Pleistocene climate changes, and not agricultural spread, accounts for range expansion and admixture in the dominant grassland species Lolium perenne L.

BlancoPastor, José Luis; Manel, Stéphanie; Barre, Philippe; Roschanski, Anna M.; Willner, Evelin; Dehmer, Klaus J.; Hegarty, Matthew; Muylle, Hilde; Ruttink, Tom; Roldán-Ruiz, Isabel; Ledauphin, Thomas; Escobar-Gutiérrez, Abraham; Sampoux, Jean-Paul

Published in: Journal of Biogeography
DOI: 10.1111/jbi.13587
Publication date: 2019

Citation for published version (APA):
BlancoPastor, J. L., Manel, S., Barre, P., Roschanski, A. M., Willner, E., Dehmer, K. J., Hegarty, M., Muylle, H., Ruttink, T., Roldán-Ruiz, I., Ledauphin, T., Escobar-Gutiérrez, A., & Sampoux, J-P. (2019). Pleistocene climate changes, and not agricultural spread, accounts for range expansion and admixture in the dominant grassland species Lolium perenne L. Journal of Biogeography. https://doi.org/10.1111/jbi.13587
Pleistocene climate changes, and not agricultural spread, account for range expansion and admixture in the dominant grassland species *Lolium perenne* L.

Running title: Phylogeography of perennial ryegrass

J. L. Blanco-Pastor¹, S. Manel², P. Barre¹, A.M. Roschanski³, E. Willner³, K. J. Dehmer³, M. Hegarty⁴, H. Muylle⁵, T. Ruttink⁵, I. Roldán-Ruiz⁵, ⁶, T. Ledauphin¹, A. Escobar-Gutierrez¹, J. P. Sampoux¹

1. INRA, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F), 86600 Lusignan, France;
2. EPHE, PSL Research University, CNRS, UM, SupAgro, IRD, INRA, UMR 5175 CEFE, F-34293 Montpellier, France;
3. Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Inselstr. 3, 23999 Malchow/Poel, Germany;
4. Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, Ceredigion, UK;
5. Flanders Research Institute for Agriculture, Fisheries and Food (ILVO) - Plant Sciences Unit, Caritasstraat 39, 9090 Melle, Belgium;
6. Ghent University, Faculty of Sciences, Dept. of Plant Biotechnology and Bioinformatics, Technologiepark 927, 9052 Ghent, Belgium.

**Corresponding author:**

J. L. Blanco-Pastor

INRA, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F)

Le Chêne - RD 150 CS 80006

86600 Lusignan, France

Phone: +33(0)549556031

Email: jose-luis.blanco-pastor@inra.fr

**Handling Editor**

Hans-Peter Comes
Acknowledgements

J. L. Blanco-Pastor has received the support of the European Commission (EC) in the framework of the Marie-Curie FP7 COFUND People Program, through the award of an AgreeSkills+ fellowship (grant agreement nº 609398). This work was funded in the frame of the project GrassLandscape awarded by the 2014 FACCE-JPI ERA-NET+ call Climate Smart Agriculture. Funding was granted by the EC (grant agreement nº 618105), by the Agence Nationale de la Recherche (ANR) and the Institut National de la Recherche Agronomique (INRA – métaprogramme ACCAF) in France, the Biotechnology and Biological Sciences Research Council (BBSRC) in the United-Kingdom, the Bundesantalt für Landwirtschaft und Ernährung (BLE) in Germany. The computational resources (Stevin Supercomputer Infrastructure) and services used for genotype calling were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government – department EWI.

The authors thank the curators from the genebanks that provided perennial ryegrass seed samples for the needs of the project and staff from European agronomic research institutes who contributed to in situ collections in 2015. Perennial ryegrass is one of the plant species covered under the Multilateral System of the International Treaty on Plant Genetic Resources for Food and Agriculture. All genetic materials used in this study were made available to the authors after signature of a Standard Material Transfer Agreement (SMTA) by the provider and the recipient. Implementation and signature of a SMTA provides compliance with the provisions of the Nagoya Protocol for parties wishing to provide and receive genetic material under the Multilaterial System. The authors declare no conflict of interest.

Abstract

Aim: Grasslands have been pivotal in the development of herbivore breeding since the Neolithic and still represent the most widespread agricultural land use across Europe. However, it remains unclear whether the current large-scale genetic variation of plant species found in natural grasslands of Europe is the result of human activities or natural processes.

Location: Europe.

Taxon: *Lolium perenne* L. (perennial ryegrass).
Methods: We reconstructed the phylogeographic history of *L. perenne*, a dominant grassland species, using 481 natural populations, including 11 populations of closely related taxa. We combined Genotyping-by-Sequencing (GBS) and pool-Sequencing (pool-Seq) to obtain high-quality allele frequency calls of ~500 k SNP loci. We performed genetic structure analyses and demographic reconstructions based on the site frequency spectrum (SFS). We additionally used the same genotyping protocol to assess the genomic diversity of a set of 32 cultivars representative of the *L. perenne* cultivars widely used for forage purposes.

Results: Expansion across Europe took place during the Würm glaciation (12-110 kya), a cooling period that decreased the dominance of trees in favour of grasses. Splits and admixtures in *L. perenne* fit historical climate changes in the Mediterranean basin. The development of agriculture in Europe (7-3.5 kya), that caused an increase in the abundance of grasslands, did not have an effect on the demographic patterns of *L. perenne*. We found that most modern cultivars are closely related to natural diversity from North-Western Europe. Thus, modern cultivars do not represent the wide genetic variation found in natural populations.

Main conclusions: Demographic events in *L. perenne* can be explained by the changing climatic conditions during the Pleistocene. Natural populations maintain a wide genomic variability at continental scale that has been minimally exploited by recent breeding activities. This variability constitutes valuable standing genetic variation for future adaptation of grasslands to climate change, safeguarding the agricultural services they provide.

Keywords: Europe, Genetic diversity, Grasslands, Perennial ryegrass, Phylogeography, Quaternary, Cultivar, Genotyping-by-Sequencing, Pool-seq, Site frequency spectrum.

Introduction
Worldwide, grasslands constitute the most extensive natural and semi-natural habitat types and are an integral part of agricultural landscapes. They have been essential for the maintenance of biodiversity, carbon sequestration and biogeochemistry of soils across the last millennia (Tilman et al., 1996; Jones & Donnelly, 2004; Hejcman et al., 2013). Grasslands support biodiversity and a variety of ecosystem services that have gained renewed interest and value to society (Werling et al., 2014). In Europe, they are the most widespread agricultural
land-use covering 45% of the total agricultural area (Eurostat, 2017). Most of these grasslands are permanent natural or semi-natural grasslands (84%; Leclère et al., 2016) and are composed of plant species and populations that may have evolved under natural environmental constraints and farming usages since the earliest stages of herbivore domestication and agriculture, i.e. about 10 kya in the Fertile Crescent (Hejcman et al., 2013). The rationalisation of farming technics in the 19th and 20th centuries has promoted the inclusion of temporary grasslands (or meadows) into crop rotation systems. These temporary meadows are sown for a period of one to five years and are usually managed in a fairly intensive way. Their acreage in Europe remains less important than that of permanent grasslands (9.76 and 56.9 million ha in EU-27, respectively) (Huyghe et al., 2014), but it is likely to increase in the coming decades because of the recognised positive impact of temporary meadows on the sustainability of crop rotation systems (e.g. Crème et al., 2018; Viaud et al., 2018).

Since the 1960s, agronomic research centres have developed breeding programs to release improved cultivars of forage species, that ensure high production of forage of good quality (Sampoux et al., 2011). New permanent or temporary grasslands are almost exclusively sown with such cultivars. The release of grassland species cultivars improved for forage performances has contributed to the expansion of the acreage of sown meadows during the last 50 years, whereas the acreage of natural and semi-natural permanent grasslands has continuously decreased (Chapman, 1992). In the EU-6 (first six members of the European Union), losses of permanent grasslands are estimated at about 30% (7 million ha) between 1967 and 2007 whereas losses in EU-15 are estimated at 15% (10.5 million ha) during the last 50 years (Peeters, 2012).

The maintenance of plant genetic diversity in grassland species can be of major importance not only in terms of adaptive potential but also in terms of productivity and ecosystem services (Crutsinger et al., 2006; Hughes et al., 2008). It has been shown that intra-population genetic diversity may favour the temporal stability of production in grasslands (Prieto et al., 2015). Large-scale ecotype variation in grassland species genetic diversity is also likely to contribute to the resilience of grassland production along climatic and other environmental gradients, as suggested for the dominant grass species L. perenne by investigations of Balfourier & Charmet (1991a) and Monestiez et al. (1994). Given the reduction in natural grasslands, and the increase in sown meadows during the last decades, there is a risk of
genetic impoverishment of grassland species across agricultural landscapes that may lead to
losses of adaptive potential and/or production capability.

Reconstructions from the pollen record revealed the extensive presence of open habitat (i.e. grasslands) in Europe since the pre-Holocene (> 12 ka ago –kya) (Pérez-Obiol & Julià, 1994; Kuneš et al., 2015). Grasslands in central-eastern Europe, for example, constituted up to 30% of the vegetation during the early-Holocene and their relative abundance has been maintained to the present (Kuneš et al., 2015). Although grasslands were widespread long before agricultural practices became established in Europe ca. 8 kya (Zohary et al., 1988; Hejcman et al., 2013; Giesecke et al., 2017), activities of early agricultural communities have been important for shaping current European vegetation and maintaining open land (Feurdean et al., 2015). Pollen records clearly show that the area dominated by grasses in Europe has increased since 4 kya, suggesting that the conversion of forest landscapes into grasslands was associated with agricultural activity (Giesecke et al., 2017). However, the effect of this human transformation of the European landscape on the natural diversity and population structure of grassland species is unknown. More specifically, little is known about the relative importance of natural and human-mediated expansions to explain the current distribution of grassland species genetic variation in natural and semi-natural grasslands. Additionally, the extent to which the increased use of cultivars has eroded genetic diversity within natural grasslands remains unevaluated. We used L. perenne as a model to investigate these questions because this species is the most widely sown forage grass species in temperate regions (Humphreys et al., 2010), has received extensive breeding effort during the last five decades (Humphreys et al., 2010; Sampoux et al., 2011) and is also one of the most abundant grass species in natural grasslands across Europe and the Fertile Crescent.

One complicating factor for the study of L. perenne is the fact that the genus Lolium comprises nine species (Terrell & Ekrem, 1968; Charmet et al., 1996) which diverged only recently (ca. 4.1 Ma ago; Inda et al., 2014) and that most of them can naturally intercross. This explains the controversial phylogenetic relationships between L. perenne and close relatives observed in the literature (Charmet et al., 1997; Catalán et al., 2004; Inda et al., 2014). A previous study based on chloroplast DNA (cpDNA) polymorphisms suggested that natural L. perenne populations could have been introduced to Europe following human migration routes from the Fertile Crescent as a weed of cereals (Balfourier et al., 2000). More recently, a study based on the analysis of 2185 transcript-anchored single nucleotide polymorphisms (SNPs) suggested that L. perenne was subjected to repeated population
expansion and contraction during the Quaternary glaciations (Blackmore et al., 2015).

However, these studies did not estimate the time of demographic events, which is required to firmly discard either of the two hypothesis.

Thanks to the advances in sequencing technologies (Emerson et al., 2010; Garrick et al., 2015), assessing fine-scale phylogenetic, phylogeographic and hybridisation patterns among recently diverged lineages is now an achievable task. Here, we present a comprehensive study of the genomic diversity of natural L. perenne populations and reconstruct historical demographic events that contributed to shape this diversity, using an extensive set of populations sampled in natural and semi-natural grasslands across most of the natural distribution range of the species. Better knowledge of the distribution of the natural genomic variation of L. perenne across its area of primary expansion would indeed provide essential information to guide the conservation strategy for this major grassland species in a context of reduction of permanent natural grasslands. This would also contribute to guide the informed use of natural populations as genetic resources for plant breeding in this species. We used genomic data to evaluate the two competing hypotheses, that large-scale genetic variation in L. perenne could be explained either by Pleistocene climate changes or by more recent human-mediated expansion. More specific objectives were: (i) to investigate the relationships between L. perenne and close relatives; (ii) to reconstruct and date the main demographic processes that occurred along the evolutionary history of L. perenne and (iii) to gain insight into the origin of L. perenne cultivars and trace the use of natural genetic resources in this species by modern plant breeding.

**Materials and Methods**

**Plant Material and Genotyping**

We obtained a batch of seeds from 476 accessions of L. perenne and related taxa maintained as seed lots in the genebanks of agronomic research institutes from different countries. These seed lots were made as to represent the genetic diversity existing within each of 476 natural populations sampled in natural and semi-natural grasslands across Europe and the Near-East (See Appendix S1 and Table S1 in Appendix S3). Monestiez et al. (1994) analysed spatial autocorrelation patterns in phenotypic traits of natural L. perenne and identified two main multivariate spatial structures with 120 and 300km ranges, respectively. The 120km range was interpreted as a result of isolation-by-distance (gene flow) and the 300km one as a signal of selection imposed by environmental factors. In accordance with these findings, we set 120km as the minimum distance between collection sites of neighbouring populations. The set...
of genebank accessions was complemented with 44 additional *Lolium* natural populations sampled *in situ* in 2015 and 32 diploid *L. perenne* cultivars representing the broad range of cultivars released for forage usage in various countries of Europe and New Zealand during the last five decades.

Accessions from genebanks and populations newly collected in 2015 were grown in an experimental garden which enabled to perform a taxonomic assessment and a flow cytometry control of the ploidy level of these materials. Natural populations of *L. perenne* and related taxa are indeed expected to be diploid whereas some *L. perenne* cultivars, not used in this study, are artificial tetraploids (Beddows, 1967; Nair, 2004; Humphreys et al., 2010). 48 genebank accessions were however not present in the experimental garden because of insufficient seed availability. After sequencing and bioinformatics processing (see below), we finally used three different accession sets for downstream analyses (see Phylogeny and population filter, Appendix S1): i) 470 monophyletic *L. perenne* natural diploid populations (*L. perenne* set), ii) the *L. perenne* set plus 11 diploid populations from other taxa (*Lolium* set) and iii) 32 *L. perenne* diploid cultivars (cultivar set). The 11 additional populations of the *Lolium* set were five populations of *L. multiflorum*, two of *L. rigidum*, two of *L. temulentum* and two of *Festuca pratensis* (outgroup). Previous studies acknowledged the close phylogenetic relationship between the *Lolium* genus and broad-leaved fescues from the *Festuca* genus including *Festuca pratensis* (Catalán et al., 2004). Note that *L. temulentum* is an autogamous taxon whereas the three others are allogamous like *L. perenne*. For the sake of reliability of downstream analyses, natural populations not grown in the experimental garden were neither included in the *L. perenne* set nor in the *Lolium* set, except 12 ones. The latter showed clear sister genomic relationships with some other natural *L. perenne* diploid populations grown in the experimental garden and were thus included as such in the *L. perenne* set.

We used GBS pool-Seq (Byrne et al., 2013) to determine genome-wide allele frequencies of natural populations and cultivars in a cost-effective manner (Schlötterer et al., 2014). We used a large number of individuals (c.a. 300) per population in order to obtain unbiased allele frequency estimations (Sham et al., 2002; Lynch et al., 2014; Schlötterer et al., 2014; Fracassetti et al., 2015; Rode et al., 2018). Genotyping a large set of *L. perenne* populations (470) was considered important to assess fine spatial distribution of the natural diversity of this species across Europe. The full GBS nuclear dataset (after SNP calling and filtering) contained population alternative allele frequencies (AAF) for a total of 507,583 SNP loci.
sequenced in at least 70% of the initial 552 entries (see Supplementary Methods in Appendix S1); the overall percentage of missing data was 10.25%. Chloroplast DNA (cpDNA) HiPlex amplicons were also designed and sequenced yielding 49 cpDNA SNP loci. We set up a SNP cpDNA dataset containing AAFs for these 49 cpDNA SNP loci and 30 additional GBS cpDNA SNP loci; this dataset had an overall 12.55% missing data. Detailed information on the DNA extraction, library preparations, sequence processing and data filtering can be found in the Supplementary Methods in Appendix S1.

Phylogeography and past demography

Population structure

To investigate the presence of genetic clusters in the L. perenne set, we applied the Discriminant Analysis of Principal Components (DAPC, Jombart et al., 2010) to the table of nuclear AAFs of the 470 populations of the L. perenne set. DAPC analysis details are shown in Appendix S1. We also computed Fst between clusters with the R package ‘StAMPP’ (Pembleton et al., 2013). Additionally, we calculated the expected heterozygosity (He) of populations and computed the He average value and standard deviation of each cluster. DAPC analyses were similarly carried out on the table of 79 cpDNA SNP loci AAFs of the Lolium set (Pool-GBS and HiPlex cpDNA data, see Appendix S1).

Isolation by distance (IBD) versus isolation by environment (IBE)

To test the effect of IBD and IBE on the Nei genetic distances (nuclear dataset) between the L. perenne set populations, we computed a geographical distance matrix with the R package ‘geosphere’ (Hijmans et al., 2012) and an Euclidean bioclimatic (environmental) distance matrix using 19 bioclimatic data layers (sensu WorldClim database, bio1-bio19; http://www.bioclim.org). Normalized matrices were used to assess IBD and IBE with the Multiple regression on distance matrices (MRM) function (Legendre et al., 1994; Wang, 2013) implemented in the R package ‘ecodist’ (Goslee & Urban, 2007).

Splits and admixture

Setting up demographic models requires prior information about the relationships between the considered populations. Splits and admixture analyses as implemented in TREE MIX v.1.13 (Pickrell & Pritchard, 2012) provide the prior information necessary to set up alternative demographic models (main tree topologies and migration directions). For TREE MIX analyses, we generated reduced tables of the nuclear dataset for the L. perenne and Lolium sets in which allele frequencies of populations were averaged for each cluster of L. perenne natural
diversity and for each of the other taxa. We ran two TreeMix analyses. The first analysis was applied to *L. perenne* clusters plus the other taxa: *F. pratensis, L. multiflorum, L. temulentum* and *L. rigidum*, i.e. to the *Lolium* set. The second analysis was applied to clusters of *L. perenne* natural diversity only, i.e. to the *L. perenne* set. For both TreeMix analyses, we first inferred a maximum likelihood (ML) tree without admixture and then ran 1000 standard bootstrap replicates to obtain statistical support for the non-admixed tree topology. Bootstrap trees were summarized in a 50% majority-rule consensus tree with SUMTrees v4.2.0 (Sukumaran & Holder, 2015). Finally, for both analyses, we fitted one to six migration events in the tree and displayed a tree graph for each number of migration events. All migration events between *L. multiflorum, L. rigidum* and *L. perenne* observed in alternative *Lolium* set TreeMix models (TMMs) and all migration events between *L. perenne* clusters observed in alternative *L. perenne* set TMMs were considered for the generation of demographic models with $\delta a\delta i$ (see below).

**Demography**

To further investigate the hybridisation patterns between the *Lolium* taxa included in our study and the demographic history of *L. perenne*, we used the program $\delta a\delta i$ (Gutenkunst et al., 2009). $\delta a\delta i$ compares the site frequency spectrum (SFS) expected under custom demographic models to that observed with actual frequency data. SFSs are simulated with a diffusion approach which is limited to a three-taxon phylogeny in $\delta a\delta i$. Comparisons of $\delta a\delta i$ models were made independently for the *L. perenne* and the *Lolium* sets. For the construction of $\delta a\delta i$ models (see Fig. S6 in Appendix S2), we used alternative migration/hybridisation scenarios as obtained from TreeMix. First, we investigated alternative scenarios of gene flow among *L. rigidum, L. multiflorum* and *L. perenne* as shown in the *Lolium* set TMMs 1-6 (see below). Second, we investigated demography and patterns of gene flow within *L. perenne*, as displayed in the *L. perenne* set TMMs 1-6 (see below). For the first analysis, we averaged nuclear AAFs in all three species. For the second one, we averaged nuclear AAFs of populations from clusters 1 to 5 and considered clusters 6 and 7 as independent lineages (as displayed in the main topology of *L. perenne* set TMMs 1-6, see below). For both analyses, we used *F. pratensis* as outgroup to establish the ancestral state for each SNP. Some $\delta a\delta i$ models were based on a non-equilibrium demography (cluster splits with a period of isolation before admixture, as assumed in TreeMix) whereas some others were based on an equilibrium demography (fixed migration structure since divergence). We also considered the
possibility of a linear growth for some models versus a constant size for others. We used the log L-BFGS-B optimisation method to fit parameters for each model. A total of 30 independent runs were used for the optimisation of each model. Each run started from a different randomly perturbed starting position and included a maximum of 20 iterations. The best diffusion fit to the observed SFS was chosen when the likelihood was the highest among the runs. Fitted models were ranked according to the Akaike information criterion (AIC) to account for the variable number of model parameters.

To set confidence intervals (CIs) for parameter estimates of the best-fit demographic models, we generated 100 datasets by non-parametric bootstrapping. Bootstrap replicates were re-optimized in $\delta a\delta i$ to estimate the parameter uncertainties. CIs were calculated as $E \pm 1.96\sigma$, where $E$ is the ML parameter estimate and $\sigma$ is the standard deviation of the parameter estimate across the bootstrap replicates. To assess the time of demographic events, it is necessary to incorporate the average generation time of taxa. $L. perenne$ requires vernalisation to flower (Thiele et al., 2009). After seed germination and emergence followed by first winter vernalisation, the seed production is often fairly small because of limited tillering. Seed production then peaks after second and third winter vernalisation. In favourable environmental conditions, some perennial ryegrass clones can live for more than 10 years (Beddows, 1967). Ramet density decreases gradually after the third year, but a small seed production can last in late years of life of clones. Considering these facts, we assumed an average of three years generation time (3y/gen) for $L. perenne$. $L. rigidum$ is an annual taxon and $L. multiflorum$ includes mostly annual but also biennial types (Terrell & Ekrem, 1968). We assumed one year generation time (1y/gen) for these two taxa. Using these generation times and a known average mutation rate of 6.03E-9 substitutions per site per generation for Poaceae (De La Torre et al., 2017), we converted effective population sizes (Ni) and Time (Ti) parameters to (breeding) individuals and years.

**Origin of cultivars**

We predicted the cluster membership of cultivars by adding them as supplementary populations in the DAPC analysis of the nuclear AAFs of the $L. perenne$ set considered as “training data”. We transformed the allele frequencies of these supplementary entries using the centring and scaling metrics of the “training data” and determined their position onto the discriminant axes.
Results

Genetic structure in *L. perenne* natural diversity

The k-means algorithm analysis of the nuclear AAFs of the *L. perenne* set detected $K = 7$ as the optimal number of genetic clusters as given by the Bayesian Information Criterion (BIC) (Fig. 1 and Fig. S2 in Appendix S2). The genetic clusters exhibited a low level of admixture and a clear geographical structure (Fig. 1a): cluster 1 fits to the South-Eastern Europe – Near East region (black colour in Fig. 1), cluster 2 to Eastern Europe (orange), cluster 3 to North-Eastern Europe (light blue), cluster 4 to Northern Iberia – Southern France (green), cluster 5 to North-Western Europe (pink), cluster 6 to Corsica – Sardinia (yellow) and cluster 7 to Northern Italy (red). Admixed populations were mainly assigned to cluster 3 and cluster 5 (Fig. 1b-1c). Additional DAPC analyses performed with $K = 8$-10 detected seven clusters with the same geographic structure as mentioned above (results not shown). Cluster 8 was formed by populations with an admixture signal between clusters 3 and 5 that were located between the distribution areas of these two clusters. Cluster 9 was also formed by a small number (11) of admixed populations with ancestry from cluster 3 and cluster 5 that were distributed in South-Eastern Europe and the Near East; we interpreted that these populations were not native from this area but imported from Western Europe and recently sown. Cluster 10 was formed by populations from cluster 1 that were located in the North-Western part of cluster 1 distribution. Because cluster 8 and cluster 9 reflected admixture and inconsistent geographical distribution (cluster 9) rather than divergence signal (reproductive isolation), we considered that $K = 7$ as given by the BIC was appropriate for downstream analyses and further discussion. With $K = 7$, the first DAPC axis was strongly correlated with longitude ($r = 0.810$, p-value < 0.001) and the second axis with latitude ($r = 0.662$, p-value < 0.001), indicating different and independent directions of differentiation along these two geographical dimensions (Fig. S3a-S3b in Appendix S2). Genetic differentiation among clusters was small, with *Fst* values ranging from 0.0152 (cluster 4-cluster 5) to 0.0776 (cluster 3-cluster 6) (Table S2 in Appendix S3). Average expected heterozygosity (*He*) values within populations ranged from 0.54 (cluster 4) to 0.47 (cluster 6) with standard deviation (STD) within clusters ranging from 5.0E-4 (cluster 5) to 1.1E-2 (cluster 1) (Fig. 2d). Cluster 1 had a remarkably high *He* STD; this cluster indeed contains a set of populations with high heterozygosity together with populations with very low heterozygosity.

The MRM analysis (Fig. 3) revealed that both IBD and IBE played a significant, but moderate, role in genetic differentiation of *L. perenne* (IBD model: $r^2 = 0.185$, p-value <
IBE model: \( r^2 = 0.112 \), p-value < 0.001; Fig. 3a-3b). IBD was more important than IBE to explain the genetic distances between populations (IBD + IBE model: \( r^2 = 0.201 \), \( \beta_D = 0.280 \), p-value < 0.001; \( \beta_E = 0.142 \), p-value < 0.001; Fig. 3c). However, geographical and environmental distances showed moderate correlation (\( r = 0.531 \), \( r^2 = 0.282 \), p-value < 0.001, Fig. 3d).

The DAPC analysis carried out with 79 cpDNA loci on the Lolium set (\( L. \) perenne and other taxa) revealed neither population geographical structure in \( L. \) perenne nor differentiation among the different Lolium taxa. The k-means algorithm failed to find genetic clusters. We chose \( K = 4 \) considering as optimal \( K \) the highest possible number of clusters that showed non-admixed populations. cpDNA clusters comprised members of the different Lolium taxa, with clusters 1 and 2 showing a higher presence in Mediterranean areas (Fig. S5).

**Splits and admixture**

The likelihood of the Lolium set TREE MIX models (TMMs) increased almost linearly with no clear stabilisation of likelihood values as the number of admixture edges increased (Fig. S6a). When adding two or more admixture edges (M2-M6), the topology of the main tree was rearranged and cluster 6 became the most basal lineage. The ln-likelihood of the Lolium set TMM increased by 591.37 from M0 to M1, 779.78 from M1 to M2 and 337.28 from M2 to M3. All additional edges (M4-M6 models) increased the ln-likelihood by less than 212. It is important to take into consideration that the increase of likelihood values does not mean that the model is closer to the true network. This is because the addition of admixture edges can never reduce the likelihood (Pickrell & Pritchard, 2012).

We analysed the \( L. \) perenne set with TREE MIX using cluster 6 as outgroup as displayed in the Lolium set TMMs 2-6 (Fig. S6b). With the addition of admixture edges, the likelihood of the \( L. \) perenne TMMs clearly increased from M0 to M2 and to a lower extent from M2 to M6 (Fig. S6b). The ln-likelihood of the \( L. \) perenne set TMM increased by 13232.72 from M0 to M1, 5208.19 from M1 to M2 and 984.35 from M2 to M3. All additional edges (M4-M6 models) increased the ln-likelihood by less than 653.

**Demography**

\( L. \) perenne model comparison (Fig. S7a) revealed that the best fit to our observed SFS was model C (Fig. 2a and Table S3 and Table S4 in Appendix S3). Positive or negative residuals of the model (normalized differences between model and data) indicate that the model predicts too many or too few alleles in a given cell of the two-population SFS, respectively.
Residuals of the best model C showed a normal distribution with a zero mean value (not shown), indicating an appropriate fit of the model to the data. ML parameter values and their 95% confidence interval (CI) obtained from non-parametric bootstrapping are shown in Table S5 in Appendix S3 (ML values also shown in Fig. 2c-2e). The best model C detected an ancestral population split into two lineages (L. rigidum and the ancestor of L. perenne and L. multiflorum) that was dated 397 kya (95% CI 238-555 kya, 1y/gen) or 1.19 Mya (95% CI 715-1666 kya, 3y/gen). A subsequent split followed by a period of isolation between L. perenne and L. multiflorum occurred 380 kya (95% CI 235-525 kya, 1y/gen) or 1.14 Mya (95% CI 706-1576 kya, 3y/gen). Then migration from L. rigidum to L. perenne and from L. multiflorum to L. perenne with constant population size in L. perenne started 366 kya (95% CI 235-497 kya, 1y/gen) or 1.10 Mya (95% CI 704-1492 kya, 3y/gen).

Model comparison using the L. perenne set (Fig. S7b) revealed that the best fit between the predicted and observed site frequency spectrum (SFS) (i.e. the highest maximum composite likelihood) was obtained with model F (Fig. 2b and Table S6 and Table S7 in Appendix S3). Residuals from the best model F followed a normal distribution with a zero mean value (not shown), indicating an appropriate fit of the model to the data. Maximum likelihood parameter values of the best model F and their 95% confidence interval (CI) obtained from non-parametric bootstrapping are shown in Table S8 in Appendix S3. The best model F showed an ancestral population split into two lineages (cluster 6 –Corsica-Sardinia– and the ancestor of remaining clusters) that was dated 174 kya (95% CI 49 – 300 kya, 3y/gen). Then a second split followed by a period of isolation between ancestor of clusters 1-5 (from Western Europe to Near East) and cluster 7 (Northern Italy) and a linear population growth of clusters 1-5 starting 72 kya (95% CI 31 – 112 kya, 3y/gen). Finally, migration from cluster 7 to clusters 1-5 and from clusters 1-5 to cluster 6 started 56 kya (95% CI 31 – 81 kya, 3y/gen).

Additionally, we ran an alternative model to model F in δaδi, in which the latter two admixture events did not coincide in time. This new model did not fit as well as the base model F (results not shown) suggesting similar timing for the two admixture events.

Origin of cultivars

The predicted membership of the 32 cultivars bred for forage usage on the L. perenne set DAPC (Fig. 4) showed that 25 out of these 32 cultivars were assigned to cluster 5 (North-Western Europe), three to cluster 7 (Northern Italy), two to cluster 2 (Central-Eastern Europe), one to cluster 3 (North-Eastern Europe) and one to cluster 4 (South-Western Europe). Most of cultivars assigned to cluster 5 were genetically very similar to the natural
populations of this cluster. Cultivars assigned to the other clusters were highly admixed, with high membership probabilities with cluster 5 except the cultivar Medea. A DAPC analysis with all populations did not separate cultivars from natural populations (not shown). This suggests that the genetic origin of cultivars is not restricted to a single source, despite the major contribution of cluster 5.

**Discussion**

Our analyses of the genomic diversity, structure, and past demography of *L. perenne* reveal that despite its extensive use, there remains a regional genetic structure of extant natural populations that was shaped by demographic events predating the onset of agriculture. The impact of these events is still visible today even if the wide presence of permanent grasslands in landscapes, hosting the natural diversity of *L. perenne*, is mostly a result of the development of agriculture during the last millennia.

According to the fossil record, the Late Glacial (ca. 13-10 kya) vegetation in Europe was dominated by herbaceous communities including a large proportion of grasses (Giesecke et al., 2017). Later on, during the Holocene (from ca. 10 kya onwards), and particularly during the Holocene Climatic Optimum (9-5 kya), Europe became dominated by dense forests of temperate deciduous trees and conifers (Furdean et al., 2015; Giesecke et al., 2017). But forests were never fully closed, enabling the persistence of grasslands throughout the Holocene (Hejcman et al., 2013). For example, small-scale steppe grasslands of natural origin were present in forest-rich areas of Central Europe before the onset of agriculture in the early Neolithic (ca. 5.5 kya) (Hejcman et al., 2013). In the late Holocene (last 3.5 kya), an increase in the abundance of grasses, accompanied by a reduction of forested areas is captured in the fossil record, reflecting the development and expansion of agriculture in this area of Europe (Hejcman et al., 2013; Giesecke et al., 2017). So far, the impact of this anthropogenic transformation of the European landscapes on the genetic diversity of key grassland species has been scarcely documented. The processes involved in the genesis of grasslands in agricultural landscapes are not precisely known. The unconscious selection of grazing-tolerant species has certainly played a major role and has been pivotal in the domestication of large herbivores. However, the importance of conscious human actions (such as voluntary seed harvesting and sowing of grasslands) remains largely unknown, and no fossil record specific for *L. perenne* or other major grass species has been reported at documented archaeological sites. Nonetheless, we can ascertain that the development of specific practices for the management and production of grasslands occurred during the Roman Empire (Hooper &
Ash, 1935; Ash et al., 1941) and that intensive grassland cultivation may not have taken place before the appearance of the first scythes during the 7th–6th century BC (Hejcman et al., 2013).

Our results reveal incongruence between main cpDNA and nuclear DNA signals that could be explained by insufficient data in the cpDNA matrix, but also by different mutation rates among marker types, incomplete sorting of cpDNA polymorphisms and/or by plastid capture within and between Lolium taxa. Our results show that nuclear DNA better captured and retained signatures of L. perenne phylogeographic history. The seven nuclear DNA genetic clusters represent genetic discontinuities that can be attributed to reduced gene flow at geographical barriers such as the Mediterranean Sea (cluster 6), the Alps (clusters 2, 3, 4, 5, 7) and the Carpathians (clusters 1, 2, 3). These clusters also showed consistency when displayed on the first two principal axes of a PCA on nuclear AAFs (Fig. S4 in Appendix S2).

Not only geographical barriers but also geographical and environmental distances should account for the genetic differentiation in L. perenne (Fig. 3). Nevertheless, because geographical and environmental distances showed moderate correlation (Fig. 3d), it remains difficult to discriminate which of the two factors was the most important to explain genetic differentiation between populations and clusters. We observed genetic clines in our data, but IBD and IBE explained only a limited proportion of genetic variation. As such, the seven clusters detected by the DAPC analysis on nuclear AAFs represent evolutionary coherent lineages that were interpreted as meta-populations geographically differentiated by reduced gene flow at geographical barriers (European mountain ranges, see Fig. 1). These seven clusters were then considered as independent entities for downstream analyses.

Because there is evidence of genetic clines combined with the genetic discontinuities represented by DAPC clusters (Fig. 1d, Fig.3 and Table S2 in Appendix S3), and because TREE MIX and $\delta a \delta i$ inferences about migration are based on these barely differentiated clusters, migration rates might have been partly overestimated in our analyses or confounded with clinal variation. However, it should be noted that the migration edges in our best TREE MIX models connected non-sister clusters (see Fig. 2b). Furthermore, our $\delta a \delta i$ analyses favoured a model with a period of isolation before migration over models based on an equilibrium demography; the latter would have had the best fit according to a pure clinal variation.
The origin of *L. perenne* and its close relatives (*L. rigidum* and *L. multiflorum*) inferred by our analyses is partially congruent with previous studies (Balfourier et al., 1998; Catalán et al., 2004). This includes the early divergence of *L. rigidum* and the sister relationship between *L. multiflorum* and *L. perenne*. However, we discovered that relationships between these three taxa are more complex than previously proposed, with *L. perenne* receiving genes from *L. rigidum* and *L. multiflorum* after divergence (Fig. 2a–2c). According to our demographic reconstructions, main divergence events among *Lolium* taxa took place during the Pleistocene glaciations, long before the onset of agriculture and main migrations of agricultural communities across Europe. This is in agreement with previous inferences from molecular dating of the grass genera *Festuca* and *Lolium* (Inda et al., 2014). Early gene flow from *L. rigidum* (in the Near East) and *L. multiflorum* (in the North Italy) to *L. perenne* (Fig. 2a–2c) is dated 366 kya (1y/gen) or 1.19 Mya (3y/gen). Indeed, in the frame of controlled experiments, it has been demonstrated that *L. multiflorum* and *L. rigidum* are completely interfertile with *L. perenne* (Terrell, 1966). A close relationship between ancestral *L. perenne* populations from the Near East and *L. rigidum* was previously inferred by Balfourier et al. (1998). In addition, it is known that both *L. perenne* and *L. multiflorum* have been present in the northern plains of Italy since the late Middle Ages (Casler, 2006). The likely long term coexistence of *L. perenne* and *L. rigidum* in the Near East and *L. perenne* and *L. multiflorum* in the Italian plains, together with their ability to intercross, may explain the introgression from *L. rigidum* and *L. multiflorum* to *L. perenne* in these areas (Fig. 2a–2c). Contrary to the hypothesis of Casler (2006), who assumed that *L. multiflorum* originated from human-mediated selection in some *L. perenne* strains in Northern Italy, our results indicate much earlier divergence between these two species in agreement with Inda et al. (2014) (see Fig. 2a and Table S3 and Table S4 in Appendix S3).

The first event reconstructed from the best *L. perenne* δaδi model (Fig. 2b) is a split between cluster 6 (Corsica-Sardinia) and the ancestor of remaining clusters occurring 174 kya (95% CI 49 – 300 kya, 3y/gen) (Fig. 2c-2e1 and Table S8 in Appendix S3). During the last glaciations (Würm 115-11.7 kya; Riss 347-128 kya), and particularly during sea level low stands (ca. 20, 140 and 260 kya), Corsica and Sardinia were connected *inter se* and joined to Tuscany through the Tuscan Archipelago (Rohling et al., 1998; Lambeck & Chappell, 2001; Rabineau et al., 2006), implying that differentiation in these islands followed by rising sea levels during inter/post-glacials could have resulted in the origin of cluster 6 (Fig. 2e1). The Corsica channel might have acted as a barrier for gene flow between *L. perenne* populations from...
Corsica and Sardinia and Italian refuge lineages already present there (as shown by the date of ancestral admixture from *L. multiflorum*, Fig. 2a, Fig. 2c and Table S5 in Appendix S3). The next event is the split between cluster 7 (Northern Italy) and ancestor of clusters 1-5, followed by the expansion of two separate evolutionary lineages colonising northwards via Western and Central-Eastern routes, respectively. This is supported by the TreeMix model with highest likelihood compatible with the best $\delta a\delta i$ model (see Fig. 2B) and admixture signal revealed by the DAPC analysis (Fig. 1). On the other hand, heterozygosity values (Fig. 2d) and second best TreeMix model (compatible with the best $\delta a\delta i$ model, M2 see Fig. S6) support a postglacial expansion through the West and Central East and next towards Eastern Europe (clockwise movement around the Alps). Note that reduced heterozygosity in cluster 6 and cluster 7 (source populations) is likely linked to a reduced effective size in these clusters. We favour the interpretation of two separate evolutionary lineages through the West and Central East, given the higher likelihood of the associated TreeMix model but the alternative scenario cannot be ruled out completely, and is also depicted in Fig. S11 (Appendix S2). The split between cluster 7 and clusters 1 to 5 and the start of the range expansion of clusters 1-5 are dated 72 kya (Fig. 2b, Fig. 2c and Table S8 in Appendix S3). The split and subsequent start of range expansion overlap with a glacial period (Würm glaciation 12- 110 kya) (Fig. 2e2). During the Würm glaciation, the Alps might have acted as a barrier to gene flow in *L. perenne* as previously suggested by Balfourier et al. (1998, 2000) and Blackmore et al. (2015). In continental Europe, the continuous cooling during that period might have affected the dominance of tree species in favour of herbs, possibly including grass species such as *L. perenne*. This is supported by the abundance estimates traced in the pollen record for different vegetation types during the Late Glacial (Giesecke et al., 2017). During the expansion of *L. perenne*, the ancestor of clusters 1, 2 and 3 might have contacted ancient *L. perenne* genotypes already located in Eastern Mediterranean areas that had previously admixed with *L. rigidum* (cluster 1, see Fig 2a and Fig. 2c). This is further supported by the DAPC analysis with $K = 10$ that identified an additional cluster in the contact zone with cluster 2 (probably as an effect of clinal variation after range expansion), and also by the analysis of genetic diversity ($H_e$) of *L. perenne* clusters (Fig. 2d) which showed that cluster 1 exhibits the most variable within-population genetic diversity. Cluster 1 may have been formed by the mixture between ancient highly diverse populations located in South-Eastern refugia and more recent immigrant populations from the ancestor of clusters 1, 2 and 3, the latter including low genetic diversity due to allele surfing in expanding populations (Edmonds et al., 2004; Excoffier & Ray, 2008). For later stages of the Würm glaciation from 56 kya until current
time, the TREEMIX analyses and best δaδi demographic model suggest migration from cluster 7 to clusters 1 and 2 and from cluster 1 to cluster 6 (Fig. 2b, Fig. 2c). During this period, and particularly 20 kya, the Mediterranean sea level reached an estimated maximum drop of 149 meters with respect to current level (Fig. 2e) (Lambeck & Chappell, 2001; Rabineau et al., 2006). This may have facilitated dispersal through land bridges in both the Adriatic Sea (from Northern Italy cluster 7 to Central Europe cluster 2) and the Ligurian Sea (from South-Eastern Europe – but also possibly Central-Southern Italy – cluster 1 to Corsica-Sardinia cluster 6). *L. perenne* natural populations do exist in Central and Southern Italy (Balfourier & Charmet, 1991b, 1994; Balfourier et al., 1998) but we did not have the opportunity to include populations from this region in our study. Balfourier et al. (1998) analysed the genetic structure of 120 wild populations of *L. perenne*, including populations from Central-Southern Italy, using allelic frequencies from 12 polymorphic isozyme loci. These authors found that populations from Central-Southern Italy were genetically closer to populations of Corsica-Sardinia, Eastern Europe and Eastern Mediterranean than to populations of Western and Northern Europe. Our TREEMIX and δaδi analyses (Fig. 2b and Fig. 2c) combined with results of Balfourier et al. (1998) support a possible connection between South-Eastern Europe - Near East cluster 1 and Corsica-Sardinia cluster 6 via Central-Southern Italy. Since the origin and expansion of domesticated plants in the Old World, trade, wars and nomadism caused extensive movements of people and livestock across Europe, possibly favouring transport of diaspores of grassland species among European regions. Nevertheless, all expansion and admixture events in *L. perenne* recovered by our δaδi analyses predate the origin of agriculture in Europe, even if taking into account the CI of our time estimates (see Fig. 2e and Table S5 and Table S8 in Appendix S3). The results we report reveal that main genetic signals of colonisation and admixture observed in *L. perenne* can be explained by climatic events predating the transformation of landscapes by human activities in Europe. A more limited study previously reported similar conclusions for *Festuca pratensis* (Fjellheim et al., 2006). *F. pratensis* and *L. perenne*, two major grass species of European grasslands, may have experienced similar evolutionary histories. They may exemplify an evolutionary trend shared with other European grassland species, in which current patterns in natural genetic diversity were shaped during range expansions of the last glacial period and not significantly disturbed by agricultural expansion.

The first detailed written records on intensification of grassland management date from the Roman Empire (Hooper & Ash, 1935; Ash et al., 1941). However, the real decline of wild
grasslands and the large-scale enlargement of hay meadows started in many European regions
during the 18th century (Semelová et al., 2008). This process may have involved increased use
of grass seeds harvested from local natural strains to re-sow pastures, but likely without direct
selection on phenotypic traits except sufficient seed production. Modern selection of *L. perenne*
for forage production based on the agronomic testing of progeny performance began
in 1919 in the United Kingdom and after World War II in other European countries, Northern
America and New Zealand (Humphreys et al., 2010; Sampoux et al., 2011). Recurrent
selection in plant breeding germplasm has so far resulted in the release of more than 1000 *L.
perenne* cultivars improved for forage production (European Commission, 2015). More
recently, since the 1960s, similar selection methods have also been implemented to release *L.
perenne* cultivars improved for turf usage (Sampoux et al., 2013). We genotyped 32 cultivars
representing a large diversity of the *L. perenne* cultivars bred for forage usage in various
countries of Europe and New Zealand. 25 out of the 32 genotyped cultivars were assigned to
the cluster of North-Western Europe (cluster 5) and were thus likely bred from germplasm of
this area. Cultivars assigned to other clusters, except *Medea*, also contained a significant part
of diversity from this cluster 5. An interesting pattern is observed in those cultivars created by
IBERS (Aberystwyth, UK). *Aurora*, a cultivar with high water soluble carbohydrate (WSC)
content bred from a Swiss ecotype (Faville et al., 2004) showed 60% membership of cluster 7
(Northern Italy). Other more recent IBERS cultivars developed from *Aurora* (*Aberdart*,
*Aberavon*, *Aberstar* and *Abermagic*) showed a decreasing content of genetic material from
cluster 7 in favour of cluster 5. This case exemplifies the strong bias towards North-Western
Europe cluster 5 in the generation of modern cultivars. Most probably, this trend is
predominantly due to the fact that modern breeding of *L. perenne* started in this part of the
world using local diversity.

We showed that *L. perenne* cultivars likely use only a small proportion of the natural genetic
diversity existing across its natural distribution range. This natural diversity thus represents a
valuable genetic resource that should be safeguarded in genebanks, but also much more
efficiently in the diverse natural and semi-natural permanent grasslands from which it
originates. However, since the mid-20th century, there has been a continuous trend of
reduction in the acreage of natural and semi-natural permanent grasslands in Europe.
Especially in intensive agricultural landscapes, permanent grasslands have tended to be
ploughed and replaced by rotations of annual crops. Rotations may indeed include temporary
meadows sown with cultivars from modern breeding in regions where agriculture combines
cash crops and cattle breeding. This practice may not only reduce the natural diversity of perennial ryegrass by extinction of natural populations but also by the expansion of North-Western European genotypes (typically found in cultivars) into other European regions through gene flow from cultivars to natural populations.

Conclusions

Demographic reconstructions assessing the SFS from pool-Seq GBS performed on a high number of populations allowed us to trace the origins of the genomic diversity of *L. perenne* at an unprecedented level of detail. The current extent of grasslands across Europe has been mainly determined by human activities. However, our results indicate that the spatial distribution of the natural genome-wide diversity of *L. perenne* has not been significantly disturbed after more than two millennia of intensive agriculture. The current *L. perenne* natural populations still maintain the genomic diversity that has allowed the species to persist during the Quaternary climatic fluctuations. Modern plant breeding has likely used only a small part of the genomic diversity of *L. perenne* naturally distributed across Europe and surroundings, and thus has likely taken limited advantage of the adaptive diversity and phenotypic variability of the species. To date, this wide natural diversity remains available but it is threatened with extinction and should be preserved. Indeed, it may constitute a valuable genetic resource for plant breeding to meet emerging agricultural challenges such as adaptation to anthropogenic climate change.

Figure legends and embedded figures

Fig. 1. Genetic structure of 470 *L. perenne* natural populations sampled across Europe and the Near East (*L. perenne* set) based on the DAPC (Discriminant Analysis of Principal Components) of population allele frequencies at 507,583 nuclear genome SNP loci. (a) Geographical distribution of genetic clusters. Bar charts (b and c) indicate Population Membership Probabilities (PMPs) to genetic clusters. (b) Representation of PMPs from all populations. (c) Zoom into PMPs of those populations having less than 90% of membership
probability to a single cluster. (d) Scatter plot of the first two discriminant axes.
Fig. 2. Phylogeography of *L. perenne*. (a) and (b) From left to right, schematic of the best demographic model estimated with $\delta a \delta i$ (best fit between the predicted and observed site frequency spectrum – SFS), heatmap representations of the joint SFS expected under the best model and the observed joint SFS, TREE MIX model with highest likelihood compatible with the best $\delta a \delta i$ model. (a) Analysis of *L. perenne* and related taxa natural populations, (b) Analysis of *L. perenne* natural populations only. (c) Geographical representation of the inferred demographic history of *L. perenne* in Europe and surroundings. Colours represent distribution of DAPC genetic clusters as in Fig. 1, dates represent ML values as obtained from the best $\delta a \delta i$ model. (d) Mean and standard deviation of expected population heterozygosity for each *L. perenne* genetic cluster. (e1-e3) Representation of sea level variation (m) across the last 400 kya obtained from Western Mediterranean paleopositions during glacial maxima of Quaternary Glacial cycles (from Rabineau et al., 2006), with superimposed time parameter values: ML (opaque colours) and 95% CI (transparent colours) of demographic events under the best *L. perenne* $\delta a \delta i$ model (model F). In (e1-e3) in light blue colour it is also displayed the time estimate for the first signs of domesticated plants in the Old World *as per* Zohary et al. (1988).
Fig. 3. Multiple regression on distance matrices (MRM) analysis performed on *L. perenne* natural populations from Europe and the Near East. Scatterplots show patterns of (a) IBD, (b) IBE, (c) combined patterns of IBD + IBE and (d) the relationship between environmental and geographical distances.

Fig. 4. Prediction of DAPC (Discriminant Analysis of Principal Components) membership probability to clusters of *L. perenne* natural diversity for a set of 32 *L. perenne* cultivars bred for forage usage and released within the last five decades in Europe, New Zealand and Australia. (a) Membership probability with respect to a DAPC constructed with 470 *L. perenne* natural populations. (b) Scatter plot of the first two discriminant axes of the DAPC with *L. perenne* natural populations displayed as transparent filled circles (same positions as on Fig. 1) and cultivars superimposed as opaque filled squares.
References

Ash, H.B., Forster, E.S., & Heffner, E. (1941) *Columella. On agriculture*. Cambridge. Mass.

Balfourier, F. & Charmet, G. (1991a) Relationships between agronomic characters and ecogeographical factors in a collection of French perennial ryegrass populations. *Agronomie*, 11, 645–657.

Balfourier, F. & Charmet, G. (1991b) Spaced plant evaluation of Mediterranean germplasm collections of perennial ryegrass. *Euphytica*, 57, 57–66.

Balfourier, F. & Charmet, G. (1994) Geographical patterns of isozyme variation in mediterranean populations of perennial ryegrass. *Heredity*, 72, 55–63.
Balfourier, F., Charmet, G., & Ravel, C. (1998) Genetic differentiation within and between natural populations of perennial and annual ryegrass (Lolium perenne and L. rigidum). 
Heredity, 81, 100–110.

Balfourier, F., Imbert, C., & Charmet, G. (2000) Evidence for phylogeographic structure in Lolium species related to the spread of agriculture in Europe. A cpDNA study. 
Theoretical and Applied Genetics, 101, 131–138.

Beddows, A.R. (1967) Biological Flora of the British Isles. Lolium perenne L. Journal of Ecology, 55, 567–587.

Blackmore, T., Thomas, I., McMahon, R., Powell, W., & Hegarty, M. (2015) Genetic–geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (Lolium perenne) using array-based SNP genotyping. Theoretical and applied genetics, 128, 1917–1932.

Byrne, S., Czaban, A., Studer, B., Panitz, F., Bendixen, C., & Asp, T. (2013) Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing. Plos One, 8, e57438.

Casler, M.D. (2006) Perennial grasses for turf, sport and amenity uses: Evolution of form, function and fitness for human benefit. Journal of Agricultural Science, 144, 189–203.

Catalán, P., Torrecilla, P., López Rodríguez, J.A.Á., & Olmstead, R.G. (2004) Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and trnL-F sequences. Molecular Phylogenetics and Evolution, 31, 517–541.

Chapman, G.P. (1992) Grass evolution and domestication. Cambridge University Press, Cambridge, UK.

Charmet, G., Balfourier, F., & Chatard, V. (1996) Taxonomic relationships and interspecific hybridization in the genus Lolium (grasses). Genetic Resources and Crop Evolution, 43, 319–327.

Charmet, G., Ravel, C., & Balfourier, F. (1997) Phylogenetic analysis in the Festuca-Lolium complex using molecular markers and ITS rDNA. Theoretical and Applied Genetics, 94, 1038–1046.

Crème, A., Rumpel, C., Le Roux, X., Romian, A., Lan, T., & Chabbi, A. (2018) Ley grassland under temperate climate had a legacy effect on soil organic matter quantity.
biogeochemical signature and microbial activities. *Soil Biology and Biochemistry*, 122, 203–210.

Crutsinger, G.M., Collins, M.D., Fordyce, J.A., Gompert, Z., Nice, C.C., & Sanders, N.J. (2006) Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science*, 313, 966–968.

Edmonds, C.A., Lillie, A.S., & Cavalli-Sforza, L.L. (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 975–9.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E., & Holzapfel, C.M. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, 107, 16196–16200.

European Commission (2015) Available at: http://ec.europa.eu/food/plant/plant_propagation_material/plant_variety_catalogues_data bases/search/public/index.cfm?event=SearchForm&ctl_type=A%0Ahttp://ec.europa.eu/food/plant/plant_propagation_material/plant_variety_catalogues_databases/search/public/.

Eurostat (2017) Agriculture, forestry and fishery statistics. Publications Office of the European Union, Luxembourg.

Excoffier, L. & Ray, N. (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, 23, 347–351.

Faville, M.J., Vecchies, A.C., Schreiber, M., Drayton, M.C., Hughes, L.J., Jones, E.S., Guthridge, K.M., Smith, K.F., Sawbridge, T., Spangenberg, G.C., Bryan, G.T., & Forster, J.W. (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics*, 110, 12–32.

Feurdean, A., Marinova, E., Nielsen, A.B., Liakka, J., Veres, D., Hutchinson, S.M., Braun, M., Timar-Gabor, A., Astalos, C., Mosburgger, V., & Hickler, T. (2015) Origin of the forest steppe and exceptional grassland diversity in Transylvania (central-eastern Europe). *Journal of Biogeography*, 42, 951–963.

Fjellheim, S., Rognli, O.A., Fosnes, K., & Brochmann, C. (2006) Phylogeographical history
of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography*, 33, 1470–1478.

Fracassetti, M., Griffin, P.C., & Willi, Y. (2015) Validation of Pooled Whole-Genome Re-Sequencing in *Arabidopsis lyrata*. *Plos One*, 10, e0140462.

Garrick, R.C., Bonatelli, I.A.S., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., Rice, E., Satler, J.D., Symula, R.E., Thomé, M.T.C., & Carstens, B.C. (2015) The evolution of phylogeographic data sets. *Molecular Ecology*, 24, 1164–1171.

Giesecke, T., Brewer, S., Finsinger, W., Leydet, M., & Bradshaw, R.H.W. (2017) Patterns and dynamics of European vegetation change over the last 15,000 years. *Journal of Biogeography*, 44, 1441–1456.

Goslee, S.C. & Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22, 1–19.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., & Bustamante, C.D. (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5, e1000695.

Hejcman, M., Hejcmanová, P., Pavlů, V., & Beneš, J. (2013) Origin and history of grasslands in Central Europe—a review. *Grass and Forage Science*, 68, 345–363.

Hijmans, R.J., Williams, E., & Vennes, C. (2012) geosphere: Spherical Trigonometry. R package version 1.2–28. *package geosphere*, https://CRAN.R-project.org/package=geosphere.

Hooper, W.D. & Ash, H.B. (1935) Cato and Varro on agriculture. *Loeb Classical Library No. 283*. Harvard: Harvard University Press.

Hughes, A.R., Inouye, B.D., Johnson, M.T.J.J., Underwood, N., & Vellend, M. (2008) Ecological consequences of genetic diversity. *Ecology Letters*, 11, 609–623.

Humphreys, M., Feuerstein, U., Vandewalle, M., & Baert, J. (2010) Ryegrasses. *Fodder Crops and Amenity Grasses* pp. 211–260. Springer New York, New York, NY.

Huyghe, C., De Vliegher, A., Van Gils, B., & Peeters, A. (2014) *Grasslands and herbivore production in Europe and effects of common policies*. Editions Quae, Versailles, France.

Inda, L.A., Sanmartín, I., Buerki, S., & Catalán, P. (2014) Mediterranean origin and Miocene-
Holocene Old World diversification of meadow fescues and ryegrasses (*Festuca* subgenus *Schedonorus* and *Lolium*). *Journal of Biogeography,* 41, 600–614.

Jombart, T., Devillard, S., Balloux, F., et al. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* 11, 94.

Jones, M.B. & Donnelly, A. (2004) Carbon sequestration in temperate grassland ecosystems and the influence of management, climate and elevated CO2. *New Phytologist,* 164, 423–439.

Kuneš, P., Svobodová-Svitavská, H., Kolář, J., Hajnalová, M., Abraham, V., Macek, M., Tkáč, P., & Szabó, P. (2015) The origin of grasslands in the temperate forest zone of east-central Europe: Long-term legacy of climate and human impact. *Quaternary Science Reviews,* 116, 15–27.

De La Torre, A.R., Li, Z., Van de Peer, Y., & Ingvarsson, P.K. (2017) Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular Biology and Evolution,* 34, 1363–1377.

Lambeck, K. & Chappell, J. (2001) Sea Level Change Through the Last Glacial Cycle. *Science,* 292, 679–686.

Leclère, D., Valin, H., Frank, S., & Havlík, P. (2016) Assessing the land use change effects of using EU grassland for biofuel production. Task 4b of tender ENER/C1/2013-412.

ECOFYS Netherlands B.V. European Commission.

Legendre, P., Lapointe, F., & Casgrain, P. (1994) Modeling brain evolution from behavior: a permutational regression approach. *Evolution,* 48, 1487–1499.

Lynch, M., Bost, D., Wilson, S., Maruki, T., & Harrison, S. (2014) Population-Genetic Inference from Pooled-Sequencing Data. *Genome Biology and Evolution,* 6, 1210–1218.

Monestiez, P., Goulard, M., & Charmet, G. (1994) Geostatistics for spatial genetic structures: study of wild populations of perennial ryegrass. *Theoretical and applied genetics,* 88, 33–41.

Nair, R.M. (2004) Developing tetraploid perennial ryegrass (*Lolium perenne* L.) populations. *New Zealand Journal of Agricultural Research,* 47, 45–49.
Peeters, A. (2012) Past and future of European grasslands. The challenge of the CAP towards 2020. *Grassland Science in Europe, 17*, 17–32.

Pembleton, L.W., Cogan, N.O.I., & Forster, J.W. (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources, 13*, 946–952.

Pérez-Obiol, R. & Julià, R. (1994) Climatic Change on the Iberian Peninsula Recorded in a 30,000-Yr Pollen Record from Lake Banyoles. *Quaternary Research, 41*, 91–98.

Pickrell, J.K. & Pritchard, J.K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics, 8*, e1002967.

Prieto, I., Violle, C., Barre, P., Durand, J.L., Ghesquiere, M., & Litrico, I. (2015) Complementary effects of species and genetic diversity on productivity and stability of sown grasslands. *Nature Plants, 1*, 15033.

Rabineau, M., Berné, S., Olivet, J.L., Aslanian, D., Guillocheau, F., & Joseph, P. (2006) Paleo sea levels reconsidered from direct observation of paleoshoreline position during Glacial Maxima (for the last 500,000 yr). *Earth and Planetary Science Letters, 252*, 119–137.

Rode, N.O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J., & Gay, L. (2018) How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Molecular Ecology Resources, 18*, 194–203.

Rohling, E.J., Fenton, M., Jorissen, F.J., Bertrand, P., Ganssen, G., & Caulet, J.P. (1998) Magnitudes of sea-level lowstands of the past 500,000 years. *Nature, 394*, 162–165.

Sampoux, J.P., Baudouin, P., Bayle, B., Béguier, V., Bourdon, P., Chosson, J.F., de Bruijn, K., Deneufbourg, F., Galbrun, C., Ghesquière, M., Noël, D., Tharel, B., & Viguié, A. (2013) Breeding perennial ryegrass (*Lolium perenne* L.) for turf usage: An assessment of genetic improvements in cultivars released in Europe, 1974-2004. *Grass and Forage Science, 68*, 33–48.

Sampoux, J.P., Baudouin, P., Bayle, B., Béguier, V., Bourdon, P., Chosson, J.F., Deneufbourg, F., Galbrun, C., Ghesquière, M., Noël, D., Pietraszek, W., Tharel, B., & Viguié, A. (2011) Breeding perennial grasses for forage usage: An experimental assessment of trait changes in diploid perennial ryegrass (*Lolium perenne* L.) cultivars released in the last four decades. *Field Crops Research, 123*, 117–129.
Schlötterer, C., Tobler, R., Kofler, R., Nolte, V., Schlotterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014) Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics, 15*, 749–763.

Semelová, V., Hejcman, M., Pavlů, V., Vacek, S., & Podrázský, V. (2008) The Grass Garden in the Giant Mts. (Czech Republic): Residual effect of long-term fertilization after 62 years. *Agriculture, Ecosystems and Environment, 123*, 337–342.

Sham, P., Bader, J.S., Craig, I., O’Donovan, M., & Owen, M. (2002) DNA pooling: A tool for large-scale association studies. *Nature Reviews Genetics, 3*, 862–871.

Sukumaran, J. & Holder, M.T. (2015) SumTrees: phylogenetic tree summarization. 4.0. 0. *Program and documentation available from the authors at*: https://github.com/jeetsukumaran/Dendrophy.

Terrell, E.E. (1966) Taxonomic implications of genetics in ryegrasses (*Lolium*). *The Botanical Review, 32*, 138–164.

Terrell, E.E. & Ekrem, T. (1968) *A taxonomic revision of the genus Lolium*. US Dept. of Agriculture, Washington, D.C.

Thiele, J., Jørgensen, R.B., & Hauser, T.P. (2009) Flowering does not decrease vegetative competitiveness of *Lolium perenne*. *Basic and Applied Ecology, 10*, 340–348.

Tilman, D., Wedin, D., & Knops, J. (1996) Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature, 379*, 718–720.

Viaud, V., Santillán-Carvantes, P., Akkal-Corfini, N., Le Guillou, C., Prévost-Bouré, N.C., Ranjard, L., & Menasseri-Aubry, S. (2018) Landscape-scale analysis of cropping system effects on soil quality in a context of crop-livestock farming. *Agriculture, Ecosystems & Environment, 265*, 166–177.

Wang, I.J. (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution, 67*, 3403–3411.

Werling, B.P., Dickson, T.L., Isaacs, R., Gaines, H., Gratton, C., Gross, K.L., Liere, H., Malmstrom, C.M., Meehan, T.D., Ruan, L., Robertson, B.A., Robertson, G.P., Schmidt, T.M., Schrottenboer, A.C., Teal, T.K., Wilson, J.K., & Landis, D.A. (2014) Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy
landscapes. *Proceedings of the National Academy of Sciences*, **111**, 1652–1657.

Zohary, D., Hopf, M., & Weiss, E. (1988) Domestication of plants in the old world: the origin and spread of cultivated plants in west Asia, Europe, and the Nile Valley. *Clarendon Press*, 243.

**Biosketch**

José Luis Blanco Pastor is a molecular ecologist. Specifically his research is focused on the genetics of plant adaptation to climate. He is particularly interested in the transfer of ecology and evolutionary biology research towards the agricultural sector. JLBP, SM, PB, EW, KJD, MH, TR, IRR and JPS designed research; JLBP analysed data; JPS, AMR, EW, KJD, MH, HM, TR and TL collected data; JLBP, SM, PB, AEG and JPS interpreted results; JLBP and JPS wrote the manuscript with feedback from SM, TR and IRR. All authors participated in the edition of the manuscript.

**Data accessibility Statement**

The genetic data reported in this study are available in the NCBI Short Read Archive (SRA) database through accession SRP136600.
Supporting Information

Appendix S1. Supplementary Methods:

Plant Material and Genotyping

DAPC analysis

Appendix S2. Supplementary Figures:

Fig. S1. Histogram of mean read depth for each locus across populations.

Fig. S2. Value of BIC for each K number of clusters as obtained with the k-means algorithm implemented in the R package ‘adegenet’ applied to the Lolium perenne set (natural populations of L. perenne).

Fig. S3. Geographical patterns of genetic differentiation in L. perenne. Scatterplot of longitude vs principal component 1 and of latitude vs principal component 2 after a PCA on allele frequencies of L. perenne natural populations.

Fig. S4. Scatterplot of the first two principal axes from a PCA of the 470 natural populations from the L. perenne set with DAPC cluster assignments.

Fig. S5. Genetic structure of the Lolium set (natural populations of L. perenne and related taxa) based on the DAPC analysis of 79 cpDNA SNPs.

Fig. S6. Lolium set (natural populations of L. perenne and related taxa) and L. perenne set (natural populations of L. perenne only) TREEMix models.

Fig. S7. Schematic representation of Lolium set and L. perenne set δαδι models.

Fig. S8. 50% Majority-rule consensus tree of 552 accessions of Lolium perenne and related taxa (all genotyped accessions).

Fig. S9. Scatterplot of the first two principal axes from a PCA of 552 accessions of Lolium perenne and related taxa (all genotyped accessions).

Fig. S10. Geographical distribution of NA values across L. perenne natural populations (L. perenne set).

Fig. S11. Alternative scenario for the range expansion in L. perenne. It shows a postglacial expansion through the West and Central East and next towards Eastern Europe (clockwise movement around the Alps).
| Table | Description |
|-------|-------------|
| S1    | Accessions from *Lolium perenne* and related taxa used in the study. |
| S2    | *Fst* statistics between seven clusters identified with the k-means algorithm implemented in adegenet of the *L. perenne* set (natural populations of *L. perenne*). |
| S3    | Parameter estimates and statistics from the fitting of 12 alternative $\delta a_0 \delta i$ interspecific models of gene flow using the *Lolium* set (natural populations of *L. perenne* and related taxa). |
| S4    | Parameter estimates and statistics from the fitting of 12 alternative $\delta a_0 \delta i$ interspecific models of gene flow using the *Lolium* set with $Ni$ and $Ti$ values converted to numbers of individuals and years, respectively. |
| S5    | Maximum Likelihood parameter estimates and non-parametric bootstrap 95% confidence interval of the best *Lolium* set gene flow model. |
| S6    | Parameter estimates and statistics from the fitting of 12 alternative $\delta a_0 \delta i$ demographic models using the *L. perenne* set. |
| S7    | Parameter estimates and statistics from the fitting of 12 alternative $\delta a_0 \delta i$ demographic models using the *L. perenne* set with $Ni$ and $Ti$ values converted to numbers of individuals and years, respectively. |
| S8    | Maximum Likelihood parameter estimates and non-parametric bootstrap 95% confidence interval of the best *L. perenne* set demographic model. |
| S9    | cpDNA primer sequences used for HiPlex amplicon sequencing. |