



RESEARCH PAPER

Open Access

Provenance variation and seed sourcing for sessile oak (*Quercus petraea* (Matt.) Liebl.) in France



Quentin Girard¹, Alexis Ducouso^{2*} , Clothilde Boubée de Gramont¹, Jean Marc Louvet², Patrick Reynet^{2^}, Brigitte Musch¹ and Antoine Kremer^{2*} 

Abstract

Key message: Sessile oak (*Quercus petraea* (Matt.) Liebl.) provenance variation was assessed in a multisite test based on traits of economic and ecological relevance in France. While climatic drivers generated genetic clines at a range-wide scale, provenance variation in France was mainly shaped by past silvicultural regimes. We developed a multitrait approach to facilitate decision-making for seed sourcing. A set of provenance clusters is proposed, supporting recommendations for plantation programmes.

Context: Among broadleaves, sessile oak (*Quercus petraea*) is likely to spread in the context of current climate change and is increasingly planted in France. Seed sourcing is of the utmost importance for ensuring plantation success and adaptation. The selection of appropriate seed sources is highly challenging when the future climate conditions of plantation areas are uncertain.

Aims: We aimed at identifying drivers of provenance variation in *Q. petraea* and to build provenance clusters based on traits of adaptive and economic value, to ultimately support decision-making in seed sourcing.

Methods: We analysed a multisite provenance test established 30 years ago and comprising a large collection of *Q. petraea* provenances by performing phenotypic assessments of survival, growth, phenology, and stem-quality traits. We analysed climate-trait correlations at a range-wide scale and used multivariate statistics [multivariate mixed models, principal component analysis (PCA)] and classification methods [hierarchical clustering analysis (HCA), K-means method] to generate an overall clustering of french provenances.

Results: Provenance effects were highly significant regardless of the trait considered, whereas interaction effects between provenance and other experimental sources of variation were minor compared to provenance and environmental variance. There was limited variation between provenances collected in the same forest in comparison to origins of different forests. We found sharp temperature-driven genetic clines for growth and phenology-related traits at a range-wide scale. The multitrait classification approach grouped the French provenances into 11 clusters, with the members of each cluster having similar trait values. Overall, the cluster composition of provenances poorly matched the provenance regions identified on ecological grounds, but rather mirrored the silvicultural regimes implemented in the source stands in the past.

Handling Editor: Ricardo Alia

* Correspondence: alexis.ducouso@inrae.fr; antoine.kremer@inrae.fr

[^]Patrick Reynet is deceased.

²UMR BIOGECO, INRAE, Université de Bordeaux, 69 Route d'Arcachon, 33612 Cestas, France

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: Two of the provenance clusters (comprising a total of 34 provenances) were identified as potentially useful sources of reproductive material. We recommend mixing seeds of different provenances from a given cluster to ensure the maintenance of diversity and to enhance adaptability to future climatic conditions.

Keywords: *Quercus petraea*, Provenance, Climate, Classification methods, Seed sourcing

1 Introduction

The broadleaved tree species of temperate Europe include sessile oak (*Quercus petraea* (Matt.) Liebl.), which is likely to expand its range in the current context of climate change (Mette, Dolos et al. 2013; Stimm, Heym et al. 2021). This species is found throughout the continent of Europe, from Spain to Southern Sweden, Norway, and Russia (Camus 1938; Le Hardy de Beaulieu and Lamant 2006). In most central European countries, *Q. petraea* is a key economic and ecological resource (Hanewinkel, Cullmann et al. 2013; Eaton, Caudullo et al. 2016). Given current trends towards an expansion of the range of *Q. petraea* (Truffaut, Chancerel et al. 2017), concerns have been raised about meeting seed needs for plantations. Recent studies have shown that seed yields are increasing with climate change, potentially facilitating natural regeneration (Caignard, Kremer et al. 2017). Natural regeneration is currently advised for stand renewal, whenever sufficient seeds are available (Jarret 2004; Sardin 2008). If seed yields are insufficient or a new area is being forested, plantation approaches are used. According to French national statistics, more than 3.5 million sessile oak seedlings were produced in French nurseries in 2017 for field plantation, making *Q. petraea* the leading broadleaved species planted in the country (Anonymous 2017). Seed sourcing has, thus, become a matter of the utmost importance for ensuring afforestation success. The selection of appropriate seed sources is challenging when the future climate conditions of the plantation area are uncertain. In forestry, the traditional approach has been to use as much locally sourced seed as possible, particularly for species displaying local adaptation. However, there has been a move in recent years towards anticipating future climate change and selecting seeds from a current geographic origin with climate conditions matching the future climatic conditions predicted for the plantation area (assisted migration) (Saenz-Romero, O'Neill et al. 2021). Provenance tests are a prerequisite for seed sourcing, as they provide invaluable clues to local adaptation and insights into the consequences of transferring seed sources. Provenance research in *Quercus petraea* has a long history (see Kleinschmit 1993 for a review). However, as the seed yield is

irregular and acorns cannot be kept for more than one year, most provenance experiments comprise only a limited number of provenances of narrow geographic origin (Kleinschmit and Svolba 1994; Liesebach and Stephan 2000; Jensen and Hansen 2008; Burianek, Benedikova et al. 2011; Wilkinson, Eaton et al. 2017). Some provenance experiments have been performed at the juvenile stage, but such experiments provide limited information for seed sourcing (Liepe 1993; Deans and Harvey 1996; Alberto, Bouffier et al. 2011). Here, we made use of a comprehensive collection of provenances, assembling material over successive years due to irregular seed harvests, and including material originating from classified and non-classified seed stands. The whole experiment, which has now been running for about 26 years, was set up over a period of 4 years, at four sites. This experiment has already been used to achieve targeted objectives with a limited number of provenances (Ducousso, Guyon et al. 1996; Kremer, Petit et al. 1998; Ducousso, Louvet et al. 2005; Sinclair, Stone et al. 2015; Lobo, Torres-Ruiz et al. 2018; Pihain, Gerhold et al. 2019; Sáenz-Romero, Kremer et al. 2019; Torres-Ruiz, Kremer et al. 2019; Bert, Lebourgeois et al. 2020; Matyas 2021). This range-wide collection allowed us to address not only the ecological and historical drivers but also the human drivers that may have shaped the extant distribution of genetic variation. Indeed, the sampled populations are spread over wide ecological gradients and originate from different source populations. They are also part of forests that underwent longstanding contrasting silvicultural regimes. Finally, our ultimate aim was to provide practical recommendations for seed sourcing of *Quercus petraea* in France. We adopted a compromise strategy, with a balance between local sources (local adaptation) and transferred sources (assisted migration), by building clusters with multiple provenances, and recommending clusters for seed sourcing in France. The rationale underlying this strategy is the mixing of seeds of different provenances from a given cluster to ensure the maintenance of diversity and to enhance adaptability to future climatic conditions. Methodological efforts were thus mostly focused on the building of clusters of provenances based on traits of adaptive and economic

value, through a combination of statistical and empirical approaches.

2 Materials and methods

2.1 Plant material

The study material was a collection of *Quercus petraea* provenances collected across Europe, from Spain to Scandinavia and from Ireland to Georgia (Kremer and Ducouso 2017; Ducouso, Ehrenmann et al. 2022). A collection unit corresponding to a provenance consisted of one or multiple forest compartments larger than 15 ha in which the species of interest was the only white oak present. Acorns were harvested from 50 spots evenly distributed throughout the collection unit and bulked into a single seed lot. The experiment focused mainly on *Q. petraea*, but a minor collection effort was devoted to *Q. robur*, which was included in the study for purely comparative purposes. Finally, collections were not restricted to seed stands or phenotypically outstanding stands. We also ensured the inclusion of provenances from the ecological and geographic margins of the distribution of the species, to obtain a representative sample of the genetic variation of the species. Seed yields are spatially and temporally heterogeneous in oaks, so acorn harvests were performed in four different years with sufficient seed crops (1986, 1987, 1989, and 1992) to ensure the construction of a range-wide collection. There were, therefore, four yearly sets of provenances. In this way, we were able to harvest the same provenance (i.e. from the same collection unit) in different years, thereby maintaining genetic connectivity across the different yearly sets. These provenances repeatedly collected over several years are referred to hereafter as “crossover” provenances. In a few forests, we also collected provenances from different collection units (within the same set, or in different sets), which we refer to as “twinned” provenances. In total, 127 provenances were harvested, including 74 French provenances (70 *Q. petraea* and 4 *Q. robur*), 27 “crossover populations” (common to several sets), and 20 “twinned” provenances (Additional file 1: Table S1a and Figure S1). The design also included four French *Q. robur* provenances. For the sake of clarity, a “provenance” is considered to be the geographic origin of a seed lot. The four yearly sets collected in 1986, 1987, 1989, and 1992 are referred to hereafter as sets 1, 2, 4, and 5 (Ducouso, Ehrenmann et al. 2022 Tables 3 and 4 therein). The 70 *Q. petraea* provenances comprised 60 seed stands from 17 provenance regions, with a total of 19 provenance regions in France (Colin, Ducouso et al. 2017; Additional file 1: Table S1a). In addition to the French collection, provenances from other countries of the natural distribution were included in the collection among which 38 are used in this study to analyse provenance-climate relationships (Additional

file 1: Table S1b). The maternal lineage of all provenances was known, thanks to the prior genotyping of chloroplast haplotypes (Kremer, Kleinschmit et al. 2002; Petit, Csaikl et al. 2002). It was, therefore, possible to compare provenance trait values between haplotypic groups.

2.1.1 Plantation and experimental layout

Seed lots were heat-treated to prevent *Ciboria batschiana* contamination during seed storage. Seedlings were raised in the Guéméné-Penfao state nursery in North-West France (latitude: 47.631° N; longitude: 1.892° W), in an experimental layout with five randomized complete blocks. After 3 years in the nursery, the seedlings of each yearly set were planted in a common garden experiment replicated in four national forests located along a geographic gradient running from the west to the east of France (Additional file 1: Figure S1, Table 3 in Ducouso, Ehrenmann et al. 2022). The conditions at the sites of three of the forests (Petite Charnie, Vincence, Sillegny) were favourable for *Q. petraea*, whereas those in the fourth forest (Vierzon) were much harsher, due to the podzolic-like soil.

With very few exceptions, all the french provenances were established at all four sites (forests), in four yearly plantation sets corresponding to the four yearly collection sets. However, provenances from other countries were not systematically replicated in the four sites due to the limitation of material. Hence the overall analysis was restricted to the french material, while the climate-trait correlation analysis was limited to only one site (Sillegny) but for all provenances installed in Sillegny (Additional file 1: Tables S1a and b). The experiment thus comprised 16 tests (4 yearly sets replicated at four sites). The four yearly sets of seeds collected in 1986, 1987, 1989, and 1992 were used to generate saplings that were planted in the early months of 1990, 1991, 1993, and 1995, labelled sets 1, 2, 4, and 5 (Fig. 2 in Ducouso, Ehrenmann et al. 2022) for a full description of the experimental layout). In what follows, the sources of variation due to the year of seed collection, and to the year of plantation are thus confounded, and are designated in this study as effects due to the “yearly set”.

At each site, the four yearly sets were planted as four adjacent units. The full experimental layout is described in Figure 2 and Table 4 of Ducouso, Ehrenmann et al. (2022). Each test was subdivided into complete macroblocks, corresponding to homogeneous site conditions. The number of macroblocks ranged from 5 to 10, depending on the test. Each macroblock comprised several randomized incomplete microblocks, each consisting of eight plots of different provenances. Each plot consisted of 24 trees of the same provenance, distributed over four rows, with six trees per row. As rows were separated by 4 m,

and there was 1.75 m between trees within a row, the plots were almost square (12 m * 10.5 m). Within a macroblock, an individual provenance was represented by at least two or three replicated plots. Within a test, each provenance was represented by 120 to 384 trees, on average. Given the complexity of the overall design, and to facilitate the construction of the statistical models for data analysis (Sections 2.2.1 to 2.2.3), Fig. 2 in Ducouso, Ehrenmann et al. (2022) helps to illustrate the different factorial and nested levels of variation considered in the experimental layout.

2.1.2 Phenotypic assessments and climatic data

Total height (HT) was assessed 4, 10, and 20 years after plantation, together with girth (CIRC) at breast height (1.3 m). Phenological observations were conducted during the spring of the third growing season. The developmental stage of the apical bud (DEB) was scored on a scale ranging from 0 to 5 (0 = dormant bud, 1 = swollen bud, 2 = bursting bud, 3 = leaf expansion, 4 = free leaf, 5 = elongating internodes). Phenological observations were conducted at a single time point in each test, usually when bud score variation was largest. Survival (SURV) was also recorded at each time point at which phenotypic observations were made. Stem shape (SHP) was assessed in years 10 and 20, with a scoring ranging from 0 to 10 (Additional file 1: Figure S2). The stem shape score took into account the straightness of the stem, the occurrence of forks, and the number and distribution of branches on the stem, as illustrated in Additional file 1: Figure S2. The experimental layout and phenotypic assessments are described in more detail in the companion data paper (Ducouso, Ehrenmann et al. 2022).

Climatic data for each provenance were extracted from the WorldClim database (Hijmans, Cameron et al., 2005), which were downscaled at 1-km spatial resolution and used in a previous range-wide investigation (Saenz-Romero, Lamy et al. 2017). We considered mean values of climatic data over the period 1950–2000 to gauge the climate at the provenance origin. Physiological meaningful climate variables were selected to account for the seasonal and annual balance of thermal energy and moisture (Saenz-Romero, Lamy et al. 2017 Supplemental Table S3 therein).

2.2 Statistical analysis

We implemented several linear mixed models, at different scales, to disentangle provenance effects from the other sources of variation. The ultimate goal was to generate a provenance classification that could be used for the selection of optimal seed sources. The mixed models were generated with R software (Munoz and Sanchez 2018), and are described in Sections 2.2.1 to 2.2.3. In all models, we considered provenance and quantitative covariables as fixed effects, and the components of the

experimental layout (from yearly set to plot) as random effects. In all the models, the β , γ , and θ terms (see model [1] to [4]) are model parameters and ε denotes the residuals. We present the full models here. In some instances, they were simplified by removing some of the random effects, if the inclusion of these effects did not significantly decrease the AIC (Akaike information criterion) (Akaike 1973).

2.2.1 Models for traits with a continuous distribution

The use of linear models assumes the independence of residual terms. In the context of forest trees, this requires checking that there are no spatial and competition effects, particularly between neighbouring trees. We used the “breedR” package (Munoz and Sanchez 2018), which is particularly suitable for the analysis of forest genetic tests conducted in common gardens, to account for spatial and competition effects.

Within each test (Fig. 2 in Ducouso, Ehrenmann et al. 2022) and for each trait of interest for a given tree (Y_{ijklm}), we implemented the following linear model accounting for the different sources of variation:

$$Y_{ijklm} = \beta_0 + \beta_1 P_i + \beta_2 B_j + \beta_3 Pl_{ijk} + \sum_u X_{ijklm}^{(u)} + Spline_l + \varepsilon_{ijklm} \quad (1)$$

where P is the provenance, B is the microblock, Pl is the plot, X is the value of a trait (other than Y) used as an explanatory variable, ε is the residual effect, and $Spline$ is the component of the residual due to the position of the tree.

The main objective of the model [1] was to estimate the value of $Spline$, which accounts for local spatial effects related to the position of the tree. In all subsequent calculations, Y values were corrected by removing the corresponding $Spline$ estimate.

For potential competition effects, we used spatial statistics in BreedR to construct variograms to scale the variance of a trait according to the distance between trees (Additional file 1: Figure S3). Variograms displaying shifts over short distances indicate probable competition effects between neighbouring trees. In the case of HT4, HT10, and CIRC10, the variograms were rather flat for all tests, suggesting that there was no competition between the neighbouring trees in a plot (Additional file 1: Figure S3a). As a result, the following model was used to analyse the data:

$$Y'_{ijklmn} = \gamma_0 + \gamma_1 P_i + \gamma_2 Set_j + \gamma_3 S_k + \gamma_4 T_{jk} + \gamma_5 B_{jkl} + \gamma_6 Pl_{ijklm} + \varepsilon_{ijklmn} \quad (2)$$

where Y' is the phenotypic value (model [1]) for a tree corrected for the $Spline$ effect, P is the provenance, Set is the yearly set, S is the site, T is the test, B is the microblock, and Pl is the plot.

We further restricted the analysis to the largest trees in each plot (in most cases, between 8 to 12 trees) to limit the bias due to the variation of survival among plots.

For CIRC20, local competition was detected based on the shape of the variogram (Additional file 1: Figure S3b). We accounted for competition effects, by amending the model [2] in two ways: we first restricted the analysis to yearly girth increments from years 10 to 20 (aCIRC10-20), rather than using CIRC20, and we then explicitly introduced additional variables accounting for competition with neighbouring trees into the model (Perot, Goreaud et al. 2010): with the two neighbours on the same row, the two neighbours in the same column, and with the four neighbours in adjacent rows and columns. Thus, model [2] becomes:

$$Y'_{ijklmn} = W + \gamma_8 Cr_{ijklmn} + \gamma_9 Cl_{ijklmn} + \gamma_{10} Crl_{ijklmn} + \varepsilon_{ijklmn} \quad (3)$$

where W is the first six terms of model [2] and Cr is the mean basal area of the two adjacent trees in the same row, Cl is the mean basal area of the two adjacent trees in the same column, and $Cr l$ is the mean basal area of the four trees located in adjacent rows and columns. Y' is the annual increase in girth between years 10 and 20 (aCIRC10-20). Adjustment for the competition effect in this way flattened the variogram of aCIRC10-20 (data not shown).

2.2.2 Models for traits with discrete distributions

Traits following binomial and polynomial distributions could not be adjusted with the spatial correction derived from the splines function of BreedR. We did not, therefore, use model [1] to account for spatial effects for traits with discrete distributions. Instead, spatial effects were mostly accounted for by microblock effects.

2.2.2.1 Survival Survival at 10 years was modelled with a binomial logistic model similar to [2] but with Y replaced by $\text{logit}(\text{pr}(Y = 1))$, and the logit function of the probability that Y_{ijklmn} , the survival at 10 years of tree $_{ijklmn}$, is equal to 1 (i.e. the tree is alive). This approach was possible because tree mortality was not correlated with competition between neighbouring trees, so survival was independent for neighbouring trees. All the trees in the plots were used for survival analysis with the logit function.

2.2.2.2 Stem shape Stem shape is of the utmost importance in forestry. A preliminary analysis showed that tree dominance, assessed on the basis of height-girth covariation, had an effect on the shape of the stem. There was a clear trend towards better shape quality in dominant trees (Additional file 1: Figure S4). Stem shape, as

recorded in this study, was a factorial ordered variable. We therefore used a polytomous ordinal model to analyse its variation. This model resembled model [2] but also included adjustment for HT and CIRC:

$$\begin{aligned} \text{logit}(\text{pr}(Y_{ijklmn} \leq q)) = & \theta_q + \theta_1 P_i + \theta_2 \text{Set}_j \\ & + \theta_3 S_k + \theta_4 T_{jk} + \theta_5 B_{kl} \\ & + \theta_6 Pl_{ijklm} + \theta_8 HT_{ijklmn} \\ & + \theta_9 CIRC_{ijklmn} + \varepsilon_{ijklmn} \quad (4) \end{aligned}$$

where $\text{logit}(\text{pr}(Y_{ijklmn} \leq q))$ is the logit of the probability that Y_{ijklmn} (stem shape score at 10 or 20 years after plantation (SHP10, SHP20) of tree $_{ijklmn}$) is equal to or lower than score q . Higher $\text{logit}(\text{pr}(Y_{ijklmn} \leq q))$ values were considered to indicate a better tree shape. θ_q is the threshold parameter of the score q . All the other terms are the same as in the model [2].

All trees with recorded girth and height values were used for the analysis of SHP10. As stem shape at 20 years was recorded only at the Sillegny test site, for all sets, the analysis of SHP20 was restricted to this test site. Finally, as extreme stem shape scores (0 or 10) were extremely rare, their frequencies were bulked with the nearest scores.

2.2.2.3 Bud burst Bud burst date is an important adaptive trait in oaks, as it may be correlated with various components of fitness, such as flowering, or the length of the growing season. Bud burst score three years after plantation (DEB3) was also analysed as a factorial ordered variable. We therefore used the same model as in [4] (where Y_{ijklmn} was the DEB3 value for tree $_{ijklmn}$). We did not include the effects of height and girth in the model, and all trees were used for the analysis. Thus, in this model, higher $\text{logit}(\text{pr}(Y_{ijklmn} \leq q))$ values indicated an earlier bud burst.

2.2.3 Implementation of the models

Models [1] to [4] were implemented at two different levels of the whole analysis and according to the distribution of the trait analysed (Table 1).

The first level of analysis was an analysis of spatial effects (model [1]). This analysis was conducted separately for each of the 16 tests (Fig. 2 in Ducouso, Ehrenmann et al. 2022). For continuous variables, this made it possible to correct the raw data for spatial effects.

The second level of analysis was a global analysis including all the terms of models [2, 3], and [4]. This global analysis extracted provenance effects (γ_I and θ_I values), which were subsequently used for provenance classification.

For the sake of completeness, the provenance * test, provenance * site, provenance * yearly set interaction terms were also considered in models [2, 3], and [4].

2.2.4 Multivariate analysis and provenance classification

Provenance effects (γ_I and θ_I values) corresponding to the different traits were integrated into a principal component analysis (PCA) to investigate multidimensional variation within the total data set. The objective of the PCA was to extract composite variables (PC, principal components) accounting for the largest proportion of multitrait variation between provenances, and to prevent the overexpression of particular traits in the next classification step. PCA was, thus, used to screen variables for their contribution to the overall variation between provenances. The selected variables were used to calculate Euclidean distances between all pairs of provenances. The matrix of these Euclidean distances was then used for hierarchical cluster analysis (HCA). HCA was performed with the agglomerative clustering algorithm, which assembles provenances in a stepwise manner to form clusters on the basis of similarity, until all provenances form a single cluster. The clustering generated by HCA was optimized by K means analysis (Kassambara, 2017). This involved assigning K clusters of provenances and reallocating provenances to clusters such that the sum of distances of the provenances to the centroid of the cluster reached a minimum value.

3 Results

3.1 Mixed models and estimation of provenance values

The spatial analysis (model [1]) for continuous traits showed that the *Spline* effects absorbed most of the spatial variation. Hence, the microblock became a negligible factor in models [2] and [3] and was discarded in the analysis of provenance effects (Table 2). Given the competition effects (Additional file 1: Figure S3) and the influence of tree size on tree shape (Additional file 1: Figure S4), corrections were made for aCIRC10-20, SHP10 and SHP20, according to models [3] and [4]. The mixed models also made it possible to account for random effects of the different sources of variation due to

the experimental layout (Table 2). For dimensional continuous traits (HT and CIRC), the main sources of variation were test and plot, in addition to the residuals (mostly within-plot effects). For bud burst (DEB3), the dominant sources of variation were test and microblock, whereas for stem shape (SHP10), the dominant sources of variation were site and test. Large effects of yearly sets on survival were detected, suggesting a possible impact of yearly differences of stress events after plantation. Finally, the effects of interactions between provenance and experimental sources of variation (test, site, or set) were limited, and these effects were therefore discarded from the models for most discrete traits (Table 2).

For all traits, we implemented all the linear models ([2] to [4]) both with and without the introduction of provenance as a fixed effect. We compared the Akaike information criterion (AIC) values between the two implementations of each model. The inclusion of provenance effects systematically decreased the AIC (by more than 2 AIC units), whether considering *Q. petraea* and *Q. robur* provenances together or separately. In addition, an analysis of variance showed that provenance effects were highly significant for all traits (F test, $p < 0.001$). The variation between provenances is illustrated in Additional file 1: Table S2. Height growth (HT10) ranged from 458 cm (provenance Parc Saint Quentin) to 369 cm (provenance Vouillé). Earlier-flushing provenances were located in the South of France (Grésigne and Vachères) and late-flushing provenances were found in the North East of France (provenance Sturzelbronn, Saint Jean, and Haguenau1). Pedunculate oak provenances (Fontainebleau2, Compiègne1 and Mees) also flushed later. The ranking of provenances for stem shape varied slightly from years 10 to 20. The provenances for which stem shape was best at age 20 were Blois2, and La Haie Rénaut. Finally, SURV10 varied little between provenances, as survival exceeded 90% for all provenances.

Table 1 Trait definition and mixed models used to estimate provenance effects

Trait	Acronym	Model
Annual girth increment between 10 and 20 years	aCIRC10-20	[1] + [3]
Girth at 10 years	CIRC10	[1] + [2] continuous trait
Bud burst (score) at 3 years	DEB3	[4]
Total height at 4 years	HT4	[1] + [2] continuous trait
Total height at 10 years	HT10	[1] + [2] continuous trait
Total height at 20 years ^a	HT20	[1] + [2] continuous trait
Survival at 10 years	SURV10	[2] discrete trait
Stem shape at 10 years	SHP10	[4]
Stem shape at 20 years	SHP20	[4]

The time points indicated are the number of years after plantation

^aHT20 was only assessed in the Sillegny plantation and used for the trait-climate correlations

Table 2 Variances due to random effects in the mixed models. Variances are expressed relative to the total variance of all random terms

Source of variation	aCIRC10-20	CIRC10	DEB3	HT4	HT10	SURV10	SHP10	SHP20
Set	0.005 ^a	0.144 ^a	0.029 ^a	0.148 ^a	0.140 ^a	0.367 ^a	0.000	0.10 ^a
Site	0.054 ^a	0.108 ^a	0.080 ^a	0.018 ^a	0.041 ^a	0.180 ^a	0.313 ^a	NE
Test	0.116 ^a	0.104 ^a	0.434 ^a	0.350 ^a	0.239 ^a	0.037 ^a	0.434 ^a	NE
Microblock	NE	NE	0.224 ^a	NE	NE	0.135 ^a	0.107 ^a	0.69 ^a
Plot	0.129 ^a	0.141 ^a	0.196 ^a	0.127 ^a	0.162 ^a	0.282 ^a	0.122 ^a	0.21 ^a
Prov*Test	0.002	0.014 ^a	0.038 ^a	0.008 ^a	0.013 ^a	NE	0.024 ^a	NE
Prov*site	0.007 ^a	0.001 ^a	NE	0.003 ^a	0.001 ^a	NE	NE	NE
Prov*Set	0.002	0.014 ^a	NE	0.014 ^a	0.009 ^a	NE	NE	0.000
Residual	0.684	0.474	NA	0.332	0.395	NA	NA	NA

NE: effects not introduced into the model (see text)

NA: residual variances were not available for discrete traits

^aSource of variation contributes significantly to the variance according to the AIC criterion (the inclusion of the source of variation in the model reduces by more than 2 units the value of AIC in comparison to the complete model)

3.2 Correlation of provenance mean values and climatic variables

To account for the ecological drivers of provenance differentiation we computed Pearson correlations between provenance trait values and climatic data, using the data of the Sillegny site, that comprised the largest number of provenances, and where HT20 was recorded as well. Using the data of Sillegny allowed also to compute trait-climate correlations over the whole set of provenances including origins from other countries than France (Additional file 1: Table S1b), thus encompassing a larger range of climate variation. Calculations were done with two samples: the overall sample of provenances, and a restricted sample discarding provenances of Georgia and Ireland that distorted the range of climatic variables and traits (Table 3). We found significant clinal genetic variations between trait values and climatic data for growth and bud burst (Table 3); we illustrate some of the clines (Figs. 1, 2, and 3) for cases where correlations were significant for both samples. Growth traits (aCIRC10_20 and HT20) were positively correlated with temperature (Figs. 1 and 2) and, to a lesser extent, negatively with precipitation, resulting in higher growth for provenances originating from climates with stronger unbalance between heat and moisture (Table 3). The timing of apical bud burst (DEB3) was correlated with higher temperature (spring and annual temperature) and warmer growing season (SSD5, GSSD5), and with the mean diurnal range of temperature (MDR, Fig. 3): early flushing provenances originate from climates with large diurnal ranges. Survival and stem shape did not exhibit congruent correlation with climatic variables between the two samples, except for SURV10-MDR.

3.3 Clustering of the provenances

Before performing a clustering analysis, we screened the variables to prevent highly correlated traits from making inflated contributions to the overall between-provenance variation. PCA based on the provenance effects shown in Additional file 1: Table S2 was performed to check for correlation between the eight traits. The first three principal components accounted for more than 75% of the total variation between provenances (Additional file 1: Figure S5). Component 1 (40% of the variation) was related mostly to growth traits (HT10, HT4, and CIR10), which were also highly correlated with each other. Component 2 (20% of the variation) was correlated with stem shape (SHP10 and SHP20), whereas component 3 was related to bud burst (DEB3) and girth increment (aCIRC10-20). Finally, MORT10 displayed little variation between provenances. The correlation circles shown in Additional file 1: Figure S5 were used to screen the traits for the subsequent clustering analysis, which was limited to the following five traits — HT10, aCIRC10-20, SHP10, DEB3 and SHP20 — on the basis of the PCA analysis.

HCA assembled provenances into clusters according to their similarity based on the Euclidean distances calculated with the five selected traits. After several attempts with different numbers K of clusters, we set K to 13 for the K-means optimization procedure (clusters “a” to “m” in Additional file 1: Figure S6). An empirical examination of the 13 clusters was performed, in which cluster compositions were compared in terms of trait values. The HCA and K-means methods are based on Euclidean distance. We intended to “return” to the original trait values to provide biological support for the final clustering. A comparison of trait mean values and their confidence intervals within and between the 13 clusters (Additional file 1: Figure S7) and the assignments of provenance to the clusters by K-means

Table 3 Pearson correlation coefficients between provenance trait values and geographic and climatic variables (values indicated by superscript ^a are significant at $p < 0.01$); values indicated by superscript ^b are significant at $p < 0.05$)

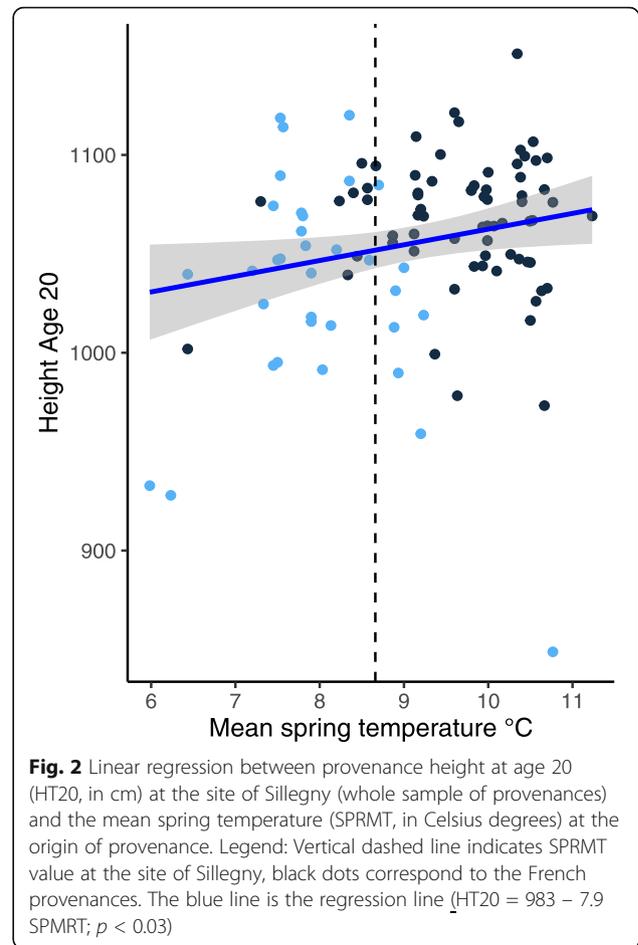
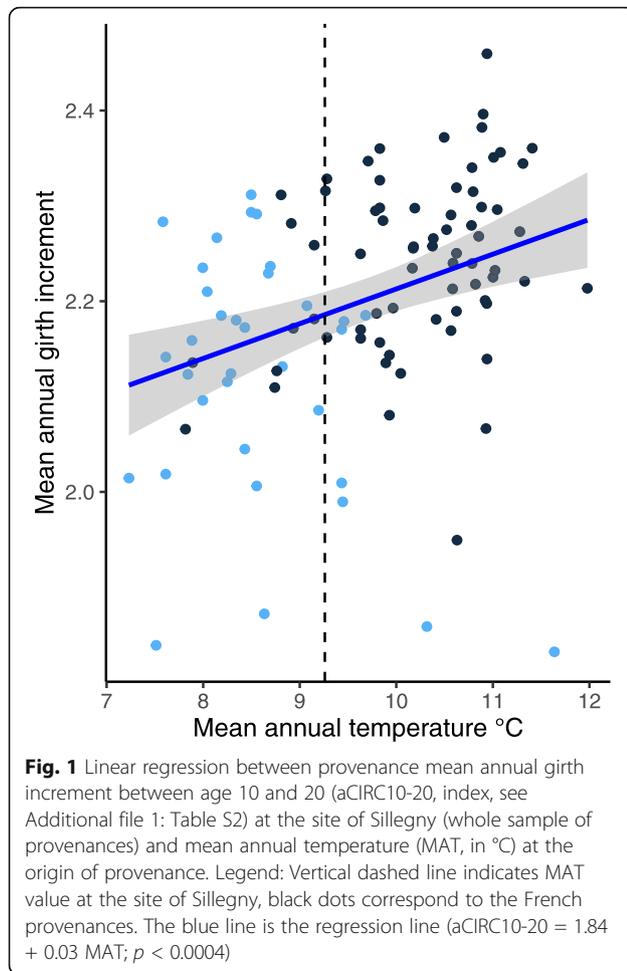
		aCIR10_20		DEB3		HT20		SURV10		SHP20	
		WS ^c	RS ^d	WS	RS	WS	RS	WS	RS	WS	RS
Geographic coordinates											
LONG	Longitude	-0.22	-0.27 ^a	0.16	0.11	-0.39 ^a	-0.30 ^a	-0.16	-0.05	0.37 ^b	-0.07
LAT	Latitude	-0.36 ^a	-0.35 ^a	-0.51 ^a	-0.50 ^a	-0.10	-0.13	0.29 ^a	0.24 ^b	0.28 ^a	0.14
ALT	Altitude	-0.05	-0.09	0.37 ^a	0.35 ^a	-0.12	-0.09	-0.23 ^b	-0.20	-0.21 ^b	-0.10
Temperature and heat related variables											
MAT	Mean annual temperature	0.35 ^a	0.43 ^a	0.16	0.16	0.15	0.24	-0.05	-0.03	-0.12	-0.05
SPRMT	Mean spring temperature	0.42 ^a	0.44 ^a	0.26 ^a	0.23 ^b	0.22 ^b	0.26 ^b	-0.09	-0.05	-0.16	-0.10
DD5	Year degree days >5°C	0.36 ^a	0.40 ^a	0.30 ^a	0.27 ^a	0.07	0.15	-0.13	-0.05	-0.29 ^a	-0.13
GSD5	Growing season degree days >5°C	0.35 ^a	0.36 ^a	0.36 ^a	0.31 ^a	0.05	0.13	-0.15	-0.05	-0.36 ^a	-0.18
MDR	Mean diurnal temperature range ^e	0.43 ^a	0.32 ^a	0.53 ^a	0.48 ^a	0.10	0.03	-0.38 ^a	-0.32 ^a	-0.29 ^a	-0.18
ISO	Isothermality ^f	0.29 ^a	0.43 ^a	0.01	0.09	0.25 ^a	0.27 ^a	-0.05	-0.15	0.18	0.06
SEA	Temperature seasonality ^g	-0.03	-0.26 ^b	0.29 ^a	0.19	-0.19 ^b	-0.27 ^a	-0.16	-0.01	-0.35 ^a	-0.18
Precipitation and moisture related traits											
MAP	Annual precipitation	-0.28 ^a	0.02	-0.14	0.10	-0.17	-0.02	0.08	-0.13	0.08	-0.05
GSP	Growing season precipitation	-0.33 ^a	0.18	0.00	0.11	-0.28 ^a	-0.16	0.02	-0.05	-0.20 ^b	-0.18
SPRP	Spring precipitation	-0.15	-0.03	0.21 ^b	0.28 ^a	-0.24 ^b	-0.11	-0.10	-0.12	-0.29 ^a	-0.17
SMRP	Summer precipitation	-0.42 ^a	-0.29 ^a	-0.22 ^b	0.09	-0.26 ^a	-0.18	0.14	0.04	-0.05	-0.15
PRATIO	GSP/MAP	-0.10	-0.27 ^a	0.11	0.01	-0.18	-0.21 ^b	-0.02	0.12	-0.32 ^a	-0.16
PSEA	Precipitation seasonality ^h	-0.46 ^a	-0.32 ^a	-0.05	0.05	-0.54 ^a	-0.40 ^a	-0.01	-0.03	-0.23 ^b	-0.10
Balance between heat and moisture											
AHM	Annual heat moisture index ⁱ	0.33 ^a	0.17	0.13	-0.03	0.18	0.11	-0.04	0.11	-0.11	-0.01
ADI	Annual dryness index ^j	0.32 ^a	0.15	0.19	0.04	0.13	0.07	-0.06	0.10	-0.18	-0.04
GSDI	Growing season dryness index ^k	0.42 ^a	0.29 ^a	0.15	0.04	0.27 ^a	0.19	-0.09	-0.00	0.03	0.08

^cWhole sample of provenances^dRestricted sample (Provenances from Georgia and Ireland omitted)^eMean of monthly (maximum temperature–minimum temperature)^fMean diurnal range/ annual temperature range (in %)^gStandard deviation of monthly temperatures (x100)^hCoefficient of variation of monthly precipitationsⁱ(MAT+10)/(MAP/1000)^jSqrt(DD5)/MAP^kSqrt(GSD5)/GSP

optimization (Additional file 1: Figure S6) led to the merging of three of the original clusters (a, b and c) into one larger cluster (A). The provenance composition of the remaining clusters was conserved (d to m, converted into D to M). End-stacked branches (clusters) of the dendrogram were then grouped by minimizing the differences between the provenances belonging to the clusters (Fig. 4).

We facilitated biological interpretation of the clusters by plotting the provenance trait values nested within each cluster (Fig. 5). Between-cluster variation was greater for DEB3 and HT10 than for the other three traits. Using these violin diagrams, we focused our attention on clusters A, E, F, G, and H, which contained the provenances with the highest values of aCIR10-20,

SHP10, and SHP20, which might be of primary importance for the choice of seed sources. The ranking of the clusters, for these three traits, decreased in a continuous manner from A to H, excluding D. However, there were noticeable differences for the other two traits (HT10 and DEB3): cluster E displayed greater height growth (HT10), G had a lower level of height growth, F flushed earlier, and H flushed later (DEB3). Clusters L and M consisted of provenances with the highest scores for shape at 10 years, but with poor growth (HT10 and aCIR10-20). Clusters K and J had small circumference increments, late budburst and poor shapes, despite having above-average height growth. The provenances belonging to these clusters were *Quercus robur* provenances, with the exception of Parc Saint Quentin.



Provenances of cluster I had lower levels of height growth and average values for the other traits. Cluster D consisted of a single provenance, Vachères, which was separated from all other provenances due to its extreme trait values (very early budburst, very low circumference increment and height, average shape).

Finally, the various clusters did not correspond to clear-cut geographic groups (Fig. 6). Geographic proximities were observed for clusters A, F, and H, but the other clusters group together provenances with very diverse geographic origins (E, G, and L).

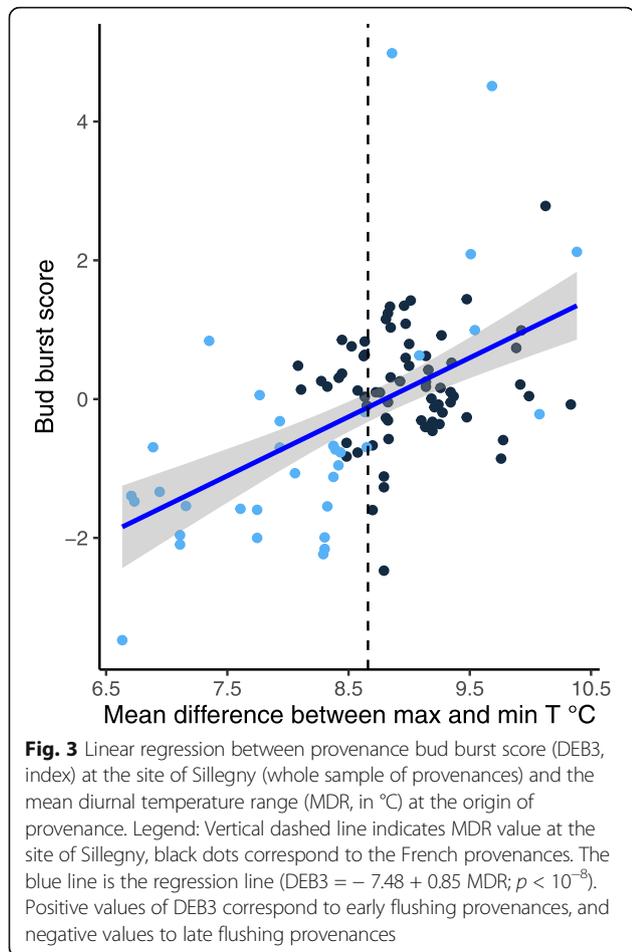
3.4 Patterns of variation between provenances

We examined in more detail whether other sources of provenance variation should be accounted for in seed sourcing. Three sources of variation were examined:

- Variation between provenances of the same forest
- Variation within and between provenance regions
- Variation within and between maternal lineages

Twinned provenances are collected from within the same forest, and are separated by no more than a few kilometres at maximum. Comparisons of twinned provenances can be used to explore the spatial extent of the provenance effect, which is of important practical value for seed collection. Among the eleven pairs of twinned provenances studied here, we found very low levels of phenotypic trait variation between twinned provenances in general, with only a few exceptions (Compiègne, Fontainebleau, and Bride) (Additional file 1: Figure S8). For Compiègne, one population came from a pedunculate oak stand, whereas the other came from a sessile oak stand. The reasons for the observed differences between twinned provenances at Fontainebleau and Bride remain unclear, as the ecological features of the different compartments in which the seeds were collected are very similar (substrate, soil depth, altitude, etc., data not shown). Some of the twinned provenances were probably planted and originated from different source materials (see Section 4).

The 70 *Q. petraea* provenances came from 17 different provenance regions (Additional file 1: Table

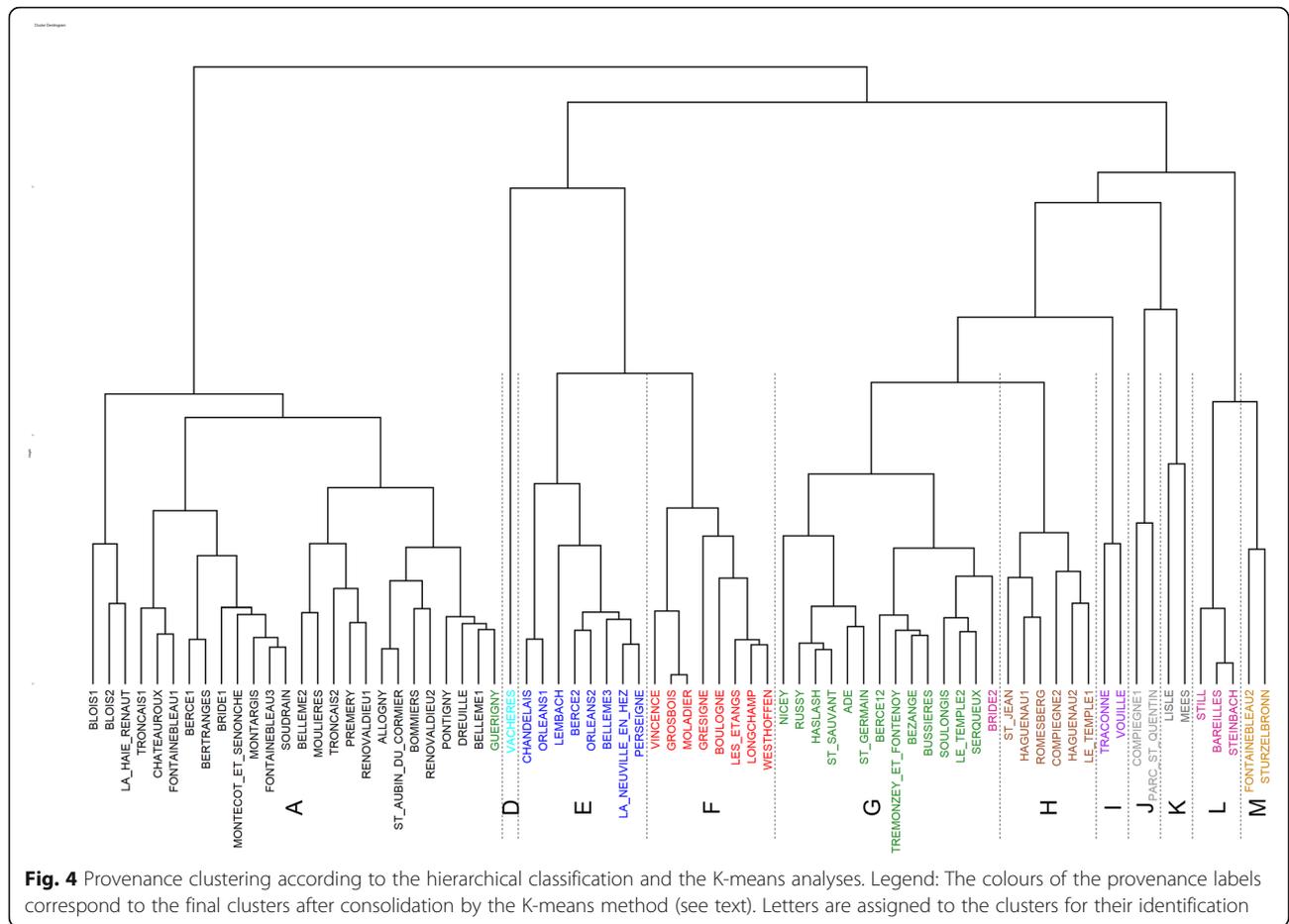


S1a). We investigated whether there were significant differences between provenance regions. Significant differences were found only for growth (aCIR10_20), between provenance regions QPE104 and 106 (highest growth) and QPE204 and 212 (lowest growth) (Additional file 1: Table S3 and Figure S9). For two traits (SHP10 and DEB3), more variation was observed between provenances within the same provenance region than between provenance regions.

We also explored whether there were differences between provenances bearing different chloroplast haplotypes. This analysis was restricted to provenances in which all trees shared the same haplotype. Polymorphic haplotypic provenances were discarded from the analysis. Haplotypes of the three main maternal lineages (Atlantic, Balkan and Italian) were represented in the provenance sample, making it possible to compare haplotype groups. The mean trait values of provenances did not differ significantly between haplotype groups (Additional file 1: Figure S10 and Table S4).

4 Discussion

We performed a multitrait and multisite analysis of genetic variation between provenances of *Q. petraea* in France, with the main aim of providing recommendations for seed sourcing. Our investigations confirmed genetic differentiation for adaptive traits highlighting clear genetic clines triggered by climatic gradients across the distribution of the species. This was the case for the time of bud burst which follows a temperature cline confirming earlier reports along either latitudinal (Ducouso, Guyon et al. 1996) or altitudinal gradients (Vitasse, Delzon et al. 2009b). Our results are the first to report stronger genetic positive clines of growth with temperature than with precipitation in *Quercus petraea* (Table 3). Both phenology and growth clines are likely the result of directional diversifying selection along temperature gradients since the establishment of the populations. Our study was performed after a European wide investigation as part of the same experiment (but at multiple testing sites) concluded that the species displayed substantial resilience to environmental changes, thanks to the genetic and plastic responses of adaptive traits (Sáenz-Romero, Lamy et al. 2017; Sáenz-Romero, Kremer et al. 2019). For seed sourcing purposes, we concentrated on a narrower geographic range, but focused on traits of economic and ecological value considered important for plantation. The assessments were performed at a young age for such a long-lived species, but the selected traits are highly relevant from an ecological standpoint for management purposes. Growth (height and diameter) after 10 years provides a reliable evaluation of the capacity of trees to survive the critical juvenile stage in which they are exposed to competition from weeds and game damage. Stem shape at 20 years is of predictive value for subsequent stem quality and stiffness; large stocks of juvenile trees with a good shape ultimately resulting in tall, high-quality forest stands result in high forest stand of improved quality (Jarret, 2004). Finally, assessments of the date of bud burst, which displays strong genetic stability across ages (Vitasse, Delzon et al. 2009a), should help to prevent exposures to late frosts, spring drought or leaf pathogens, by making it possible to discard trees that flush extremely early or extremely late (Deans and Harvey 1996; Dantec, Ducasse et al. 2015). We found significant differences between provenances for all traits, and we grouped the provenances into 11 clusters of potential utility for seed sourcing, using multivariate approaches. Our results open up two important areas of discussion. We will first consider cluster composition with respect to

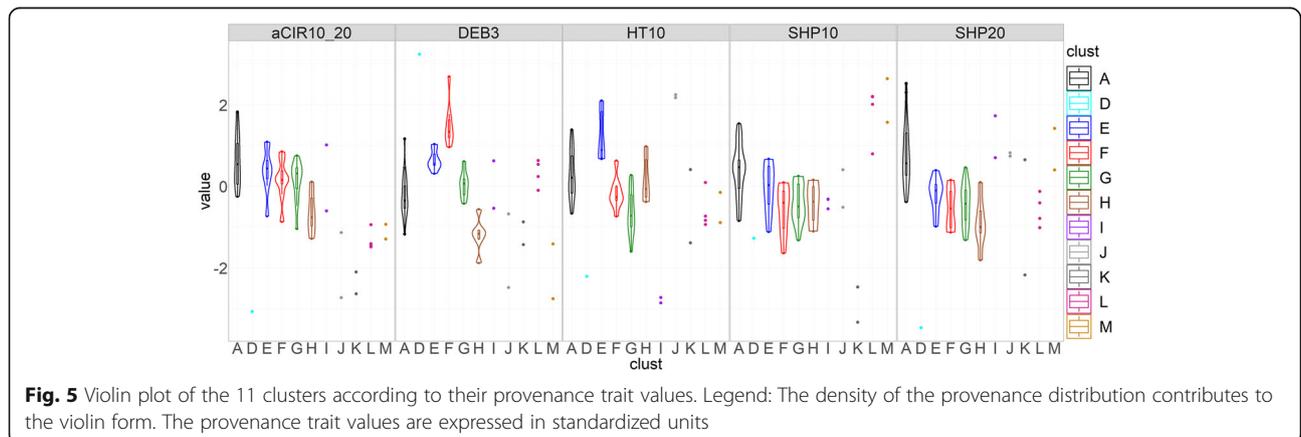


other classifications of stands and silvicultural practices. We will then discuss targeted recommendations for seed sourcing that can be inferred from our results.

4.1 Provenance clustering

Our analysis of twinned provenances clarified the spatial envelope of a “provenance”. We found that there was

less variation between seed lots collected from the same forest than between seed lots collected from different forests. However, this conclusion may be valid only for forests in which all the compartments were regenerated by natural means. The prior historical introduction of foreign material, which may not always be known, may affect this conclusion, due to genetic differentiation of the introduced material from the local population or



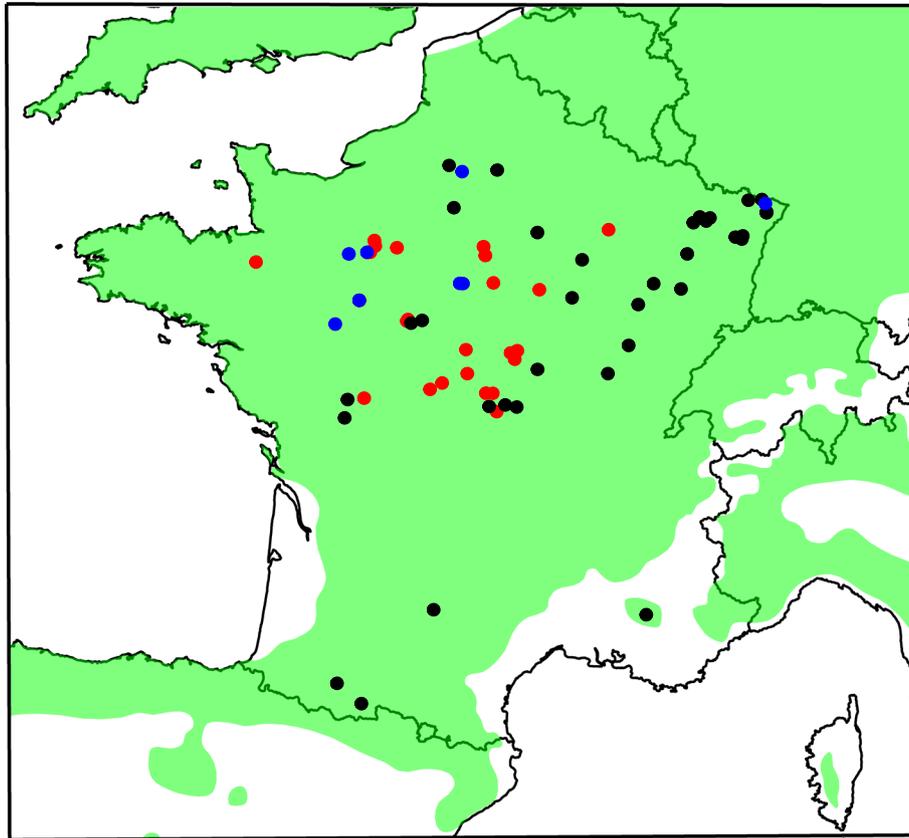


Fig. 6 Geographic distribution of the provenances of the different clusters. Legend: Red dots correspond to provenances belonging to cluster A. Blue dots correspond to provenances belonging to cluster E, and black dots correspond to provenances belonging to any of the remaining clusters. Green area corresponds to the distribution of *Quercus petraea* according to Caudullo, Welk et al.(2017)

genetic pollution resulting from interbreeding between introduced and local stands (Geburek and Myking 2018). This was probably the situation at Fontainebleau and Bride, for which we found differences between seed lots from the same forest. Such conclusions have important practical consequences for seed collection within a so-called “seed stand” within a given forest, as seed yields may be heterogeneous within a given forest. However, we draw attention to the temporal genetic differentiation between old and younger stands of sessile oak, due to the different selection pressures operating between the late Little Ice Age (before 1850) and Anthropocene warming (post-1850), suggesting to preferentially harvest seed in stands born after the Little Ice Age (Saleh, Chen et al. 2022). Twinned “temporal” provenances from the same forest were not sampled, a priori, during the collection of provenances, and it was not possible to investigate temporal differentiation in this study.

Overall, the cluster composition of provenances poorly matched the provenance regions identified on ecological grounds. Finally, the clusters (especially clusters A, F, and H) matched the geographic distribution to some

extent (Fig. 6), despite some provenances from the same geographic area being assigned to different clusters. An illustration of this situation is provided by provenances from North East France (Bride1, Lembach, Westhoffen, Haslach, Haguenu1, Steinbach, Sturzelbronn). These seven provenances were assigned to seven different clusters, despite their geographic proximity (they are less than 80 km apart). One probable reason for their assignment to different clusters may be the more frequent use of plantation approaches to renew oak stands in North East France, due to more irregular seed yields in this region (Verger and Ducouso 2009). The German forestry administration implemented intensive oak afforestation, by direct seeding or plantation, in Alsace in the late nineteenth century, to repair the considerable damage to these forests incurred during the 30 years’ war (Dion 1970; Degron and Husson 1999). However, the geographic origin of the material introduced during this period is unknown, and genetic marker analysis revealed no strong differentiation, suggesting probable transfers over short distances in the Rhine Valley (Neophytou, Gartner et al. 2015). For the clusters matching

geographic distribution, climatic and ecological proximity may be the source of similarity. However, silvicultural practices may also have contributed to clustering, particularly for clusters A and E. Most of the provenances of clusters A and E are located in the Seine, Loire and Allier valleys (RénoValdieu, Bellème, Bercé, Blois, Fontainebleau, Montargis, Bertranges, Tronçais), where high-forest silvicultural treatment is a long-standing tradition (Ducouso, Louvet et al. 2004; Valadon, 2009), at least since Colbert's royal "ordonnance" in 1669. We recently showed that high-forest practices generate genetic changes favouring faster-growing trees with straighter stems, even in a single generation (Alexandre, Truffaut et al. 2019; Alexandre, Truffaut et al. 2020). Conversely, forests in which silvicultural treatments have switched, at different times between coppicing under forest and high forest with standard trees may have been assigned to different clusters in our study.

4.2 Clustering for seed sourcing

Given the traits assessed, clusters A and E contain candidate provenances for seed sourcing in operational forestry. Provenances from these two clusters display high levels of juvenile growth and are, thus, better able to resist competition and to overcome game damage, and to achieve an acceptable stem shape. The main difference between the provenances of clusters A and E is the slightly earlier flushing and faster early growth of cluster E (Fig. 6).

Clusters A and E are part of the continuum of clusters A-E-F-G-H displaying only very small differences in growth (aCIR10_20) and shape (SHP10 and SHP20). However, these traits may vary considerably between extreme provenances within a cluster, even if these differences are not statistically significant. In addition, the provenances of clusters F-G-H differ from those of clusters A and E for other traits. The provenances of cluster F flush much earlier, potentially exposing them to more frequent damage due to biotic or abiotic agents. Conversely, the provenances of cluster H flush later and are more like *Q. robur* provenances, raising questions about possible genetic admixture in provenances belonging to this cluster. Provenances from the remaining clusters had extreme values for some traits, raising concerns about their use as seed sources. The provenances of cluster I have slow initial growth. The provenances of clusters L and M display poor growth after 10 years and provenances from cluster M flush very late.

Clusters J and K contain very few provenances, and most are *Q. robur* provenances. Finally, cluster D contains only one provenance (Vachères), located at the southern margin of the species distribution in France and with extreme trait values (very early flushing, poor growth, and poor stem shape). This analysis

of the clusters obtained leads us to recommend the use of provenances from clusters A and E as potential seed sources. Provenances from clusters F, G, and H may be alternative candidates, provided that earlier or later flushing is not detrimental at the intended plantation site. The inclusion of multiples provenances within a cluster should also help to secure seed procurement regardless of heterogeneity in seed yields. Overall, our results confirm and reinforce earlier predictions based on juvenile traits in the same experimental plantations (Ducouso, Louvet et al. 2004).

Where could these recommendations be implemented? Our results were obtained at test sites within the central zone of sessile oak cultivation in France, between the Loire and Rhine rivers and encompassing areas of maritime to semi-continental climates. Each of the recommended clusters includes provenances with broad west-to-east (Saint Aubin du Cormier, Lembach), and north-to-south (Perseigne, Moulières) distributions. Previous genetic surveys conducted with protein markers (Zanetto and Kremer 1995; Le Corre, Roussel et al. 1998) or whole-genome sequences (Leroy, Louvet et al. 2020) showed that genetic differentiation within this rather large geographic unit is weak, providing additional genetic support for clusters A and E. In addition, our results also indicate that the effects of interactions between provenance and other experimental sources of variation (site, set, test) were weaker than provenance effects (Table 2). Based on this reasoning, we conclude that provenances from clusters A and E (or alternatively F, G, and H) are suitable for use throughout most of the territory of France, excluding the southern margins of the species distribution. We would also recommend the mixing of different sources from the same clusters for seedling procurement for plantations. The use of mixtures within clusters would improve the tolerance of ongoing climate change, by exposing trees of different origins to natural selection. Other currently proposed strategies are based on a climate-oriented approach for seed sourcing based on species distribution models, and thus recommend targeted assisted migration. Such reasoning, based exclusively on climatic data, ignores biotic interactions and may provide biased estimates of the niche of the species, ultimately leading to the misdirection of transfers (Boiffin, Badeau et al. 2017). The mixing of seeds from different provenances should be seen here as an alternative approach, enhancing adaptability to future climatic conditions and maintaining diversity.

5 Conclusion

We propose a collection of 34 provenances belonging to two clusters for seed sourcing, for the central part of the distribution range of *Q. petraea* in France, and we recommend the mixing of seeds from different

provenances. We acknowledge that our conclusions are based on assessments performed in current climate conditions, but we anticipate their applicability during the next few decades, given the overall resilient response of sessile oak populations observed over a larger climatic range of provenance tests (Sáenz-Romero, Lamy et al. 2017; Sáenz-Romero, Kremer et al. 2019). Our recommendations should reinforce resilience and increase diversity, based on the provenance composition of the clusters. The future amendment of these recommendations based on ongoing assessments in established provenance tests remains possible.

6 Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13595-022-01140-0>.

Additional file 1: Table S1a. Origin and passport data of the French *Q. petraea* provenances. **Table S1b.** Additional provenances used for the trait-climate correlation analysis in site Sillegny. **Table S2.** Estimates of provenance values derived from the linear models. **Table S3.** Results of the ANOVAs comparing trait values among provenance regions. **Table S4.** Results of the ANOVAs comparing provenance traits values among cytotype groups. **Figure S1.** Geographic distribution of the provenances and provenance tests. Black dots indicate the location of the provenance tests (from West to East: La Petite Charrie, Vierzon, Vincence and Sillegny). Red dots indicate the origin of the *Q. petraea* provenances. Blue dots indicate the origin of *Q. robur* provenances. Green area corresponds to the distribution of *Quercus petraea* according to Caudullo, Welk et al. (2017). **Figure S2.** Stem shape scoring procedure used for the assessments of stem shape at ages 10 and 20. **Figure S3.** Variograms for various continuous traits (set 1 at the Petite Charrie site). The empirical isotropic variogram is constructed on the basis of the variance of trait differences (deviance) between trees separated by distance h . All pairs of trees are considered, regardless of the direction of their distance. The deviance is calculated as follows:

$$\gamma h = 12VZu - Zvasafunctionofh$$
a Variogram of girth at 10 years (CIRC10). The y axis is in mm^2 . **b** Variogram of girth at 20 years (CIRC20). The y axis is in cm^2 . **Figure S4.** Relationship between the tree shape, height (HT10) and girth (CIRC10). This graph illustrates how the scoring of tree shape is influenced by the dimensions of the tree: higher trees and trees with wider girths tend to have higher tree shape scores. The graph is a biplot of single-tree phenotypic values of HT10 (in cm), and CIRC10 (in mm) recorded for set 1 at the Sillegny site (Test 4) and labelled according to SHP10. Similar observations were made in other tests and also for SHP20. **Figure S5.** Correlation circles for the eight traits according to the first three components of the principal component analysis. The directions of the traits were plotted pairwise along the first three principal components (1:2, 1:3, 2:3). Traits oriented in the same direction are highly correlated. Principal component values are expressed in standardized units. **Figure S6.** Dendrogram of the 74 provenances resulting in the hierarchical clustering analysis (HCA). Lower case letters label the 13 clusters resulting in the HCA analysis. **Figure S7.** Profile of the provenance means and their confidence intervals between the 13 clusters. Trait values (y axis) are expressed in standardized units. **Figure S8.** Comparison of twinned provenances. This graph represents the distribution of the twinned provenances of each forest according to the two first principal components of the PCA (Dim1 and Dim 2 see text). The red dots correspond to the twinned provenances linked to their clusters. Upper case letters indicate the final cluster labels. Dim1 and Dim2 are expressed in standardized units. **Figure S9.** Comparison of provenances trait values according to the region of provenances. Mean values (black dots) and range of confidence intervals (bars) are plotted for the different provenances, which are clustered according to the region of provenances (first line on the top of the figure). Trait values are expressed in standardized units. **Figure S10.** Comparison of provenances trait values according to their chloroplast haplotype. Mean values (black dots) and range of confidence intervals (bars) are plotted for the different provenances, which are clustered

according to their chloroplast haplotype (first line on the top of the figure). Haplotypes 10, 11, and 12 belong to the Atlantic lineage (lineage B), haplotype 7 belongs to the Balkanic lineage (lineage A), haplotype 1 belongs to the Italian lineage (lineage C), and haplotype 21 to the South Iberian lineage (lineage D). Trait values are expressed in standardized units.

Acknowledgements

We thank the successive executive officers of INRAE and ONF for their continuous support for this long term experiment, especially Michel Buffet of ONF who was instrumental for the launching of the sessile oak provenance research in the late eighties. We thank the staff of each local ONF unit hosting a plantation, the personnel of the ONF Research and development units of Dole, Nancy and Orléans, the technicians of BIOGECO, and the personnel of the INRAE Experimental Units of Nancy (JEFL, Unite Experimentale Forestiere de Lorraine), Orléans (Unité expérimentale de Génétique et de Biomasse Forestière), Bordeaux (Unité Expérimentale Forêt Pierroton), for their assistance during the installation and management of the plantation and for the contributions to the field assessments of the trees.

The provenance tests are part of the INRAE experimental network of forest tree common gardens, GEN4X and of INRAE-ONF White Oak Genetic Resources Network. These networks belong to the research infrastructure IN-SYLVA France (https://www6.inrae.fr/in-sylva-france_eng/Services/In-Situ/GEN4X-Network). The GEN4X network's metadata are archived at: <https://doi.org/10.15454/50RS8C>.

Authors' contributions

Conceptualization: Antoine Kremer and Alexis Ducouso. Coordination and monitoring of data: Alexis Ducouso, Jean Marc Louvet, Patrick Reynet, and Brigitte Musch. Data analysis: Quentin Girard and Clothilde Boubée de Gramont. Writing of the paper: Quentin Girard and Antoine Kremer. Review and editing: All authors. The authors read and approved the final manuscript.

Funding

The *Quercus petraea* provenance test network was supported since its installation by a joint cooperation agreement between INRAE and ONF (Office National des Forêts). We are grateful to the European Union for their financial contribution at the installation of the experiment (Project MA2B/0022 and FAIROAK (FAIR1 PL95-0297)).

Availability of data and materials

The datasets generated and/or analyzed during the current study are described in a companion Data Paper, also published in *Annals of Forest Science* (2022), and are available in the Data INRAE repository, <https://doi.org/10.15454/838U9L>.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare no conflict of interest.

Author details

¹UMR BIOFORA, INRAE, ONF, 2163 Avenue de la pomme de pin, 45075 Orléans, France. ²UMR BIOGECO, INRAE, Université de Bordeaux, 69 Route d'Arcachon, 33612 Cestas, France.

Received: 6 July 2021 Accepted: 1 April 2022

Published online: 16 June 2022

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. Petrovand F. Caski, eds., Akademiai Kiado, Budapest. *Proceed Sec Int Sympos Inform Theory* 1973:267–281
- Alberto F, Bouffier L, Louvet JM, Lamy JB, Delzon S, Kremer A (2011) Adaptive responses for seed and leaf phenology in natural populations of sessile oak

- along an altitudinal gradient. *J Evol Biol* 24(7):1442–1454. <https://doi.org/10.1111/j.1420-9101.2011.02277.x>
- Alexandre H, Truffaut L, Ducouso A, Louvet JM, Nepveu G, Torres-Ruiz JM, Lagane F, Firmat C, Musch B, Delzon S, Kremer A (2019) In situ estimation of genetic variation of functional and ecological traits in *Q. petraea* and *Q. robur*. *Tree Genet Genomes* 16:32
- Alexandre H, Truffaut L, Klein E, Ducouso A, Chancerel E, Lesur I, Dencausse B, Louvet JM, Nepveu G, Torres-Ruiz JM, Lagane F, Musch B, Delzon S, Kremer A (2020) How does contemporary selection shape oak phenotypes? *Evol Appl* 13(10):2772–2790. <https://doi.org/10.1111/eva.13082>
- Anonymous (2017) Résultats de l'enquête statistique annuelle MAAF/IRSTEA sur les ventes en France de plants forestiers pour la campagne de plantation 2015-2016. In: 2017-229 NdSD. Ministère de l'Agriculture, p 32. <https://info.agriculture.gouv.fr/gedei/site/bo-agri/instruction-2017-229/telechargement>
- Bert D, Lebourgeois F, Ponton S, Musch B, Ducouso A (2020) Which oak provenances for the 22nd century in Western Europe? *Dendroclimatology in common gardens*. *Plos One* 15:6. <https://doi.org/10.1371/journal.pone.0234583>
- Boiffin J, Badeau V, Breda N (2017) Species distribution models may misdirect assisted migration: insights from the introduction of Douglas-fir to Europe. *Ecol Appl* 27(2):446–457. <https://doi.org/10.1002/eap.1448>
- Burianek V, Benedikova M, Kylesakova J (2011) Evaluation of twenty years old pedunculate and sessile oak provenance trial. *J Forest Sci* 57(4):153–169. <https://doi.org/10.17221/117/2010-JFS>
- Caignard T, Kremer A, Firmat C, Nicolas M, Venner S, Delzon S (2017) Increasing spring temperatures favor oak seed production in temperate areas. *Sci Rep* 7(1):8555. <https://doi.org/10.1038/s41598-017-09172-7>
- Camus A (1938) Les chênes. Monographie du genre *Quercus*. Tome II. Genre *Quercus*. Sous-genre *Euquercus* (Section *Lepidobalanus* et *Macrobalanus*), Editions Paul Lechevalier, Paris
- Caudullo G, Welk E, San-Miguel-Ayanz J (2017) Chorological maps for the main European woody species. *Data Brief* 12:662–666. <https://doi.org/10.1016/j.dib.2017.05.007>
- Collin E, Ducouso A, Ricodeau N (2017). Fiche de conseils d'utilisation des MFR de chêne sessile par régions de provenance. Version du 12/07/2017. <https://agriculture.gouv.fr/graines-et-plants-forestiers-conseils-dutilisation-des-provenances-et-varietes-forestieres>.
- Dantec CF, Ducasse H, Capdevielle X, Fabreguettes O, Delzon S, Desprez-Loustau ML (2015) Escape of spring frost and disease through phenological variations in oak populations along elevation gradients. *J Ecol* 103(4):1044–1056. <https://doi.org/10.1111/1365-2745.12403>
- Deans JD, Harvey FJ (1996) Frost hardness of 16 European provenances of Sessile oak growing in Scotland. *Forestry* 69(1):5–11. <https://doi.org/10.1093/forestry/69.1.5>
- Degron R, Husson JP (1999) Les forêts d'Alsace-Lorraine (1871-1918): les contrastes de l'héritage allemand, reflets de l'annexion. Histoire, économie et société 1999(18):593–611
- Dion J (1970) Les forêts de la France du Nord-Est. *Revue géographique de l'Est* 10(3):155–277. <https://doi.org/10.3406/rgeest.1970.1170>
- Ducouso A, Ehrenmann F, Girard Q, Lamy JB, Louvet JM, Reynet P, Musch B, Kremer A (2022) Long-term and large-scale *Quercus petraea* population survey conducted in provenance tests installed in France. [Data paper]. *Annals of Forest Science* (Accepted). <https://doi.org/10.1186/s13595-022-01141-z>
- Ducouso A, Guyon JP, Kremer A (1996) Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl.). *Annales Des Sciences Forestieres* 53(2-3):775–782. <https://doi.org/10.1051/forest:19960253>
- Ducouso A, Louvet JM, Faucher M, Legroux P, Jarret P, Kremer A (2004) Révision des régions de provenances et peuplements sélectionnés: le cas des chênes sessile et pédonculé. *Rendez-vous techniques ONF Hors Série N°1*:33-42
- Ducouso A, Louvet JM, Jarret P, Kremer A (2005) Geographic variations of sessile oak in french provenance tests. In: Rogers R, Ducouso A, Kanazashi A (eds) Proceedings of the Joint Meeting of IUFRO Working Group "Genetics of oaks and Improvement and silviculture of oaks". *Forestry and Forest Products Research Institute (FFPRI), Tsukuba*, pp 128–138
- Eaton E, Caudullo G, Oliveira S, de Rigo D (2016) *Quercus robur* and *Quercus petraea* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A (eds) *European atlas of forest trees*. Publication Office of the European Union, Luxembourg, pp 160–163
- Geburek T, Myking T (2018) Evolutionary consequences of historic anthropogenic impacts on forest trees in Europe. *For Ecol Manage* 422:23–32. <https://doi.org/10.1016/j.foreco.2018.03.055>
- Hanewinkel M, Cullmann DA, Schelhaas MJ, Nabuurs GJ, Zimmermann NE (2013) Climate change may cause severe loss in the economic value of European forest land. *Nature Climate Change* 3(3):203–216. <https://doi.org/10.1038/nclimate1687>
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated surfaces for global land areas. *Int J Climatol* 25(15):1965–1978. <https://doi.org/10.1002/joc.1276>
- Jarret P (2004) Chêne atlantique, Lavoisier
- Jensen JS, Hansen JK (2008) Geographical variation in phenology of *Quercus petraea* (Matt) Liebl and *Quercus robur* L oak grown in a greenhouse. *Scand J Forest Res* 23:179–188. <https://doi.org/10.1080/02827580801995331>
- Kassambara A (2017) Practical guide to cluster analysis in R. Published by STHDA (Statistical Tools For High-Throughput Data Analysis <http://www.sthda.com>)
- Kleinschmit J (1993) Intraspecific variation of growth and adaptive traits in European oak species. *Annales Des Sciences Forestieres* 50(Supplement): S166–S185. <https://doi.org/10.1051/forest:19930716>
- Kleinschmit J, Svobla J (1994) Intraspecific variation of growth and stem form in *Quercus robur* and *Quercus petraea*. In: Commission E (ed) Inter- and intra-specific variation in European oaks: evolutionary applications and practical consequences. European Commission, Brussels, pp 217–238
- Kremer A, Ducouso A (2017). The french series of *Quercus petraea* provenance tests. http://provenances.pierroton.inra.fr/provenance_tests/protocols?page=protocols
- Kremer A, Kleinschmit J, Cottrell J, Cundall EP, Deans JD, Ducouso A, König AO, Lowe AJ, Munro RC, Petit RJ, Stephan BR (2002) Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *For Ecol Manage* 156(1-3):75–87. [https://doi.org/10.1016/S0378-1127\(01\)00635-1](https://doi.org/10.1016/S0378-1127(01)00635-1)
- Kremer A, Petit RJ, Ducouso A (1998) Structure of gene diversity, geneflow, and gene conservation in *Quercus petraea*. In: Turok J, Kremer A, De Vries S (eds) First Eurofor Meeting on Social Broadleaves. IPGRI, Rome, pp 133–145
- Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity* 80(4):464–473. <https://doi.org/10.1038/sj.hdy.6883130>
- Le Hardy de Beaulieu A, Lamant T (2006) Guide illustré des chênes. Tome 1, Editions du 8ième, Paris
- Leroy T, Louvet JM, Lalanne C, Le Provost G, Labadie K, Aury JM, Delzon S, Plomion C, Kremer A (2020) Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytol* 226:1171–1182
- Liepe K (1993) Growth chamber trial on frost hardness and field trial on flushing of sessile oak (*Quercus petraea* Liebl.). *Ann Forest Sci* 50(Supplement):208–214. <https://doi.org/10.1051/forest:19930719>
- Lieseback M, Stephan BR (2000) Development of provenances of *Quercus petraea* and *Quercus robur* from acorns to six year old plants in relation to species specific traits. *Glasnik za Sumske Pokuse* 37:413–423
- Lobo A, Torres-Ruiz JM, Burlett R, Lemaire C, Parise C, Francioni C, Truffaut L, Tomaskova I, Hansen JK, Kjaer ED, Kremer A, Delzon S (2018) Assessing inter- and intraspecific variability of xylem vulnerability to embolism in oaks. *For Ecol Manage* 424:53–61. <https://doi.org/10.1016/j.foreco.2018.04.031>
- Matyas C (2021) Adaptive pattern of phenotypic plasticity and inherent growth reveal the potential for assisted transfer in sessile oak (*Quercus petraea*). *For Ecol Manage* 482:118832. <https://doi.org/10.1016/j.foreco.2020.118832>
- Mette T, Dolos K, Meinardus C, Brauning A, Reineking B, Blaschke M, Pretzsch H, Beierkuhnlein C, Gohlke A, Wellstein C (2013) Climatic turning point for beech and oak under climate change in Central Europe. *Ecosphere* 4(12): art145. <https://doi.org/10.1890/es13-00115.1>
- Munoz F, L Sanchez (2018) breedR: Statistical methods for Forest Genetic Resources Analysis. R package version 0.12-5. <https://github.com/famuvie/breedR>
- Neophytou C, Gartner SM, Vargas-Gaete R, Michiels HG (2015) Genetic variation of Central European oaks: shaped by evolutionary factors and human intervention? *Tree Genet Genomes* 11. <https://doi.org/10.1007/s11295-015-0905-7>
- Perot T, Goureaud F, Ginisty C, Dhote JF (2010) A model bridging distance-dependent and distance-independent tree models to simulate the growth of mixed forests. *Ann Forest Sci* 67. <https://doi.org/10.1051/forest/2010004>

- Petit RJ, UM Csaikl, S Bordacs, K Burg, E Coart, J Cottrell, B van Dam, JD Deans, S Dumolin-Lapegue, S Fineschi, R Finkeldey, A Gillies, I Glaz, PG Goicoechea, JS Jensen, AO König, AJ Lowe, SF Madsen, G Matyas, RC Munro, M Olalde, MH Pemonge, F Popescu, D Slade, H Tabbener, D Turchini, SGM de Vries, B Ziegenhagen, A Kremer (2002) Chloroplast DNA variation in European white oaks - Phylogeography and patterns of diversity based on data from over 2600 populations. *For Ecol Manage* 156:5-26. doi: [https://doi.org/10.1016/s0378-1127\(01\)00645-4](https://doi.org/10.1016/s0378-1127(01)00645-4)
- Pihain M, Gerhold P, Ducouso A, Prinzing A (2019) Evolutionary response to coexistence with close relatives: increased resistance against specialist herbivores without cost for climatic-stress resistance. *Ecol Lett* 22(8):1285–1296. <https://doi.org/10.1111/ele.13285>
- Sáenz-Romero C, Kremer A, Nagy L, Újvári-Jármay É, Ducouso A, Kóczán-Horváth A, Hansen JK, Mátyás C (2019) Common garden comparisons confirm inherited differences in sensitivity to climate change between forest tree species. *PeerJ* 7:e6213. <https://doi.org/10.7717/peerj.6213>
- Sáenz-Romero C, Lamy JB, Ducouso A, Musch B, Ehrenmann F, Delzon S, Cavers S, Chalupka W, Dagdas S, Hansen JK, Lee SJ, Liesebach M, Rau HM, Psomas A, Schneck V, Steiner W, Zimmermann NE, Kremer A (2017) Adaptive and plastic responses of *Quercus petraea* populations to climate across Europe. *Glob Chang Biol* 23(7):2831–2847. <https://doi.org/10.1111/gcb.13576>
- Sáenz-Romero C, O'Neill G, Aitken S, Lindig-Cisneros R (2021) Assisted migration field tests in Canada and Mexico: lessons, limitations and challenges. *Forests* 12(1):9. <https://doi.org/10.3390/f12010009>
- Saleh D, JC Leplé, T Leroy, L Truffaut, B Dencausse, C Lalanne, K Labadie, I Lesur, D Bert, F Lagane, F Morneau, JM Aury, C Plomion, M Lascoux, A Kremer (2022) Genome-wide evolutionary response of European oaks during the Anthropocene. *Evolution Letters* 6:1-4–20 <https://doi.org/10.1002/evl3.269>
- Sardin T (2008) Guide des sylvicultures. In: Chênaies continentales. Office National des Forêts, Paris
- Sinclair FH, Stone GN, Nicholls JA, Cavers S, Gibbs M, Butterill P, Wagner S, Ducouso A, Gerber S, Petit RJ, Kremer A, Schoenrogge K (2015) Impacts of local adaptation of forest trees on associations with herbivorous insects: implications for adaptive forest management. *Evol Appl* 8(10):972–987. <https://doi.org/10.1111/eva.12329>
- Stimm K, Heym M, Uhl E, Tretter S, Pretzsch H (2021) Height growth-related competitiveness of oak (*Quercus petraea* (Matt.) Liebl. and *Quercus robur* L.) under climate change in Central Europe. Is silvicultural assistance still required in mixed species stands? *For Ecol Manage* 482:118780
- Torres-Ruiz JM, Kremer A, Carins-Murphy MR, Brodribb TJ, Lamarque LJ, Truffaut L, Bonne F, Ducouso A, Delzon S (2019) Genetic differentiation in functional traits among European sessile oak populations. *Tree Physiology* 39:1736–1749. <https://doi.org/10.1093/treephys/tpz090>
- Truffaut L, Chancerel E, Ducouso A, Dupouey JL, Badeau V, Ehrenmann F, Kremer A (2017) Fine-scale species distribution changes in a mixed oak stand over two successive generations. *New Phytol* 215(1):126–139. <https://doi.org/10.1111/nph.14561>
- Valadon A (2009) Effets des interventions sylvicoles sur la diversité génétique des arbres forestiers. Office National des Forêts, Paris
- Verger S, Ducouso A (2009) Conserver les ressources génétiques du chêne sessile en France: pourquoi, comment ? *Rendez-Vous Techniques ONF* 23-24: 55–63
- Vitasse Y, Delzon S, Bresson C, Michalet R, Kremer A (2009b) Altitudinal differentiation in growth and phenology of temperate-zone tree species growing in a common garden. *Can J For Res* 39(7):1259–1269. <https://doi.org/10.1139/X09-054>
- Vitasse Y, Delzon S, Dufrene E, Pontailier JY, Louvet JM, Kremer A, Michalet R (2009a) Leaf phenology sensitivity to temperature in European trees: do within-species populations exhibit similar responses? *Agric For Meteorol* 149(5):735–744. <https://doi.org/10.1016/j.agrformet.2008.10.019>
- Wilkinson M, Eaton EL, Morison JIL (2017) Variation in the date of budburst in *Quercus robur* and *Q-petraea* across a range of provenances grown in Southern England. *Eur J For Res* 136(1):1–12. <https://doi.org/10.1007/s10342-016-0998-z>
- Zanetto A, Kremer A (1995) Geographic structure of gene diversity in *Quercus petraea* (Matt.) Liebl. 1. Monol Patterns Variation *Heredity* 75(5):506–517. <https://doi.org/10.1038/hdy.1995.167>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

