2	9.55	0	*1	#3	1
136	T016	Pt	2	22.95	0	*11	#16	0
138	T018	Pt	2	3.56	0	*11	#16	0
15	A004	Pa	1	NA	0	A022	A012	0
16	A005	Pa	1	NA	0	A022	C064	1
19	A011	Pc	1	NA	0	A008	A012	1
22	A038	Pa	1	NA	0	A051	A012	0
28	A062	Pa	1	NA	0	*7	#2	0
30	A088	Pc	1	NA	0	*4	#3	1
32	A129	Pa	1	NA	0	A051	#8	0
35	A134	Pa	1	NA	0	A051	C064	0
36	A135	Pa	1	NA	0	*8	#7	0
37	A136	Pa	1	NA	0	A022	A012	1
38	A137	Pa	1	NA	0	A051	#8	0
39	A138	Pa	1	NA	0	A051	#5	0
40	A139	Pa	1	NA	0	A051	#8	0

Table S2 Continued

MLL code	Code of exemplary ramet	Species	Number of ramets	Extension (km)	# offspring (COLONY)	Parent A identity	Parent B identity	# offspring (FRANz)
41	A140	Pa	1	NA	0	A022	#10	0
43	A143	Pa	1	NA	0	A022	#10	0
44	A144	Pa	1	NA	0	*9	A141	0
45	A145	Pa	1	NA	0	*4	#5	0
46	A146	Pa	1	NA	0	A051	#9	0
48	A149	Pa	1	NA	0	*6	A147	0
50	A151	Pa	1	NA	0	A051	A012	0
51	A183	Pc	1	NA	0	A008	A012	0
54	A202	Pa	1	NA	0	A051	#1	0
60	A222	Pa	1	NA	0	A217	A012	1
64	A253	Pa	1	NA	0	*6	#1	1
66	A263	Pa	1	NA	0	A051	C064	0
67	A266	Pa	1	NA	0	*9	#11	0
69	A277	Pc	1	NA	0	*6	#12	0
70	A281	Pa	1	NA	0	*6	#9	0
71	A282	Pa	1	NA	0	A051	#13	0
72	A283	Pa	1	NA	0	*8	#2	0
75	A289	Pa	1	NA	0	C084	A012	0
76	A296	Pa	1	NA	0	*6	#12	1
77	A302	Pa	1	NA	0	*7	#12	0
78	A303	Pa	1	NA	0	*6	#14	0
79	A307	Pa	1	NA	0	A220	#15	0
81	A311	Pc	1	NA	0	A008	A012	1
85	A318	Pa	1	NA	0	A217	A012	0
89	A329	Pa	1	NA	0	A321	#1	0
91	A343	Pa	1	NA	0	A051	A012	0
94	A361	Pa	1	NA	0	A051	A373	0
96	A363	Pa	1	NA	0	A051	#2	0
97	A367	Pa	1	NA	0	A051	A012	0
98	A368	Pa	1	NA	0	A051	#1	0
99	A369	Pc	1	NA	0	*1	#3	0
100	A370	Pa	1	NA	0	A287	A286	0
104	A384	Pa	1	NA	0	*3	A012	0
3	AD171	Pa	1	NA	0	C084	A012	0
5	AD203	Pa	1	NA	0	A022	A012	0
7	AD221	Pa	1	NA	0	A022	A012	1
8	AD222	Pa	1	NA	0	A022	A012	0
10	AD231	Pa	1	NA	0	A022	C064	0
13	AD251	Pa	1	NA	0	A022	#14	0
1	AD91	Pt	1	NA	0	*11	#16	0
109	C010	Pc	1	NA	0	A008	A012	0
114	C050	Pc	1	NA	0	A008	C064	1
115	C052	Pa	1	NA	0	*12	#9	0
119	C060	Pc	1	NA	0	*1	A012	0
122	C067	Pt	1	NA	0	*11	#16	0
123	C068	Pt	1	NA	0	*11	#19	0
124	C078	Pa	1	NA	0	*2	A012	0
127	C101	Pa	1	NA	0	*6	C064	0
130	T006	Pt	1	NA	0	*11	T014	0
131	T007	Pt	1	NA	0	*11	#16	0
132	T008	Pt	1	NA	0	*11	#16	0
137	T017	Pt	1	NA	0	*11	#19	0
141	T024	Pt	1	NA	0	T010	T003	0

NA: Not applicable, as only one ramet was found.

Prefixed star (*) in the column 'Parent A identity' indicates that the potential parent in not is the sample

Prefixed hash (#) in the column 'Parent B identity' indicates that the potential parent is not in the sample

Universidad de **Valladolid**