Article:
Macaya-Sanz, D., Heuertz, M., Lindtke, D. et al. (3 more authors) (2016) Causes and consequences of large clonal assemblies in a poplar hybrid zone. Molecular Ecology. ISSN 0962-1083

https://doi.org/10.1111/mec.13850

Received Date : 13-Mar-2015

Revised Date : 02-Sep-2016

Accepted Date : 06-Sep-2016

Article type : Original Article

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

David Macaya-Sanz[1,2], Myriam Heuertz[3], Dorothea Lindtke[4,5], Giovanni G. Vendramin[6], Christian Lexer[4,7], Santiago C. González-Martínez[1,3*]

[1] Department of Forest Ecology and Genetics, INIA-Forest Research Centre, 28040 Madrid, Spain.

[2] Department of Biology, West Virginia University, Morgantown WV 26505, USA.

[3] BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France.

[4] Unit of Ecology and Evolution, Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland.

[5] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

[6] Institute of Biosciences and Bioresources, National Research Council, 50019 Sesto Fiorentino (Florence), Italy.

[7] Department of Botany and Biodiversity Research, Faculty of Life Sciences, University of Vienna, A-1030 Vienna, Austria.

*Corresponding author:

Santiago C. González-Martínez

BIOGECO, INRA, 69 route d'Arcachon, 33610 Cestas, France

Ph: +33 557122757, E-mail: santiago.gonzalez-martinez@pierroton.inra.fr

Running head: Large clonal assemblies in poplar

Key words: clonality, hybrid zone, specific ancestry, Populus spp., Iberian Peninsula

Abstract

Asexual reproduction is a common and fundamental mode of reproduction in plants. Although persistence in adverse conditions underlies most known cases of clonal dominance, proximal genetic drivers remain unclear, in particular for populations dominated by a few large clones. In this paper, we studied a clonal population of the riparian tree *Populus alba* in the Douro river basin (northwestern Iberian Peninsula) where it hybridizes with *P. tremula*, a species that grows in highly contrasted ecological conditions. We used 73 nuclear microsatellites to test whether genomic background (species ancestry) is a relevant cause of clonal success, and to assess the evolutionary consequences of clonal dominance by a few genets. Additional Genotyping-by-Sequencing (GBS) data were produced to estimate the age of the largest clones. We found 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 produced. Interestingly, large clones possessed two genomic regions introgressed from *P. tremula*, which may have favored their spread under stressful environmental conditions. At the population level, the spread of large genets was

accompanied by an overall ancient (>0.1 Myr) but soft decline of effective population size. Despite this decrease, and the high clonality and dominance of sexual reproduction by large clones, the Douro hybrid zone still displays considerable genetic diversity and low inbreeding. This suggests that, even in extreme cases as in the Douro, asexual and sexual dominance of a few large, geographically-extended individuals does not threaten population survival.

Introduction

Asexual reproduction is common in plant species, affecting ~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). Assurance of plant persistence in unpredictable or recently-colonized environments has been proposed as the main explanation for the prevalence of clonality over sexual reproduction (Vallejo-Marin *et al.* 2010). Once established in a population, clonality can also be favored by functional specialization among ramets and better access to constrained resources (also called 'foraging') (Vallejo-Marin *et al.* 2010; and references therein). In addition, asexual reproduction can limit sexual reproductive output due to competition of (normally vigorous) asexual sprouts with sexual propagules, accumulation of somatic mutational load on fertility traits and increased geitonogamy (i.e. mating between ramets; Ally *et al.* 2010; Barrett 2002; Eckert 2001). Moreover, the existence of large clonal assemblies will exacerbate these processes. Hence, sexual recruitment limitation is both a cause and a consequence of increased clonality and may create a detrimental feedback loop, ultimately resulting in loss of sexual function (Honnay & Bossuyt 2005).

Across evolutionary time scales, Rice (2002) found medium-term propagation success of clonal species, but also frequent extinction of phylogenetic branches derived from them. This is in accordance with theoretical models predicting selection to favor sex and recombination over clonality in the long term (Otto & Lenormand 2002). However, the evolution of highly clonal plant taxa, such as some seagrasses (Arnaud-Haond *et al*. 2012 and reference therein) or tree species from the Salicaceae family, seems to oppose this finding. Within the Salicaceae, *Populus* species (e.g. *Populus alba* and *P. tremuloides*; Brundu *et al.* 2008; Slavov & Zhelev 2010) contain the largest natural terrestrial clones currently known (Mock *et al.* 2008; and references therein). In these species, the prevalence of clonality is variable depending on the population and geographical range. Clonal size can be estimated using the number of observed ramets or the spatial distribution of clones. Typical Pareto distributions of clone size, with few large and many small genets, have been reported in most plant species and populations that exhibit a clonal structure (Arnaud-Haond *et al.* 2007). Nevertheless, extremely uneven size distributions, such as those found in *Populus* (e.g. Santos-del-Blanco *et al*. 2013 for *P. alba* in central Spain), are uncommon.

The spread of particular genets may depend on extrinsic factors, such as ecological conditions that favor specific adaptive traits, but also on intrinsic genetic drivers such as the genetic ancestry of competing clones. The role of ecological conditions on clonal structure has been extensively reported. For instance, in *Populus*, several studies have pointed to the level of moisture and flood control by human-made infrastructures as factors affecting the rate of clonality and effective sexual recruitment (Gonzalez *et al.* 2010a for *P.alba*; Slavov *et al.* 2010; Vonlanthen *et al.* 2010). Because there is genetic variation for traits involved in asexual reproduction, as shown in greenhouse experiments (e.g. Stenvall *et al.* 2005; Yu *et al.* 2001), some genotypes should have better capacity for clonal spread under specific

environmental conditions. Clones with higher heterozygosity, in particular those introgressed by closely-related species, could display higher fitness due to heterosis and/or heterozygote advantage (Charlesworth & Willis 2009; Hedrick 2012). Conversely, introgression of genetically distinct populations/species may result in reduced fitness through breakdown of co-adapted gene complexes (i.e. outbreeding depression, Lynch 1991). However, the role of genomic ancestry on clonal spread has never been thoroughly assessed in natural populations.

The effects of prolonged asexual reproduction on populations, if any, are not clear. On the one hand, competition among genets eventually reduces genotypic richness and may impact on genetic diversity. Even though the effect of shrinking genotypic richness on genetic diversity is disputed (Balloux *et al.* 2003; Mock *et al.* 2008), it is a fact that loss of genotypes and their specific genetic combinations will lead to a reduction in functional diversity and may increase inbreeding. Concurrently, the uneven number of ramets per clone in populations that exhibit a few clones of large size could generate a mating bias towards larger clones, leading to a long-term decrease of effective population size and increased genetic drift. Nevertheless, in most situations, a higher rate of clonality (evaluated as the proportion between clonal and sexual reproduction) does not affect the asymptotic trend of inbreeding (F_{IS}), but only slows down the process to reach Hardy-Weinberg equilibrium (Reichel *et al*. 2016). On the other hand, assuming that larger clones are also older, the lack of selection against deleterious mutations on sexual function may have eroded the sexual capability of large clones, neutralizing their plausible sexual dominance. Considering the plausible co-occurrence of these two opposing processes, it would be helpful to gain empirical knowledge on the effects of clonality and the inherent reduction of

genotypic richness on sexual mating bias and potential long-term consequences for genetic diversity, effective population size and inbreeding.

Natural hybrid zones of the highly clonal European poplars of section *Populus* (*P. alba*, *P. tremula* and their hybrids) are outstanding natural laboratories to study the influence of genomic background on adaptation, clonal spread, and reproductive isolation (van Loo *et al*. 2008; Stölting *et al*. 2013; Christe *et al*. 2016). These species display different ecological habits, with *P. alba* being a riparian lowland warm-temperate tree while *P. tremula* grows in its Mediterranean range at locations with boreal climate (high-altitude hillsides). Distinguishing *P. alba* from *P. tremula* is easy, based especially on the shape and pubescence of leaves and petioles; they are also distinguished by overall shape, bark and many other morphological features. However, ascertainment of hybrids is not always unambiguous, especially for second (F2 and BC1 backcrosses) and later generations, since hybrids do not usually show intermediate morphology for most traits, due to, for example, transgressive segregation or hybrid vigor (Rieseberg & Carney 1998). Intriguingly, clonality is highly variable across European poplar hybrid zones, from being relatively low in Central Europe (largest clone size < 200 m; van Loo *et al.* 2008) to very high, with large clones spreading over 150 km in the Iberian Peninsula (Santos-del-Blanco *et al.* 2013), providing room for comparative analyses in natural populations. While ecological causes of differences in clonal spread have been investigated in different *Populus* species (Slavov *et al.* 2010; Vonlanthen *et al.* 2010), the role of genomic ancestry (in particular hybridization and introgression) remains unexplored despite early evidence of fitness differences between hybrids and parental species in natural populations (e.g. Schweitzer *et al*. 2002; van Loo *et al.* 2008) and agronomic trials (e. g. Marron *et al.* 2010).

In this work, we focus on a European poplar hybrid zone (located in the Douro river valley, northwestern Iberian Peninsula) that is characterized by large and geographically extended clonal assemblies (Santos-del-Blanco *et al.* 2013) and low levels of hybridization. Multi-locus genotypes from 73 nuclear microsatellite loci were used to identify clones and determine their age, as well as to discern their genetic background (as in Lindtke *et al.* 2012). In addition, 0.507-1.249 Gbp (giga base pairs) of sequence data were obtained from the two largest clones by means of Genotyping-by-Sequencing to estimate their age with higher precision. We then tested the following hypotheses: (i) genomic background (species ancestry) is associated with clonal success; and (ii) ancient large clones have higher sexual reproductive success, ultimately affecting overall population demography and/or increasing inbreeding. Our study provides novel insights into the demographic trajectories of highly clonal populations and, in particular, is relevant to the management and conservation of riparian forests in the face of environmental change.

Material and Methods

Study site and sample collection

The Douro poplar hybrid zone occurs in the riparian forests that border the middle course of the Douro River and its tributaries, a river system in the Iberian Northern Plateau which flows westward to the Atlantic Ocean. Because of its ecological value and relatively low level of human intervention, the area has been proposed as a site of Community importance by the European Commission (codes: ES4170083 and ES4120068). Different *Populus* taxa grow naturally in this region (*Populus alba*, *P. tremula, P.* × *canescens* and *P. nigra*).

Leaf tissue was collected from 533 poplar trees, whose positions were recorded using a GPS device. The sampling comprised the two pure species of section *Populus* (*P. alba* and *P. tremula*) and their natural hybrid (*P.* × *canescens*). *Populus nigra* was not included in the study because it belongs to section *Aigeiros,* which does not naturally hybridize with section *Populus*. *Populus alba* (*N*=360) and *P.* × *canescens* (*N*=145) samples were collected mainly along ~125 km of the hybrid zone, while *P. tremula* samples (*N*=28) were collected in stands located in the mountains nearby, where it is more abundant (Figure 1). Following previous studies (Macaya-Sanz *et al.* 2012), sampling was designed to cover a large geographical area while avoiding overrepresentation of local genets. Sampled stems (ramets) were spaced at intervals of least 100 meters in the core area (Douro middle course and the lower courses of major tributaries), and at intervals of at least 1,000 meters elsewhere. Despite the uneven sampling effort, all known stands in the surveyed area were represented in our sampling.

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 microsatellite (SSR) markers (see Table S1 in Supporting Information; Lexer *et al.* 2005). Once Multi-Locus Genotypes (MLGs) and Multi-Locus Lineages (MLLs; i.e. those ignoring somatic mutations) were resolved (see below), at least one sample of each MLL was genotyped with an additional set of 53 SSR markers spaced evenly along the poplar genome (137 samples and 73 markers total; Table S1 in Supporting Information).

The first set of markers (20 SSRs) was amplified following the protocols of Macaya-Sanz *et al.* (2012). PCR fragment analysis was carried out with a 4300 DNA Analyzer (Li-Cor Biosciences, Lincoln, NE, USA), using internal standards to facilitate allele 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.

Genotyping-by-Sequencing (GBS)

To estimate the age of the two largest clones (MLL006 and MLL009, see *Results*) with higher precision, we genotyped nine ramets of each using Genotyping-by-Sequencing (GBS). One ramet of each clone was genotyped twice to control for technical error. DNA was digested using the MSeI and EcoRI enzymes and the libraries were prepared following the protocol described by Parchman *et al.* (2012) with slight modifications. Then, libraries were sequenced in a 150-cycle single-end run (V3 chemistry) on a MiSeq Illumina sequencer.

Sequences obtained by GBS were processed independently for each clone using the package Stacks version 1.4 (Catchen *et al.* 2011, 2013). After standard quality filtering, sequences were trimmed and cropped using Trimmomatic version 0.36 (Bolger *et al.* 2014). We split the reads into two sets of 60-bp length that were pipelined independently, yet merged for subsequent statistical analyses (see Table S2 in Supporting Information for number of reads and coverage per sample). The rationale for analyzing two sets of 60-bp fragments instead of only one set of longer fragments was to include more overall sequence data, given that Stacks needs all fragments to be trimmed and cropped to a unique length.

Moreover, it reduced the number of haplotype mismatches for subsequent analyses. We did not use the available *P. trichocarpa* or *P. tremula* reference genome sequences to assembly reads, given that the high presence of orthologous and paralogous genes, even in related species, would have induced spurious alignments potentially hindering the detection of true somatic mutations. Besides, given that the ten samples (including the technical replicate) belonged to the same clone, we expected high levels of similarity that would facilitate stack formation even without a reference genome. To control for PCR and MiSeq sequencer errors, the depth of stacks (*m* parameters) was set high (to six for the first set of 60 bp and to four for the second one). The other parameters (*M*, *N*, *max_locus_stacks*) were set to two and the deleveraging and removal algorithms were enabled to control the merging of haplotypes from paralogous loci. After these steps, 1.249 Gbp of aligned sequence (20,810 stacks) for MLL009 and 0.507 Gbp of aligned sequence (8,451 stacks) for MLL006 were available for further analyses.

Multilocus genotypes (MLGs) and multilocus lineages (MLLs)

The first set of 20 SSR loci allowed for initial MLG assignment, as provided by GIMLET software (Valiere 2002). A further analysis of divergence among MLGs permitted to collapse ramets with somatic variants into sexually derived genets (MLL assignment), following Arnaud-Haond *et al.* (2007). The rationale of this assignment method rests in the fact that the divergence among MLGs derived from asexual reproduction should be much lower (i.e. involving only few somatic mutations) than that among sexually-produced MLGs (i.e. after recombination of two distinct genomes). When computing pairwise genetic distances between MLGs, a bimodal distribution is expected, where the smaller peak at shorter genetic distances represents those MLG pairs that are differentiated by only a few somatic

mutations and that can thus be grouped within the same MLL (Figure S1 in Supporting Information). Finally, the geographical distribution of MLLs and MLGs was plotted to examine visually the spatial distribution of clones. Geographical representation of the whole population and the largest clonal assemblies was carried out with a GIS (ArcMap version 9.2; ESRI, Redlands, CA, USA).

Age of large clones

The number of somatic mutations within a genet is correlated with its age. Thus, given a sufficient number of molecular markers, genetic divergence within genets can be used to estimate clone age (assuming neutrality). Seventy-three nuclear microsatellite loci were used to estimate the age of the two largest clones, MLL006 and MLL009, based on seven and eight ramets with somatic mutations, respectively. In addition, a GBS run was conducted on nine ramets plus a technical replicate for each of these two clones to obtain complementary age estimates using a larger set of markers. We did not provide age estimates based on clone size (estimated as maximum geographic distance between ramets) because expansion of clones in a riparian ecosystem is likely anisometric. Moreover, long-distance translocation of propagules by water has been documented in poplars (Barsoum *et al.* 2004).

Somatic mutations were counted directly for SSRs and the ancestral state was identified based on allele frequencies. For GBS data, the stacks for each clone were screened for low-frequency polymorphisms. Then, only those SNPs with a frequency below or equal to 0.1 based on at least 5 ramets (to account for missing data and allele dropout; Andrews et al. 2016) were considered probable somatic mutations. Besides, to reduce the possibility of calling a spurious SNP produced by the merging of paralogous sequence, only

stacks with a single SNP were considered. The number of mismatches in low frequency SNPs (presence *vs.* absence in genotypes), as defined above, between technical replicates was used as a baseline to control for spurious calls. Then, the number of somatic mutations in the genet was calculated by taking the maximum number of mismatches in low frequency SNPs found among all the pairwise comparisons between ramets, and subtracting from it the baseline calculated from the technical replicates. Note that this number represents a conservative estimate, as we have established stringent conditions to detect somatic mutations and, moreover, missing data have probably prevented some somatic mutations from being detected in specific pairwise comparisons.

The accumulation of mutations since the common ancestor in different ramet lineages within a MLL is a Poisson process (Thomson *et al.* 2000). Hence, the number of independent somatic mutations should follow a Poisson distribution. We note that this number must be estimated counting each somatic mutation just once, even when they may be present in several ramets. The expected number of somatic mutations is then equal to $\mu_{loc} \times TMRCA$ (Thomson *et al.* 2000), where μ_{loc} is the mutation rate scaled to twice the number of loci and *TMRCA* the time to the most recent common ancestor, i.e. an approximation of the clone age (Thomson *et al.* 2000). Mutation rate per year for SSRs was obtained from MSVAR (see below) considering a generation time of 40 years (as in Macaya-Sanz *et al*. 2012). For sequences obtained by GBS, we used a previous estimation for *P. tremula* ($2.5E^{-9}$ per year; Ingvarsson 2008). An Infinite Allele Mutation (IAM) model was assumed, which is reasonable, considering the much larger number of loci than mutations for both SSRs and GBS and the low probability of a somatic mutation occurring twice at the same site.

Genomic background of large clones and locus-specific ancestry

Genome-wide ancestry for each MLL was estimated by computing admixture proportions (*Q*) using an admixture model with correlated allele frequencies in STRUCTURE vs. 2.3.2.1 (Pritchard *et al*. 2000; Falush *et al.* 2003), and by applying the linkage model to obtain ancestry estimates per locus, i.e. locus-specific ancestries (LSAs). As this model incorporates information from linked markers for ancestry estimation, only 72 markers with known genomic position were used (listed in Table S1). Setting the number of clusters (*K*) to two, the probabilities that zero, one, or both alleles at each locus have *P. alba* ancestry (ss1, ss2+ss3, or ss4) were estimated. Following Lexer *et al*. (2010) and Lindtke *et al*. (2012), 'specific homozygosity' (ss1 – ss4) and 'interspecific heterozygosity' ([ss2+ss3] – [ss1+ss4]) were calculated and plotted using R (R Development Core Team 2009). Briefly, 'specific homozygosity' refers to the statistical support for both allele copies at a locus having ancestry from one of the species, *P. alba* or *P. tremula*, whereas 'interspecific heterozygosity' refers to the support for each allele copy at a locus having ancestry from the same vs. different parental species (see details in Lindtke *et al.* 2012).

To identify loci with unusual introgression in the ten largest clones but not the remaining MLLs, LSA was computed as (ss2+ss3)/2 + ss4 (so that interspecific heterozygosity is taken into account; values ranging from 0 to 1). Loci with unusual introgression were identified by first assigning ranks for LSA to MLLs, separately for each locus (rank function in R, taking average ranks for ties). Then, for each locus, ranks were summed for the ten largest clones (either in terms of number of ramets or spatial extension), and for the remaining clones. Loci with the 5% highest or lowest values in large clones but not so in the remaining clones were determined as outliers. This procedure was repeated ten times leaving one large clone out at each time (jackknife resampling), and only loci identified in all

subsets were considered as robust outliers. As LSA scores can be affected by uncertainty in genotype, e.g. due to low genetic differentiation between parental species or missing data, the analysis was repeated (but without the jackknife resampling) using discrete values of the ancestral genotype instead of continuous LSA (LSAGs; i.e. scores of 0, 0.5, or 1 for the three possible genotypes given their maximum posterior probability). Finally, to determine the expected number of outlier loci in the ten largest clones, 1,000 permutations of clone size (keeping genotypes) were generated.

Mating success of large clones

Parental relationships among MLLs were estimated using COLONY 2.0 (Jones & Wang 2010) and FRANz 2.0.0 (Riester *et al.* 2009, 2010). While the COLONY approach employs a maximum likelihood method to assess relatedness, not considering in any way the actual genet size, FRANz considers ramet number as prior information. Three parallel medium-length runs were carried out on COLONY using the full-likelihood method with medium precision and without allele frequency updating, and considering allelic dropout and marker genotyping error rates of 0.01 (and no inbreeding as the species are dioecious). 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. All genets were considered as offspring and candidate mothers and fathers at the same time, as the gender of most genets was unknown. Since the age of genets was not known either, directionality of parentage relationships (i.e. which is the parent and which is the offspring) could not be directly inferred. Nevertheless, the focus of these analyses, i.e. the identification of parents with several offspring, should not suffer from this issue. For FRANz runs, we considered a maximum number of fathers in the population of 1,000,000 (to simulate infinite) and genotyping error of 0.01, and ran 20,000

iterations, including ramet number as a prior. Both pedigree reconstruction methods were computed based on a subset of 56 loci, discarding those that were physically linked or had low polymorphism (Table S1 in Supporting Information). Finally, we tested for correlations between the number of sexual offspring and genet size (measured as number of ramets and as spatial extension, i.e. geographical maximum distance between ramets) using Spearman rank tests.

Clonal population structure and demographic history

To evaluate clonal diversity, we calculated the following population 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) (Arnaud-Haond *et al.* 2007). Population genetic descriptors were computed at MLL level using the full set of 73 SSRs. SPAGeDi 1.3 software (Hardy & 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 (F_{ST}) between *P. alba* and *P.* × *canescens*. HP-Rare 1.0 (Kalinowski 2005) was used to calculate total and rarified allelic richness, and rarified number of private alleles.

Demographic history was inferred using MSVAR vs 1.3 software on MLLs and assuming neutrality of molecular markers (Beaumont 1999; Storz & Beaumont 2002). MSVAR produces more reliable inferences when the population size change is large, fluctuations are ancient and the population size declines (Girod *et al.* 2011), which was our case (see Results). Both linear and exponential models of population size change were considered. Changes during short periods are prone to be proportional to population size, thus exponential curves are expected to fit better, while changes through longer periods are more related to environmental or evolutionary shifts, which often behave linearly

(Beaumont 1999). Only SSRs with perfect repeats under a Stepwise Mutation Model (44 loci; listed in Table S1 in Supporting Information) and one ramet per MLL were used in MSVAR runs. Five independent MCMC simulations were performed for each scenario, with $4.5E^9$ iterations each, of which $2E^9$ 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 were set to $1E^4$ and starting mutation rate was set to $1E^{-4}$ for all loci; starting time since decline or expansion was set to $3E^6$. Prior distributions were all rectangular (i.e. uniform distribution) with the following moments (provided in decimal logarithmic scale, as in the input of the software): both population size distributions had a mean of 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 to 40 years, following Macaya-Sanz *et al*. (2012). We allowed the program to update the values of the starting parameters. Finally, the CODA package (Plummer *et al.* 2006) in the R environment (R Development Core Team 2009) was used to summarize MCMC outputs and evaluate chain convergence using the Gelman-Rubin statistic (Gelman & Rubin 1992).

Results

MLGs, MLLs and age of large clones

Nuclear microsatellites resolved 132 multilocus genotypes (MLGs) among 533 samples in the Douro poplar hybrid zone, of which 95 were of *P. alba* (out of 360 samples), 20 were hybrids (admixture proportion *Q* between 0.1 and 0.9; out of 145 samples), and 17 were of *P. tremula* (out of 28 samples). When putative somatic mutations were taken into account,

only 82 multilocus lineages (MLLs) were recognized in *P. alba* and 13 in the hybrid individuals, while each MLG was assigned to a different MLL in *P. tremula*. The low number of MLLs was due to the presence of a few large clones, extending over dozens to more than a hundred kilometers (Table 1 and Figure 1, and Figure S2 in Supporting Information).

Considering the accumulation of somatic mutations since the common ancestor and mutation rates, the age of MLL009 (the largest clone) was estimated to 7,117 (CIs: 3,553-12,734) years using SSRs. The age estimation of MLL006 (the second largest clone) was estimated to 4,414 (CIs: 1,620-9,608) years. Based on GBS data, age estimates using the same method were 4,828 years (CIs: 2,691-8,416 years) for MLL009 and 7,438 years (CIs: 3,375-14,500 years) for MLL006, based on 1.249 Gbp and 0.507 Gbp of aligned sequence, respectively. Although age estimates differ between markers (but notice that CIs overlap for both clones), we can conclude that at least some of the living clones today are several millennia old.

Genomic background of large clones and locus-specific ancestry

Genomic background of *P.* × *canescens* corresponded mainly to first generation hybrids (F1) and backcrosses to *P. alba*, although the exact hybrid generation was difficult to determine. Many *P. alba* individuals displayed weak introgression from *P. tremula*, represented by a soft cline from putatively first-generation backcrosses to pure *P. alba* (Figure 2). In contrast, only few genets of *P. tremula* (3 out of 17) showed introgression from *P. alba*. Unidirectional introgression of this type has been repeatedly reported in poplar hybrid zones.

Four loci showed extreme LSA scores in the ten largest clones, but not in the rest of the MLLs (two loci for the ten clones with the largest spatial extension and three loci for clones with the largest number of ramets; one locus common to both sets). Three of these

loci remained robust outliers after jackknife resampling, but only two loci, GCPM1274 (chromosome I) and ORPM374 (chromosome VIII), additionally had extreme LSAG scores. However, observing this number of loci with extreme LSA scores in large clones only can be expected by chance, as shown by permutations of clone size while keeping genotypes (Figure S3 in Supporting Information). Thus, although GCPM1274 and ORPM374 can be regarded as robust outliers for increased *P. tremula* ancestry in large clones, the presence of outliers should also be expected by chance, without selection acting on these loci.

Locus GCPM1274 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 (based on *P. trichocarpa* genome assembly version 3). Locus ORPM374 is also located close (<10 kb) to two genes: one annotated as an ATP-citrate synthase, involved in Krebs' cycle, and the other with a domain similar to an iron-sulfur binding protein C terminal. This motif is found in genes involved in the circadian regulation in cyanobacteria. Interestingly, genes encoding a phytochrome and an Iojap protein, related to protein synthesis down-regulation during starvation, are also located nearby (<15 kb) to locus ORPM374.

Mating success of large clones

Parental assignments by COLONY and FRANz were highly correlated (Pearson's correlation coefficient of 0.95). Large clones had a higher number of sexual offspring than smaller clones in the population, with the exception of MLL006 (a putative F1-hybrid; see Figure 2), whose number of offspring was comparatively small in relation with its size (Table 1). Considering COLONY parental assignments (Table S3), the four largest clones (3.6% of the MLLs), with more than ten ramets each, were the parents of 60 offspring (57.1%).

Spearman's rank tests showed significant correlations between clone size (estimated as number of ramets) and offspring number, both considering all genets (ρ = 0.377; P < 0.001) and the ten largest ones (ρ = 0.768; P < 0.01; Figure S4 in Supporting Information). However, correlations were not significant when clone size was estimated as the maximum distance between ramets (all genets: ρ = 0.088; P = 0.58; ten largest genets: ρ = 0.540; P = 0.11; Figure S4). For many individuals (40 clones, i.e. 35.7%), no parent was found among the sampled trees, probably because the parents were located outside of the sampling area, were overlooked by our sampling scheme, or were already dead.

Clonal population structure and demographic history

Due to the occurrence of very large clones, genotypic richness was low and clonal structure uneven in the Douro hybrid zone (R = 0.228 and V = 0.532 for *P. alba*), despite appreciable levels of genetic diversity and allelic richness (H_E = 0.405 and A' = 2.61 for *P. alba*; Table 2). Population inbreeding was also low but significantly different from zero (F_{IS} = 0.051 and 0.066 in *P. alba* and hybrids, respectively; Table 2). Demographic analyses using MSVAR (one ramet per MLL) showed a long-term decline in population effective size (Table 3, Figure S5 in Supporting Information). The exponential decline model fitted better with observed data than the linear model (AIC of 23,776 vs 24,992). Nonetheless, the comparatively small difference between AIC values indicates that neither of the models is significantly superior. Both models indicated a soft but persistent population size decline during the last hundreds of thousands of years, with current effective population size reduced to about one tenth of the ancestral size. Nevertheless, current effective size was still considerable (~2,240 for the exponential model; Table 3). These calculations considered a generation time of 40 years. Given the 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 cannot be excluded.

Discussion

Origin of large and ancient clonal assemblies in the Douro basin

A few large, ancient and widespread clones, together with many small ones with more restricted geographical distribution, characterized the Douro hybrid zone. The two largest genets were a male pure *P. alba* (MLL009) and a female hybrid (MLL006) (189 and 124 ramets, respectively), and their origin may date back to several thousand years (at least 3-4 millennia at 95% probability). Large and widespread clones have been repeatedly reported in *Populus* species (e.g. Ally *et al.* 2008; Brundu *et al.* 2008), including for the Douro area (Santos-del-Blanco *et al.* 2013), but their age has only rarely been estimated, except for the *P. tremuloides* 'Pando' clone reported to be almost ten thousand years old (Mock *et al.* 2008). Outside *Populus* spp., age estimates of particularly old clones are more frequent: May *et al.* (2009) claimed an age in excess of 13,000 years for a *Quercus palmeri* clone living in Southern California and clones of the bushes *Lomatia tasmanica* and *Larrea tridentate* were estimated to be 43,600 and 11,700 years old, respectively (Lynch *et al.* 1998; Vasek 1980).

The Douro hybrid zone is located in a region (the Castilian Plateau) where human activity dates back to ancient times. Nevertheless, the spread of *Populus* clones in the region by humans is an unlikely explanation. Although riverbank reforestation with poplars has surely been conducted in historic times, the estimated age of some of the large clones (over 2,000 years) exceeds the times when social civilization could have fostered this regional

expansion, during the Romanization of the Iberian Peninsula. Moreover, there is not any associated economical use that could have contributed to ancient human-mediated spread of *Populus* clones in the region. Exclusive root-sucker expansion is not a sufficient alternative explanation to human-mediated spread, given the wide extension of large clones (close to 160 km for the largest, MLL006). However, the translocation of twigs by water currents is known to be a successful long-distance propagation mechanism in species of genera *Salix* and *Populus* (Barsoum *et al.* 2004; and references therein). The Douro River and its tributaries are large watercourses in the study area, and the constant flow of water can transport propagules across long distances. Large birds, such as the locally abundant storks, could also have effectively contributed to translocate twigs (used for nesting), although, to our knowledge, no study has assessed the magnitude of this dispersal mode yet.

Several processes may have created opportunities for expansion of current large clones in detriment of sexual reproduction. First, environmental fluctuations, which are common in riparian habitats, could have hampered sexual reproduction more than asexual propagation, which depends less on ecological factors. In addition, clonal reproduction is often more abundant in range margins and more stressful environments (Honnay & Bossuyt 2005; Kawecki 2008; Silvertown 2008), as is the case in the Douro basin (as compared to central European poplar hybrid zones; see below). Second, ancient geological events, such as the transition from the Pliocene to the Pleistocene glaciations, may have promoted clonal spread in the Douro hybrid zone (Macaya-Sanz *et al.* 2012). *Populus alba* evolved during the late Tertiary and early Quaternary, under more humid subtropical conditions (Eckenwalder 1996). Climate desiccation and cold temperatures could therefore 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). Moreover, during colder periods, geographical isolation may have imposed limitations on gene flow among isolated genets. Third, in more recent times, human activities have certainly affected the ecology of *P. alba* in this region, which is highly susceptible to natural disturbances. Indeed, river flood regulation is directly involved in failure of sexual recruitment in *P. alba* (Gonzalez *et al.* 2010b), as well as in other *Populus* species (e.g. *P. nigra*; Barsoum 2001).

The role of genomic background in clonal success

Hybrid clones were neither overrepresented among large clones (Table 1; two hybrids out of ten large clones in a population of 13 hybrids out of 112 MLLs; Fisher's exact test: $P = 0.236$) nor significantly larger than pure *P. alba* (average size estimated as maximum distance between ramets: hybrids, 2016 ± 1333 (SE) m; *P. alba*, 939 ± 334 (SE) m; Kruskal-Wallis test: $P = 0.775$; two largest clones, MLL006 and MLL009, excluded), suggesting a lack of impact of overall genomic background on clonal spread. However, more frequent asexual reproduction in hybrids has been reported in other case studies (e.g. *Narcissus*; Marques *et al.* 2011), including some in *Populus* (Schweitzer *et al.* 2002; Van Loo *et al.* 2008), and, given the limited number of hybrids in our study, this result should be considered with caution.

Large clones, however, displayed two genomic regions with unusual genetic ancestry (when evaluating collectively the ten largest clones, see *Results*). Locus GCPM1274 showed strong introgression by one *P. tremula* allele, even in otherwise pure *P. alba* backgrounds (e.g. MLL002). GCPM1274 also displayed highly distorted genotype proportions in a controlled backcross of a F1 hybrid to *P. alba*, with segregation of the *P. tremula* alleles dominating (Macaya-Sanz *et al.* 2011), and lies close (< 10kb) to candidate genes for adaptive traits. The poplar telomeric region of chromosome I adjacent to GCPM1274 is rich

in NBS class resistance genes (Kohler *et al.* 2008) and in expressed small RNAs (Klevebring *et al.* 2009). Moreover, this locus is located in close proximity to a glutaredoxin (GRX) gene. Plant glutaredoxins have pivotal roles in plant redox biology (Rouhier 2010) and are associated with increased tolerance to drought (e.g. Guo *et al.* 2010). Locus ORPM374 is located in a gene-rich arm of chromosome VIII, in close proximity to four annotated genes, two of them involved in circadian cycling, and the other two in cell basal energy regulation.

Our findings are limited to a low number of highly-successful clones and genomic regions (not higher than those expected by chance, as shown by permutation tests), and thus are difficult to generalize. Nonetheless, they suggest that localized regions of specific ancestry, rather than overall genomic background, may play a role in clonal spread under stressful environmental conditions. Similar results have been reported in sunflower where regions of unusual heterozygosity in hybrids underlie increased salt-tolerance (Lexer *et al*. 2004). Also, recent results on Central European hybrid zones of *P. alba* and *P. tremula* point to important effects of heterospecific ancestry on hybrid survival and persistence (Christe *et al*. 2016).

Population effects of large, geographically-extended clones

Large, geographically-extended clones dominated sexual reproduction in the Douro hybrid zone, as shown by a significant correlation between clone size and number of descendants (Figure S4 in Supporting Information). Sexual dominance results from both greater opportunities for mating each season (due to the larger 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 depauperation of pollen quality and quantity. Furthermore, in cases of very large clones, as found here, it is expected that the time needed for development of sterility will be somewhat longer, as mutations need to accumulate throughout most ramets. (ii) At the genet level, uneven sexual reproduction reduces the effective population size as it increases the variance of reproductive success (Balloux *et al.* 2003), thus potentially raising genetic drift and affecting population demography (see below). Interestingly, in the Douro hybrid zone, although significantly higher than zero, population-level inbreeding (F_{IS} = 0.051) was lower than in other (less-clonal) European poplar hybrid zones (F_{IS} = 0.143-0.173) and we did not observe significant reductions of genetic diversity (see Table 2). Thus, the sexual dominance of the largest clones in this region does not seem to be pervasive enough to increase inbreeding or reduce genetic diversity at the population level, even after several thousand years of clonal reproduction and despite the existence of large, geographically-extended clones. This is in agreement with mathematical models and simulations showing departures from Hardy-Weinberg equilibrium only for populations with extremely rare sexual reproduction or in transient states (Balloux *et al*. 2003; Halkett *et al*. 2005; Reichel *et al*. 2016). Nevertheless, Bayesian simulations with MSVAR showed a long-term population size decline in the Douro hybrid zone. High levels of genetic diversity and low inbreeding at the population level (Table 2) suggest that long-term effective population size decline is due to outcompeting of other genets by larger clones rather than to increased inbreeding due to their sexual dominance. According to Eriksson's Initial Seedling Recruitment model (Eriksson

1997), genotype richness in clonal populations is expected to decrease as the stand ages and some clones die, due to the vegetative expansion of the remaining clones. This pattern has frequently been observed in herbaceous plants (Barsoum *et al.* 2004; and references therein) and explains the usual left-skewed shape of the ramet number per genet distribution (i.e. many small and a few large genets; Arnaud-Haond *et al.* 2007).

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, until a new balance is established. In general, rates of clonality can increase or decrease, but some authors point also to extreme situations where equilibrium is not recovered. Honnay & Bossuyt (2005) argued that environmental conditions precluding effective sexual reproduction can move a population to a tipping point for irreversible extinction of sexual function. Our results suggest that effective population size is contracting in the Douro poplar hybrid zone, with a few large clones dominating sexual reproduction. However, we did not find significant levels of inbreeding or substantial losses of genetic diversity, suggesting that population persistence is not threatened by demographic decline (see above).

New ecological conditions fostered by climate change could either increase or prevent asexual reproduction, shifting the Douro population to a new equilibrium. On-going processes such as increasing river regulation, growing aridity and temperature, or the introduction of exotic invasive plants and pathogens will likely impede sexual reproduction more severely than asexual reproduction and could push this population to the verge of sex extinction (as observed in Sardinian *P. alba*; Brundu *et al.* 2008), thus enhancing demographic decline and reducing population 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). Alternatively, new ecological conditions could also challenge the large clones that have thrived in former ecological conditions, opening new niches to smaller or new genets, and re-balancing the genotype numbers. Within this perspective, climatic changes and other ecological disturbances may indeed help to prevent the long-term dominance of few genets in clonal long-living organisms, in a similar fashion to (although on a different scale than) short-term ecological oscillations maintaining standing genetic variation in natural populations.

Acknowledgements

We would like to acknowledge the help and support during lab and field work provided by Mike Fay, Jeffrey Joseph, Carmen García Barriga, Fátima Torrico, Jose Antonio Mancha Núñez, Javier Arenillas Cristóbal, Jesús Pastor Aceves, Jorge García Moya, Manuel Núñez Sanz, Ángel Vicente Blázquez, Luis Santos-del-Blanco, Juan Andrés Oria de Rueda and Nuria Alba. DMS was supported by a FPI fellowship (BES-2006-13132) of the Spanish Ministry of Science and Education. MH was supported by a 'Ramón y Cajal' fellowship (RYC2009-04537) of the Spanish Ministry of Economy and Competitiveness (MINECO). SCGM and MH acknowledge the support of Marie Curie Intra European Fellowships within the 7th European Community Framework Programme (contract numbers 328146 and 329088, respectively). This work has been funded by the Spanish National Plan for Science, project ReproFor (AGL2005-07440-C02-01), and by grants from the U.K. Natural Environment Research Council (NERC) and Swiss National Science Foundation (SNF) to CL.

References

Alberti AP, Díaz MV, Chao RB (2004) Pleistocene glaciation in Spain. In: *Developments in Quaternary Sciences* (eds. Ehlers J, Gibbard PL), pp. 389-394. Elsevier, Amsterdam, The Netherlands.

Ally D, Ritland K, Otto SP (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, Ritland K, Otto SP (2010) Aging in a long-lived clonal tree. *Plos Biology* **8**.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81-92.

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology* **16**, 5115-5139.

Arnaud-Haond S, Duarte CM, Diaz-Almela E *et al.* (2012) Implications of extreme life span in clonal organisms: millenary clones in meadows of the threatened seagrass *Posidonia oceanica*. *PLoS One* **7**, e30454. 10.1371/journal.pone.0030454.

Balloux F, Lehmann L, de Meeus T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* **164**, 1635-1644.

Barrett SCH (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, Muller E, Skot L (2004) Variations in levels of clonality among *Populus nigra* L. stands of different ages. *Evolutionary Ecology* **18**, 601-624.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013-2029.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

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.

Castiglione S, Cicatelli A, Lupi R *et al.* (2010) Genetic structure and introgression in riparian populations of *Populus alba* L. *Plant Biosystems* **144**, 656-668.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783-796.

Catchen J, Amores A, Hohenlohe P *et al.* (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171-182.

Catchen J, Hohenlohe P, Bassham S *et al.* (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Christe C, Stölting KN, Bresadola L *et al.* (2016) Selection against recombinant hybrids maintains reproductive isolation in hybridizing Populus species despite F1 fertility and recurrent gene flow. *Molecular Ecology* **25**, 2482-2498.

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.

Eckert CG (2001) The loss of sex in clonal plants. *Evolutionary Ecology* **15**, 501-520.
Eriksson O (1997) Clonal life histories and the evolution of seed recruitment. In: *The Ecology and Evolution of Clonal Plants* (eds. De Kroon H, Van Groenendael J), pp. 211-226. Backhuys Publishers, Leiden, The Netherlands.
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.
Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457-472.
Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the msvar method. *Genetics* **188**, 165-287.
Gonzalez E, Antonio Comin F, Muller E (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, Gonzalez-Sanchis M, Cabezas A *et al.* (2010b) Recent changes in the riparian forest of a large regulated Mediterranean river: implications for management. *Environmental Management* **45**, 669-681.
Guo YS, Huang CJ, Xie Y *et al.* (2010) A tomato glutaredoxin gene SlGRX1 regulates plant responses to oxidative, drought and salt stresses. *Planta* **232**, 1499-1509.
Halkett F, Simon J-C, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology and Evolution* **20**, 194-201.
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.
Hedrick PW (2012) What is the evidence for heterozygote advantage selection? *Trends in Ecology & Evolution* **27**, 698-704.
Honnay O, Bossuyt B (2005) Prolonged clonal growth: escape route or route to extinction? *Oikos* **108**, 427-432.
Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* **180**, 329-340.
Jones OR, Wang JL (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**, 551-555.
Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187-189.
Kawecki TJ (2008) Adaptation to marginal habitats. *Annual Review of Ecology Evolution and Systematics* **39**, 321-342.
Klekowski EJ (1997) Somatic mutation theory of clonality. In: *The Ecology and Evolution of Clonal Plants* (eds. de Kroon H, van Groenendael J), pp. 227-241. Backhuys, Leiden, The Netherlands.
Klevebring D, Street NR, Fahlgren N *et al.* (2009) Genome-wide profiling of *Populus* small RNAs. *Bmc Genomics* **10**.
Klimes L, Klimesova J, Hendriks R, van Groenendael J (1997) Clonal plant architecture: a comparative analysis of form and function. In: *The Ecology and Evolution of Clonal Plants* (eds. de Kroon H, van Groenendael J), pp. 1-29. Backhuys, Leiden, The Netherlands.

Kohler A, Rinaldi C, Duplessis S *et al.* (2008) Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Molecular Biology* **66**, 619-636.
Lexer C, Fay MF, Joseph JA *et al.* (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, Joseph JA, van Loo M *et al.* (2010) Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**, 699-712.
Lexer C, Lai Z, Rieseberg LH (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, Buerkle CA, Barbara T *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.
Lynch M (1991) The genetic interpretation of inbreeding and outbreeding depression. *Evolution* **45**, 622-629.
Lynch AJJ, Barnes RW, Cambecedes J, Vaillancourt RE (1998) Genetic evidence that *Lomatia tasmanica* (Proteaceae) is an ancient clone. *Australian Journal of Botany* **46**, 25-33.
Macaya-Sanz D, Heuertz M, Lopez-de-Heredia U *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, Suter L, Joseph J *et al.* (2011) Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species. *Heredity* **107**, 478-486.
Marques I, Feliner GN, Martins-Loucao MA, Aguilar JF (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, Storme V, Dillen SY *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.
May MR, Provance MC, Sanders AC *et al.* (2009) A Pleistocene clone of Palmer's oak persisting in Southern California. *PLoS One* **4**, e8346. doi:10.1371/journal.pone.0008346.
Mock KE, Rowe CA, Hooten MB *et al.* (2008) Clonal dynamics in western North American aspen (*Populus tremuloides*). *Molecular Ecology* **17**, 4827-4844.
Otto SP, Lenormand T (2002) Resolving the paradox of sex and recombination. *Nature Reviews Genetics* **3**, 252-261.
Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**, 2991-3005.
Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7-11.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
Reichel K, Masson J-P, Malrieu F, Arnaud-Haond S, Stoeckel S (2016) Rare sex or out of reach equilibrium? The dynamics of F_{IS} in partially clonal organisms. *BMC Genetics* **17**, 76.
Rice WR (2002) Experimental tests of the adaptive significance of sexual recombination. *Nature Reviews Genetics* **3**, 241-251.

Rieseberg LH, Carney SE (1998) Plant hybridization. *New phytologist* **140**, 599-624.
Riester M, Stadler PF, Klemm K (2009) FRANz: reconstruction of wild multi-generation pedigrees. *Bioinformatics* **25**, 2134-2139.
Riester M, Stadler PF, Klemm K (2010) Reconstruction of pedigrees in clonal plant populations. *Theoretical Population Biology* **78**, 109-117.
Rouhier N (2010) Plant glutaredoxins: pivotal players in redox biology and iron-sulphur centre assembly. *New Phytologist* **186**, 365-372.
Sachs T (2001) Developmental processes and the evolution of plant clonality. *Evolutionary Ecology* **15**, 485-500.
Santos-del-Blanco L, de-Lucas A, González-Martínez S *et al.* (2013) Extensive clonal assemblies in *Populus alba* and *Populus* x *canescens* from the Iberian Peninsula. *Tree Genetics & Genomes* **9**, 499-510.
Schweitzer JA, Martinsen GD, Whitham TG (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 GT, Leonardi S, Adams WT, Strauss SH, DiFazio SP (2010) Population substructure in continuous and fragmented stands of *Populus trichocarpa*. *Heredity* **105**, 348-357.
Slavov GT, Zhelev P (2010) Salient biological features, systematics, and genetic variation of *Populus*. In: *Genetics and Genomics of* Populus (eds. Jansson S, Bhalerao RP, Groover AT), pp. 15-38. Springer Science+Business Media, LLC, New York.
Stenvall N, Haapala T, Aarlahti S, Pulkkinen P (2005) The effect of soil temperature and light on sprouting and rooting of root cuttings of hybrid aspen clones. *Canadian Journal of Forest Research* **35**, 2671-2678.
Stölting KN, Nipper R, Lindtke D *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.
Storz JF, Beaumont MA (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, Pritchard JK, Shen PD *et al.* (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.
Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**, 377-379.
Vallejo-Marin M, Dorken ME, Barrett SCH (2010) The ecological and evolutionary consequences of clonality for plant mating. *Annual Review of Ecology, Evolution, and Systematics* **41**, 193-213.
van Loo M, Joseph JA, Heinze B *et al.* (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.
Vasek FC (1980) Creosote bush - Long-lived clones in the Mojave Desert. *American Journal of Botany* **67**, 246-255.
Vonlanthen B, Zhang XM, Bruelheide H (2010) Clonal structure and genetic diversity of three desert phreatophytes. *American Journal of Botany* **97**, 234-242.

Yu QB, Mantyla N, Salonen M (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.

Data Accessibility:

Microsatellite and GBS data, as well as the scripts used for analyzing GBS data, were deposited in the Dryad repository with doi: 10.5061/dryad.mp745.

Author Contributions:

DMS performed the research and wrote the first draft of the manuscript. GGV produced the GBS data, DL and MH contributed to the data analyses. SCGM and CL designed and supervised the research, and wrote parts of the manuscript. All authors revised the manuscript and suggested improvements.

Figure legends:

Figure 1 Location of individuals sampled within the Douro poplar hybrid zone including the main river course, its major tributaries and nearby mountains. Circles represent *Populus alba*, triangles *P. × canescens*, and stars *P. tremula*. Red symbols represent large MLLs (>10 ramets), while black symbols represent small ones (≤10 ramets). Two enlarged windows of the core sampling area (A and B) are also provided to facilitate individual identification.

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: the ten largest clones, *P. tremula*, *P. × canescens*, *P. alba.* (a) Genotypes with high scores for interspecific heterozygosity are in blue, while low scores are in brown. In (b), high and low

scores for specific homozygosity, i.e. *P. tremula* and *P. alba* ancestry, are indicated in blue and brown, respectively. Loci are ordered following their genomic position, within chromosomes. Two genomic regions with increased introgression of *P. tremula* alleles in the ten clones (GCPM1274 in chromosome I and ORPM374 in chromosome VIII) are indicated with a green box. Note that chromosome XVII is not represented. STRUCTURE *Q* values (ancestry proportions) are also shown (in the side graph).

Table 1 Representative clones (MLLs) found in the Douro poplar hybrid zone, including large clones. *N*: Number of ramets; *Extension*: longest distance between ramets within clones; *Q*: ancestry proportion based on 72 SSRs (*Q* = 1 for pure *P. alba*); *Offspring*: number of descendants detected by either COLONY or FRANz software (see main text for details); N_{off}: average number of ramets for MLL's offspring (assigned by COLONY). Pa: *P. alba*; Pc: *P. × canescens*. NA: not applicable. The last two rows report averaged values and standard deviation (within parentheses) for the rest of *P. alba* and *P. × canescens* MLLs.

MLL	Species	*N*	Extension (km)	*Q*	Offspring		
					COLONY	N_{off}	FRANz
MLL009	Pa	189	99.5	0.99	29	1.93	21
MLL006	Pc	124	158.6	0.48	7	1.71	2
MLL025	Pa	26	74.6	0.99	23	1.52	23
MLL073	Pa	17	5.6	0.97	1	1.00	2
MLL002	Pa	8	22.6	0.91	0	NA	0
MLL074	Pa	7	10.7	0.92	1	1.00	1
MLL011	Pa	6	4.1	0.99	13	1.23	8
MLL049	Pa	5	17.5	0.99	0	NA	0
MLL057	Pc	5	0.6	0.84	0	NA	2
MLL083	Pa	5	2.6	1.00	0	NA	2

MLL126	Pa	3	4.9	0.93	5	4.60	2
MLL086	Pa	2	1.5	1.00	5	2.20	4
MLL053	Pa	2	5.2	0.98	0	NA	4
MLL111	Pa	2	17.1	0.94	0	NA	2
MLL058	Pa	1	NA	1.00	3	1.33	2
MLL030	Pa	1	NA	0.96	0	NA	1
MLL120	Pa	1	NA	0.91	8	1.25	4
Rest of *P. alba*		1.3 (0.5)	NA	0.98 (0.02)	0.08 (0.32)	NA	0.22 (0.42)
Rest of *P. × canescens*		1.5 (0.7)	NA	0.69 (0.10)	0.00 (0.00)	NA	0.55 (0.52)

Table 2 Clonal structure and genetic diversity in the Douro poplar hybrid zone compared to other European poplar hybrid zones studied to date, recomputed from van Loo *et al.* (2008) and Lexer *et al.* (2010) for Danube clonal structure and population genetics parameters, respectively, and Castiglione *et al.* (2010) for Ticino hybrid zone (all parameters). Population genetic statistics are based on MLLs only. NA: Not available. Bold font indicates tests significantly different from zero (P<0.01).

		Douro			Danube			Ticino		
		P. alba	*P. × canescens*	*P. tremula*	*P. alba*	*P. × canescens*	*P. tremula*	*P. alba*	*P. × canescens*	*P. tremula*
Clonal structure										
Number of ramets	*N*	360	145	28	222	185	NA	23	26	NA
Number of MLGs	*G'*	95	20	17	NA	NA	NA	NA	NA	NA
Number of MLLs	*G*	82	13	17	169	123	NA	11	22	NA
Genotypic richness	*R*	0.228	0.090	0.593	0.760	0.663	NA	0.455	0.840	NA
Simpson evenness	*V*	0.532	0.112	0.617	0.891	0.911	NA	NA	NA	NA
Log-scaled Pareto distribution										
Additive inverse of the	β	0.081	0.012	0.899	2.203	1.581	NA	NA	NA	NA

slope										
Coefficient of determination	r^2	0.722	0.374	0.988	0.911	0.935	NA	NA	NA	NA
Genetic parameters										
Allelic richness	A	4.92	3.78	4.38	7.18	12.91	7.18	4.72	8.89	5.17
Rarefacted allelic richness[a]	A'	2.61	3.11	3.29	3.46	3.99	3.43	2.80	3.67	3.01
Number of private alleles[b]	A_p'	0.65	0.52	1.53	1.03	1.07	1.81	0.91	0.64	0.95
Genetic diversity	H_E	0.405	0.551	0.517	0.508	0.596	0.533	0.428	0.590	0.479
Inbreeding coefficient[c]	F_{IS}	**0.051**	**0.066**	**0.215**	**0.143**	**0.180**	**0.205**	**0.173**	**0.124**	**0.269**

[a] Rarefied to 10 chromosomes.
[b] Considering also *P. tremula* alleles.
[c] Computed using within-species allele frequencies as reference.

Table 3 Posterior density of main demographic parameters, calculated by MSVAR using 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, the mutation-scaled population size, 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	$2.254E^{-6}$	$6.998E^{-5}$	$2.203E^{-3}$	$1.991E^{-6}$	$6.109E^{-5}$	$1.888E^{-3}$
t_1	$2.911E^{4}$	$1.019E^{6}$	$3.648E^{7}$	$1.148E^{4}$	$3.908E^{5}$	$1.352E^{7}$

Supporting Information:

Table S1 Nuclear microsatellites, including locus code, chromosome and genetic position, and the marker subsets used in the different data analyses.

Table S2 Basic statistics per sample for GBS data: number of reads, coverage and number of stacks.

Table S3 General information on MLLs sampled, including the code of an exemplary ramet, the species (Pa for *P. alba*, Pc for *P. × canescens*, and Pt for *P. tremula*), the number of ramets, the spatial extension of the MLL, the number of offspring assigned by COLONY, the identity of both parents assigned by COLONY, and the number of offspring assigned by FRANz.

Figure S1 Distribution of pairwise genetic similarity between MLGs, measured as percent of identical alleles. Note that the distribution has more than two modes due to the population including two different species. The mode with less percentage of similarity corresponds to the pairwise comparisons between species, the mode in the middle to the pairwise comparisons within species, and the small mode with higher genetic similarity represents the comparison between MLGs within the same MLL.

Figure S2 Geographic distribution of ramets for clones MLL009 (a), MLL006 (b), MLL025 (c) and MLL073 (d). Color key indicates the genotype (i.e. MLG) of each ramet.

Figure S3 Distribution of expected number of loci with extreme LSA scores in a subset of 10 clones, but without extreme values in the remaining clones, as identified by 1,000 permutations (without replacement). Gray bars indicate lower 5% and upper 95% quantiles of the distribution. The dotted and dashed lines show the observed number of loci with extreme LSA scores in the 10 largest clones in terms of spatial extension and number of ramets, respectively.

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.