

Aberystwyth University

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

Blanco-Pastor, 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):

Blanco-Pastor, 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, 46(7), 1451-1465. https://doi.org/10.1111/jbi.13587

Document License CC BY-NC

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or You may not further distribute the material or use it for any profit-making activity or commercial gain
 You may not further distribute the material or use it for any profit-making activity or commercial gain

- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400 email: is@aber.ac.uk

1	Pleistocene climate changes, and not agricultural spread, account for range
2	expansion and admixture in the dominant grassland species Lolium perenne
3	<i>L</i> .
4	Running title: Phylogeography of perennial ryegrass
5	J. L. Blanco-Pastor ¹ , S. Manel ² , P. Barre ¹ , A.M. Roschanski ³ , E. Willner ³ , K. J. Dehmer ³ , M.
6	Hegarty ⁴ , H. Muylle ⁵ , T. Ruttink ⁵ , I. Roldán-Ruiz ^{5, 6} , T. Ledauphin ¹ , A. Escobar-Gutierrez ¹ , J.
7	P. Sampoux ¹
8	
9	1. INRA, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F), 86600 Lusignan, France;
10	2. EPHE, PSL Research University, CNRS, UM, SupAgro, IRD, INRA, UMR 5175
11	CEFE, F-34293 Montpellier, France;
12	3. Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Inselstr. 3, 23999
13	Malchow/Poel, Germany;
14	4. Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth
15	University, Aberystwyth, Ceredigion, UK;
16	5. Flanders Research Institute for Agriculture, Fisheries and Food (ILVO) - Plant
17	Sciences Unit, Caritasstraat 39, 9090 Melle, Belgium;
18	6. Ghent University, Faculty of Sciences, Dept. of Plant Biotechnology and
19	Bioinformatics, Technologiepark 927, 9052 Ghent, Belgium.
20	
21	Corresponding author:
22	J. L. Blanco-Pastor
23	INRA, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F)
24	Le Chêne - RD 150 CS 80006
25	86600 Lusignan, France
26	Phone: +33(0)549556031
27	Email: jose-luis.blanco-pastor@inra.fr
28	
29	Handling Editor
20	Hong Datas Comos

30 Hans-Peter Comes

31 Acknowledgements

32 J. L. Blanco-Pastor has received the support of the European Commission (EC) in the

- 33 framework of the Marie-Curie FP7 COFUND People Program, through the award of an
- 34 AgreenSkills+ fellowship (grant agreement n° 609398). This work was funded in the frame of
- 35 the project GrassLandscape awarded by the 2014 FACCE-JPI ERA-NET+ call Climate Smart
- 36 Agriculture. Funding was granted by the EC (grant agreement n° 618105), by the Agence
- 37 Nationale de la Recherche (ANR) and the Institut National de la Recherche Agronomique
- 38 (INRA métaprogramme ACCAF) in France, the Biotechnology and Biological Sciences
- 39 Research Council (BBSRC) in the United-Kingdom, the Bundesantalt für Landwirtschaft und
- 40 Ernährung (BLE) in Germany. The computational resources (Stevin Supercomputer
- 41 Infrastructure) and services used for genotype calling were provided by the VSC (Flemish
- 42 Supercomputer Center), funded by Ghent University, FWO and the Flemish Government –
- 43 department EWI.

44 The authors thank the curators from the genebanks that provided perennial ryegrass seed

- 45 samples for the needs of the project and staff from European agronomic research institutes
- 46 who contributed to in situ collections in 2015. Perennial ryegrass is one of the plant species
- 47 covered under the Multilateral System of the International Treaty on Plant Genetic Resources
- 48 for Food and Agriculture. All genetic materials used in this study were made available to the
- 49 authors after signature of a Standard Material Transfer Agreement (SMTA) by the provider
- 50 and the recipient. Implementation and signature of a SMTA provides compliance with the
- 51 provisions of the Nagoya Protocol for parties wishing to provide and receive genetic material
- 52 under the Multilateral System. The authors declare no conflict of interest.
- 53

54 Abstract

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

59 **Location**: Europe.

60 **Taxon**: *Lolium perenne* L. (perennial ryegrass).

61 **Methods:** We reconstructed the phylogeographic history of *L. perenne*, a dominant grassland

- 62 species, using 481 natural populations, including 11 populations of closely related taxa. We
- 63 combined Genotyping-by-Sequencing (GBS) and pool-Sequencing (pool-Seq) to obtain high-
- 64 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
- 67 cultivars representative of the *L. perenne* cultivars widely used for forage purposes.
- 68 **Results:** Expansion across Europe took place during the Würm glaciation (12-110 kya), a
- 69 cooling period that decreased the dominance of trees in favour of grasses. Splits and
- 70 admixtures in *L. perenne* fit historical climate changes in the Mediterranean basin. The
- 71 development of agriculture in Europe (7-3.5 kya), that caused an increase in the abundance of
- 72 grasslands, did not have an effect on the demographic patterns of *L. perenne*. We found that
- 73 most modern cultivars are closely related to natural diversity from North-Western Europe.
- 74 Thus, modern cultivars do not represent the wide genetic variation found in natural
- 75 populations.
- 76 Main conclusions: Demographic events in *L. perenne* can be explained by the changing
- 77 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
- 80 climate change, safeguarding the agricultural services they provide.
- 81
- 82 Keywords: Europe, Genetic diversity, Grasslands, Perennial ryegrass, Phylogeography,
- 83 Quaternary, Cultivar, Genotyping-by-Sequencing, Pool-seq, Site frequency spectrum.
- 84

85 Introduction

- 86 Worldwide, grasslands constitute the most extensive natural and semi-natural habitat types
- 87 and are an integral part of agricultural landscapes. They have been essential for the
- 88 maintenance of biodiversity, carbon sequestration and biogeochemistry of soils across the last
- 89 millennia (Tilman et al., 1996; Jones & Donnelly, 2004; Hejcman et al., 2013). Grasslands
- 90 support biodiversity and a variety of ecosystem services that have gained renewed interest and
- 91 value to society (Werling et al., 2014). In Europe, they are the most widespread agricultural

92 land-use covering 45% of the total agricultural area (Eurostat, 2017). Most of these grasslands

93 are permanent natural or semi-natural grasslands (84%; Leclère et al., 2016) and are

94 composed of plant species and populations that may have evolved under natural

95 environmental constraints and farming usages since the earliest stages of herbivore

96 domestication and agriculture, i.e. about 10 kya in the Fertile Crescent (Hejcman et al., 2013).

97 The rationalisation of farming technics in the 19th and 20th centuries has promoted the

98 inclusion of temporary grasslands (or meadows) into crop rotation systems. These temporary

99 meadows are sown for a period of one to five years and are usually managed in a fairly

100 intensive way. Their acreage in Europe remains less important than that of permanent

101 grasslands (9.76 and 56.9 million ha in EU-27, respectively) (Huyghe et al., 2014), but it is

102 likely to increase in the coming decades because of the recognised positive impact of

103 temporary meadows on the sustainability of crop rotation systems (e.g. Crème et al., 2018;

104 Viaud et al., 2018).

105 Since the 1960s, agronomic research centres have developed breeding programs to release 106 improved cultivars of forage species, that ensure high production of forage of good quality 107 (Sampoux et al., 2011). New permanent or temporary grasslands are almost exclusively sown 108 with such cultivars. The release of grassland species cultivars improved for forage 109 performances has contributed to the expansion of the acreage of sown meadows during the 110 last 50 years, whereas the acreage of natural and semi-natural permanent grasslands has 111 continuously decreased (Chapman, 1992). In the EU-6 (first six members of the European 112 Union), losses of permanent grasslands are estimated at about 30% (7 million ha) between 113 1967 and 2007 whereas losses in EU-15 are estimated at 15% (10.5 million ha) during the last 114 50 years (Peeters, 2012).

115 The maintenance of plant genetic diversity in grassland species can be of major importance 116 not only in terms of adaptive potential but also in terms of productivity and ecosystem 117 services (Crutsinger et al., 2006; Hughes et al., 2008). It has been shown that intra-population 118 genetic diversity may favour the temporal stability of production in grasslands (Prieto et al., 119 2015). Large-scale ecotype variation in grassland species genetic diversity is also likely to 120 contribute to the resilience of grassland production along climatic and other environmental 121 gradients, as suggested for the dominant grass species L. perenne by investigations of 122 Balfourier & Charmet (1991a) and Monestiez et al. (1994). Given the reduction in natural 123 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.

126 Reconstructions from the pollen record revealed the extensive presence of open habitat (*i.e.* 127 grasslands) in Europe since the pre-Holocene (> 12 ka ago –kya) (Pèrez-Obiol & Julià, 1994; 128 Kuneš et al., 2015). Grasslands in central-eastern Europe, for example, constituted up to 30% 129 of the vegetation during the early-Holocene and their relative abundance has been maintained 130 to the present (Kuneš et al., 2015). Although grasslands were widespread long before 131 agricultural practices became established in Europe ca. 8 kya (Zohary et al., 1988; Hejcman et 132 al., 2013; Giesecke et al., 2017), activities of early agricultural communities have been 133 important for shaping current European vegetation and maintaining open land (Feurdean et 134 al., 2015). Pollen records clearly show that the area dominated by grasses in Europe has 135 increased since 4 kya, suggesting that the conversion of forest landscapes into grasslands was 136 associated with agricultural activity (Giesecke et al., 2017). However, the effect of this 137 human transformation of the European landscape on the natural diversity and population 138 structure of grassland species is unknown. More specifically, little is known about the relative 139 importance of natural and human-mediated expansions to explain the current distribution of 140 grassland species genetic variation in natural and semi-natural grasslands. Additionally, the 141 extent to which the increased use of cultivars has eroded genetic diversity within natural 142 grasslands remains unevaluated. We used L. perenne as a model to investigate these questions 143 because this species is the most widely sown forage grass species in temperate regions 144 (Humphreys et al., 2010), has received extensive breeding effort during the last five decades 145 (Humphreys et al., 2010; Sampoux et al., 2011) and is also one of the most abundant grass 146 species in natural grasslands across Europe and the Fertile Crescent.

147 One complicating factor for the study of L. perenne is the fact that the genus Lolium 148 comprises nine species (Terrell & Ekrem, 1968; Charmet et al., 1996) which diverged only 149 recently (ca. 4.1 Ma ago; Inda et al., 2014) and that most of them can naturally intercross. 150 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., 151 152 2014). A previous study based on chloroplast DNA (cpDNA) polymorphisms suggested that 153 natural L. perenne populations could have been introduced to Europe following human 154 migration routes from the Fertile Crescent as a weed of cereals (Balfourier et al., 2000). More 155 recently, a study based on the analysis of 2185 transcript-anchored single nucleotide 156 polymorphisms (SNPs) suggested that L. perenne was subjected to repeated population

- 157 expansion and contraction during the Quaternary glaciations (Blackmore et al., 2015).
- 158 However, these studies did not estimate the time of demographic events, which is required to
- 159 firmly discard either of the two hypothesis.
- 160 Thanks to the advances in sequencing technologies (Emerson et al., 2010; Garrick et al., 161 2015), assessing fine-scale phylogenetic, phylogeographic and hybridisation patterns among 162 recently diverged lineages is now an achievable task. Here, we present a comprehensive study 163 of the genomic diversity of natural L. perenne populations and reconstruct historical 164 demographic events that contributed to shape this diversity, using an extensive set of 165 populations sampled in natural and semi-natural grasslands across most of the natural 166 distribution range of the species. Better knowledge of the distribution of the natural genomic 167 variation of L. perenne across its area of primary expansion would indeed provide essential 168 information to guide the conservation strategy for this major grassland species in a context of 169 reduction of permanent natural grasslands. This would also contribute to guide the informed 170 use of natural populations as genetic resources for plant breeding in this species. We used 171 genomic data to evaluate the two competing hypotheses, that large-scale genetic variation in 172 L. perenne could be explained either by Pleistocene climate changes or by more recent 173 human-mediated expansion. More specific objectives were: (i) to investigate the relationships 174 between L. perenne and close relatives; (ii) to reconstruct and date the main demographic 175 processes that occurred along the evolutionary history of L. perenne and (iii) to gain insight 176 into the origin of L. perenne cultivars and trace the use of natural genetic resources in this 177 species by modern plant breeding.
- 178 Materials and Methods

179 Plant Material and Genotyping

180 We obtained a batch of seeds from 476 accessions of L. perenne and related taxa maintained 181 as seed lots in the genebanks of agronomic research institutes from different countries. These 182 seed lots were made as to represent the genetic diversity existing within each of 476 natural 183 populations sampled in natural and semi-natural grasslands across Europe and the Near-East 184 (See Appendix S1 and Table S1 in Appendix S3). Monestiez et al. (1994) analysed spatial 185 autocorrelation patterns in phenotypic traits of natural L. perenne and identified two main 186 multivariate spatial structures with 120 and 300km ranges, respectively. The 120km range 187 was interpreted as a result of isolation-by-distance (gene flow) and the 300km one as a signal 188 of selection imposed by environmental factors. In accordance with these findings, we set 189 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.

194 Accessions from genebanks and populations newly collected in 2015 were grown in an 195 experimental garden which enabled to perform a taxonomic assessment and a flow cytometry 196 control of the ploidy level of these materials. Natural populations of *L. perenne* and related 197 taxa are indeed expected to be diploid whereas some L. perenne cultivars, not used in this 198 study, are artificial tetraploids (Beddows, 1967; Nair, 2004; Humphreys et al., 2010). 48 199 genebank accessions were however not present in the experimental garden because of 200 insufficient seed availability. After sequencing and bioinformatics processing (see below), we 201 finally used three different accession sets for downstream analyses (see Phylogeny and 202 population filter, Appendix S1): i) 470 monophyletic L. perenne natural diploid populations 203 (L. perenne set), ii) the L. perenne set plus 11 diploid populations from other taxa (Lolium set) 204 and iii) 32 L. perenne diploid cultivars (cultivar set). The 11 addditional populations of the 205 Lolium set were five populations of L. multiflorum, two of L. rigidum, two of L. temulentum 206 and two of *Festuca pratensis* (outgroup). Previous studies acknowledged the close 207 phylogenetic relationship between the Lolium genus and broad-leaved fescues from the 208 Festuca genus including Festuca pratensis (Catalán et al., 2004). Note that L. temulentum is 209 an autogamous taxon whereas the three others are allogamous like *L. perenne*. For the sake of 210 reliability of downstream analyses, natural populations not grown in the experimental garden 211 were neither included in the *L. perenne* set nor in the *Lolium* set, except 12 ones. The latter 212 showed clear sister genomic relationships with some other natural L. perenne diploid 213 populations grown in the experimental garden and were thus included as such in the L.

214 *perenne* set.

215 We used GBS pool-Seq (Byrne et al., 2013) to determine genome-wide allele frequencies of

216 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

218 frequency estimations (Sham et al., 2002; Lynch et al., 2014; Schlötterer et al., 2014;

219 Fracassetti et al., 2015; Rode et al., 2018). Genotyping a large set of *L. perenne* populations

220 (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)

222 contained population alternative allele frequencies (AAFs) 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
- 226 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.
- 230 Phylogeography and past demography
- 231 Population structure
- 232 To investigate the presence of genetic clusters in the *L. perenne* set, we applied the
- 233 Discriminant Analysis of Principal Components (DAPC, Jombart et al., 2010) to the table of
- 234 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'
- 236 (Pembleton et al., 2013). Additionally, we calculated the expected heterozygosity (*He*) of
- 237 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*
- 239 set (Pool-GBS and HiPlex cpDNA data, see Appendix S1).
- 240 Isolation by distance (IBD) *versus* isolation by environment (IBE)
- 241 To test the effect of IBD and IBE on the Nei genetic distances (nuclear dataset) between the
- 242 L. perenne set populations, we computed a geographical distance matrix with the R package
- ²⁴³ 'geosphere' (Hijmans et al., 2012) and an Euclidean bioclimatic (environmental) distance
- 244 matrix using 19 bioclimatic data layers (*sensu* WorldClim database, bio1-bio19;
- 245 <u>http://www.bioclim.org</u>). Normalized matrices were used to assess IBD and IBE with the
- 246 Multiple regression on distance matrices (MRM) function (Legendre et al., 1994; Wang,
- 247 2013) implemented in the R package 'ecodist' (Goslee & Urban, 2007).
- 248 Splits and admixture
- 249 Setting up demographic models requires prior information about the relationships between the
- considered populations. Splits and admixture analyses as implemented in TREEMIX v.1.13
- 251 (Pickrell & Pritchard, 2012) provide the prior information necessary to set up alternative
- 252 demographic models (main tree topologies and migration directions). For TREEMIX analyses,
- 253 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 255 256 applied to L. perenne clusters plus the other taxa: F. pratensis, L. multiflorum, L. temulentum 257 and L. rigidum, i.e. to the Lolium set. The second analysis was applied to clusters of L. 258 perenne natural diversity only, *i.e.* to the *L. perenne* set. For both TREEMIX analyses, we first 259 inferred a maximum likelihood (ML) tree without admixture and then ran 1000 standard 260 bootstrap replicates to obtain statistical support for the non-admixed tree topology. Bootstrap 261 trees were summarized in a 50% majority-rule consensus tree with SUMTREES v4.2.0 262 (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 263 264 events between L. multiflorum, L. rigidum and L. perenne observed in alternative Lolium set 265 TREEMIX models (TMMs) and all migration events between L. perenne clusters observed in 266 alternative L. perenne set TMMs were considered for the generation of demographic models 267 with $\delta a \delta i$ (see below).

268

269 Demography

270 To further investigate the hybridisation patterns between the Lolium taxa included in our 271 study and the demographic history of L. perenne, we used the program $\delta a \delta i$ (Gutenkunst et 272 al., 2009). $\delta a \delta i$ compares the site frequency spectrum (SFS) expected under custom 273 demographic models to that observed with actual frequency data. SFSs are simulated with a 274 diffusion approach which is limited to a three-taxon phylogeny in $\delta a \delta i$. Comparisons of $\delta a \delta i$ 275 models were made independently for the *L. perenne* and the *Lolium* sets. For the construction 276 of $\delta a \delta i$ models (see Fig. S6 in Appendix S2), we used alternative migration/hybridisation 277 scenarios as obtained from TREEMIX. First, we investigated alternative scenarios of gene flow 278 among L. rigidum, L. multiflorum and L. perenne as shown in the Lolium set TMMs 1-6 (see 279 below). Second, we investigated demography and patterns of gene flow within L. perenne, as 280 displayed in the L. perenne set TMMs 1-6 (see below). For the first analysis, we averaged 281 nuclear AAFs in all three species. For the second one, we averaged nuclear AAFs of 282 populations from clusters 1 to 5 and considered clusters 6 and 7 as independent lineages (as 283 displayed in the main topology of L. perenne set TMMs 1-6, see below). For both analyses, 284 we used F. pratensis as outgroup to establish the ancestral state for each SNP. Some $\delta a \delta i$ 285 models were based on a non-equilibrium demography (cluster splits with a period of isolation 286 before admixture, as assumed in TREEMIX) whereas some others were based on an 287 equilibrium demography (fixed migration structure since divergence). We also considered the

288 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

- 290 independent runs were used for the optimisation of each model. Each run started from a
- 291 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
- 293 the runs. Fitted models were ranked according to the Akaike information criterion (AIC) to
- account for the variable number of model parameters.
- 295 To set confidence intervals (CIs) for parameter estimates of the best-fit demographic models, 296 we generated 100 datasets by non-parametric bootstrapping. Bootstrap replicates were re-297 optimized in $\delta a \delta i$ to estimate the parameter uncertainties. CIs were calculated as $E \pm 1.96\sigma$, 298 where E is the ML parameter estimate and σ is the standard deviation of the parameter 299 estimate across the bootstrap replicates. To assess the time of demographic events, it is 300 necessary to incorporate the average generation time of taxa. L. perenne requires vernalisation 301 to flower (Thiele et al., 2009). After seed germination and emergence followed by first winter 302 vernalisation, the seed production is often fairly small because of limited tillering. Seed 303 production then peaks after second and third winter vernalistion. In favourable environmental 304 conditions, some perennial ryegrass clones can live for more than 10 years (Beddows, 1967). 305 Ramet density decreases gradually after the third year, but a small seed production can last in 306 late years of life of clones. Considering these facts, we assumed an average of three years 307 generation time (3y/gen) for L. perenne. L. rigidum is an annual taxon and L. multiflorum 308 includes mostly annual but also biennial types (Terrell & Ekrem, 1968). We assumed one year 309 generation time (1y/gen) for these two taxa. Using these generation times and a known 310 average mutation rate of 6.03E-9 substitutions per site per generation for Poaceae (De La 311 Torre et al., 2017), we converted effective population sizes (Ni) and Time (Ti) parameters to 312 (breeding) individuals and years.

313 Origin of cultivars

- 314 We predicted the cluster membership of cultivars by adding them as supplementary
- 315 populations in the DAPC analysis of the nuclear AAFs of the *L. perenne* set considered as
- 316 "training data". We transformed the allele frequencies of these supplementary entries using
- 317 the centring and scaling metrics of the "training data" and determined their position onto the
- 318 discriminant axes.

319 Results

320 Genetic structure in *L. perenne* natural diversity

321 The k-means algorithm analysis of the nuclear AAFs of the *L. perenne* set detected K = 7 as 322 the optimal number of genetic clusters as given by the Bayesian Information Criterion (BIC) 323 (Fig. 1 and Fig. S2 in Appendix S2). The genetic clusters exhibited a low level of admixture 324 and a clear geographical structure (Fig. 1a): cluster 1 fits to the South-Eastern Europe – Near 325 East region (black colour in Fig. 1), cluster 2 to Eastern Europe (orange), cluster 3 to North-326 Eastern Europe (light blue), cluster 4 to Northern Iberia – Southern France (green), cluster 5 327 to North-Western Europe (pink), cluster 6 to Corsica - Sardinia (yellow) and cluster 7 to 328 Northern Italy (red). Admixed populations were mainly assigned to cluster 3 and cluster 5 329 (Fig. 1b-1c). Additional DAPC analyses performed with K = 8-10 detected seven clusters 330 with the same geographic structure as mentioned above (results not shown). Cluster 8 was 331 formed by populations with an admixture signal between clusters 3 and 5 that were located 332 between the distribution areas of these two clusters. Cluster 9 was also formed by a small 333 number (11) of admixed populations with ancestry from cluster 3 and cluster 5 that were 334 distributed in South-Eastern Europe and the Near East; we interpreted that these populations 335 were not native from this area but imported from Western Europe and recently sown. Cluster 336 10 was formed by populations from cluster 1 that were located in the North-Western part of 337 cluster 1 distribution. Because cluster 8 and cluster 9 reflected admixture and inconsistent 338 geographical distribution (cluster 9) rather than divergence signal (reproductive isolation), we 339 considered that K = 7 as given by the BIC was appropriate for downstream analyses and 340 further discussion. With K = 7, the first DAPC axis was strongly correlated with longitude (r= 341 0.810 p-value < 0.001) and the second axis with latitude (r= 0.662 p-value < 0.001), 342 indicating different and independent directions of differentiation along these two geographical 343 dimensions (Fig. S3a-S3b in Appendix S2). Genetic differentiation among clusters was small, 344 with Fst values ranging from 0.0152 (cluster 4-cluster 5) to 0.0776 (cluster 3-cluster 6) (Table 345 S2 in Appendix S3). Average expected heterozygosity (*He*) values within populations ranged 346 from 0.54 (cluster 4) to 0.47 (cluster 6) with standard deviation (STD) within clusters ranging 347 from 5.0E-4 (cluster 5) to 1.1E-2 (cluster 1) (Fig. 2d). Cluster 1 had a remarkably high He 348 STD; this cluster indeed contains a set of populations with high heterozygosity together with 349 populations with very low heterozygosity.

350 The MRM analysis (Fig. 3) revealed that both IBD and IBE played a significant, but

351 moderate, role in genetic differentiation of *L. perenne* (IBD model: $r^2 = 0.185$, p-value <

- 352 0.001; IBE model: $r^2 = 0.112$, p-value < 0.001; Fig. 3a-3b). IBD was more important than IBE
- 353 to explain the genetic distances between populations (IBD + IBE model: $r^2 = 0.201$, $\beta_D =$
- 354 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, r² = 0.282, p-value < 0.001,
- 356 Fig. 3d).
- 357 The DAPC analysis carried out with 79 cpDNA loci on the *Lolium* set (*L. perenne* and other
- 358 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
- 360 chose K = 4 considering as optimal K the highest possible number of clusters that showed
- 361 non-admixed populations. cpDNA clusters comprised members of the different *Lolium* taxa,
- 362 with clusters 1 and 2 showing a higher presence in Mediterranean areas (Fig. S5).

363 Splits and admixture

- 364 The likelihood of the Lolium set TREEMIX models (TMMs) increased almost linearly with no 365 clear stabilisation of likelihood values as the number of admixture edges increased (Fig. S6a). 366 When adding two or more admixture edges (M2-M6), the topology of the main tree was 367 rearranged and cluster 6 became the most basal lineage. The ln-likelihood of the Lolium set 368 TMM increased by 591.37 from M0 to M1, 779.78 from M1 to M2 and 337.28 from M2 to 369 M3. All additional edges (M4-M6 models) increased the In-likelihood by less than 212. It is 370 important to take into consideration that the increase of likelihood values does not mean that 371 the model is closer to the true network. This is because the addition of admixture edges can 372 never reduce the likelihood (Pickrell & Pritchard, 2012).
- 373 We analysed the *L. perenne* set with TREEMIX using cluster 6 as outgroup as displayed in the
- 374 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
- 376 (Fig. S6b). The In-likelihood of the *L. perenne* set TMM increased by 13232.72 from M0 to
- 377 M1, 5208.19 from M1 to M2 and 984.35 from M2 to M3. All additional edges (M4-M6
- 378 models) increased the ln-likelihood by less than 653.

379 Demography

- 380 *Lolium* set model comparison (Fig. S7a) revealed that the best fit to our observed SFS was
- 381 model C (Fig. 2a and Table S3 and Table S4 in Appendix S3). Positive or negative residuals
- 382 of the model (normalized differences between model and data) indicate that the model
- 383 predicts too many or too few alleles in a given cell of the two-population SFS, respectively.

- 384 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
- 386 95% confidence interval (CI) obtained from non-parametric bootstrapping are shown in Table
- 387 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.*
- 389 *multiflorum*) that was dated 397 kya (95% CI 238-555 kya, 1y/gen) or 1.19 Mya (95% CI
- 390 715-1666 kya, 3y/gen). A subsequent split followed by a period of isolation between *L*.
- 391 perenne and L. multiflorum occurred 380 kya (95% CI 235-525 kya, 1y/gen) or 1.14 Mya
- 392 (95% CI 706-1576 kya, 3y/gen). Then migration from *L. rigidum* to *L. perenne* and from *L.*
- 393 multiflorum to L. perenne with constant population size in L. perenne started 366 kya (95% CI
- 394 235-497 kya, 1y/gen) or 1.10 Mya (95% CI 704-1492 kya, 3y/gen).

395 Model comparison using the L. perenne set (Fig. S7b) revealed that the best fit between the 396 predicted and observed site frequency spectrum (SFS) (*i.e.* the highest maximum composite 397 likelihood) was obtained with model F (Fig. 2b and Table S6 and Table S7 in Appendix S3). 398 Residuals from the best model F followed a normal distribution with a zero mean value (not 399 shown), indicating an appropriate fit of the model to the data. Maximum likelihood parameter 400 values of the best model F and their 95% confidence interval (CI) obtained from non-401 parametric bootstrapping are shown in Table S8 in Appendix S3. The best model F showed an 402 ancestral population split into two lineages (cluster 6 - Corsica-Sardinia- and the ancestor of 403 remaining clusters) that was dated 174 kya (95% CI 49 - 300 kya, 3y/gen). Then a second

- 404 split followed by a period of isolation between ancestor of clusters 1-5 (from Western Europe
- 405 to Near East) and cluster 7 (Northern Italy) and a linear population growth of clusters 1-5
- 406 starting 72 kya (95% CI 31 112 kya, 3y/gen). Finally, migration from cluster 7 to clusters 1-
- 407 5 and from clusters 1-5 to cluster 6 started 56 kya (95% CI 31 81 kya, 3y/gen).
- 408 Additionally, we ran an alternative model to model F in $\delta a \delta i$, in which the latter two
- 409 admixture events did not coincide in time. This new model did not fit as well as the base
- 410 model F (results not shown) suggesting similar timing for the two admixture events.

411 Origin of cultivars

- 412 The predicted membership of the 32 cultivars bred for forage usage on the *L. perenne* set
- 413 DAPC (Fig. 4) showed that 25 out of these 32 cultivars were assigned to cluster 5 (North-
- 414 Western Europe), three to cluster 7 (Northern Italy), two to cluster 2 (Central-Eastern
- 415 Europe), one to cluster 3 (North-Eastern Europe) and one to cluster 4 (South-Western
- 416 Europe). Most of cultivars assigned to cluster 5 were genetically very similar to the natural

- 417 populations of this cluster. Cultivars assigned to the other clusters were highly admixed, with
- 418 high membership probabilities with cluster 5 except the cultivar *Medea*. A DAPC analysis
- 419 with all populations did not separate cultivars from natural populations (not shown). This
- 420 suggests that the genetic origin of cultivars is not restricted to a single source, despite the
- 421 major contribution of cluster 5.

422 **Discussion**

432

- 423 Our analyses of the genomic diversity, structure, and past demography of *L. perenne* reveal
- 424 that despite its extensive use, there remains a regional genetic structure of extant natural
- 425 populations that was shaped by demographic events predating the onset of agriculture. The
- 426 impact of these events is still visible today even if the wide presence of permanent grasslands
- 427 in landscapes, hosting the natural diversity of *L. perenne*, is mostly a result of the
- 428 development of agriculture during the last millennia.
- 429 According to the fossil record, the Late Glacial (*ca.* 13-10 kya) vegetation in Europe was
- 430 dominated by herbaceous communities including a large proportion of grasses (Giesecke et
- 431 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

- 433 temperate deciduous trees and conifers (Feurdean et al., 2015; Giesecke et al., 2017). But
- 434 forests were never fully closed, enabling the persistence of grasslands throughout the
- 435 Holocene (Hejcman et al., 2013). For example, small-scale steppe grasslands of natural origin
- 436 were present in forest-rich areas of Central Europe before the onset of agriculture in the early
- 437 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
- 439 fossil record, reflecting the development and expansion of agriculture in this area of Europe
- 440 (Hejcman et al., 2013; Giesecke et al., 2017). So far, the impact of this anthropogenic
- 441 transformation of the European landscapes on the genetic diversity of key grassland species
- 442 has been scarcely documented. The processes involved in the genesis of grasslands in
- 443 agricultural landscapes are not precisely known. The unconscious selection of grazing-tolerant
- 444 species has certainly played a major role and has been pivotal in the domestication of large
- 445 herbivores. However, the importance of conscious human actions (such as voluntary seed
- 446 harvesting and sowing of grasslands) remains largely unknown, and no fossil record specific
- 447 for *L. perenne* or other major grass species has been reported at documented archaeological
- 448 sites. Nonetheless, we can ascertain that the development of specific practices for the
- 449 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).

453 Our results reveal incongruence between main cpDNA and nuclear DNA signals that could be 454 explained by insufficient data in the cpDNA matrix, but also by different mutation rates 455 among marker types, incomplete sorting of cpDNA polymorphisms and/or by plastid capture 456 within and between Lolium taxa. Our results show that nuclear DNA better captured and 457 retained signatures of L. perenne phylogeographic history. The seven nuclear DNA genetic 458 clusters represent genetic discontinuities that can be attributed to reduced gene flow at 459 geographical barriers such as the Mediterranean Sea (cluster 6), the Alps (clusters 2, 3, 4, 5, 7) 460 and the Carpathians (clusters 1, 2, 3). These clusters also showed consistency when displayed 461 on the first two principal axes of a PCA on nuclear AAFs (Fig. S4 in Appendix S2). 462 Not only geographical barriers but also geographical and environmental distances should account for the genetic differentiation in L. perenne (Fig. 3). Nevertheless, because 463 464 geographical and environmental distances showed moderate correlation (Fig. 3d), it remains 465 difficult to discriminate which of the two factors was the most important to explain genetic 466 differentiation between populations and clusters. We observed genetic clines in our data, but 467 IBD and IBE explained only a limited proportion of genetic variation. As such, the seven 468 clusters detected by the DAPC analysis on nuclear AAFs represent evolutionary coherent 469 lineages that were interpreted as meta-populations geographically differentiated by reduced 470 gene flow at geographical barriers (European mountain ranges, see Fig. 1). These seven 471 clusters were then considered as independent entities for downstream analyses. 472 Because there is evidence of genetic clines combined with the genetic discontinuities 473 represented by DAPC clusters (Fig. 1d, Fig.3 and Table S2 in Appendix S3), and because 474 TREEMIX and $\delta a \delta i$ inferences about migration are based on these barely differentiated

475 clusters, migration rates might have been partly overestimated in our analyses or confounded

476 with clinal variation. However, it should be noted that the migration edges in our best

477 TREEMIX models connected non-sister clusters (see Fig. 2b). Furthermore, our $\delta a \delta i$ analyses

478 favoured a model with a period of isolation before migration over models based on an

479 equilibrium demography; the latter would have had the best fit according to a pure clinal

480 variation.

481 The origin of L. perenne and its close relatives (L. rigidum and L. multiflorum) inferred by our 482 analyses is partially congruent with previous studies (Balfourier et al., 1998; Catalán et al., 483 2004). This includes the early divergence of L. rigidum and the sister relationship between L. 484 *multiflorum* and *L. perenne*. However, we discovered that relationships between these three 485 taxa are more complex than previously proposed, with L. perenne receiving genes from L. 486 rigidum and L. multiflorum after divergence (Fig. 2a- 2c). According to our demographic 487 reconstructions, main divergence events among Lolium taxa took place during the Pleistocene 488 glaciations, long before the onset of agriculture and main migrations of agricultural 489 communities across Europe. This is in agreement with previous inferences from molecular 490 dating of the grass genera *Festuca* and *Lolium* (Inda et al., 2014). Early gene flow from L. 491 rigidum (in the Near East) and L. multiflorum (in the North Italy) to L. perenne (Fig. 2a-2c) is 492 dated 366 kya (1y/gen) or 1.19 Mya (3y/gen). Indeed, in the frame of controlled experiments, 493 it has been demonstrated that L. multiflorum and L. rigidum are completely interfertile with L. 494 perenne (Terrell, 1966). A close relationship between ancestral L. perenne populations from 495 the Near East and L. rigidum was previously inferred by Balfourier et al. (1998). In addition, 496 it is known that both L. perenne and L. multiflorum have been present in the northern plains of 497 Italy since the late Middle Ages (Casler, 2006). The likely long term coexistence of L. 498 perenne and L. rigidum in the Near East and L. perenne and L. multiflorum in the Italian 499 plains, together with their ability to intercross, may explain the introgression from L. rigidum 500 and L. multiflorum to L. perenne in these areas (Fig. 2a-2c). Contrary to the hypothesis of 501 Casler (2006), who assumed that L. multiflorum originated from human-mediated selection in 502 some L. perenne strains in Northern Italy, our results indicate much earlier divergence 503 between these two species in agreement with Inda et al. (2014) (see Fig. 2a and Table S3 and 504 Table S4 in Appendix S3).

505 The first event reconstructed from the best L. perenne $\delta a \delta i$ model (Fig. 2b) is a split between 506 cluster 6 (Corsica-Sardinia) and the ancestor of remaining clusters occurring 174 kya (95% CI 507 49-300 kya, 3y/gen) (Fig. 2c-2e1 and Table S8 in Appendix S3). During the last glaciations 508 (Würm 115-11.7 kya; Riss 347-128 kya), and particularly during sea level low stands (ca. 20, 509 140 and 260 kya), Corsica and Sardinia were connected *inter se* and joined to Tuscany 510 through the Tuscan Archipelago (Rohling et al., 1998; Lambeck & Chappell, 2001; Rabineau 511 et al., 2006), implying that differentiation in these islands followed by rising sea levels during 512 inter/post-glacials could have resulted in the origin of cluster 6 (Fig. 2e1). The Corsica

513 channel might have acted as a barrier for gene flow between *L. perenne* populations from

514 Corsica and Sardinia and Italian refuge lineages already present there (as shown by the date of 515 ancestral admixture from L. multiflorum, Fig. 2a, Fig. 2c and Table S5 in Appendix S3). The 516 next event is the split between cluster 7 (Northern Italy) and ancestor of clusters 1-5, followed 517 by the expansion of two separate evolutionary lineages colonising northwards via Western 518 and Central-Eastern routes, respectively. This is supported by the TREEMIX model with 519 highest likelihood compatible with the best $\delta a \delta i$ model (see Fig. 2B) and admixture signal 520 revealed by the DAPC analysis (Fig. 1). On the other hand, heterozygosity values (Fig. 2d) 521 and second best TREEMIX model (compatible with the best $\delta a \delta i$ model, M2 see Fig. S6) 522 support a postglacial expansion through the West and Central East and next towards Eastern 523 Europe (clockwise movement around the Alps). Note that reduced heterozygosity in cluster 6 524 and cluster 7 (source populations) is likely linked to a reduced effective size in these clusters. 525 We favour the interpretation of two separate evolutionary lineages through the West and 526 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 527 528 split between cluster 7 and clusters 1 to 5 and the start of the range expansion of clusters 1-5 529 are dated 72 kya (Fig. 2b, Fig. 2c and Table S8 in Appendix S3). The split and subsequent 530 start of range expansion overlap with a glacial period (Würm glaciation 12-110 kya) (Fig. 531 2e2). During the Würm glaciation, the Alps might have acted as a barrier to gene flow in L. 532 perenne as previously suggested by Balfourier et al. (1998, 2000) and Blackmore et al. 533 (2015). In continental Europe, the continuous cooling during that period might have affected 534 the dominance of tree species in favour of herbs, possibly including grass species such as L. 535 *perenne*. This is supported by the abundance estimates traced in the pollen record for different 536 vegetation types during the Late Glacial (Giesecke et al., 2017). During the expansion of L. 537 perenne, the ancestor of clusters 1, 2 and 3 might have contacted ancient L. perenne 538 genotypes already located in Eastern Mediterranean areas that had previously admixed with L. 539 rigidum (cluster 1, see Fig 2a and Fig. 2c). This is further supported by the DAPC analysis 540 with K = 10 that identified an additional cluster in the contact zone with cluster 2 (probably as 541 an effect of clinal variation after range expansion), and also by the analysis of genetic 542 diversity (He) of L. perenne clusters (Fig. 2d) which showed that cluster 1 exhibits the most 543 variable within-population genetic diversity. Cluster 1 may have been formed by the mixture 544 between ancient highly diverse populations located in South-Eastern refugia and more recent 545 immigrant populations from the ancestor of clusters 1, 2 and 3, the latter including low 546 genetic diversity due to allele surfing in expanding populations (Edmonds et al., 2004; 547 Excoffier & Ray, 2008). For later stages of the Würm glaciation from 56 kya until current

548 time, the TREEMIX analyses and best $\delta a \delta i$ demographic model suggest migration from cluster 549 7 to clusters 1 and 2 and from cluster 1 to cluster 6 (Fig. 2b, Fig. 2c). During this period, and 550 particularly 20 kya, the Mediterranean sea level reached an estimated maximum drop of 149 551 meters with respect to current level (Fig. 2e3) (Lambeck & Chappell, 2001; Rabineau et al., 552 2006). This may have facilitated dispersal through land bridges in both the Adriatic Sea (from 553 Northern Italy cluster 7 to Central Europe cluster 2) and the Ligurian Sea (from South-Eastern 554 Europe -- but also possibly Central-Southern Italy- cluster 1 to Corsica-Sardinia cluster 6). L. 555 perenne natural populations do exist in Central and Southern Italy (Balfourier & Charmet, 556 1991b, 1994; Balfourier et al., 1998) but we did not have the opportunity to include 557 populations from this region in our study. Balfourier et al. (1998) analysed the genetic 558 structure of 120 wild populations of L. perenne, including populations from Central-Southern 559 Italy, using allelic frequencies from 12 polymorphic isozyme loci. These authors found that 560 populations from Central-Southern Italy were genetically closer to populations of Corsica-561 Sardinia, Eastern Europe and Eastern Mediterranean than to populations of Western and 562 Northern Europe. Our TREEMIX and $\delta a \delta i$ analyses (Fig. 2b and Fig. 2c) combined with results 563 of Balfourier et al. (1998) support a possible connection between South-Eastern Europe -564 Near East cluster 1 and Corsica-Sardinia cluster 6 via Central-Southern Italy.

565 Since the origin and expansion of domesticated plants in the Old World, trade, wars and 566 nomadism caused extensive movements of people and livestock across Europe, possibly 567 favouring transport of diaspores of grassland species among European regions. Nevertheless, 568 all expansion and admixture events in L. perenne recovered by our $\delta a \delta i$ analyses predate the 569 origin of agriculture in Europe, even if taking into account the CI of our time estimates (see 570 Fig. 2e and Table S5 and Table S8 in Appendix S3). The results we report reveal that main 571 genetic signals of colonisation and admixture observed in L. perenne can be explained by 572 climatic events predating the transformation of landscapes by human activities in Europe. A 573 more limited study previously reported similar conclusions for Festuca pratensis (Fjellheim et 574 al., 2006). F. pratensis and L. perenne, two major grass species of European grasslands, may 575 have experienced similar evolutionary histories. They may exemplify an evolutionary trend 576 shared with other European grassland species, in which current patterns in natural genetic 577 diversity were shaped during range expansions of the last glacial period and not significantly 578 disturbed by agricultural expansion.

579 The first detailed written records on intensification of grassland management date from the580 Roman Empire (Hooper & Ash, 1935; Ash et al., 1941). However, the real decline of wild

581 grasslands and the large-scale enlargement of hay meadows started in many European regions 582 during the 18th century (Semelová et al., 2008). This process may have involved increased use 583 of grass seeds harvested from local natural strains to re-sow pastures, but likely without direct 584 selection on phenotypic traits except sufficient seed production. Modern selection of L. 585 *perenne* for forage production based on the agronomic testing of progeny performance began 586 in 1919 in the United Kingdom and after World War II in other European countries, Northern 587 America and New Zealand (Humphreys et al., 2010; Sampoux et al., 2011). Recurrent 588 selection in plant breeding germplasm has so far resulted in the release of more than 1000 L. 589 perenne cultivars improved for forage production (European Commission, 2015). More 590 recently, since the 1960s, similar selection methods have also been implemented to release L. 591 perenne cultivars improved for turf usage (Sampoux et al., 2013). We genotyped 32 cultivars 592 representing a large diversity of the L. perenne cultivars bred for forage usage in various 593 countries of Europe and New Zealand. 25 out of the 32 genotyped cultivars were assigned to 594 the cluster of North-Western Europe (cluster 5) and were thus likely bred from germplasm of 595 this area. Cultivars assigned to other clusters, except Medea, also contained a significant part 596 of diversity from this cluster 5. An interesting pattern is observed in those cultivars created by 597 IBERS (Aberystwyth, UK). Aurora, a cultivar with high water soluble carbohydrate (WSC) 598 content bred from a Swiss ecotype (Faville et al., 2004) showed 60 % membership of cluster 7 599 (Northern Italy). Other more recent IBERS cultivars developed from Aurora (Aberdart, 600 Aberavon, Aberstar and Abermagic) showed a decreasing content of genetic material from 601 cluster 7 in favour of cluster 5. This case exemplifies the strong bias towards North-Western 602 Europe cluster 5 in the generation of modern cultivars. Most probably, this trend is 603 predominantly due to the fact that modern breeding of L. perenne started in this part of the 604 world using local diversity.

605 We showed that L. perenne cultivars likely use only a small proportion of the natural genetic 606 diversity existing across its natural distribution range. This natural diversity thus represents a 607 valuable genetic resource that should be safeguarded in genebanks, but also much more 608 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 609 610 reduction in the acreage of natural and semi-natural permanent grasslands in Europe. 611 Especially in intensive agricultural landscapes, permanent grasslands have tended to be 612 ploughed and replaced by rotations of annual crops. Rotations may indeed include temporary

613 meadows sown with cultivars from modern breeding in regions where agriculture combines

- 614 cash crops and cattle breeding. This practice may not only reduce the natural diversity of
- 615 perennial ryegrass by extinction of natural populations but also by the expansion of North-
- 616 Western European genotypes (typically found in cultivars) into other European regions
- 617 through gene flow from cultivars to natural populations.

618 Conclusions

619 Demographic reconstructions assessing the SFS from pool-Seq GBS performed on a high 620 number of populations allowed us to trace the origins of the genomic diversity of L. perenne 621 at an unprecedented level of detail. The current extent of grasslands across Europe has been 622 mainly determined by human activities. However, our results indicate that the spatial 623 distribution of the natural genome-wide diversity of L. perenne has not been significantly 624 disturbed after more than two millennia of intensive agriculture. The current L. perenne 625 natural populations still maintain the genomic diversity that has allowed the species to persist 626 during the Quaternary climatic fluctuations. Modern plant breeding has likely used only a 627 small part of the genomic diversity of *L. perenne* naturally distributed across Europe and 628 surroundings, and thus has likely taken limited advantage of the adaptive diversity and 629 phenotypic variability of the species. To date, this wide natural diversity remains available but 630 it is threatened with extinction and should be preserved. Indeed, it may constitute a valuable 631 genetic resource for plant breeding to meet emerging agricultural challenges such as 632 adaptation to anthropogenic climate change.

633

634 **Figure legends and embedded figures**

635 Fig. 1. Genetic structure of 470 *L. perenne* natural populations sampled across Europe and the

- 636 Near East (L. perenne set) based on the DAPC (Discriminant Analysis of Principal
- 637 Components) of population allele frequencies at 507,583 nuclear genome SNP loci. (a)
- 638 Geographical distribution of genetic clusters. Bar charts (b and c) indicate Population
- 639 Membership Probabilities (PMPs) to genetic clusters. (b) Representation of PMPs from all
- 640 populations. (c) Zoom into PMPs of those populations having less than 90% of membership



641 probability to a single cluster. (d) Scatter plot of the first two discriminant axes.

645 Fig. 2. Phylogeography of *L. perenne*. (a) and (b) From left to right, schematic of the best 646 demographic model estimated with $\delta a \delta i$ (best fit between the predicted and observed site 647 frequency spectrum – SFS), heatmap representations of the joint SFS expected under the best 648 model and the observed joint SFS, TREEMIX model with highest likelihood compatible with 649 the best $\delta a \delta i$ model. (a) Analysis of *L. perenne* and related taxa natural populations, (b) 650 Analysis of *L. perenne* natural populations only. (c) Geographical representation of the 651 inferred demographic history of L. perenne in Europe and surroundings. Colours represent 652 distribution of DAPC genetic clusters as in Fig. 1, dates represent ML values as obtained from 653 the best $\delta a \delta i$ model. (d) Mean and standard deviation of expected population heterozygosity 654 for each L. perenne genetic cluster. (e1-e3) Representation of sea level variation (m) across 655 the last 400 kya obtained from Western Mediterranean paleopositions during glacial maxima 656 of Quaternary Glacial cycles (from Rabineau et al., 2006), with superimposed time parameter 657 values: ML (opaque colours) and 95% CI (transparent colours) of demographic events under 658 the best *L. perenne δaδi* 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 etal. (1988).



- 663 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.



668

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.



678 References

676

- Ash, H.B., Forster, E.S., & Heffner, E. (1941) Columella. On agriculture. Cambridge. Mass.
- 680 Balfourier, F. & Charmet, G. (1991a) Relationships between agronomic characters and
- ecogeographical factors in a collection of French perennial ryegrass populations. *Agronomie*, **11**, 645–657.
- 683 Balfourier, F. & Charmet, G. (1991b) Spaced plant evaluation of Mediterranean germplasm
- 684 collections of perennial ryegrass. *Euphytica*, **57**, 57–66.
- 685 Balfourier, F. & Charmet, G. (1994) Geographical patterns of isozyme variation in
- 686 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.
- 695 Blackmore, T., Thomas, I., McMahon, R., Powell, W., & Hegarty, M. (2015) Genetic-
- 696 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.
- 707 Chapman, G.P. (1992) *Grass evolution and domestication*. Cambridge University Press,
 708 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.
- 715 Crème, A., Rumpel, C., Le Roux, X., Romian, A., Lan, T., & Chabbi, A. (2018) Ley
 716 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.
- 725 Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E.,
- 726& Holzapfel, C.M. (2010) Resolving postglacial phylogeography using high-throughput
- sequencing. *Proceedings of the National Academy of Sciences*, **107**, 16196–16200.
- 728 European Commission (2015) Available at:
- 729 http://ec.europa.eu/food/plant/plant_propagation_material/plant_variety_catalogues_data
- 730 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
 /.
- Furostat (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.
- 737 Faville, M.J., Vecchies, A.C., Schreiber, M., Drayton, M.C., Hughes, L.J., Jones, E.S.,
- 738 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
- 740 in perennial ryegrass (Lolium perenne L.). Theoretical and Applied Genetics, 110, 12–
- 741 32.
- 742 Feurdean, A., Marinova, E., Nielsen, A.B., Liakka, J., Veres, D., Hutchinson, S.M., Braun,
- 743 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.
- 746 Fjellheim, S., Rognli, O.A., Fosnes, K., & Brochmann, C. (2006) Phylogeographical history

- 747 of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast
 748 DNA sequences. *Journal of Biogeography*, **33**, 1470–1478.
- Fracassetti, M., Griffin, P.C., & Willi, Y. (2015) Validation of Pooled Whole-Genome ReSequencing 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.
- 775 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.
- 784 Kuneš, P., Svobodová-Svitavská, H., Kolář, J., Hajnalová, M., Abraham, V., Macek, M.,

785 Tkáč, P., & Szabó, P. (2015) The origin of grasslands in the temperate forest zone of
786 east-central Europe: Long-term legacy of climate and human impact. *Quaternary Science*787 *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.
- Pèrez-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.
- 814 Prieto, I., Violle, C., Barre, P., Durand, J.L., Ghesquiere, M., & Litrico, I. (2015)
- 815 Complementary effects of species and genetic diversity on productivity and stability of 816 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.
- 820 Rode, N.O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J., & Gay, L. (2018) How to
- 821 optimize the precision of allele and haplotype frequency estimates using pooled822 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.
- 825 Sampoux, J.P., Baudouin, P., Bayle, B., Béguier, V., Bourdon, P., Chosson, J.F., de Bruijn,
- 826 K., Deneufbourg, F., Galbrun, C., Ghesquière, M., Noël, D., Tharel, B., & Viguié, A.
- 827 (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.
- 830 Sampoux, J.P., Baudouin, P., Bayle, B., Béguier, V., Bourdon, P., Chosson, J.F.,
- 831 Deneufbourg, F., Galbrun, C., Ghesquière, M., Noël, D., Pietraszek, W., Tharel, B., &
- 832 Viguié, A. (2011) Breeding perennial grasses for forage usage: An experimental
- 833 assessment of trait changes in diploid perennial ryegrass (Lolium perenne L.) cultivars
- released in the last four decades. *Field Crops Research*, **123**, 117–129.

- 835 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.
- 838 Semelová, V., Hejcman, M., Pavlů, V., Vacek, S., & Podrázský, V. (2008) The Grass Garden
- 839 in the Giant Mts. (Czech Republic): Residual effect of long-term fertilization after 62
 840 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.
- 843 Sukumaran, J. & Holder, M.T. (2015) SumTrees: phylogenetic tree summarization. 4.0. 0.
- 844 *Program and documentation available from the authors at:*, https://github.
- 845 com/jeetsukumaran/Dendrophy.
- 846 Terrell, E.E. (1966) Taxonomic implications of genetics in ryegrasses (*Lolium*). *The Botanical*847 *Review*, **32**, 138–164.
- 848 Terrell, E.E. & Ekrem, T. (1968) *A taxonomic revision of the genus Lolium*. US Dept. of
 849 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., Santillàn-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.
- 861 Werling, B.P., Dickson, T.L., Isaacs, R., Gaines, H., Gratton, C., Gross, K.L., Liere, H.,
- 862 Malmstrom, C.M., Meehan, T.D., Ruan, L., Robertson, B.A., Robertson, G.P., Schmidt,
- 863 T.M., Schrotenboer, A.C., Teal, T.K., Wilson, J.K., & Landis, D.A. (2014) Perennial
- grasslands enhance biodiversity and multiple ecosystem services in bioenergy

865

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 NileValley. *Clarendon Press*, 243.

869

870 Biosketch

871 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,

874 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

S76 JPS wrote the manuscript with feedback from SM, TR and IRR. All authors participated in the

877 edition of the manuscript.

878 Data accessibility Statement

879 The genetic data reported in this study are available in the NCBI Short Read Archive (SRA)

880 database through accession SRP136600.

- 882 Supporting Information
- 883 Appendix S1. Supplementary Methods:
- 884 Plant Material and Genotyping
- 885 DAPC analysis
- 886 Appendix S2. Supplementary Figures:
- Fig. S1. Histogram of mean read depth for each locus across populations.
- 888 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
- 890 populations of *L. perenne*).
- Fig. S3. Geographical patterns of genetic differentiation in *L. perenne*. Scatterplot of
- 892 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
- 899 (natural populations of *L. perenne* only) TREEMIX models.
- 900 Fig. S7. Schematic representation of *Lolium* set and *L. perenne* set δaδi models.
- 901 Fig. S8. 50% Majority-rule consensus tree of 552 accessions of *Lolium perenne* and related
- 902 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).
- 907 Fig. S11. Alternative scenario for the range expansion in *L. perenne*. It shows a postglacial
- 908 expansion through the West and Central East and next towards Eastern Europe (clockwise909 movement around the Alps).

- 910 Appendix S3. Supplementary Tables:
- 911 Table S1. Accessions from *Lolium perenne* and related taxa used in the study.
- 912 Table S2. *Fst* statistics between seven clusters identified with the k-means algorithm
- 913 implemented in adegenet of the *L. perenne* set (natural populations of *L. perenne*).
- 914 Table S3. Parameter estimates and statistics from the fitting of 12 alternative $\delta a \delta i$
- 915 interspecific models of gene flow using the *Lolium* set (natural populations of *L. perenne* and
- 916 related taxa).
- 917 Table S4. Parameter estimates and statistics from the fitting of 12 alternative $\delta a \delta i$
- 918 interspecific models of gene flow using the Lolium set with Ni and Ti values converted to
- 919 numbers of individuals and years, respectively.
- 920 Table S5. Maximum Likelihood parameter estimates and non-parametric bootstrap 95%
- 921 confidence interval of the best *Lolium* set gene flow model.
- 922 Table S6. Parameter estimates and statistics from the fitting of 12 alternative $\delta a \delta i$
- 923 demographic models using the *L. perenne* set.
- 924 Table S7. Parameter estimates and statistics from the fitting of 12 alternative $\delta a \delta i$
- 925 demographic models using the *L. perenne* set with *Ni* and *Ti* values converted to numbers of
- 926 individuals and years, respectively.
- 927 Table S8. Maximum Likelihood parameter estimates and non-parametric bootstrap 95%
- 928 confidence interval of the best *L. perenne* set demographic model.
- Table S9. cpDNA primer sequences used for HiPlex amplicon sequencing.