A comprehensive strategy for the conservation of forest tree genetic diversity: an example with the protected Pinus nigra subsp. salzmannii (Dunal) Franco in France

C. Scotti-Saintagne, A. de Sousa Rodrigues, Anne Roig, B. Fady

TITLE

A comprehensive strategy for the conservation of forest tree genetic diversity: an example with the protected *Pinus nigra* subsp. *salzmannii* (Dunal) Franco in France.

Published in Conservation Genetics: https://doi.org/10.1007/s10592-023-01581-8

AUTHOR INFORMATION

C. Scotti-Saintagne, A. de Sousa Rodrigues, A. Roig, B. Fady

INRAE, Ecologie des Forêts Méditerranéennes (URFM), Avignon, France

Corresponding author: B. Fady (bruno.fady@inrae.fr)

ORCID numbers:

C. Scotti-Saintagne: 0000-0001-6891-7315

B. Fady: 0000-0003-2379-7617

ABSTRACT

Genetic diversity is essential to evolution and a recognized target of conservation. When threats are high and populations are small, in-situ gene conservation needs to be reinforced with an ex-situ approach, where a genetically representative sample of the target taxon is safeguarded in a favorable environment. The fragmented habitat of *Pinus nigra* subsp. *salzmannii* (Dunal) Franco 1943 (Salzmann's pine) is threatened by wildfires. In France, gene flow from massive plantations of other subspecies of black pines threatens its genetic diversity. Here, using individual tree genotypic data at thirteen microsatellite loci, we identified differentiated lineages for in-situ gene conservation. Discriminating between autochthonous and hybrid trees, we proposed a method for the creation of an ex-situ core collection. We confirmed that Salzmann's pine is an original genetic lineage within the western European and Mediterranean black pine subspecies. We identified five genetic groups in France that can serve as the basis for in-situ gene conservation. Maximizing overall genetic diversity while maintaining among population diversity, we identified 80 native and non-hybridized trees that can form the basis of a representative ex-situ core collection. Our cost-effective methods combining in-situ and ex-situ conservation can be easily applied to many forest tree species.

Keywords (4-6)

Genetic structure, core collection, protected area, microsatellites,

INTRODUCTION

Genetic diversity is essential to evolution as natural selection can act on genetic diversity to adapt populations to environmental change (Lande & Shannon 1996, Jump & Peñuelas 2005). Genetic diversity is thus a recognized objective and target of conservation and sustainable use in the text of the 1992 Convention on Biological Diversity (CBD) of the United Nations. With human-induced environmental change dramatically altering biodiversity and disrupting local adaptation and community assemblages in many habitats, conserving genetic diversity is increasingly seen as an insurance policy for continued population adaptation, resilience and long-term survival (Fady et al. 2020). The International Union for the Conservation of Nature IUCN recognizes that genetic diversity should be better considered in conservation planning and that genetic diversity conservation as a goal in itself warrants the status of protected area for forest trees (IUCN 2016, 2021). In practice, however, genetic diversity conservation is rarely a goal of conservation planning and indicators of genetic diversity and evolutionary potential are yet to be implemented for monitoring biodiversity trends in natural populations (Hoban et al. 2020, Forester et al. 2022). The most recent Global Biodiversity Outlook concluded that CBD Aichi Target 13, Safeguarding genetic diversity, is far from being reached (Secretariat of the Convention on Biological Diversity 2020).

The gold standard for maintaining evolutionary potential is in-situ conservation of genetic diversity where natural selection can shift phenotypic optima as environmental changes occur. This is widely applicable to forest tree species (Potter et al. 2017). In-situ gene conservation requires that evolutionary lineages are identified and that, in each lineage, representative populations are selected, in which management for genetic diversity can be applied (Palsbøll et al. 2007). Areas protected for genetic diversity, or gene conservation units (GCUs), need to be sufficiently large to maintain adaptive potential over generations, while avoiding the risk of drift and inbreeding or, conversely, of gene swamping and outbreeding depression (Koskela et al. 2013).

Forest tree species are particularly suitable for this in-situ strategy. Native populations are often genetically diverse, with large effective population size and experiencing significant long-distance gene flow (Alberto et al. 2013). In some economically important species, there may be long-lasting human impacts on genetic diversity due to exploitation (fragmentation) and seed transfer (lineage mixing) that need to be identified prior to the implementation of GCUs (Raffl et al. 2018). Areas protected for genetic diversity are also increasingly found relevant as a resource for assisted migration, where pre-adapted genotypes are collected and moved to environments where they do not yet occur within or outside their current geographic distribution range (Aitken & Whitlock 2013, Schueler et al. 2014).

When threats increase and populations are small, an ex-situ approach is advisable, where a representative sample of the threatened population is safeguarded in a favorable environment (McGowan et al. 2017, Potter et al. 2017), either as a fixed copy collection (e.g. seeds in a cold chamber) or as an evolving collection under a new environment (e.g. tree archive plantation where the seeds produced are replanted). Constructing such a core collection requires that genetic diversity should be maximized while constraining the number of individuals and taking into account population genetic structure and diversity (Gapare et al. 2008, Hoban & Schlarbaum 2014, Hoban 2019).

Pinus nigra subsp. *salzmannii* (Dunal) Franco 1943 (Salzmann's pine), is one of the subspecies of the widespread European black pine (*Pinus nigra* J.F. Arnold, 1785). Salzmann's pine forests are recognized as a priority habitat in Europe under Habitats Council

Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. In France, where it reaches its northern distribution edge, its habitat has certainly shrunk over time, potentially due to increased dryness and fires during the second part of the Holocene, along with agriculture, grazing and cutting for the mining industry over centuries (Vernet 2006). It is currently and increasingly threatened by wildfires as climate warms. Natural populations are often surrounded by large plantations of other sub-species of black pine (*Pinus nigra* subsp. *nigra* Arnold, the Austrian black pine, and *Pinus nigra* subsp. *laricio* Palib. ex Maire, 1928, the Corsican black pine) which have been used since the mid-19th century for agricultural land reclamation, erosion control and timber production (Combes 1989). Habitat size estimates are 5 000 ha nationally for *P.n. salzmannii* while black pine plantations cover an estimated 300 000 ha. Divergence among black pine lineages is recent and dated to the late Pleistocene - early Holocene (Scotti-Saintagne et al. 2019). As there are no documented zygotic or gametic barriers to fertilization between sub-species, hybridization with *P.n. laricio* and *P.n. nigra* is possible and recognized as a potential threat to autochthonous *P.n. salzmannii* populations, particularly in France (Isajev et al. 2005, Fady et al. 2016, Scotti-Saintagne et al. 2018).

Here, we describe the steps undertaken for the development and implementation of a comprehensive program for the conservation of genetic diversity of *P.n. salzmanni*, a forest tree species whose census size is low, whose habitat is threatened but which still has sizable forests where natural regeneration is possible and encouraged by management agencies. We consider that this strategy applies broadly to widespread forest tree species, across distribution ranges and particularly at range edges, whether they are fragmented or not.

For this, our specific objectives are the following:

1- We test whether *P.n. salzmanni* is an original lineage within the western European and Mediterranean *Pinus nigra* sub species.
2- Sampling autochthonous diversity in France, we test that fragmented forests correspond to differentiated lineages and we identify representative regions for in-situ gene conservation.
3- Using individual tree genotypic data from all French lineages, we propose a method for the creation of an ex-situ core collection, representative of genetic originality and lineage diversity, excluding non- autochthonous and hybrid individuals, and maximizing allelic diversity.

MATERIAL AND METHODS

Sampling of autochthonous material

P.n. salzmannii leaf samples were collected during the period 2008 - 2012 from seven forest areas in two regions of south-eastern France: (1) Conflent in the Pyrenees and (2) St-Guilhem, Ardeche, Col d'Uglas, Tour-de-Palau and Gorges du Tarn in the southern Massif Central - Cevennes region (Table 1, Figure 1, Table S1 in Online Resource ESM01.xls). These forests sample the entire distribution of the French Salzmann's pine stands identified by relevant public and private forest management agencies as likely to be autochthonous. Black pine autochthony is a difficult issue to solve in France as intraspecific exotic material has been extensively used in plantations since 1860 when forest restoration laws were passed and implemented for erosion control and pasture land reclamation (Combes 1989). Here, we considered that individuals born before 1860 were very likely to be autochthonous. Thus, based on phenotype, all potentially old-enough trees were cored, their tree rings counted, and only those that passed the age threshold were sampled (Figure 2). As a result, out of the 4500 trees cored, 580 were retained for genotyping as likely autochthonous individuals (Table S1 in

Online Resource ESM01.xls). Detailed information on the genotyped and phenotyped individuals are available in Fady et al. (2023).

As pre-1860 plantations cannot be entirely excluded (Flaugère 1925), we selected a subset of trees older than 250 years, established between 1542 and 1758, as our baseline genetic reference for the endemic *P.n. salzmannii* (Figure 2). We chose these trees from a single population, the one with the highest number of very old trees, St-Guilhem. These 49 samples were used for comparison with the most commonly planted black pines in France (see testing step 1 below): *P.n. laricio* (49 samples from Corsica in France and Calabria in Italy) and *P.n. nigra* (49 samples from Austria, Romania and Serbia) (Table S2 in Online Resource ESM01.xls).

Genetic data acquisition and analysis of genetic diversity within population

For genotyping, we used the 13 nuclear microsatellite markers developed by Giovannelli et al. (2017) and their DNA extraction and scoring protocols.

To test for the presence of null alleles in our 13-marker dataset, we used *INEST* 2.2 (Chybicki & Burczyk 2009) and its Bayesian approach (IIM) by running $5x10^5$ Markov chain Monte Carlo iterations, of which the first $5x10^4$ were discarded as a burn-in phase. We compared all available model combinations, from the full model (nfb, including null alleles (n), genotyping failures (f), and inbreeding (b)), to the one-parameter models (n, f, or b). The best-fitting model was indicated by the lowest value of the deviation information criterion (DIC). Null allele frequencies were estimated for each locus as the mode of the posterior distribution. We also estimated null allele frequencies using the method of Dempster et al. (1977) and generated a null allele corrected dataset in *FreeNA* (Chapuis and Estoup 2007) using 1,000 replicates.

Within population diversity indices were calculated using *Spagedi* 1.5 (Hardy & Vekemans 2002): number of alleles (Na), number of rarefied alleles (Nae), gene diversity corrected for sample size (Nei 1978) (He), observed heterozygosity (Ho), inbreeding coefficient (Fi). The two-sided p-values of Fi were obtained after 10000 randomization of gene copies among individuals.

Differentiation and identification of evolutionary lineages

Blind approach to explore population genetic structure

To test for the presence of a population genetic structure without *a priori* on the geographic origin of the individuals, we used the popular software *STRUCTURE* v2.3 (Pritchard et al. 2000). It uses a Bayesian clustering analysis to group individuals into K groups according to their genotype and maximizing Hardy-Weinberg equilibrium. As among subspecies divergence is recent in black pine (Scotti-Saintagne et al. 2019), we used an admixture model, for which individuals can have a mixed ancestry and thus belong to several groups with different membership rates. We carried out fifteen independent iterations for each of the values of K, exploring values of K from 1 to 10, with a burn-in of 50,000 steps followed by 500,000 Markov chain Monte Carlo replications for each iteration. We relied on the mean likelihood L(K) and rate of variation (ΔK) values between successive values of K as described by Evanno et al. (2005) and provided by *StructureHarvester* (Earl and von Holdt 2012) to estimate the most probable number of K for each analysis. The results of the fifteen iterations were then compiled and visualized using the *CLUMPAK* web application (Kopelman et al. 2015) which uses the

CLUMPP software (Jakobsson & Rosenberg 2007) for compilation and the *DISTRUCT* software (Rosenberg 2004) for graphical visualization.

Testing if Salzmann's pine is a true taxonomic entity – step 1

To test whether endemic Salzmann pines represent an original taxonomic entity, distinct from other neighboring black pines, we ran *STRUCTURE* v2.3 (Pritchard et al. 2000) using three populations of an equal size of 49 individuals each and *P.n. salzmannii* individuals older than 250 years (see "sampling of autochthonous material" above and Figure 2, and Table S2 in Online Resource ESM01.xls). Assuming no gene flow from exotic plantations, we considered the level of "non-Salzmann" genetic admixture observed in these old Salzmann's pines as a background value of historical origin.

Testing the presence of sub-structure in Salzmann's pine and identifying potential hybrids – step 2

We re-analyzed the three populations of reference used in step 1 with all other *P.n. salzmannii* samples. In total 678 trees were analyzed using *STRUCTURE* v2.3 (Pritchard et al. 2000). We considered the maximum value of the membership coefficient Q of "non-Salzmann" observed in individuals of the Salzmann pine reference population as a threshold value above which admixture is probably of anthropogenic origin.

Genetic diversity among populations

The distribution of genetic diversity among the sampled populations was explored using a principal correspondence analysis (PCA) performed on the allelic frequencies at the individual level using the *dudi.pca* function of the *ade4* package in R (Dray et al. 2007). Variables were standardized around a mean of zero with a standard deviation of one.

The genetic differentiation for each pair of populations was estimated in several ways. The F_{st} of Weir (1996) was estimated using and without using the *ENA* correction for the presence of null allele as described in Chapuis and Estoup (2007). We also estimated Hedrick's G'_{st} (2005) and Jost's D (2008), as unbiased alternatives to F_{st} and G_{st}, unaffected by within-population diversity values. Their significance was tested with a test of 999 permutations using *GenAlEx* (Smouse & Peakall 2012).

Construction of a core collection

The core collection was built using the *Core Hunter v3* R-package (Thachuk et al. 2009; De Beukelaer et al. 2018) with a double objective. The first was to maximize the representation of the alleles present in the full collection, reaching 100 % of allele coverage (CV criterion, Odong et al. 2013). The second was to obtain a uniform sampling of the genetic diversity and the genetic structure observed in the full collection. For this, we minimized the average distance between each individual (from the full collection) and the closest selected individual (accession) for the core collection (Average accession-to-nearest-entry distance (AN criterion), Odong et al. 2013). In the first step we explored the minimum and maximum ranges of AN and CV for different core collection sizes. We identified the minimum size of the core collection as the size to achieve 100% allele coverage (CV). Working from this minimum size based on the CV criterion (14% of all individuals, see results), we then designed two additional core collections, with 20% and 30% of all available individuals, to increase the AN criterion. Since the two objectives (CV & AN) can be achieved from different combination of clones, we

performed several replicates until the same clonal combinations were drawn five times. The genetic diversity (He and Ho) and the genetic structure of the core collections obtained were compared to the full data set. The R scripts we created for this step are available upon request.

RESULTS

Genetic diversity within populations

As individuals from Tour-de-Palau that were not hybrids clustered mostly with individuals from St-Guilhem (Table S3 in Online Resource ESM01.xls, and see below), all following analyses were done grouping the two populations as one, St-Guilhem. More than 80% of the tests performed did not detect null alleles or only at low frequencies (<0.08) (Table S4 & S5 in Online Resource ESM01.xls). The threshold of 8% null allele frequency was chosen as the threshold below which population genetic parameters may be mostly unbiased (Oddou-Muratorio et al. 2009). The two methods tested (*FreeNA* & *INEst*) gave similar results. Among the 13 loci used, only locus PtTX4001 presented null allele frequencies higher than 0.08 in half of Salzmann's pine populations with both methods while null allele frequencies were non-significant at eight loci, and non-significant overall in all populations over all loci. Although the presence of null alleles slightly underestimates expected and observed heterozygosity (Table S6 in Online Resource ESM01.xls), we decided to use raw data uncorrected for the presence of null alleles. This is because null alleles were not present in most populations, and because our analyses showed that other factors such as inbreeding and amplification failure, and not just null alleles, could explain deviations from Hardy-Weinberg equilibrium (Table S4 in Online Resource ESM01.xls). Also, although possible at population level, deciding which homozygote individuals in a population should be assigned a null allele is impossible.

The genetic diversity (He) of *Pinus nigra salzmannii* was similar to that observed in *P.n. nigra* and *P.n. laricio* (0.73, 0.72 and 0.72, respectively). The rarefied number of alleles (Nae, sample size of 44) varied from 5.64 (Col d'Uglas) to 9.71 (*P.n. nigra*). The coefficient of inbreeding, Fis, was always positive indicating a deficit of heterozygotes. The deficit was especially strong in the two populations Col d'Uglas and Gorges du Tarn (0.14 and 0.13, respectively) which were also those displaying the lowest rarefied number of alleles (Table 2).

Differentiation and identification of evolutionary lineages

Testing if Salzmann's pine is a true taxonomic entity

The best number of genetic clusters obtained was K=3 (Figure S1 in Online Resource ESM02.doc). The three genetic groups corresponded to the three populations used as genetic references: *Pinus nigra nigra*, *P.n. laricio* and *P.n. salzmannii*. The individual assignment of the *P.n. salzmannii* individuals to their own cluster varied from 0.58 to 0.99. Whatever the cluster, a non-negligible number of individuals displayed a shared genetic background with at least one other cluster (Figure 3). The first two axes of the PCA, which explained respectively 3% and 2.6% of the total variance, also showed a clear difference between Salzmann's pines, Austrian black pines and Corsican black pines (Figure 3). The overlapping scatterplots indicated again that some trees of the different subspecies share similar allele frequencies

Testing the presence of sub-structure in Salzmann's pine and identifying potential hybrids

When all sampled trees were analyzed (step 2), the STRUCTURE results indicated that the best number of genetic clusters was K = 7 (Figure S2 in Online Resource ESM02.doc). The exotic black pines (*P.n. nigra* and *P.n. laricio*) formed a single genetic cluster while Salzmann's pines were grouped into six different clusters: St-Guilhem (two clusters with admixed individuals), Col d'Uglas, Gorges-du-Tarn, Conflent and Ardeche. The trees from Tour-de-Palau appeared as a mix of different genetic clusters: over 50% were hybrids with the non-Salzmann group, while the remaining clustered with St-Guilhem, Gorges du Tarn, or result from a mixture of Salzmann populations (Figure 4 and Table S3 in Online Resource ESM01.xls). K = 4 was also statistically likely, indicating a genetic proximity between the Gorges du Tarn and Conflent populations one the hand, and Col d'Uglas and Ardeche populations on the other hand (Figure S2 in Online Resource ESM02.doc). Although K = 4 is meaningful from a phylogeographic point of view, we used the more detailed K = 7 structure to proceed with our analysis.

In order to set an admixture threshold above which trees were of likely anthropogenic origin (due to RTM plantations) we observed the distribution of the membership coefficient Q corresponding to the "*nigra-laricio*" group in the St-Guilhem Salzmann's pine trees older than 250 years. The membership coefficient Q corresponding to the "*nigra-laricio*" group was of 0.046 on average in the St-Guilhem trees older than 250 years. We used the maximum value of Q observed in this 250-year-old group of trees (Q = 0.38) as a threshold value above which admixture is likely due to anthropogenic factors. We finally identified 12 trees (among the 580 *P.n. salzmannii* analyzed) with a probable anthropogenic origin ("nigra-laricio" Q between 0.384 and 0.91, Table S3 in Online Resource ESM01.xls). Six trees were from Tour-de-Palau, three from St-Guilhem and three from Ardeche (Table S7 in Online Resource ESM01.xls).

The first two axes of the PCA, which explained respectively 1.8% and 1.5% of the total variance, separated Salzmann black pines from the other two black pines along the x-axis (Figure 4). Within the Salzmann black pine group, the St-Guilhem population was central and clustered, while the Col d'Uglas, Gorges du Tarn and Conflent populations were scattered. The distribution of trees older than 250 years in St-Guilhem overlapped that of the younger trees.

Genetic differentiation indices

Differentiation indices (Hedrick's G'_{st} and Jost's D) were significant for all provenance pairs (Table 3 and Table S8 in Online Resource ESM01.xls). We did not detect genetic differentiation between the 49 trees of St-Guilhem older than 250 years and the other trees from St-Guilhem. The highest differentiation values were observed between the Col d'Uglas population and all the others. The highest differentiation within the *P.n. salzmannii* group was between Col d'Uglas and Gorges du Tarn (Table 3), often showing F_{st} values higher than those between populations from different subspecies (F_{st} between 0.03 and 0.054).

Core collection

To keep all alleles observed in the full Salzmann's pine data set (208 alleles), the minimum number of clones required in the core collection was 80 (14% of the full dataset, Figure S3 in Online Resource ESM02.doc). The selected clones optimized both the allele coverage CV (100%) and the representativeness of the genetic structure with AN equal to 0.44. We also constructed core collections representing respectively 20% (117 clones) and 30% (170 clones) of the full dataset. While allelic coverage remained at 100%, the AN target was slightly optimized (0.40 and 0.34 respectively). The list of the selected clones is given in table S9 in Online Resource ESM01.xls. The genetic diversity (He) of each core collection is given in Table

S10 in Online Resource ESM01.xls. While we observed a deficit of heterozygotes in the complete collection as well as in the 20% and the 30% collections, conversely, an excess of heterozygotes was observed in the 14% collection (Figure S4 in online Resource ESM2.doc). The relative proportion of each genetic group in the complete collection was maintained in all three core collections (Figure S5 in online Resource ESM2.doc).

DISCUSSION

Pinus nigra salzmannii is an original genetic lineage within the western European and Mediterranean black pine subspecies. It appears clearly differentiated from *P.n. nigra* and *P.n. laricio*. Our results confirm several studies (Aguinagalde et al. 1996) and the botanical nomenclature proposed by leading botanical initiatives such as the World Flora Online (WFO, www.worldfloraonline.org) which identifies *P.n. salzmanni* as a recognized taxon at subspecies level within the *Pinus nigra* group. Many autochthonous individuals appear as of mixed ancestry, indicating porous boundaries between genetic lineages that were explained by a recent (late Pleistocene) evolutionary divergence and incomplete lineage sorting (Scotti-Saintagne et al. 2019). Excess admixture above this baseline porosity indicates hybridization of anthropic origin, including some dated before the onset of large reforestation programs in 1860, a threshold date before which possible anthropic effects on observed patterns of genetic diversity can be generally ruled out. However, historical archives mention some plantations of *P.n. laricio* around 1830 in the Cevennes, south eastern Massif Central (Faugère 1925).

Populations in France, which are at the very northern edge of the distribution of Salzmann's pine in western Europe, belong to three genetic groups sub-structured into six clearly identified clusters of individuals. While St Guilhem is made of two genetic groups, all other groups form geographically identified populations, some located within short distances (25-125 km) of one another. The most isolated populations were also those displaying the lowest within-population genetic diversity (Col d'Uglas, notably). While unsurprising as low census size and isolation are classically linked to low genetic diversity (although numerous counter-examples exist, particularly within the conifers, see for example Eliades et al. 2011, Wachowiak et al. 2011 or Sękiewicz et al. 2020), it is a clear sign of threat to their persistence as their habitat might continue shrinking with global climate change. The five lineages are probably recent, as with all black pine lineages (Scotti-Saintagne et al. 2019), and it can be speculated that recurrent deforestation during the late Holocene (Roiron et al. 2013) has increased genetic differentiation overall and loss of allelic diversity via genetic drift in the smallest populations.

Our results have consequences for conservation. Our relatively small dataset of thirteen SSR loci can be used for the management of protected areas where Salzmann's pine is present and the target of conservation management. Phenotypic traits in black pines often overlap among subspecies and botanical identifications can be far from straightforward (Isajev et al. 2005). Undesirable individuals in the context of conservation of autochthonous biodiversity, whether belonging to another subspecies or lineage than the local resource, or being a hybrid, can now be identified by genotyping using our set of thirteen SSR loci.

The five genetic groups identified in France can be the basis for in-situ conservation, where five different gene conservation units (GCU) need to be identified and implemented in natural forests for optimum genetic diversity and differentiation coverage overall (Koskela et al. 2013, Potter et al. 2017). As one GCU already exists for the Ardeche group (Bois d'Abeau), and St-Guilhem is a large population currently located in a biodiversity protected area, priority should be placed on designing and implementing GCUs in the small populations of Gorges du Tarn and especially the highly differentiated Col d'Uglas which also experiences significant genetic

drift. Pending further investigations, Conflent might also serve as a GCU for the Spanish side of the Pyrenees as genetic diversity in the Pyrenees tends to be shared between the French and Spanish sides for multiple species, while differentiation occurs along the east-west axis of the mountain (Scotti-Saintagne et al. 2021). In the Conflent, however, the issue of hybridization must be resolved as plantations with Austrian black pine and Corsican black pine abound.

In the European Union, *Pinus nigra* and its subspecies are regulated under Council Directive 1999/105/EC on the marketing of forest reproductive material. In most European countries, recommendations for the use of forest reproductive material in forest plantations is linked to subsidies or mandatory practices. In France, because of hybridization and introgression risks, exotic black pine forest reproductive material is not recommended (and thus cannot be subsidized for plantation) within one kilometer of natural forests in regions where *P.n. salzmannii* is autochthonous. As hybridization between lineages is not negligible, recommended good practice for the use of reproductive material in Europe should similarly discourage or ban plantations of exotic material, whether from one subspecies or another, in the close vicinity of documented autochthonous black pine forests, particularly when they are protected forests where black pine is a keystone species.

To reduce hybrid gene flow and safeguard existing resources in this threatened pine, our objective was to produce an ex-situ core collection maximizing overall genetic diversity while maintaining among population diversity, with as few trees as possible. Among the available methods to design core collections (Jeong et al. (2017) for a review) we used *Core Hunter* which proposes clearly defined criteria for measuring the quality of the core collection under different, simultaneous objectives. A full representativeness of alleles was achieved using a set of only 80 individuals (14% of the entire data set) which is rather low compared to the effective population size of 500 generally suggested for in-situ gene conservation to prevent genetic erosion and maintain an ability to adapt (Hoban et al. 2020). Although we have used a few neutral genetic markers, our sampling scheme for the core collection of Salzmann's pine in France provides for a wide representation of the existing diversity in each population, making it likely that we also incorporated ample adaptive diversity. However, using genomic scans of functional gene variants would provide an additional insurance that the core collection includes all of the adaptive diversity available.

The size of this collection can be tailored to fit the needs of conservation agencies, for example in terms of land available for planting. As grafting is possible in this pine, the core collection can be planted ex-situ. Such an ex-situ collection exists in southern France, on communal land in Saint Paul lez Durance and managed by the forest service (ONF). It can be duplicated as needed for conservation and is also intended to be used as a seed orchard where selection is focused on increased diversity rather than, classically, a group of specific traits, for diversifying the portfolio of seeds available to adapt plantation strategy to climate change (Fady et al. 2016; Hof et al. 2017).

Finally, our strategy of combining in-situ and ex-situ conservation and the methods we used to achieve this goal for a species that is widely distributed, with fragmented patches of occurrence, in which gene flow is substantial, can be easily transferred to many forest tree and, possibly, plant species. Microsatellites continue to demonstrate significant statistical power at a reasonable cost, providing a valuable tool for conservation agencies. Their sequencing (rather than length polymorphism which we used here) might prove an interesting way forward that could generate increased statistical power, if need be (Barthe et al. 2012). At a time when funds are not necessarily plentiful for protecting biodiversity and as climate change increasingly threatens forest tree species, we advocate the systematic use of a combined in-situ and ex-situ strategy. Our ex-situ strategy does not include seed banking. Installing trees on suitable land and letting them grow and reproduce is an alternate and opportune dynamic

ex-situ conservation strategy, both safeguarding threatened genetic resources outside of their environment and letting natural selection in a non-native environment generate an evolved collection of genotypes within each seed crop (EUFORGEN, 2021).

REFERENCES

Aitken SN, Whitlock MC (2013) Assisted gene flow to facilitate local adaptation to climate change. Annual Review of Ecology, Evolution, and Systematics 44:367-388.

Aguinagalde I, Llorente F, Benito C (1996) Relationships among five populations of European Black pine (Pinus nigra ARN.) using morphometric and isozyme markers. Silvae Genetica 46:1-5.

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, et al. (2013) Potential for evolutionary responses to climate change–evidence from tree populations. Global Change Biology 19(6):1645-1661.

Barthe S, Gugerli F, Barkley NA, Maggia L, Cardi C, Scotti I (2012). Always look on both sides: phylogenetic information conveyed by simple sequence repeat allele sequences. PLoS One 7(7):e40699.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Molecular and Biology and Evolution 24:621– 631.

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. Journal of Heredity 100(1):106-113.

Combes F (1989) Restauration des terrains en montagne. Du rêve à la réalité. Revue Forestière Française 41(2):91-106.

Dąbrowski MJ, Pilot M, Kruczyk M, Żmihorski M, Umer HM, Gliwicz J (2014) Reliability assessment of null allele detection: inconsistencies between and within different methods. Molecular Ecology Resources 14:361-373.

De Beukelaer H, Davenport GF, Fack V (2018) Core Hunter 3: flexible core subset selection. BMC bioinformatics 19:1-12.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: series B (methodological) 39(1):1-22.

Dray S, Dufour AB (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software 22(4):1–20.

Earl DA., VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359-361.

Eliades NGH, Gailing O, Leinemann L, Fady B, Finkeldey R (2011) High genetic diversity and significant population structure in Cedrus brevifolia Henry, a narrow endemic Mediterranean tree from Cyprus. Plant Systematics and Evolution 294:185-198.

EUFORGEN (2021) Forest Genetic Resources Strategy for Europe. European Forest Institute. Barcelona, Spain. ISBN 978-952-7426-48-7.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14(8):2611-2620.

Fady B, Brahic P, Cambon D, Giovannelli G, Musch B, Thévenet J, Turion N (2016) Le Pin de Salzmann (Pinus nigra Arnold ssp. Salzmanni (Dunal) Franco), un patrimoine naturel vivant à protéger et valoriser. Causses et Cévennes 121: 336-341. https://hal.inrae.fr/hal-02630610v1.

Fady B, Cottrell J, Ackzell L, Alía R, Muys B, Prada A, González-Martínez SC (2016) Forests and global change: what can genetics contribute to the major forest management and policy challenges of the twenty-first century? Regional Environmental Change 16(4):927-939 (https://doi.org/10.1007/s10113-015-0843-9),

Fady B, Aravanopoulos F, Benavides R, González-Martínez SC, Grivet D, Lascoux M, Lindner M, Rellstab C, Valladares F, Vinceti B (2020). Genetics to the rescue: managing forests sustainably in a changing world. Tree Genetics and Genomes 16(6):1-11 (https://doi.org/10.1007/s11295-020-01474-8).

Fady B, Scotti-Saintagne C, Roig A, Touhami N, Lingrand M, Rei F, Turion N, Gilg O, Jean F, Corréard M, Clastre P (2023). Georeferenced, dendrometric and genotypic data for 1333 Pinus nigra ssp. salzmannii trees from native populations in France and Europe. Recherche Data Gouv, V1, https://doi.org/10.57745/QTGYY8.

Flaugère A (1925) Forêts nouvelles des Cévennes. De l'acclimatation du pin laricio de Corse et des considérations qui en résultent concernant la reconstitution de la forêt climatique. Revue des Eaux et Forêts 63:201-211. https://gallica.bnf.fr/ark:/12148/bpt6k9736927z/f223.item.

Forester BR, Beever EA, Darst C, Szymanski J, Funk WC (2022) Linking evolutionary potential to extinction risk: applications and future directions. Frontiers in Ecology and the Environment 20(9):507-515.

Gapare WJ, Yanchuk AD, Aitken SN (2008) Optimal sampling strategies for capture of genetic diversity differ between core and peripheral populations of Picea sitchensis (Bong.) Carr. Conservation Genetics 9:411–418. https://doi.org/10.1007/s10592-007-9353-8.

Giovannelli G, Roig A, Spanu I, Vendramin GG, Fady B (2017) A New Set of Nuclear Microsatellites for an Ecologically and Economically Important Conifer: the European Black Pine (Pinus nigra Arn.). Plant Molecular Biology Reporter 35:379–388.

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(4):618-620.

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638.

Hoban S, Schlarbaum S (2014) Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. Biological Conservation 177:90-99.

Hoban S (2019). New guidance for ex situ gene conservation: Sampling realistic population systems and accounting for collection attrition. Biological Conservation 235:199-208.

Hoban S, Bruford M, Jackson JDU, et al. (2020) Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biological Conservation 248:108654. https://doi.org/10.1016/j.biocon.2020.108654

Hof AR, Dymond CC, Mladenoff DJ (2017) Climate change mitigation through adaptation: the effectiveness of forest diversification by novel tree planting regimes. Ecosphere 8(11):e01981.

IUCN (2016) IUCN World Conservation Congress 2016 , recommendation 104: WCC-2016-Rec-104-EN. Integrating autochthonous forest genetic diversity into protected area conservation objectives.

IUCN (2021) IUCN World Conservation Congress 2020, recommendation 109: WCC-2020-Res-093-EN. A call for increased consideration of genetic diversity in IUCN planning and actions.

Isajev V, Fady B, Semerci H, Andonowski V (2005) Technical guidelines for genetic conservation and use of European black pine (*Pinus nigra*). EUFORGEN, International Plant Genetic Resources Institute, Rome, Italy.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801-1806.

Jeong S, Kim JY, Jeong SC, Kang ST, Moon JK, Kim N (2017) GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. PLoS One 12(7):e0181420.

Jombart T (2008) *adegenet*: an R package for the multivariate analysis of genetic markers. Bioinformatics 24(11): 1403–1405

Jost L (2008) GST and its relatives do not measure differentiation. Molecular Ecology 17(18):4015-4026.

Jump AS, Peñuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters 8(9):1010-1020.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15(5):1179-1191.

Koskela J, Lefèvre F, Schüler S, et al. (2013) Translating conservation genetics into management: pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. Biological Conservation 157:39–49.

Lande R, Shannon S (1996) The role of genetic variation in adaptation and population persistence in a changing environment. Evolution 50:434-437.

McGowan PJ, Traylor-Holzer K, Leus K (2017) IUCN guidelines for determining when and how ex situ management should be used in species conservation. Conservation Letters 10(3):361-366.

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

Oddou-Muratorio S, Vendramin GG, Buiteveld J, Fady B (2009) Population estimators or progeny tests: what is the best method to assess null allele frequencies at SSR loci? Conservation Genetics 10(5):1343-1347.

Odong TL, Jansen J, Van Eeuwijk FA, van Hintum TJ (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theoretical and Applied Genetics 126:289-305.

Palsbøll PJ, Berube M, Allendorf FW (2007) Identification of management units using population genetic data. Trends in Ecology and Evolution 22(1):11-16.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945-959.

Potter KM, Jetton RM, Bower A, Jacobs DF, Man G, Hipkins VD, Westwood M (2017) Banking on the future: progress, challenges and opportunities for the genetic conservation of forest trees. New Forests 48:153-180.

Raffl H, Konrad H, Curtu LA, Geburek T (2018) Genetic evidence of human mediated, historical seed transfer from the Tyrolean Alps to the Romanian Carpathians in Larix decidua (Mill.) forests. Annals of Forest Science 75(4):1-13.

Roiron P, Chabal L, Figueiral I, Terral JF, Ali AA (2013) Palaeobiogeography of Pinus nigra Arn. subsp. salzmannii (Dunal) Franco in the north-western Mediterranean Basin: a review based on macroremains. Review of Palaeobotany and Palynology 194:1-11.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4(1):137-138.

Secretariat of the Convention on Biological Diversity (2020) Global Biodiversity Outlook 5. Montreal. https://www.cbd.int/gbo5

Schueler S, Falk W, Koskela J, et al. (2014) Vulnerability of dynamic genetic conservation units of forest trees in Europe to climate change. Global Change Biology 20(5):1498-1511.

Scotti-Saintagne C, Fady B, Giovannelli G, Turion N (2018) Le pin de Salzmann : une espèce à protéger. La Garance Voyageuse 123:7-12. https://hal.inrae.fr/hal-01894847v1.

Scotti-Saintagne C, Giovannelli G, Scotti I, Roig A, Spanu I, Vendramin GG, Guibal F, Fady B (2019) Recent, late-Pleistocene fragmentation shaped the phylogeographic structure of the European black pine (Pinus nigra Arnold). Tree Genetics and Genomes 15: 76 (https://doi.org/10.1007/s11295-019-1381-2)

Scotti-Saintagne C, Boivin T, Suez M, Musch B, Scotti I, Fady B (2021) Signature of mid-Pleistocene lineages in the European silver fir (*Abies alba* Mill.) at its geographic distribution margin. Ecology and Evolution 11:10984-10999 (https://doi.org/10.1002/ece3.7886).

Sękiewicz K, Walas Ł, Beridze B, Fennane M, Dering M (2020) High genetic diversity and low future habitat suitability: will Cupressus atlantica, endemic to the High Atlas, survive under climate change? Regional Environmental Change 20(4):132.

Smouse RPP, Peakall R (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28(19):2537-2539.

Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF (2009) Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. BMC Bioinformatics 10:1-13.

Vernet JL (2006) History of the Pinus sylvestris and Pinus nigra ssp. salzmanni forest in the Sub-Mediterranean mountains (Grands Causses, Saint-Guilhem-le-Désert, southern Massif Central, France) based on charcoal from limestone and dolomitic deposits. Vegetation History and Archaeobotany 16:23-42.

Wachowiak W, Salmela MJ, Ennos RA, Iason G, Cavers S (2011) High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (Pinus sylvestris L.) in Scotland. Heredity 106(5):775-787.

STATEMENTS & DECLARATIONS

Acknowledgements
We thank N. Turion, O. Gilg, F. Rei (INRAE UEFM, Avignon, France) for their contribution to field sampling and wood core collection and analysis. We thank D. Cambon, L. Golliard, B. Latour (ONF, Occitanie Region, France), F. Jean (INRAE URFM, Avignon, France) and D. Guillemet and A. Laurioux (Département Ardèche, France) for their contribution to field sampling. We thank F. Guibal (IMBE, Aix-en-Provence, France) for his contribution to wood core analysis. We thank P. Brahic, M. De Castro and J. Reilhan (ONF experimental nursery, Cadarache, France) and J. Thévenet (INRAE UEFM) for their contribution to grafting and multiplying the clones that will be in the core collection.

Funding
This work received financial support from ONF (the French Forest Service) via the French national program for the conservation of Salzmann's pine. We would like specially to thank D. Cambon (ONF) who was instrumental to the success of this program. We also acknowledge the financial support of the French Occitanie Region (project OcciGen).

Competing Interests
The authors have no relevant financial or non-financial interests to disclose.

Author Contributions
C. Scotti-Saintagne and B. Fady conceived, designed and coordinated the study. Material preparation, data collection, data curation and analysis were performed by C. Scotti-Saintagne, A. de Sousa Rodrigues and A. Roig. The first draft of the manuscript was written by C. Scotti-Saintagne, A. de Sousa Rodrigues and B. Fady and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability
The datasets generated and analyzed during the current study are available in the open-access repository: Georeferenced, dendrometric and genotypic data for 1333 *Pinus nigra* ssp. *salzmannii* trees from native populations in France and Europe, https://doi.org/10.57745/QTGYY8.

Figure 1: a) Occurrence of natural populations of the different *Pinus nigra* sub species in Europe (based on Isajev et al. 2005); b) Location of the French populations analyzed in this study; c) photograph of the population sampled in Ardeche.

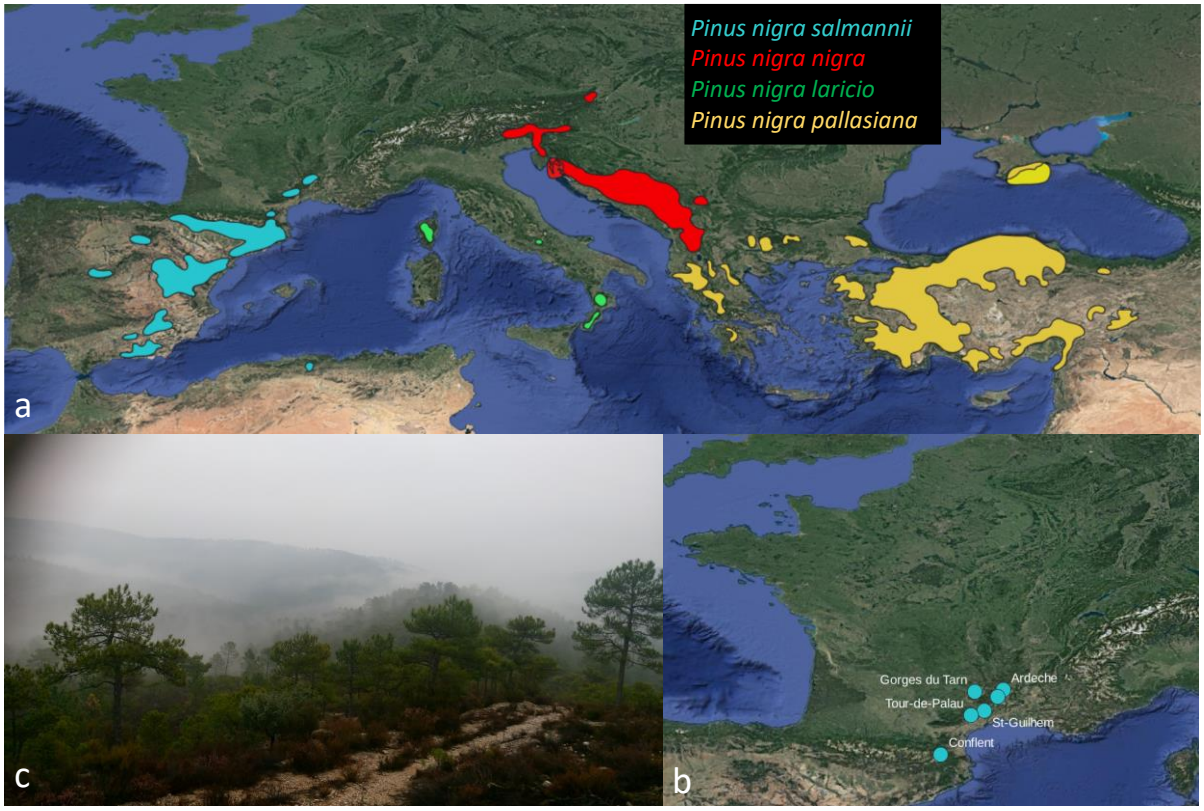


Figure 2: Age distribution of the sampled Salzmann's pines.

Genotyped trees were all older than 163 years (thus born before 1860). The 250 year-old threshold was used to select a group of 49 autochthonous trees, all from population Saint Guilhem. This group was used as the genetic reference for Salzmann's pine in France in step 1 of our analysis.

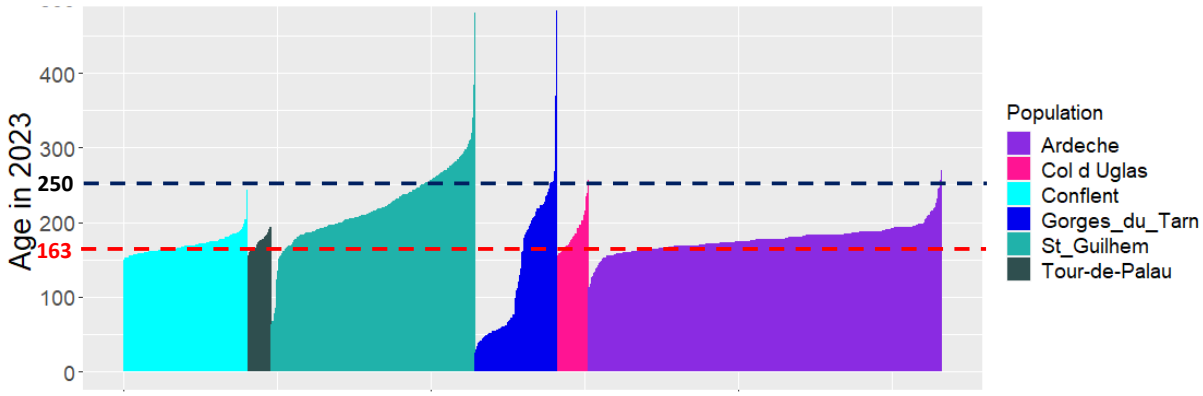


Figure 3: Structure of the genetic diversity between three sub-species of *Pinus nigra* (*nigra*, *laricio* and *salmannii*) **a**. Principal component analysis. The confidence ellipses shown are defined as the regions that contain 95% of all samples that can be drawn from the underlying Gaussian distribution. The first two axes of the PCA, which explain respectively 3% and 2.6% of the total variance, show the degree of separation between Salzmann's, Austrian and Corsican black pines. **b**. Results of the STRUCTURE analysis considering the best number of genetic groups, K=3. Each vertical bar represents the membership coefficients of one individual and each of the three colors indicates the proportion of ancestry in a particular genetic group.

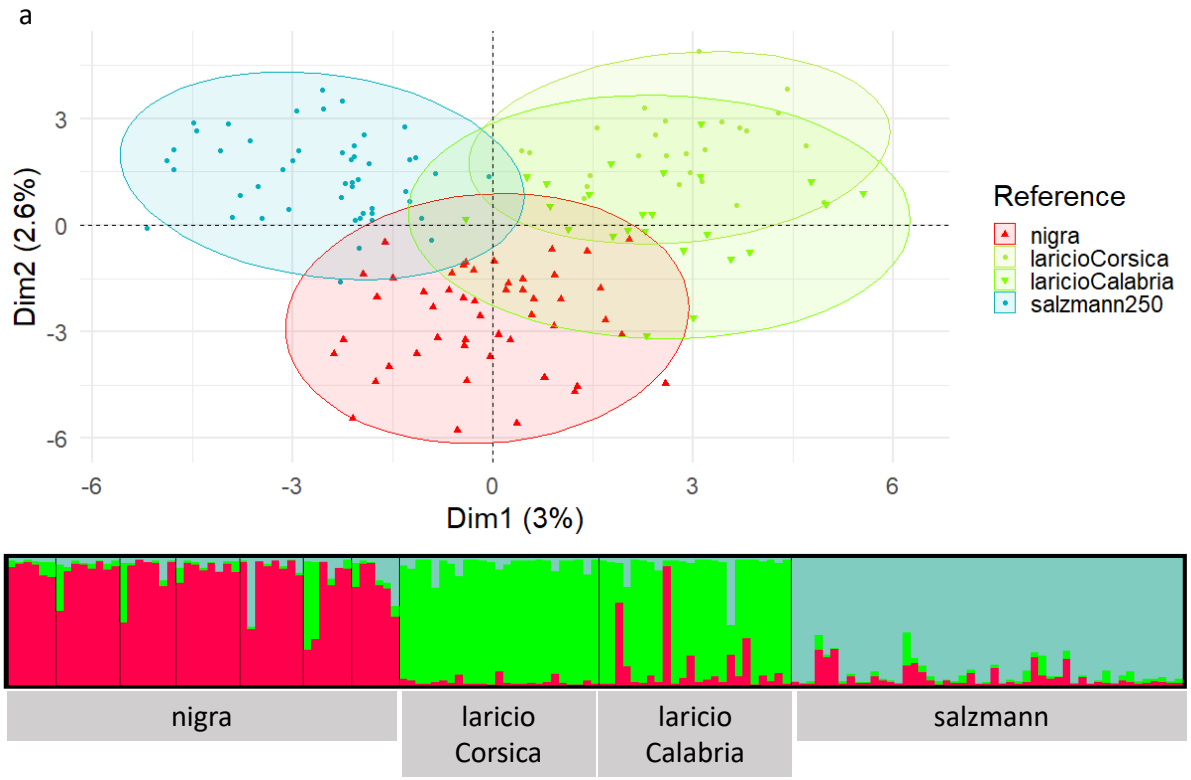
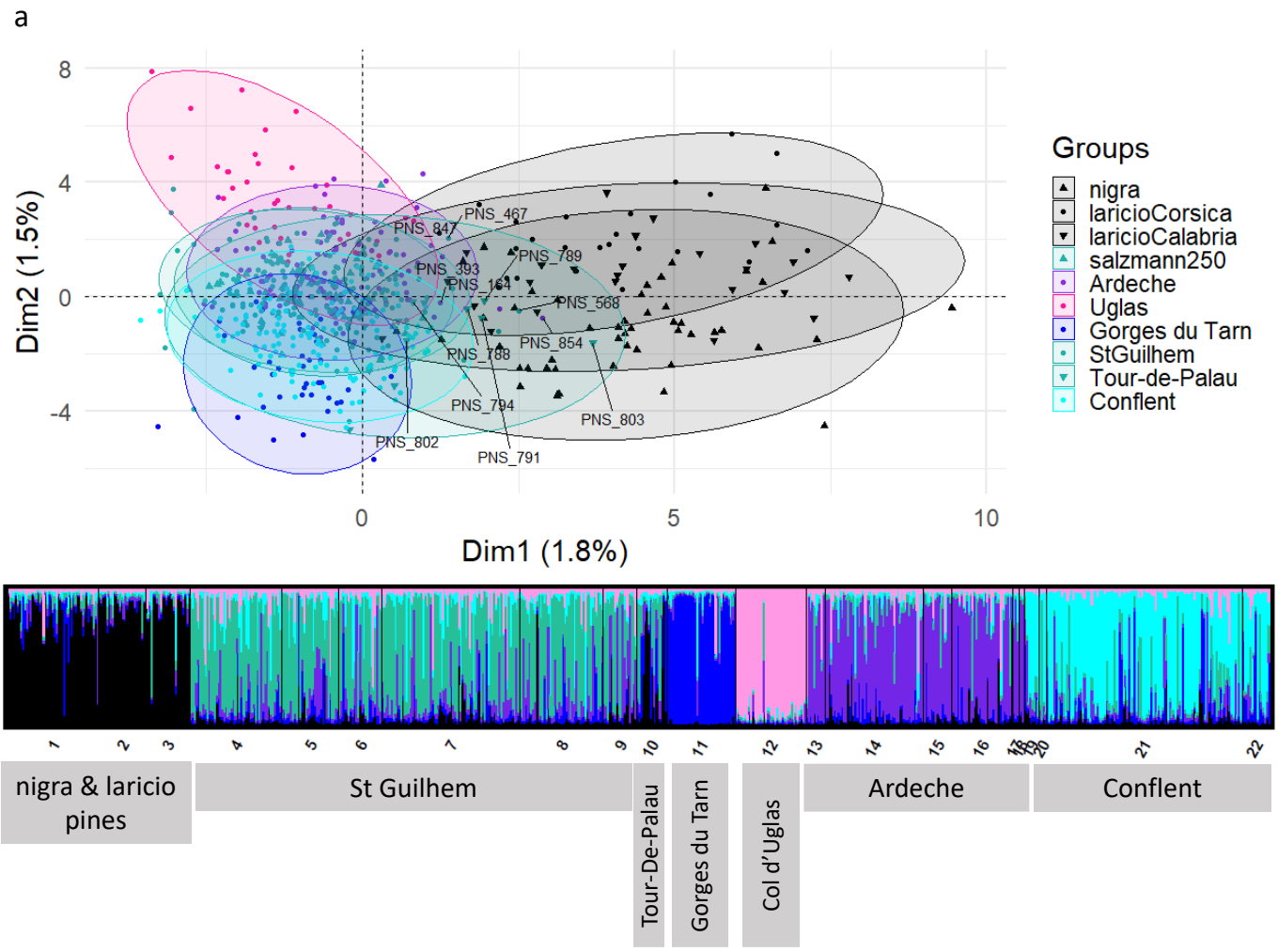


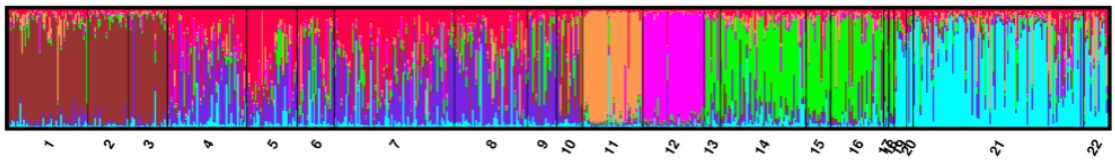
Figure 4: Results of the STRUCTURE analysis obtained from 678 *Pinus nigra* trees belonging to three sub-species used as reference (*nigra*, *laricio*, *salzmannii,* 147 trees) and 531 additional trees established before 1860 and likely belonging to *Pinus nigra salzmannii*

a. Principal component analysis. The confidence ellipses shown are defined as the regions that contain 95% of all samples that can be drawn from the underlying Gaussian distribution. The first two axes of the PCA, which explain respectively 1.8% and 1.5% of the total variance, show the degree of separation between Salzmann's, Austrian and Corsican black pines. Within the Salzmann's pine group, the Saint Guihlem population is central and clustered, while there is a splintering of the point cloud for the trees of the Col d'Uglas, the Gorges du Tarn and Conflent populations.The distribution of trees older than 250 years in Saint Guilhem overlaps with the distribution of younger trees in Saint Guilhem.Trees with a label in the PCA are considered as antropogenic hybrids (see results from STRUCTURE) and were excluded from the Core collection.

b. Results of the STRUCTURE analysis considering the best number of genetic groups, K=6. Each vertical bar represents the membership coefficients of one individual and each of the six colors indicates the proportion of ancestry in a particular genetic group. Sub-populations are identified by number, their full name is given in Table S1 in Online Resource ESM01.xls)



K=7 13/15, Mean(LnProb) = -30893.769, Mean(similarity score) = 0.988



nigra & laricio pines	St Guilhem	Tour-De-Palau	Gorges du Tarn	Col d'Uglas	Ardeche	Conflent

Table 1: Information on sampled Salzmann's pine populations: names, sample size, average, minimum and maximum date of seedling installation based on tree ring width of individual trees, geographic coordinates in decimal degrees.

Population	Sample size	Date of seedling installation			Longitude	Latitude
		Mean	**Maximum**	**Minimum**		
Conflent	132	1847	1778	1859	2.346622	42.553448
Ardeche	117	1836	1752	1859	4.0901849	44.342144
Gorges du Tarn	37	1790	1539	1843	3.2883365	44.2941212
St-Guilhem	240	1790	1539	1843	3.554148	43.7675598
Tour-de-Palau	16	1847	1837	1858	3.1849491	43.6390566
Col d'Uglas	38	1831	1766	1859	3.9204612	44.1413098
Total	**580**	**1817**	**1539**	**1859**		

Table 2: Within-population genetic diversity in *Pinus nigra salzmannii*, *nigra* and *laricio*

Population	Sample size	NA	Nae	AR (k=44)	He	Ho	Fi
nigra	49	11.92	6.04	9.71	0.716	0.658	0.085
laricio	49	11.31	5.6	9.01	0.724	0.671	0.074
St-Guilhem>250	49	10.54	5.63	8.78	0.733	0.680	0.073
Ardeche	114	12.08	6.00	8.84	0.729	0.678	0.070
Col d'Uglas	38	6.00	3.53	5.64	0.641	0.537	0.165
Gorges du Tarn	37	7.62	4.68	6.99	0.669	0.571	0.148
St-Guilhem	198	13.77	5.55	8.92	0.719	0.651	0.095
Conflent	132	12.77	6.00	8.74	0.737	0.674	0.086

NA: number of alleles

NAe: Effective number of alleles, i.e. the number of alleles at equal frequencies according to the He of the population, Nielsen et al. (2003)

AR(k=44): Allelic richness (expected number of alleles among 44 gene copies)

He: gene diversity corrected for sample size, Nei (1978)

Ho: observed heterozygosity

Fi: inbreeding coefficient (all values are highly significant, as determined by their p-value after 10000 randomizations of gene copies among individuals)

Table 3: Genetic differentiation between pairs of black pine populations - Fst values are below the diagonal, Joost's D values above the diagonal.

	nigra	Laricio Corsica	Laricio Calabria	St-Guilhem 250	Ardeche	Col d'Uglas	Gorges du Tarn	St-Guilhem	Conflent
nigra	-	0.216	0.155	0.141	0.185	0.306	0.206	0.136	0.162
Laricio Corsica	0.051	-	0.095	0.238	0.178	0.294	0.335	0.229	0.214
Laricio Calabria	0.037	0.030	-	0.221	0.171	0.325	0.297	0.197	0.196
St-Guilhem 250	0.032	0.054	0.048	-	0.074	0.149	0.112	**-0.005**	0.051
Ardeche	0.038	0.041	0.038	0.017	-	0.150	0.139	0.071	0.071
Col d'Uglas	0.074	0.078	0.080	0.040	0.038	-	0.277	0.167	0.198
Gorges du Tarn	0.050	0.082	0.070	0.030	0.034	0.076	-	0.107	0.127
St Guilhem	0.029	0.051	0.043	**0.003**	0.015	0.042	0.028	-	0.063
Conflent	0.033	0.047	0.041	0.013	0.015	0.048	0.031	0.013	-

All values are significant (p-value<=0.001) except two that appear in bold. Cells highlighted in grey emphasize comparisons involving Corsican and Austrian black pine populations.