Canonical correlations reveal adaptive loci and phenotypic responses to climate in perennial ryegrass

José Luis Blanco-pastor, Philippe Barre, Thomas Keep, Thomas Ledauphin, Abraham Escobar-Gutiérrez, Anna Maria Roschanski, Evelyn Willner, Klaus J. Dehmer, Matthew Hegarty, Hilde Muylle, et al.

DR. JOSÉ LUIS BLANCO-PASTOR (Orcid ID : 0000-0002-7708-1342)

PROF. STÉPHANIE MANEL (Orcid ID : 0000-0001-8902-6052)

Article type : Resource Article

Canonical correlations reveal adaptive loci and phenotypic responses to climate in perennial ryegrass

Short title: Climate and phenotype to detect adaptive loci

J.L. Blanco-Pastor*[1], P. Barre[1], T. Keep[1], T. Ledauphin[1], A. Escobar-Gutiérrez[1], A.M. Roschanski[2], E. Willner[2], K. J. Dehmer[2], M. Hegarty[3], H. Muylle[4], E. Veeckman[4, 5, 6], K. Vandepoele[4, 5, 7], T. Ruttink[4], I. Roldán-Ruiz[4, 6], S. Manel[8], J.P. Sampoux[1]

1. INRAE, UR4 (URP3F), Centre Nouvelle-Aquitaine-Poitiers, 86600 Lusignan, France;
2. Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Inselstr. 9, 23999 Malchow/Poel, Germany;
3. Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, SY23 3FL Ceredigion, UK;
4. Flanders Research Institute for Agriculture, Fisheries and Food (ILVO) - Plant Sciences Unit, Caritasstraat 39, 9090 Melle, Belgium;
5. Bioinformatics Institute Ghent, Ghent University, 9052 Ghent, Belgium;
6. Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium;
7. Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium;
8. CEFE, Univ. Montpellier, CNRS, EPHE-PSL University, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France.

Corresponding author
José Luis Blanco-Pastor, INRAE, UR4 (URP3F),
Centre Nouvelle-Aquitaine-Poitiers, Le Chêne – RD 150 CS 80006, 86600 Lusignan, France.
Telephone: +33549556031
Email: jose-luis.blanco-pastor@inra.fr

Abstract

Germplasm from perennial ryegrass (*Lolium perenne* L.) natural populations is useful for breeding because of its adaptation to a wide range of climates. Climate-adaptive genes can be detected from associations between genotype, phenotype and climate but an integrated framework for the analysis of these three sources of information is lacking.

We used two approaches to identify adaptive loci in perennial ryegrass and their effect on phenotypic traits. First, we combined Genome-Environment Association (GEA) and GWAS analyses. Then, we implemented a new test based on a Canonical Correlation Analysis (CANCOR) to detect adaptive loci. Furthermore, we improved the previous perennial ryegrass gene set by *de novo* gene prediction and functional annotation of 39,967 genes.

GEA-GWAS revealed eight outlier loci associated with both environmental variables and phenotypic traits. CANCOR retrieved 633 outlier loci associated with two climatic gradients, characterized by cold-dry vs mild-wet winter and long rainy season vs long summer, and pointed out traits putatively conferring adaptation at the extremes of these gradients. Our CANCOR test also revealed the presence of both polygenic and oligogenic climatic adaptations. Our gene annotation revealed that 374 of the CANCOR outlier loci were positioned within or close to a gene. Co-association networks of outlier loci revealed a potential utility of CANCOR for investigating the interaction of genes involved in polygenic adaptations.

The CANCOR test provides an integrated framework to analyze adaptive genomic diversity and phenotypic responses to environmental selection pressures that could be used to facilitate the adaptation of plant species to climate change.

Keywords: Adaptation, Agriculture, Climate Change, Ecological Genetics, Landscape Genetics, Quantitative Genetics.

Introduction

It is now widely acknowledged that droughts and heatwaves will become more frequent and more intense in Europe with climate change (Samaniego et al., 2018; Teuling, 2018) and that rising global temperature will have a profound effect on natural plant populations and crops (Mora et al., 2015; Thuiller, Lavorel, Araújo, Sykes, & Prentice, 2005; Travis, 2016). Climate change will cause an increasing number of hot summers, will lengthen the growing season at high latitudes such as in the Nordic countries and will shorten it at southern latitudes such as in the Mediterranean region (Dai, 2013). Climate extremes, changes in the length of the growing season and their interaction constitute complex challenges for biodiversity conservation and plant breeding (Savolainen, Lascoux, & Merilä, 2013).

Natural plant populations have used diverse survival strategies to adapt to a variety of climates at variable temporal and spatial scales (Brozynska, Furtado, & Henry, 2016; Godfray et al., 2010; Henry & Nevo, 2014; Hodgkin & Bordoni, 2012). However, plant breeding has only used a tiny fraction of the genetic diversity available within the entire gene pool of the species and the genetic diversity in cultivated gene pools is low compared to that harbored by natural populations (Blanco-Pastor et al., 2019; Brozynska et al., 2016; Redden et al., 2015; Warschefsky, Penmetsa, Cook, & von Wettberg, 2014). As a consequence, natural populations are one of the most critical assets to address climate change adaptation of species used in agriculture. As wild plant populations have evolved to cope with changes in their environment by means of natural selection, they constitute useful sources of diversity that can be used to improve crop resistance to extreme climatic conditions (FAO, 2015; Redden et al., 2015; Vincent et al., 2013; Warschefsky et al., 2014).

One promising strategy to create plant genotypes adapted to extreme climatic conditions is to identify loci responsible for adaptation in stress-tolerant natural populations. This strategy has become feasible thanks to recent advances in the development of genomic tools (Bansal, Lenka, & Mondal, 2014; de la Peña, Ebert, Gniffke, Hanson, & Symonds, 2011) and predictive statistical approaches (Manel et al., 2018). However, the agronomic performance of natural populations is in general lower than that of commercial varieties. It would thus be particularly challenging to create improved gene pools from natural populations that would combine adaptation to particular climatic conditions and sufficient value for cultivation. Often, a compromise needs to be found between climatic adaptation and agronomic value, as an adaptive gene might have negative pleiotropic effects on other traits or might be in linkage disequilibrium with genes of agronomic interest (i.e. linkage drag, Zamir, 2001). To account for these issues, a primary step in breeding programs that focus on adaptation to climatic threats is the identification of the genomic diversity responsible for adaptation and the documentation of its relationship with phenotypes (Shukla & Mattoo, 2013). This documentation is essential because phenotypic variation is the ultimate driver of both climate adaptation and agronomic value (Rieseberg, Widmer, Arntz, & Burke, 2002; Sampoux, Barre, & Litrico, 2014). But it is also challenging because adaptive phenotypic responses may be determined by a large number of loci of small effect (polygenic traits) that are difficult to identify with current analytical approaches (Berg & Coop, 2014; Berg et al., 2019; Pritchard & Di Rienzo, 2010; Santure & Garant, 2018; Savolainen et al., 2013).

The combination of the univariate association methods Genome-Wide Association Studies (GWAS) and Genome-Environment Association analyses (GEA) has proven to be an effective approach to reveal the genomic determinism of phenotypic traits and its relationship with climate adaptation (Anderson, Kono, Stupar, Kantar, & Morrell, 2016; Atwell et al., 2010; Contreras-Moreira et al., 2019; Fournier-Level et al., 2011; Talbot et al., 2017). These methods have however some limitations. GWAS and GEA can only detect adaptive loci whose effects are not hidden by the confounding effect of neutral genetic structure (but see Caye, Jumentier, Lepeule, & François, 2019; Frichot & François, 2015; Frichot, Schoville, Bouchard, & François, 2013; Price, Lopez, Platts, & Lasky, 2020). More importantly, they only provide a partial discovery of adaptive diversity, as local adaptation can be largely determined by coordinated shifts in allele

frequencies from multiple loci that are ignored when single-locus analyses are used (Berg & Coop, 2014; Exposito-Alonso et al., 2018; Josephs, Berg, Ross-Ibarra, & Coop, 2019). Although recent reviews have stressed the relevance of multivariate analyses that integrate environmental, genotypic and phenotypic data to uncover adaptive loci with small-effect while reducing the number of false positives (Barrett & Hoekstra, 2011; Hoban et al., 2016), a relatively small number of studies have achieved this integration so far (but see Berg & Coop, 2014; Exposito-Alonso et al., 2018). In that sense, ordination-based multivariate methods show promise as they can effectively detect multilocus selection by analyzing how groups of markers covary in response to multiple predictors (Forester, Lasky, Wagner, & Urban, 2018).

Grassland ecosystems are ubiquitous across temperate and tropical regions. They constitute the most extensive semi-natural habitat type accounting for 37% of the terrestrial land cover (Loveland et al., 2000). They are essential for the maintenance of biodiversity, for carbon sequestration and for the functioning of soil biogeochemical cycles (Hejcman, Hejcmanová, Pavlů, & Beneš, 2013; Jones & Donnelly, 2004; Tilman, Wedin, & Knops, 1996). In Europe, they cover 45% of the total agricultural area (Eurostat, 2017). Perennial ryegrass (*Lolium perenne* L.) is one of the most prevalent grass species in natural and semi-natural permanent grasslands across Europe. Its high nutritive value for herbivores and its relatively good adaptation to grazing and trampling have long been recognized. Perennial ryegrass has thus extensively been bred during the past fifty years to deliver improved commercial varieties to sow and regenerate meadows as well as to set up and repair turf areas (Sampoux et al., 2013, 2011). While domestication of major crops started ca. 10000 years ago (Zohary, Hopf, & Weiss, 2012), conscious breeding in perennial ryegrass was initiated only during the last century (Blanco-Pastor et al., 2019; Humphreys, Feuerstein, Vandewalle, & Baert, 2010; Sampoux et al., 2013, 2011). Because of this recent start of human selection, wild populations of perennial ryegrass may still contain potentially useful genetic resources that could be easily incorporated into breeding programs. With that regard, wild populations have extensively been collected in the last decades (Sampoux et al., 2014).

Perennial ryegrass natural populations colonized Europe during the Quaternary glacial cycles while adapting to a wide range of environmental conditions (Barre et al., 2017; Blanco-Pastor et al., 2019). Natural ecotypes of perennial ryegrass are today present over a wide range of climatic conditions across Europe and the Near-East (Blanco-Pastor et al., 2019). Cold, heat and drought stresses in the latitudinal and longitudinal extremes of Europe have likely led to the evolution of seasonal acclimation processes regulating climate adaptation in perennial ryegrass (Ergon et al., 2018; Thomas & James, 1999; Zhang, Fei, Arora, & Hannapel, 2010). Consequently, the extant wide natural diversity of perennial ryegrass should represent a valuable genetic resource for its adaptation to climate change (Sampoux et al., 2014). Past breeding activities in perennial ryegrass have mainly focused on improving forage yield, disease resistance and seed yield for the seed industry (Humphreys et al., 2010; Wilkins & Humphreys, 2003). In contrast, there have been fewer efforts to improve resistance to cold, heat and drought stresses (Charmet, Balfourier, Ravel, & Denis, 1993) and the specific phenotypic traits linked to climatic adaptation remain insufficiently documented (Barre et al., 2017; Ergon et al., 2018; but see Kovi et al., 2015).

We used genomic and phenotypic data from 469 perennial ryegrass natural populations collected across the natural distribution range of the species (427 genebank accessions and 42 populations collected in situ) (see Blanco-Pastor et al., 2019). We combined Genotyping-by-Sequencing (GBS) and Highly Multiplexed Amplicon Sequencing (HiPlex) pool-Seq genotyping data, extensive phenotyping characterization in three experimental gardens in France, Belgium and Germany and fine-resolution environmental data at population collection sites. We implemented two data-driven analytical approaches. First, we used GEA combined with GWAS to identify putative climate adaptive loci. In a second approach, we implemented a statistical test that used the output of a Canonical Correlation Analysis (CANCOR). The CANCOR (also abbreviated CCorA) is a multivariate analysis that reveals the co-inertia between two tables that describe the same set of observations (here the SNPs) with two different sets of possibly covarying variables (here environmental and phenotypic variables). This approach analyzed simultaneously the environment at sites of origin of populations, their phenotype assessed in experimental gardens and the allelic frequencies of populations in order to identify the environmental variables imposing selection, the adaptive phenotypic responses and the adaptive loci.

An extensively annotated gene set can help to identify climate adaptive genes and gene functions under selection. In view of that, we also improved the published perennial ryegrass gene set (Byrne et al., 2015) by *de novo* gene prediction and functional annotation of our genomic dataset. The CANCOR test and the new functional annotation provided a list of loci and molecular functions putatively linked to environmental adaptation that could be used in breeding programs to adapt perennial ryegrass to climate change.

Materials and Methods

Genetic material

We examined 469 natural populations of perennial ryegrass that were either obtained from genebanks of agronomic research institutes from multiple countries or sampled *in situ* (Fig. 1 and Table S1). They were chosen to capture the extant natural genetic diversity of perennial ryegrass across its natural distribution range (Europe and the Near East). Full description of this set of populations can be found in Blanco-Pastor *et al.* (2019) named as the '*L. perenne set*'.

Genotype data

The genetic data was generated using a Genotyping-by-Sequencing (GBS) pool-Seq protocol (Blanco-Pastor et al., 2019) based on the protocol of Byrne *et al.* (2013). We also re-sequenced from same pools 185 genomic regions of 80-140 bp positioned in, or near, 42 candidate genes putatively involved in environmental adaptation using Highly Multiplexed Amplicon Sequencing (HiPlex set) (see gene descriptions in Supporting Information, Table S2, and further information in Supporting Information, Methods S1). For the GBS and HiPlex genotyping methods, balanced leaf material from c.a. 300

individuals per population were pooled before DNA extraction. Variants were called using the draft reference genome sequence of Byrne et al. (2015). Further details are available in the Supporting Information of Blanco-Pastor *et al.* (2019). We merged the two datasets (GBS and HiPlex) for analyses and performed a stringent filter on the minor allele frequency (MAF) to reduce the proportion of low frequency alleles. We retained SNP loci if their MAF was greater than 5% in at least 10 populations. The final merged dataset comprised alternative allele frequencies (AAFs) of 189,968 SNP loci in the 469 natural populations (Data S1 in Blanco-Pastor et al., 2020). The genotype data included 7.81% missing values that were imputed by using the mean allele frequency across populations. To avoid the effect of linkage disequilibrium in outlier discovery, we calculated the kinship-corrected correlation decay with increasing base pair distance for SNP markers belonging to a same scaffold. Based on the squared correlation decay curve, and in line with results from Keep et al. (2020), we considered that two loci were linked when the correlation between their alternative allele frequencies corrected for kinship was larger than 0.4. In such case, we only kept the locus displaying the best association with a phenotypic trait (lowest p-value in independent GWAS analyses).

Environmental variables

We collected a set of 112 variables documenting environmental conditions at sites of origin of the 469 studied populations: bioclimatic indices, Climate Change Detection Indices (ETCCDI), ecophysiological indices relevant to the life cycle of perennial ryegrass and soil data derived from the European Soil Database. An exhaustive overview of the environmental variables used is provided in Supporting Information, Table S3 and Methods S2.

Phenotypic traits

For the needs of phenotyping, 385 of the 469 perennial ryegrass populations were sown in experimental gardens in three locations: Poel Island (PO) in Germany on April 2015, Melle (ME) in Belgium on October 2015 and Lusignan (LU) in France on April 2015. In each of these three locations, each population was sown in three 1m² micro-swards (small plots sown as to reach plant density similar to real grasslands) arranged in three replicated blocks. Trials were monitored until end of 2017 at PO and ME and until end of 2018 at LU. Micro-swards were cut (all aerial biomass higher than 7 cm above ground surface) regularly as to simulate common cutting regime of meadows used for green forage production or grazing. Weather conditions experienced at each trial location are displayed per season of each year in Supporting Information, Table S4. LU was characterized by severe water stress in summer. At PO, water stress was negligible in summer but cold stress was experienced during winter periods. ME was characterized by cool summer and mild winter conditions. Scores or measurements of phenotypic traits were recorded at the level of 1 m² micro-swards over all plants. A set of 145 phenotypic traits were recorded for 385 of the 469

perennial ryegrass populations. This set included traits related to vigor after sowing, morphology of plants, sward density, phenology, investment in sexual reproduction, dynamics of vegetative growth in spring, summer and autumn, regrowth after cutting, abiotic and biotic stresses related traits, dynamics of persistency, biochemistry of aerial biomass and leaf lamina traits. An exhaustive overview of the recorded phenotypic traits is provided in Supporting Information, Table S5 and Methods S3.

The GEA-GWAS approach

We performed a "triangulation" of association analyses (e.g. Talbot et al., 2017) to detect putative adaptive loci (Fig. 2a). In this approach, we looked for strong significant environment-genotype (GEA), phenotype-genotype (GWAS) and direct environment-phenotype associations to investigate whether putative adaptive loci were also potentially involved in the determinism of potentially adaptive traits.

To identify putative adaptive loci, we used a GEA linear mixed model similar to that of Yoder *et al.* (2014) (Supporting Information, Methods S4). We additionally used a GWAS linear mixed model to assess individual locus effect on a given phenotypic trait (Supporting Information, Methods S5). Both GEA and GWAS were run using the *GWAS* function of the 'rrBLUP' R package (Endelman, 2011). Among the significant loci revealed by the GEA analysis, we only considered as GEA-GWAS outlier those also significantly associated with a phenotypic trait in GWAS. We used a liberal threshold of False Discovery Rate (FDR) of 0.2 in both GEA and GWAS because a SNP needed to be significant in the two independent analyses to be considered as outlier. But we also report results using the more conservative FDR = 0.1. The final step of the "triangulation" approach was the assessment of direct correlations between environmental variables and phenotypic traits significantly associated with a same locus (direct correlations significant at p-value < 0.05).

The CANCOR test

As an alternative to investigate adaptive diversity, we implemented a Canonical Correlation Analysis (CANCOR) (Hotelling, 1936) (Fig. 2b) that analyzed simultaneously the association of genomic polymorphisms with environmental variables and phenotypic traits. The CANCOR multivariate analysis aims to reveal the co-inertia between two sets of possibly co-varying variables that describe the same set of experimental units (or observations). It looks for successive pairs of linear combinations from each set (canonical variables) that are maximally correlated (canonical correlations). Successive canonical variables in each set are constrained as to be uncorrelated. In a preliminary step, univariate regression models were implemented to regress the population alternative allele frequency (AAF) of each genotyped SNP locus on each environmental variable (values at sites of origin of populations) and on each phenotypic trait (mean values of populations). The CANCOR was then performed by considering the loci as the experimental units and the two sets of regression slopes of alternative allele frequencies, on environmental variables on the

one hand (Y, Fig. 2b) and on phenotypic traits on the other hand (X, Fig. 2b), as the two sets of input variables to analyze (Supporting Information, Methods S6).

We also ran an additional CANCOR in order to discern the general structure of correlation between environmental and phenotypic variables at the population level. In this analysis, the populations were considered as the experimental units and the value of environmental variables at sites of origin of populations on the one hand, and the population means of phenotypic traits on the other hand, as the two sets of input variables to analyze.

We ran the CANCOR analysis using the R package 'vegan' (Oksanen et al., 2018). We tested the significance of outlier loci using a χ^2 test on Mahalanobis distances following the method of Luu et al. (2017) and Capblancq at al. (2018), which we call hereafter CANCOR test (Supporting Information, Methods S6), and a locus was considered as outlier at FDR = 0.1.

To relate CANCOR outliers to putative adaptive traits and selective environmental variables, we first selected the CANCOR input variables best represented in the first two canonical dimensions, which were the only two retained by our CANCOR test. To simplify results we selected only input variables with projection norms larger than 0.95 and 0.90 in the first environmental and phenotypic canonical planes, respectively (thresholds that returned a similar number of environmental and phenotypic variables). We finally retained the corresponding environmental and phenotypic variables if their correlation with the AAFs of at least one CANCOR outlier locus was sufficiently high ($|r| > 0.5$) (see Forester et al., 2018). We also explored the variable heading date (*HEA_avg*) despite its smaller projection norm because of its known importance for adaptation. In order to relate environmental selection pressures to phenotypic responses, we identified those environmental variables sharing their position in the CANCOR first two dimensions with the highest number of phenotypic traits. We further investigated these relationships by performing a linear regression of the phenotypic trait on the environmental variable using the trait mean values of populations and the value of environmental variables at sites of origin of populations.

Co-association networks

To visualize the interaction of SNPs in the bi-dimensional space defined by the CANCOR test, we adapted the approach of Lotterhos et al. (2018). In order to account for the information from both the alternative and reference alleles, we classified CANCOR outliers into two groups according to their position in the four CANCOR quadrants of the first phenotypic canonical plane. We grouped outliers from quadrants I and III since they were expected to be associated to adaptation to the same environmental gradients. For that, we changed the value of the loadings on the CANCOR axes of SNPs in quadrants III to their symmetrical value in quadrant I (negative signs of the loadings on the axes 1 and 2 replaced by positive signs), as the position of SNPs in quadrant I or III only depends on whether the alternative or the reference allele is associated to adaptation. Similarly, we grouped outliers from quadrants II and IV and we changed the value of the loadings of SNPs in quadrants IV to their symmetrical value in quadrant II (positive sign of loadings on the

axis 1 replaced by negative sign and negative sign on the axis 2 replaced by positive sign). Then we used these modified canonical loadings to calculate a matrix of pairwise Euclidean distances between SNPs. For each of the two groups, we used undirected graph networks to visualize modules of SNPs. Nodes were connected by edges according to three different thresholds of pairwise Euclidean distances (d) (< 1, < 0.5 and < 0.1). Co-association networks were visualized using the 'igraph' R package (Csardi & Nepusz, 2006).

To demonstrate the utility of CANCOR for investigating the genetic basis of complex traits, for each of the two groups of SNPs, we performed two independent Gene Ontology (GO) enrichment analyses in the largest modules obtained with threshold of d < 0.1. The GO enrichment analyses were performed with agriGO 2.0 (Tian et al., 2017). We used the Singular Enrichment Analysis (SEA) tool with a customized annotation of GO terms obtained from the new gene prediction and functional annotation (see below) and used the Locus ID (PLAZA3.0) as reference. We applied the Fisher's exact test with the Benjamini-Hochberg FDR correction (FDR < 0.05) (Benjamini & Hochberg, 1995).

Gene prediction and functional annotation

The EVidenceModeler (EVM) (Haas et al., 2008) was used to improve completeness, without losing gene model accuracy, of the previously published set of 28,182 genes annotated on the *L. perenne* draft genome sequence (Byrne et al., 2015). For this, the annotation set was complemented with a less conservative set of gene predictions, orthology-guided transcript assemblies (Ruttink et al., 2013) and aligned proteomes of closely related species (*Brachypodium distachyon*, rice, maize and sorghum). All evidence tracks were generated using the GenomeThreader (Gremme, Brendel, Sparks, & Kurtz, 2005) with default settings and used as input for the EVM. The completeness was estimated using BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) and the PLAZA 2.5 monocots core gene families (Van Bel et al., 2012). Functional annotation making use of ontologies was generated using InterPro2GO mapping, Gene Ontology (GO) projection between orthologs and MapMan. Additionally, gene descriptions were added using AnnoMine, a homology-based text-mining approach (Van Landeghem, De Bodt, Drebert, Inzé, & Van de Peer, 2013). The gene annotation set has been made publicly available on the PLAZA comparative genomics platform version 4.5 Monocots (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_5_monocots/organism/view/Lolium+perenne).

Results

The GEA-GWAS approach

A total of 10,220 and 15,854 loci were found as outliers with the GEA analyses at FDR = 0.1 and FDR = 0.2, respectively. A total of 330 and 543 loci were found as outliers with the GWAS analyses at FDR = 0.1 and FDR = 0.2, respectively. Among these, only 18 and 49 outliers were significant in both the GEA and GWAS at FDR = 0.1 and FDR = 0.2, respectively (Supporting Information, Table S6 and Table S7).

Environmental and phenotypic variables most strongly associated with GEA-GWAS outliers at FDR = 0.1 were *bd_subsoil*, *bio.ad.27*, *bio10*, *pet_wi*, *lmts*, *oc_topsoil*, *sis_wi*, *tawc_soil*, *tr_an*; and *AHD_me16*, *AMH_po17*, *CH400h_po16*, *CHs500_me17*, *HFY_lu15*, *HFY_po15*, *HST_lu17*, *SCD_wi1516_po*, *VAC_avg*, *VAC_lu17*, respectively (Table 1 and Supporting Information, Table S7). At FDR = 0.1, phenotypic traits showing strongest association with the highest number of GEA-GWAS outliers were *AMH_po17* (autumn canopy height, 6 outliers) and *CH400h_po16* (spring canopy height 400 growing-degree-days before spike emergence, 5 outliers) (Table 2 and Supporting Information, Table. S7). We found 8 outlier loci significantly associated at FDR = 0.1 with an environmental variable and a phenotypic trait whose direct correlation was significant (p-value < 0.05), and 34 outlier loci at FDR = 0.2. These loci were located in the proximity of 3 (FDR = 0.1) and 16 (FDR = 0.2) independent known genes (Supporting Information, Table S7). The InterPro domains, the Gene Ontology and the functional annotations for these genes are provided in Table S7 and more information on the genome sequence context flanking these genes is available via PLAZA monocots 4.5. Our genotyped loci included SNPs from 185 amplicon regions positioned in, or in the proximity of, 42 candidate genes possibly involved in environmental adaptation (HiPlex set, Supporting Information, Table S2). Of these, the GEA-GWAS approach did not detect any SNP as putatively adaptive at FDR = 0.1. At FDR = 0.2, GEA-GWAS detected 6 SNPs (from two candidate genes) as putatively adaptive but only one showed direct correlation between the associated environmental and phenotypic variables (p-value < 0.05) (Supporting Information, Table S7).

The CANCOR approach

The CANCOR using the loci as experimental units (Fig. 3b-c) found the first 14 canonical correlations ('CanCorr' elements of the CCorA function of the R package 'vegan') larger than 0.9. The environmental input variables (regression slopes of AAFs on environmental variables) were highly correlated (|r| > 0.7) to the first and second environmental canonical variates for 44 and 35 environmental variables, respectively. Likewise, the phenotypic input variables (regression slopes of AAFs on phenotypic variables) were highly correlated (|r| > 0.7) to the first and the second phenotypic canonical variates for 35 and 37 phenotypic variables, respectively. In contrast, the CANCOR using populations as experimental units (Fig. 3d-f) only found the first two canonical correlations larger than 0.9. None of the correlations between environmental canonical variates and environmental variables were larger than 0.7. Only the first phenotypic canonical variate showed high correlation (|r| > 0.7) with four phenotypic traits. Populations from a same geographical origin tended to cluster together on the first phenotypic canonical plane.

With the CANCOR test, we observed that the distribution of p-values was correct (flat distribution with enrichment only for the low values) exclusively when only the first two canonical dimensions were considered (K =2) and therefore only the results with K = 2 are discussed here (see Supporting Information, Methods S6). At FDR = 0.1, the CANCOR test retrieved 633 outlier loci ("CANCOR outliers") which were located in the proximity of 158 independent known genes (Fig. 3a and Supporting Information, Fig. S2 and Table S8) among which 13 were "HiPlex" loci. The CANCOR test only found four and 10 outlier loci that

were also significant outliers in both the GEA and GWAS linear mixed models at FDR = 0.1 and FDR = 0.2, respectively (Supporting Information, Fig. S3). CANCOR outliers showing high correlation (|r| > 0.5) with environmental and phenotypic variables well represented in the first two canonical dimensions (projection norm of input variables larger than 0.95 and 0.90, respectively) are shown in Fig. 4.

The two main environmental gradients revealed by the CANCOR test are highlighted in Fig. 5. A first gradient opposed the first and third quadrants of the first environmental canonical plane with increasing winter temperature (*tnn_wi*, *txx_wi*, *tasmax_wi*, *tasmin_wi*, *bio3* and *bio6*) and precipitation during the wet season (*rx1day_au*, *sdii_au*, *sdii_sp*, *sdii_wi* and *bio.ad.20*) towards the third quadrant (Fig. 5a) (mild-wet vs cold-dry winter gradient). A second gradient opposed the second and fourth quadrants with increasing duration of summer period (*su_an*), decreasing duration of the rainy periods in autumn (*r01mm_au*) and winter (*r01mm_wi*) and increasing mean diurnal temperature range (*bio2*, *dtr_au*) towards the fourth quadrant (Fig. 5a) (long summer and high diurnal temperature range vs long rainy season and low diurnal temperature range). Note that soil properties were not evidenced to contribute to co-inertia in this CANCOR analysis, at least in the first two canonical dimensions.

In the first phenotypic canonical plane, the first quadrant was associated with canopy height during vegetative spring growth in the northernmost experimental garden (*CH300h_po16*, *CH400h_po16*) and with spike emergence date (*HEA_avg*) whereas the third quadrant was associated with winter damage (*WID_po16*) in the northernmost experimental garden (Fig. 5b). The second quadrant was associated with canopy height and canopy growth rate in summer (*SMH_me16, SGR_po17* and *SMH_po17*) and in autumn (*AGR_po17* and *AMH_po17*) in the two northern experimental gardens and with good persistency after winter in the northernmost one (*SCD_wi1617_po*) (Fig. 5b). The fourth quadrant was associated with seed production traits, namely aftermath heading (successive recurrent elongation of fertile stems) (*AHD_lu16* and *AHD_po17*) and spike density (*DST_avg*, *DST_lu17* and *resDST_lu17*), with lignin content in vegetative biomass (*ADL_10_me17* and *ADL_avg*) and with canopy growth rate in summer in the southernmost garden (*SGR_lu16*). Among the preceding traits, *HEA_avg*, *ADL_10_me17*, *ADL_avg*, *DST_avg*, *DST_lu17* and *resDST_lu17* were correlated (|r| > 0.5) to a small number of outlier loci: 5, 3, 1, 3, 20 and 21 respectively (Fig. 4b). Other traits, namely *AHD_lu16*, *AHD_po17*, *WID_po16*, *CH300h_po16*, *CH400h_po16*, *SGR_lu16*, *SGR_po17*, *SMH_po17*, *AGR_po17*, *AMH_po17* and *SCD_wi1617_po* were correlated (|r| > 0.5) to a higher number of outlier loci (30 to 289) (Fig. 4b).

An environmental variable and a phenotypic trait were considered as associated if they shared position in the first two canonical dimensions (see Fig. 5). Univariate regressions testing the association of pairs of environmental and phenotypic variables located in the same quadrant were highly significant in all cases (p-value < 0.05), with r^2 values ranging between 0.014 and 0.382. Plots of two types of regressions are displayed in Fig. 6: (i) phenotypic traits whose input variable is associated with the first or third quadrants of the first phenotypic canonical plane regressed on minimum temperature of winter period (*tnn_wi*) and (ii) phenotypic traits whose input variable is associated with the second or fourth quadrants regressed on the number of summer days (*su_an*). These regressions confirmed a clear relationship between phenotypic

means and values of the climatic variables at sites of origin of populations. Adaptation to cold stress in winter (low *tnn_wi*) was associated with high spring growth in cold conditions, late spike emergence and small damage during cold winters. Adaptation to long summer (high *su_an*), and likely to drought and heat stresses, was associated with high aftermath heading and spike density (reproductive investment), high lignin content in vegetative biomass, high growth in warm summer conditions, but low growth in cool summer and autumn conditions and low persistency after cold winter.

Among the HiPlex set, the CANCOR test detected 13 loci within three different known genes as outliers at FDR = 0.1 threshold (Supporting Information, Table S8).

Co-association networks

The two co-association analyses with threshold of $d < 1$ showed a single large module. With threshold $d < 0.5$, SNPs from both quadrants I-III and II-IV showed a single module together with three/four SNPs that were isolated or forming a small cluster. With this threshold at least three and two sub-modules could be observed in quadrants I-III and quadrants II-IV, respectively. Analyses with threshold $d < 0.1$ mainly resulted in singletons together with multiple small modules (Fig. 7).

We did not find any significantly enriched GO term in the largest module from quadrants I-III at $d < 0.1$. However, we found two significantly enriched GO terms (FDR < 0.05) in the largest module from quadrants II-IV at the same threshold: GO:0055114 (oxidation-reduction process) and GO:0016491 (oxidoreductase activity).

An improved gene annotation for identifying adaptive gene functions in *Lolium perenne*

Previous gene space completeness analysis (Veeckman *et al*., 2016) showed that the gene space was well represented in the *L. perenne* genome assembly (previously published set of 28,182 annotated genes), but that gene prediction was incomplete, as compared to BUSCO (81.6%) and PLAZA 2.5 monocots core gene families (CoreGF, 76.9%). Our additional gene annotation resulted in 39,967 consensus gene models. Gene space completeness was estimated at 92.6% (single: 89.0%, duplicated: 3.6%, fragmented: 2.5%, missing: 4.9%, no. of genes: 1440) using BUSCO and 89.4% using the PLAZA 2.5 CoreGF. This corresponded to an overall increase of completeness of more than 10% compared to the previously published gene annotation set. Functional annotation resulted in GO, InterPro and AnnoMine annotations for 23,879 *L. perenne* genes (59.8%). This final gene set was better suited for checking whether outlier loci from the CANCOR test matched candidate regions or were located in the proximity of a known gene, as it was more complete and more informative thanks to the improved functional annotation. Using the initial gene annotation set, 306 out of the 633 CANCOR outlier loci were positioned within or close to a gene, the

average distance to the closest gene was 16 kb and 93 loci were positioned on scaffolds without a gene. Using the new gene annotation set resulted in 374 CANCOR outlier loci positioned within or close to a gene, the average distance to the closest gene dropped to 9 kb and only 30 loci were positioned on scaffolds without a gene.

Discussion

A novel approach to detect genomic and phenotypic adaptive diversity and to identify environmental factors imposing selection

In our study, the combined GEA-GWAS approach was less effective than the CANCOR test in simultaneously detecting the environmental variable and phenotypic trait associated with a putative adaptive locus, even if the FDR thresholds used with GEA and GWAS were more liberal than the one used with CANCOR (GEA-GWAS: 49 outliers at FDR = 0.2 with 34 showing significant direct environment-phenotype correlation at p-value < 0.05; CANCOR: 633 outliers at FDR = 0.1). Certain climate-genotype-phenotype associations found with the GEA-GWAS approach were also found with CANCOR (4 and 10 outlier SNPs found with both GEA-GWAS and CANCOR depending on FDR thresholds used for GEA-GWAS, see Supporting Information, Results S1). But in general different associations were found with the two methods. The GEA-GWAS approach detected interesting soil-genotype-phenotype associations that were not detected by the CANCOR test, probably because the soil variables had little contribution in the first two canonical dimensions used for CANCOR outlier detection. GEA-GWAS outliers associated with soil variables were also associated with phenotypic traits describing the morphology of plants, investment in sexual reproduction, phenology or plant growth. These set of outliers could be of interest for eventual breeding programs that would aim to improve adaptation to soil features.

In most cases, a large part of the phenotypic variance remains unexplained by loci detected by GWAS (Maher, 2008), a problem that was notably encountered in perennial ryegrass (Harper et al., 2019). GWAS models control for false-positive associations due to population structure or genetic relatedness and inference statistics are corrected for multiple tests (Yu et al., 2006). But because of these corrections, they are prone to miss causal loci with small effect involved in polygenic adaptations (Josephs et al., 2019) or other adaptive loci whose allelic distribution is confounded with population structure (Atwell et al., 2010), a trend that is particularly common in natural populations (Barton, Hermisson, & Nordborg, 2019; Gienapp et al., 2017; Storz, 2005). When compared with CANCOR, the GEA-GWAS method is likely more well suited to identification of major effect loci affected by a single environmental variable and a phenotypic response dominated by a single trait. On the other hand, CANCOR is not directly based on extreme associations of genotyped loci with single environmental or phenotypic variables. And it can be expected as more powerful to detect groups of co-varying small effect loci involved in response to multivariate environment and associated with multivariate phenotypic responses.

Despite the high dimensionality of the genotype and the environment, univariate GEA methods have been the most popular approach to identify adaptive loci (Coop, Witonsky, Di Rienzo, & Pritchard, 2010; Frichot et al., 2013; e.g. Joost et al., 2007; Lasky, Forester, & Reimherr, 2017; Stucki et al., 2017; Yoder et al., 2014). Meanwhile, it has been claimed that multivariate methods are far more effective at detecting weak polygenic adaptation, as these methods can analyze the covariation between groups of loci and multiple environmental predictors (Forester et al., 2018; Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015). Many adaptive processes are indeed expected to be driven by weak polygenic effects as a result of recent selection on standing genetic variation that has not yet led to allele fixation or conditional neutrality (Berg & Coop, 2014; Le Corre & Kremer, 2012; Savolainen et al., 2013; Tiffin & Ross-Ibarra, 2014). Several multivariate ordination techniques have recently been proposed to identify adaptive loci (Capblancq et al., 2018; Forester, Jones, Joost, Landguth, & Lasky, 2015; Grivet, Sork, Westfall, & Davis, 2008; Luu et al., 2017). Although the need of integrating phenotypic data in adaptation studies has recurrently been stressed (Barrett & Hoekstra, 2011; Berg & Coop, 2014; Exposito-Alonso et al., 2018; Fournier-Level et al., 2011; Lasky et al., 2017; Steane et al., 2014), few studies investigating environmental adaptation have combined phenotypic and environmental data so far. Here, we introduced a CANCOR test to integrate both types of data in a single genome scan.

Multivariate analyses such CANCOR are not expected to be biased by collinearity. Keeping all available variables, as we did, showed that our CANCOR approach can be used in an exploratory manner and make possible to avoid the variable selection step. Environmental and phenotypic variables whose regression slopes are best correlated to the canonical variables revealed adaptive trends consistent with functional ecology expectations (Fig. 5).

The CANCOR analysis using the loci as observations revealed roughly the same patterns of association between environment and phenotype as the CANCOR analysis using populations as observations. However canonical correlations were larger with the former analysis suggesting that adaptive trends are better revealed at the genomic level than at the population level (Fig. 3b-c and Fig. 3e-f).

Adaptation to winter temperature

The CANCOR test detected 36 to 169 outlier loci highly associated (|r| >0.5) with traits reporting for spring growth in a cold winter environment (*CH300h_po16* and *CH400h_po16*) and 370 with winter damage during a cold winter (*WID_po16*). This result points to phenotypic adaptations to cold winter conditions (quadrant I, Fig. 5) that are under highly polygenic determinism (Fig. 4b and Supporting Information, Table S8). The CANCOR analysis indicated that perennial ryegrass populations from areas with low minimum winter temperatures (low *tnn_wi*) are more resistant to winter damages (*WID_po16*) than others when grown in the cold winter conditions of Northern Europe (Fig. 6a). Previous research on perennial ryegrass also showed that populations from southern Europe were the most susceptible to cold stress (Lorenzetti, Tyler, Cooper, & Breese, 1971). The CANCOR test detected only five outlier loci associated with spike

emergence date (*HEA_avg*), a trait also involved in adaptation to winter temperature (Fig. 4b and Supporting Information, Table S8). The proportion of phenotypic variance explained by these outlier loci (r^2) ranged from 0.25 to 0.33 (univariate linear model phenotype ~ outlier locus, results not shown). This is in accordance with previous results which found a few major genes involved in the determinism of spike emergence date in perennial ryegrass (Armstead et al., 2004, 2008; Keep et al., 2020; Skøt et al., 2005, 2011). These previous studies identified a major QTL explaining 70% of the trait variation in a F_2 mapping family. This QTL showed a high degree of synteny with the *Hd3* spike emergence date QTL region of rice LG6 that codes for the *Flowering locus T*. The gene prediction analysis found that the spike emergence date loci pointed out by our CANCOR test included three loci located within the *Flowering locus T* (*LpFT3* gene in the *L. perenne* genome) (Skøt et al., 2011; Veeckman, Vandepoele, Asp, Roldán-Ruiz, & Ruttink, 2016) and one locus located in its close proximity (at 272 bp downstream of the gene; Supporting Information, Table S8). Our results are thus in agreement with these previous findings and evidence that spike emergence date loci in perennial ryegrass evolved naturally along a winter temperature gradient (Fig. 6a).

The functional ecology theory tells that adaptation to climatic stresses can be provided by escape, avoidance and tolerance strategies (Levitt, 1962). A tolerance strategy notably involves a strong reduction or cessation of growth during the stress period (Gillespie & Volaire, 2017). In perennial ryegrass, the peak of spring vegetative growth occurs during a 15 days period preceding spike emergence (Roschanski et al., 2018). Late spike emergence of perennial ryegrass populations from areas with low minimum winter temperature, and likely with long winter period, corresponds to an escape strategy in which the peak of vegetative spring growth is scheduled to escape the latest period of cold stress. Our results confirm that this important adaptive feature is determined by a small number of genes. On the other hand, small winter damage in the northernmost experimental garden for populations from low minimum temperature areas indicates additional tolerance mechanisms under highly polygenic determinism, which in turn favor a strong spring vegetative growth after a cold winter.

Adaptation to summer length

The CANCOR test detected 26 and 46 outlier loci associated (|r| >0.5) with aftermath heading (*AHD_po17* and *AHD_lu16*), 3, 20 and 21 associated with spike density (*DST_avg*, *DST_lu17* and *resDST_lu17*) and 33 with growing rate in a dry summer (*SGR_lu16*). These results revealed phenotypic adaptations to long summer duration (quadrant IV, Fig. 5) that are under polygenic determinism (Fig. 4b and Supporting Information, Table S8). Sixty-one outliers in total were associated with aftermath heading or spike density and thus with investment in sexual reproduction.

Our results are consistent with previous reports indicating that perennial ryegrass populations from dry habitats recover from drought more rapidly than those from moist habitats (Norris & Thomas, 1982). The CANCOR analysis evidenced that perennial ryegrass populations from areas with long summer season

(high *su_an*), and thus with high probability of exposure to drought stress, use several functional strategies to adapt to this climatic constraint as described by Volaire (2018). A dehydration escape strategy is likely provided by investment in sexual reproduction with high aftermath heading and spike density (Berger, Palta, & Vadez, 2016). Better growing rate in a long and dry summer could be due to the growth of numerous elongating stems from aftermath heading in relation with the escape strategy, but it can also be due to some features enabling stress avoidance such as root architecture optimizing soil water extraction (Volaire, 2018).

The CANCOR test detected only two outlier loci associated (|r| >0.5) with lignin content (one associated with *ADL_10_me17* and another one with *ADL_10_me17* and *ADL_avg*) (Fig. 4b and Supporting Information, Table S8) with an r^2 ranging from 0.26 to 0.27 (univariate linear model phenotype ~ outlier locus, results not shown). This suggests that the genetic determinism of lignin content involves some large-effect genes. According to our gene prediction analysis, one of these outlier loci is located in the coding region of the *Anti-sigma-I factor RsgI6* gene which binds to and hydrolyses insoluble and soluble xylan substrates (Bahari et al., 2011), a group of hemicelluloses that is found in all cell walls of grasses (Mellerowicz & Gorshkova, 2011). High lignin content in vegetative biomass provides a high density of leaf tissues (high Leaf Dry Matter Content or LDMC), which has been reported to contribute to resistance to water stress (Garnier, Shipley, Roumet, & Laurent, 2001; Wilson, Thompson, & Hodgson, 1999).

The adaptive features of populations from long summer areas have however counterparts in less summer-stressful climates. A trade-off between summer growth in the southernmost experimental garden (*SGR_lu16*) and investment in sexual reproduction (*AHD_po17, AHD_lu16*, *DST_avg*,*DST_lu17 and resDST_lu17*) on the one hand, and summer growth (*SMH_po17* and *SGR_po17*), autumn growth (*AMH_*po17 and *AGR_po17*) and winter persistency (*SCD_wi1617_po*) in the northernmost experimental garden on the other hand, was evidenced along the summer length gradient (Fig. 6b). The outlier loci correlated to these two kinds of phenotypic features were common to some extent (Fig. 4b), suggesting that this trade-off could be partly due to antagonistic pleiotropy (Exposito-Alonso, Burbano, Bossdorf, Nielsen, & Weigel, 2019; Savolainen et al., 2013). However, the number of outlier loci found for autumn growth in the northernmost experimental garden (*AGR_po17*) (289, 45,66% of all outliers) was notably higher than the number found for summer growth in the southernmost experimental garden (*SGR_lu16*) (44, 6,95% of all outliers) (Fig. 4b).

From an ecophysiological point of view, a strong investment in seed production may have a negative impact on winter survival of vegetative tillers (low *SCD_wi1617_po*) (Barre et al., 2017). Balfourier & Charmet (1991) also found correlations between aftermath heading in perennial ryegrass natural populations and the latitude, temperature and aridity factors of their sites of origin. They observed that populations from hot and dry regions tended to invest more in seed production, while populations from cool and wet areas had more vigorous vegetative growth (higher spring and autumn growth and persistence) and less aftermath heading. This trade-off between vegetative and reproductive investments was also

pointed out as a major lever of adaptation to warming conditions in other perennial grasses (Volaire, Barkaoui, & Norton, 2014).

Lessons for perennial ryegrass breeding in the context of climate change

During the 50 past years, perennial ryegrass has been subjected to intense breeding to create cultivars for sowing meadows. Breeding efforts in Europe have been successful to strongly reduce aftermath heading and this has resulted in a correlative improvement of autumn growth and persistency in cool climate areas of Europe. A correlative increase in forage quality was also obtained including a lower fiber (notably lignin) content in vegetative biomass (Sampoux et al., 2011).

In the next decades, longer and drier summers are foreseen to occur in a large part of Europe due to anthropogenic climate change (Ergon et al., 2018). Although the length of the winter cold period is expected to shorten with milder average temperature, low temperatures events may still occur even quite late in the season (Dalmannsdottir et al., 2017; Ergon et al., 2018). Therefore, new efforts in breeding programs aiming to adapt perennial ryegrass to longer and drier summers should not be at the expense of adaptation to winter cold stresses. Our results showed that sets of different loci are involved in adaptation to long summer climate and adaptation to low winter temperatures. It should thus be possible to combine to a large extent these two kinds of adaptive features by genetic recombination.

The co-association network analysis revealed modules for both analyses of SNPs from quadrants I-III and quadrants II-IV (at $d < 0.1$, see Fig. 7). The observation of these modules point to the presence of relatively independent groups of SNPs with homogeneous trends of variation under the environmental conditions defined by these quadrants. Some of these modules are likely linked to functional genes that collectively serve a similar role of adaptation to cold winters (quadrants I-III) or long summers (quadrants II-IV). Indeed, we observed significantly enriched GO terms in the largest module of quadrants II-IV at $d < 0.1$. These GO terms were associated with redox regulation: oxidation-reduction processes (GO:0055114) and oxidoreductase activity (GO:0016491), which play an essential role in the acclimation of plants to abiotic stresses (Suzuki, Koussevitzky, Mittler, & Miller, 2012). This and other modules could be interpreted as reporting for adaptive genes with either, very close additive effects on phenotypes or with true interaction (epistasis), the latter meaning that adaptive alleles need to covary to have an effect on adaptation. All in all, co-association networks revealed a potential utility of CANCOR for investigating the interaction of adaptive loci involved in polygenic adaptations. Further experimental studies, higher density of sequencing and new progress in functional gene annotation would however be required to better understand the specific roles of adaptive genes and their interactions, and in general, the genomic architecture of environmental adaptation in perennial ryegrass.

Concluding remarks

Our pool-Seq GBS and HiPlex genotyping led to the identification of around 60.000 GBS tags of 86 bp per population (max. 80.000 GBS tags), in a genome of about 2.2 Gbp. Thus, only about 0.23% of the total genome size was effectively sequenced and markers from neighboring GBS tags were on average spaced about 40.000 bp apart. Given this fairly low GBS tag density and the expected short LD within outcrossing *L. perenne* natural populations (Blackmore et al., 2016; Keep et al., 2020), adaptive loci not in LD with GBS tags may have gone undetected. Whole genome re-sequencing would enhance the chance to detect more adaptive loci but would obviously require a higher cost for sequencing and computational power. Nevertheless, it is noteworthy that a significant part of the adaptive genetic variability has been detected with markers covering only 0.23% of the total genome size.

A statistical relationship between environment, genotype and phenotype does not constitute the unequivocal identification of adaptive loci. The identification of putatively adaptive loci should be confirmed by the implementation of empirical selection experiments testing the fitness consequences of specific alleles or of their combination (Hoban et al., 2016; Pardo-Diaz, Salazar, & Jiggins, 2015). Despite its possible limitations, our approach is distinctive at simultaneously analyzing multivariate environment, genotype and phenotype data. Because environment, genotype and phenotype are in essence mutually correlated and multi-dimensional, the ordination-based CANCOR test is a straightforward and efficient way to detect adaptive loci while at the same time identifying environmental gradients imposing selection and phenotypic traits responsible for adaptation.

Acknowledgements

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 European Commission (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. J.L. Blanco-Pastor has received the support of the EC in the framework of the Marie-Curie FP7 COFUND People Program, through the award of an AgreenSkills+ fellowship (grant agreement nº 609398). Support to J.L. Blanco-Pastor came partially from RéGàTe, a project funded by the French Ministry of Agriculture through the 2015 CASDAR program. The computational resources (Stevin Supercomputer Infrastructure) and services used for genotype calling were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University in Belgium, FWO and the Flemish Government – department EWI. The authors thank Michiel van Bel (VIB) for building the new PLAZA4.5 monocots instance that includes the novel gene set of perennial ryegrass and its functional annotations. We also thank two anonymous reviewers for their insightful comments that improved the quality of the manuscript.

Climate data was processed by Milka Radojevik and Christian Pagé (CECI, Université de Toulouse, CNRS CERFACS http://cerfacs.fr) from EURO4M-MESAN and EUMETSAT CM SAF grids.

References

Anderson, J. E., Kono, T. J. Y., Stupar, R. M., Kantar, M. B., & Morrell, P. L. (2016). Environmental association analyses identify candidates for abiotic stress tolerance in *Glycine soja*, the wild progenitor of cultivated soybeans. *G3; Genes|Genomes|Genetics*, *6*(4), 835–843. doi: 10.1534/g3.116.026914

Armstead, I. P., Turner, L. B., Farrell, M., Skøt, L., Gomez, P., Montoya, T., … Humphreys, M. O. (2004). Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the Hd3 heading-date locus in rice. *Theoretical and Applied Genetics*, *108*(5), 822–828. doi: 10.1007/s00122-003-1495-6

Armstead, I. P., Turner, L. B., Marshall, A. H., Humphreys, M. O., King, I. P., & Thorogood, D. (2008). Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics. *New Phytologist*, *178*(3), 559–571. doi: 10.1111/j.1469-8137.2008.02413.x

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, *465*(7298), 627–631. doi: 10.1038/nature08800

Bahari, L., Gilad, Y., Borovok, I., Kahel-Raifer, H., Dassa, B., Nataf, Y., … Bayer, E. A. (2011). Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. *Journal of Industrial Microbiology & Biotechnology*, *38*(7), 825–832. doi: 10.1007/s10295-010-0848-9

Balfourier, F., & Charmet, G. (1991). Relationships between agronomic characters and ecogeographical factors in a collection of French perennial ryegrass populations. *Agronomie*, *11*(8), 645–657. doi: 10.1051/agro:19910802

Bansal, K. C., Lenka, S. K., & Mondal, T. K. (2014). Genomic resources for breeding crops with enhanced abiotic stress tolerance. *Plant Breeding*, *133*(1), 1–11. doi: 10.1111/pbr.12117

668 Barre, P., Ruttink, T., Muylle, H., Lootens, P., Sampoux, J.-P., Rohde, A., … Roldán-Ruiz, I.
669 (2017). Natural diversity in vegetative and reproductive investments of perennial ryegrass
670 is shaped by the climate at the place of origin. *Grass and Forage Science*, *73*(1), 193–205.
671 doi: 10.1111/gfs.12304

672 Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic
673 level. *Nature Reviews Genetics*, *12*(11), 767–780. doi: 10.1038/nrg3015

674 Barton, N., Hermisson, J., & Nordborg, M. (2019). Why structure matters. *ELife*, *8*, e45380. doi:
675 10.7554/eLife.45380

676 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and
677 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
678 *(Methodological)*, *57.1*, 289–300. doi: 10/gfpkdx

679 Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genet*,
680 *10*(8), e1004412. doi: 10.1371/journal.pgen.1004412

681 Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., …
682 Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *ELife*,
683 *8*, e39725. doi: 10.7554/eLife.39725

684 Berger, J., Palta, J., & Vadez, V. (2016). An integrated framework for crop adaptation to dry
685 environments: responses to transient and terminal drought. *Plant Science*, *253*, 58–67. doi:
686 10.1016/j.plantsci.2016.09.007

687 Blackmore, T., Thorogood, D., Skøt, L., Mcmahon, R., Powell, W., & Hegarty, M. (2016).
688 Germplasm dynamics: the role of ecotypic diversity in shaping the patterns of genetic
689 variation in \textlessi\textgreaterLolium perenne\textless/i\textgreater. *Scientific Reports*,
690 *6*(March), 22603. doi: 10/gf4d26

691 Blanco-Pastor, José L., Barre, P., Keep, T., Ledauphin, T., Escobar-Gutiérrez, A., Roschanski, A.
692 M., … Sampoux, J.-P. (2020). *Data from: Canonical correlations reveal adaptive loci and*
693 *phenotypic responses to climate in perennial ryegrass* (Version 1, p. 591690831 bytes)
694 [Data set]. Dryad. doi: 10.5061/DRYAD.0P2NGF1XK

695 Blanco-Pastor, José Luis, Manel, S., Barre, P., Roschanski, A. M., Willner, E., Dehmer, K. J., …
696 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. doi: doi.org/10.1111/jbi.13587

Brozynska, M., Furtado, A., & Henry, R. J. (2016). Genomics of crop wild relatives: Expanding the gene pool for crop improvement. *Plant Biotechnology Journal*, *14*(4), 1070–1085. doi: 10.1111/pbi.12454

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*(3), e57438. doi: 10.1371/journal.pone.0057438

Byrne, S. L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., … Asp, T. (2015). A synteny-based draft genome sequence of the forage grass *Lolium perenne*. *Plant Journal*, *84*(4), 816–826. doi: 10.1111/tpj.13037

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. doi: 10.1111/1755-0998.12906

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution*, *36*(4), 852–860. doi: 10.1093/molbev/msz008

Charmet, G., Balfourier, F., Ravel, C., & Denis, J.-B. (1993). Genotype x environment interactions in a core collection of French perennial ryegrass populations. *Theoretical and Applied Genetics*, *86*(6), 731–736. doi: 10.1007/BF00222663

Contreras-Moreira, B., Serrano-Notivoli, R., Mohammed, N. E., Cantalapiedra, C. P., Beguería, S., Casas, A. M., & Igartua, E. (2019). Genetic association with high-resolution climate data reveals selection footprints in the genomes of barley landraces across the Iberian Peninsula. *Molecular Ecology*, *28*(8), 1994–2012. doi: 10.1111/mec.15009

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*(4), 1411–1423. doi: 10.1534/genetics.110.114819

Dai, A. (2013). Increasing drought under global warming in observations and models. *Nature Climate Change*, *3*(1), 52–58. doi: 10.1038/nclimate1633

726 Dalmannsdottir, S., Jørgensen, M., Rapacz, M., Østrem, L., Larsen, A., Rødven, R., & Rognli, O.
727 A. (2017). Cold acclimation in warmer extended autumns impairs freezing tolerance of
728 perennial ryegrass (*Lolium perenne*) and timothy (*Phleum pratense*). *Physiologia*
729 *Plantarum*, *160*(3), 266–281. doi: 10.1111/ppl.12548

730 de la Peña, R. C., Ebert, A. W., Gniffke, P. A., Hanson, P., & Symonds, R. C. (2011). Genetic
731 adjustment to changing climates: Vegetables. In *Crop adaptation to climate change* (pp.
732 396–410). Oxford, UK: Wiley-Blackwell. doi: 10.1002/9780470960929.ch27

733 Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package
734 rrBLUP. *The Plant Genome*, *4*(3), 250–255. doi: 10.3835/plantgenome2011.08.0024

735 Ergon, Å., Seddaiu, G., Korhonen, P., Virkajärvi, P., Bellocchi, G., Jørgensen, M., … Volaire, F.
736 (2018). How can forage production in Nordic and Mediterranean Europe adapt to the
737 challenges and opportunities arising from climate change? *European Journal of Agronomy*,
738 *92*, 97–106. doi: 10.1016/j.eja.2017.09.016

739 Eurostat. (2017). *Agriculture, forestry and fishery statistics* (R. Forti, Ed.). Luxembourg:
740 Publications Office of the European Union. doi: 10.2785/570022

741 Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural
742 selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*,
743 *573*(7772), 126–129. doi: 10.1038/s41586-019-1520-9

744 Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018).
745 Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis*
746 *thaliana*. *Nature Ecology and Evolution*, *2*(2), 352–358. doi: 10.1038/s41559-017-0423-0

747 FAO (Ed.). (2015). *Coping with climate change: the roles of genetic resources for food and*
748 *agriculture*. Italy, Rome: FAO.

749 Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2015). Detecting spatial
750 genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*,
751 *25*(1), 104–120. doi: 10.1111/mec.13476

752 Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for
753 detecting multilocus adaptation with multivariate genotype–environment associations.
754 *Molecular Ecology*, *27*(9), 2215–2233. doi: 10.1111/mec.14584

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science*, *334*(6052), 86–89. doi: 10.1126/science.1209271

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. doi: 10.1111/2041-210x.12382

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. doi: 10.1093/molbev/mst063

Garnier, E., Shipley, B., Roumet, C., & Laurent, G. (2001). A standardized protocol for the determination of specific leaf area and leaf dry matter content. *Functional Ecology*, *15*(5), 688–695. doi: 10.1046/j.0269-8463.2001.00563.x

Gienapp, P., Fior, S., Guillaume, F., Lasky, J. R., Sork, V. L., & Csilléry, K. (2017). Genomic quantitative genetics to study evolution in the wild. *Trends in Ecology & Evolution*, *32*(12), 897–908. doi: 10.1016/j.tree.2017.09.004

Gillespie, L. M., & Volaire, F. A. (2017). Are winter and summer dormancy symmetrical seasonal adaptive strategies? The case of temperate herbaceous perennials. *Annals of Botany*, *119*(3), 311–323. doi: 10.1093/aob/mcw264

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., … Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, *327*(5967), 812–818. doi: 10.1126/science.1185383

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, *47*(15), 965–978. doi: 10.1016/j.infsof.2005.09.005

Grivet, D., Sork, V. L., Westfall, R. D., & Davis, F. W. (2008). Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): a multivariate genetic approach to conservation planning. *Molecular Ecology*, *17*(1), 139–156. doi: 10.1111/j.1365-294x.2007.03498.x

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. doi: 10.1186/gb-2008-9-1-r7

Harper, J., De Vega, J., Swain, S., Heavens, D., Gasior, D., Thomas, A., … Armstead, I. (2019). Integrating a newly developed BAC-based physical mapping resource for *Lolium perenne* with a genome-wide association study across a *L. perenne* European ecotype collection identifies genomic contexts associated with agriculturally important traits. *Annals of Botany*, *123*(6), 977–992. doi: 10.1093/aob/mcy230

Hejcman, M., Hejcmanová, P., Pavlů, V., & Beneš, J. (2013). Origin and history of grasslands in Central Europe–a review. *Grass and Forage Science*, *68*(3), 345–363. doi: 10.1111/gfs.12066

Henry, R. J., & Nevo, E. (2014). Exploring natural selection to guide breeding for agriculture. *Plant Biotechnology Journal*, *12*(6), 655–662. doi: 10.1111/pbi.12215

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397. doi: 10.1086/688018

Hodgkin, T., & Bordoni, P. (2012). Climate change and the conservation of plant genetic resources. *Journal of Crop Improvement*, *26*(3), 329–345. doi: 10.1080/15427528.2011.609928

Hotelling, H. (1936). Relation between two sets of variates. *Biometrika*, *28*(3–4), 321–377. doi: 10.1093/biomet/28.3-4.321

Humphreys, M., Feuerstein, U., Vandewalle, M., & Baert, J. (2010). Ryegrasses. In *Fodder crops and amenity grasses* (pp. 211–260). New York, NY: Springer New York. doi: 10.1007/978-1-4419-0760-8_10

Jones, M. B., & Donnelly, A. (2004). Carbon sequestration in temperate grassland ecosystems and the influence of management, climate and elevated CO2. *New Phytologist*, *164*(3), 423–439. doi: 10.1111/j.1469-8137.2004.01201.x

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, *16*(18), 3955–3969. doi: 10.1111/j.1365-294x.2007.03442.x

Josephs, E. B., Berg, J. J., Ross-Ibarra, J., & Coop, G. (2019). Detecting adaptive differentiation in structured populations with genomic data and common gardens. *Genetics*, *211*(3), 989–1004. doi: 10.1534/genetics.118.301786

Keep, T., Sampoux, J.-P., Blanco-Pastor, J. L., Dehmer, K. J., Hegarty, M. J., Ledauphin, T., … Barre, P. (2020). High-Throughput Genome-Wide Genotyping To Optimize the Use of Natural Genetic Resources in the Grassland Species Perennial Ryegrass (Lolium perenne L.). *G3: Genes, Genomes, Genetics*, *10*(9), 3347–3364. doi: 10/gg92xh

Kovi, M. R., Fjellheim, S., Sandve, S. R., Larsen, A., Rudi, H., Asp, T., … Rognli, O. A. (2015). Population structure, genetic variation, and linkage disequilibrium in perennial ryegrass populations divergently selected for feezing tolerance. *Frontiers in Plant Science*, *6*, 1–13. doi: 10.3389/fpls.2015.00929

Lasky, J. R., Forester, B. R., & Reimherr, M. (2017). Coherent synthesis of genomic associations with phenotypes and home environments. *Molecular Ecology Resources*, *18*(1), 91–106. doi: 10.1111/1755-0998.12714

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, *21*(7), 1548–1566. doi: 10.1111/j.1365-294x.2012.05479.x

Levitt, J. (1962). *Responses of plants to environmental stresses*. Academic Press, New York. Maguire, JD.

Lorenzetti, F., Tyler, B. F., Cooper, J. P., & Breese, E. L. (1971). Cold tolerance and winter hardiness in *Lolium perenne*: I. Development of screening techniques for cold tolerance and survey of geographical variation. *The Journal of Agricultural Science*, *76*(2), 199–209. doi: 10.1017/s0021859600025545

Lotterhos, K. E., Yeaman, S., Degner, J., Aitken, S., & Hodgins, K. A. (2018). Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biology*, *19*(1), 157. doi: 10.1186/s13059-018-1545-7

839 Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W.
840 (2000). Development of a global land cover characteristics database and IGBP DISCover
841 from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*(6–7), 1303–1330.
842 doi: 10.1080/014311600210191

843 Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for
844 selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–
845 77. doi: 10.1111/1755-0998.12592

846 Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature News*,
847 *456*(7218), 18–21. doi: 10.1038/456018a

848 Manel, S. S., Andrello, M., Henry, K., Verdelet, D., Darracq, A., Guerin, P.-E. P.-E., … Devaux,
849 P. (2018). Predicting genotypes environmental range from genome-environment
850 associations. *Molecular Ecology*, *27*(13), 2823–2833. doi: 10.1111/mec.14723

851 Mellerowicz, E. J., & Gorshkova, T. A. (2011). Tensional stress generation in gelatinous fibres: a
852 review and possible mechanism based on cell-wall structure and composition. *Journal of*
853 *Experimental Botany*, *63*(2), 551–565. doi: 10.1093/jxb/err339

854 Mora, C., Caldwell, I. R., Caldwell, J. M., Fisher, M. R., Genco, B. M., & Running, S. W. (2015).
855 Suitable days for plant growth disappear under projected climate change: Potential human
856 and biotic vulnerability. *PLOS Biology*, *13*(6), e1002167. doi:
857 10.1371/journal.pbio.1002167

858 Norris, I. B., & Thomas, H. (1982). The effect of droughting on varieties and ecotypes of *Lolium*,
859 *Dactylis* and *Festuca*. *Journal of Applied Ecology*, *19*(3), 881–889. doi: 10.2307/2403290

860 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H.
861 (2018). *vegan: Community Ecology Package. R package version 2.5-2. https://cran.r-*
862 *project.org/package=vegan*.

863 Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of
864 adaptive evolution. *Methods in Ecology and Evolution*, *6*(4), 445–464. doi: 10.1111/2041-
865 210x.12324

Price, N., Lopez, L., Platts, A. E., & Lasky, J. R. (2020). In the presence of population structure: From genomics to candidate genes underlying local adaptation. *Ecology and Evolution*, *10*(4), 1889–1904. doi: 10.1002/ece3.6002

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation – not by sweeps alone. *Nature Reviews Genetics*, *11*(10), 665–667. doi: 10.1038/nrg2880

Redden, R., Yadav, S. S., Maxted, N., Dulloo, M. E., Guarino, L., & Smith, P. (2015). *Crop wild relatives and climate change*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. doi: 10.1111/mec.13322

Rieseberg, L. H., Widmer, A., Arntz, A. M., & Burke, J. M. (2002). Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences*, *99*(19), 12242–12245. doi: 10.1073/pnas.192360899

Roschanski, A. M., Barre, P., Escobar-Gutiérrez, A., Sampoux, J. P., Muylle, H., Thomas, I., … Willner, E. (2018). Patterns of spring growth and phenology in natural populations of *Lolium perenne* under contrasting field conditions. In G. Brazauskas, G. Statkevičiūtė, & K. Jonavičienė (Eds.), *Breeding Grasses and Protein Crops in the Era of Genomics* (pp. 14–19). Cham: Springer International Publishing. doi: 10.1007/978-3-319-89578-9_3

Ruttink, T., Sterck, L., Rohde, A., Bendixen, C., Rouzé, P., Asp, T., … Roldan-Ruiz, I. (2013). Orthology guided assembly in highly heterozygous crops: creating a reference transcriptome to uncover genetic diversity in *Lolium perenne*. *Plant Biotechnology Journal*, *11*(5), 605–617. doi: 10.1111/pbi.12051

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., … Marx, A. (2018). Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, *8*(5), 421–426. doi: 10.1038/s41558-018-0138-5

Sampoux, J. P., Barre, P., & Litrico, I. (2014). High-throughput genome-wide genotyping to revive the use of natural diversity in forage and turf breeding. In D. Sokolović, C. Huyghe, & J. Radović (Eds.), *Quantitative Traits Breeding for Multifunctional Grasslands and Turf* (pp. 313–323). Springer Netherlands.

Sampoux, J. P., Baudouin, P., Bayle, B., Béguier, V., Bourdon, P., Chosson, J. F., … 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*(1), 33–48. doi: 10.1111/j.1365-2494.2012.00896.x

Sampoux, J. P., Baudouin, P., Béguier, V., Bourdon, P., Chosson, J. F., Deneufbourg, F., … 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*(2), 117–129. doi: 10.1016/j.fcr.2011.05.007

Santure, A. W., & Garant, D. (2018). Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources*, *18*(4), 729–738. doi: 10.1111/1755-0998.12901

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. doi: 10.1038/nrg3522

Shukla, V., & Mattoo, A. K. (2013). Developing robust crop plants for sustaining growth and yield under adverse climatic changes. *Climate Change and Plant Abiotic Stress Tolerance*, 27–56. doi: 10.1002/9783527675265.ch02

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. doi: 10.1093/bioinformatics/btv351

Skøt, L., Humphreys, M. O., Armstead, I., Heywood, S., Skøt, K. P., Sanderson, R., … Hamilton, N. R. S. (2005). An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Molecular Breeding*, *15*(3), 233–245. doi: 10.1007/s11032-004-4824-9

Skøt, L., Sanderson, R., Thomas, A., Skøt, K., Thorogood, D., Latypova, G., … Armstead, I. (2011). Allelic variation in the perennial ryegrass FLOWERING LOCUS T gene is associated with changes in flowering time across a range of populations. *Plant Physiology*, *155*(2), 1013–1022. doi: 10.1104/pp.110.169870

Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., & Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, *23*(10), 2500–2513. doi: 10.1111/mec.12751

Storz, J. F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, *14*(3), 671–688. doi: 10.1111/j.1365-294x.2005.02437.x

Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C., … Joost, S. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, *17*(5), 1072-1089. doi: 10.1111/1755-0998.12629

Suzuki, N., Koussevitzky, S., Mittler, R., & Miller, G. (2012). ROS and redox signalling in the response of plants to abiotic stress. *Plant, Cell & Environment*, *35*(2), 259–270. doi: 10.1111/j.1365-3040.2011.02336.x

Talbot, B., Chen, T. W., Zimmerman, S., Joost, S., Eckert, A. J., Crow, T. M., … Manel, S. (2017). Combining genotype, phenotype, and environment to infer potential candidate genes. *Journal of Heredity*, *108*(2), 207–216. doi: 10.1093/jhered/esw077

Teuling, A. J. (2018, May 23). A hot future for European droughts. *Nature Climate Change*, Vol. 8, pp. 364–365. Nature Publishing Group. doi: 10.1038/s41558-018-0154-5

Thomas, H., & James, A. R. (1999). Partitioning of sugars in *Lolium perenne* (perennial ryegrass) during drought and on rewatering. *New Phytologist*, *142*(2), 295–305. doi: 10.1046/j.1469-8137.1999.00388.x

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8245–8250. doi: 10.1073/pnas.0409902102

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., … Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, *45*(W1), W122–W129. doi: 10.1093/nar/gkx382

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution*, *29*(12), 673–680. doi: 10.1016/j.tree.2014.10.004

951 Tilman, D., Wedin, D., & Knops, J. (1996). Productivity and sustainability influenced by
952 biodiversity in grassland ecosystems. *Nature*, *379*(6567), 718–720. doi: 10.1038/379718a0

953 Travis, W. R. (2016). Agricultural impacts: Mapping future crop geographies. *Nature Climate*
954 *Change*, *6*(6), 544–545. doi: 10.1038/nclimate2965

955 Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., &
956 Vandepoele, K. (2012). Dissecting plant genomes with the PLAZA comparative genomics
957 platform. *Plant Physiology*, *158*(2), 590–600. doi: 10.1104/pp.111.189514

958 Van Landeghem, S., De Bodt, S., Drebert, Z. J., Inzé, D., & Van de Peer, Y. (2013). The potential
959 of text mining in data integration and network biology for plant research: a case study on
960 Arabidopsis. *The Plant Cell*, *25*(3), 794–807. doi: 10.1105/tpc.112.108753

961 Veeckman, E., Vandepoele, K., Asp, T., Roldán-Ruiz, I., & Ruttink, T. (2016). Genomic variation
962 in the FT gene family of perennial ryegrass (*Lolium perenne*). In I. Roldán-Ruiz, J. Baert,
963 & D. Reheul (Eds.), *Breeding in a World of Scarcity* (pp. 121–126). Cham: Springer
964 International Publishing. doi: 10.1007/978-3-319-28932-8_18

965 Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., … Maxted,
966 N. (2013). A prioritized crop wild relative inventory to help underpin global food security.
967 *Biological Conservation*, *167*, 265–275. doi: 10.1016/j.biocon.2013.08.011

968 Volaire, F. (2018). A unified framework of plant adaptive strategies to drought: crossing scales
969 and disciplines. *Global Change Biology*, *24*(7), 2929–2938. doi: 10.1111/gcb.14062

970 Volaire, F., Barkaoui, K., & Norton, M. (2014, January 1). Designing resilient and sustainable
971 grasslands for a drier future: Adaptive strategies, functional traits and biotic interactions.
972 *European Journal of Agronomy*, Vol. 52, pp. 81–89. Elsevier. doi:
973 10.1016/j.eja.2013.10.002

974 Warschefsky, E., Penmetsa, R. V., Cook, D. R., & von Wettberg, E. J. (2014). Back to the wilds:
975 tapping evolutionary adaptations for resilient crops through systematic hybridization with
976 crop wild relatives. *American Journal of Botany*, *101*(10), 1791–1800. doi:
977 10.3732/ajb.1400116

978 Wilkins, P. W., & Humphreys, M. O. (2003, March). Progress in breeding perennial forage
979 grasses for temperate agriculture. *Journal of Agricultural Science*, Vol. 140, pp. 129–150.
980 Cambridge University Press. doi: 10.1017/S0021859603003058

981 Wilson, P. J., Thompson, K. E. N., & Hodgson, J. G. (1999). Specific leaf area and leaf dry matter
982 content as alternative predictors of plant strategies. *New Phytologist*, *143*(1), 155–162. doi:
983 10.1046/j.1469-8137.1999.00427.x

984 Yoder, J. B., Stanton-Geddes, J., Zhou, P., Briskine, R., Young, N. D., & Tiffin, P. (2014).
985 Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics*, *196*(4),
986 1263–1275. doi: 10.1534/genetics.113.159319

987 Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., … Buckler, E. S.
988 (2006). A unified mixed-model method for association mapping that accounts for multiple
989 levels of relatedness. *Nature Genetics*, *38*(2), 203–208. doi: 10.1038/ng1702

990 Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nature Reviews*
991 *Genetics*, *2*(12), 983–989. doi: 10.1038/35103590

992 Zhang, C., Fei, S., Arora, R., & Hannapel, D. J. (2010). Ice recrystallization inhibition proteins of
993 perennial ryegrass enhance freezing tolerance. *Planta*, *232*(1), 155–164. doi:
994 10.1007/s00425-010-1163-4

995 Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of Plants in the Old World: The origin*
996 *and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean*
997 *Basin*. New York, USA: Oxford University Press.

998

Data Accessibility

The DNA data is available in the NCBI Sequence Read Archive (BioProject PRJNA445949, Accessions SRR10243777 to SRR10244245). Supplemental data is available at https://doi.org/10.5061/dryad.0p2ngf1xk. Supplemental data includes: 1) Table S1 - Accessions from the natural diversity of perennial ryegrass used in the study; 2) Table S2 - Re-sequenced genomic regions in candidate genes putatively involved in climatic adaptation using Highly Multiplex Amplicon Sequencing (HiPlex). The table includes gene descriptions, primers and amplicons in GFF format, and primer sequences and amplicons in BED format (region between primers); 3) Table S3 - Description and values of variables reporting for the environment at sites of origin of studied populations from the natural diversity of perennial ryegrass; 4) Table S4 - Seasonal climatic conditions at the three experimental gardens (LU, ME, PO) over the duration of the experiments; 5) Table S5 - Description and values of phenotypic traits recorded on studied populations from the natural diversity of perennial ryegrass in three experimental gardens; 6) Table S6 - Number of outlier loci per environmental variable according to the GEA and GWAS univariate mixed models; 7) Table S7 - Outlier loci detected as strongly associated with environmental variables in GEA linear mixed models (FDR = 0.2) and with phenotypic traits in GWAS mixed models (FDR = 0.2); 8) Table S8 - CANCOR outlier SNP loci: Associated environmental variables and phenotypic traits and closest known gene including position, distance to outlier SNP, InterPro domain, gene ontology and functional annotation derived from gene prediction analysis; and 9) Data S1 - Genomic data: allele frequencies of 189,968 SNP loci in the 469 natural populations of *perennial ryegrass*. The remaining information that supports the findings of this study has been uploaded as a Supporting Information file: 1) Methods S1: HiPlex SNP set; 2) Methods S2: Environmental variables (climate-related variables and soil variables); 3) Methods S3: High throughput phenotyping; 4) Methods S4: GEA linear mixed models; 5) Methods S5: GWAS linear mixed models; 6) Methods S6: CANCOR test; 7) Results S1: Outlier loci detected by the CANCOR and the GEA-GWAS approaches. Code for running the CANCOR test and data files to replicate the analysis are available at https://doi.org/10.5281/zenodo.3992813.

Author Contributions

Author contributions: J.L.B.P., P.B., S.M. and J.P.S. designed research; T.K. performed the GWAS analysis, T.R., E.V. and K.V. performed the gene prediction and annotation analyses, J.L.B.P. performed all other analyses; T.K., T.L., A.E.G., A.M.R., E.W., K.J.D., M.H., H.M., T.R., E.V., I.R.R., K.V. and J.P.S. collected data; J.L.B.P., P.B., T.K., S.M. and J.P.S. interpreted results; J.L.B.P. and J.P.S. wrote the manuscript with feedback from P.B., T.K., T.L., A.E.G., K.J.D., H.M., T.R., I.R.R. and S.M.

Tables

Table 1. Environmental variables found as associated with SNP loci of perennial ryegrass by either the GEA-GWAS or CANCOR methods (FDR = 0.1 with both methods). Additional information about environmental variables is available in Supplemental Information Methods S2 and Table S3.

Environmental variable	Type	Description	Unit	Method†
bd_subsoil	Soil data	Subsoil bulk density	g cm^{-3}	GEA-GWAS
bio.ad.20	BIOCLIM derived variables	Precipitation - evapotranspiration of wettest quarter	mm	CANCOR
bio.ad.24	BIOCLIM derived variables	Evapotranspiration of wettest quarter	mm	CANCOR
bio.ad.27	BIOCLIM derived variables	Evapotranspiration of coldest quarter	mm	GEA-GWAS
bio2	BIOCLIM derived variables	Mean diurnal range	°C	CANCOR
bio3	BIOCLIM derived variables	Isothermality (*bio2* / *bio*7 × 100)	%	CANCOR
bio4	BIOCLIM derived variables	Temperature seasonality (standard deviation of average daily mean temperature per year-slice x100)	°C × 100	CANCOR
bio6	BIOCLIM derived variables	Average daily minimum temperature (*tasmin*) of coldest 14/15 days period	°C	CANCOR
bio7	BIOCLIM derived variables	Temperature Annual Range	°C	CANCOR
bio10	BIOCLIM derived variables	Mean temperature of warmest quarter	°C	GEA-GWAS
dtr_au	ETCCDI derived indices	Average daily temperature range for autumn period	°C	CANCOR
dtr_wi	ETCCDI derived indices	Average daily temperature range for winter period	°C	CANCOR
lmts	Ecophysiological indices	Length of the heat stress period	number of days	GEA-GWAS
oc_topsoil	Soil data	Topsoil organic carbon content	%	GEA-GWAS
pet_wi	Seasonal climate descriptors	Cumulated evapotranspiration for winter period	mm	GEA-GWAS
r01mm_au	ETCCDI derived indices	Count of days when precipitation ≥ 1mm for autumn period	count of days	CANCOR
r01mm_wi	ETCCDI derived indices	Count of days when precipitation ≥ 1mm for winter period	count of days	CANCOR
rx1day_au	ETCCDI derived indices	Maximum 1-day precipitation for autumn period	mm	CANCOR
sdii_au	ETCCDI derived indices	Simple precipitation intensity index for autumn period	mm	CANCOR
sdii_sp	ETCCDI derived indices	Simple precipitation intensity index for spring period	mm	CANCOR
sdii_wi	ETCCDI derived indices	Simple precipitation intensity index for winter period	mm	CANCOR

sis_wi	Seasonal climate descriptors	Average surface incident shortwave solar radiation per day for winter period	W m^{-2}	GEA-GWAS and CANCOR
su_an	ETCCDI derived indices	Number of summer days during the year	count of days	CANCOR
tasmax_wi	Seasonal climate descriptors	Average daily maximum temperature for winter period	°C	CANCOR
tasmin_wi	Seasonal climate descriptors	Average daily minimum temperature for winter period	°C	CANCOR
tawc_soil	Soil data	Total available water content from Pedo-Transfer-Function	mm	GEA-GWAS
tnn_wi	ETCCDI derived indices	Minimum value of daily minimum temperature for winter period	°C	CANCOR
tr_an	ETCCDI derived indices	Number of tropical nights during the year	count of nights	GEA-GWAS
txx_wi	ETCCDI derived indices	Maximum value of daily maximum temperature for winter period	°C	CANCOR

†GEA-GWAS: variables found significantly associated with GEA-GWAS outlier loci (FDR = 0.1 with GEA and GWAS). Only variables with strongest association with outlier SNPs are shown. CANCOR: variables with norm of regression slope projection on the first environmental canonical plane greater than 0.95 and highly correlated (|r| > 0.5) to some CANCOR outlier loci (FDR = 0.1 with the CANCOR test).

Table 2. Phenotypic traits found as potentially adaptive in natural populations of perennial ryegrass by either the GEA-GWAS or CANCOR methods (FDR = 0.1). Additional information about phenotypic traits is available in Supplemental Information Methods S3 and Table S5.

Phenotypic trait	**Exp. garden(s)**	**Record year(s)**	**Description**	**Unit**	**Method†**
ADL_10_me17	ME	2017	Acid Detergent Lignin content in aerial biomass dry matter	% dry matter	CANCOR
ADL_avg	ME	2017	Acid Detergent Lignin content in aerial biomass dry matter (average over record dates)	% dry matter	CANCOR
AGR_po17	PO	2017	Autumn growth rate	mm / growing-degree-days	CANCOR
AHD_lu16	LU	2016	Aftermath heading	1 (no fertile stem) to 9 (100% plants with fertile stems)	CANCOR

AHD_me16	ME	2016	Aftermath heading	1 (no fertile stem) to 9 (100% plants with fertile stems)	GEA-GWAS
AHD_po17	PO	2017	Aftermath heading	1 (no fertile stem) to 9 (100% plants with fertile stems)	CANCOR
AMH_po17	PO	2017	Autumn maximum height	mm	GEA-GWAS and CANCOR
CH300h_po16	PO	2016	Canopy height 300 degree days before spike emergence	mm	CANCOR
CH400h_po16	PO	2016	Canopy height 400 degree days before spike emergence	mm	GEA-GWAS and CANCOR
CHs500_me17	ME	2017	Canopy height 500 degree days after start of spring growth	mm	GEA-GWAS
DST_avg	LU, PO	2017	Spike density (average over exp. gardens)	1 (no fertile stem) to 9 (maximum density)	CANCOR
DST_lu17	LU	2017	Spike density	1 (no fertile stem) to 9 (maximum density)	CANCOR
HEA_avg	LU, PO	2016, 2017	Spike emergence (heading) date (average over exp. gardens and record years)	Growing-degree-days from start of spring growth (see Methods S6)	CANCOR‡
HFY_lu15	LU	2015	Proportion of plants heading during sowing year	1 (no fertile stem) to 9 (100% plants with fertile stems)	GEA-GWAS
HFY_po15	PO	2015	Proportion of plants heading during sowing year	1 (no fertile stem) to 9 (100% plants with fertile stems)	GEA-GWAS
HST_lu17	LU	2017	Straw height	cm	GEA-GWAS
resDST_lu17	LU	2017	Residual of regression of DST_lu17 on HEA_avg	1 (no fertile stem) to 9 (maximum density)	CANCOR
SCD_wi1516_po	PO	2015 to 2016	Soil coverage loss throughout winter 2015-2016 at PO	Difference between late and early scores, each recorded on a 1 (no living plants) to 9 (best soil coverage) scale	GEA-GWAS
SCD_wi1617_po	PO	2016 to 2017	Soil coverage loss throughout winter 2016-2017 at PO	Difference between late and early scores, each recorded on a 1 to 9 (best soil coverage) scale	CANCOR
SGR_lu16	LU	2016	Summer growth rate	mm / growing-degree-days	CANCOR

SGR_po17	PO	2017	Summer growth rate	mm / growing-degree-days	CANCOR
SMH_me16	ME	2016	Summer maximum height	mm	CANCOR
SMH_po17	PO	2017	Summer maximum height	mm	CANCOR
VAC_avg	LU, PO	2016, 2017	Vigor after cutting (average over exp. gardens and record dates)	1 (no regrowth) to 9 (strongest regrowth)	GEA-GWAS
VAC_lu17	LU	2017	Vigor after cutting (average after two cutting dates at LU in 2017)	1 (no regrowth) to 9 (strongest regrowth)	GEA-GWAS
WID_po16	PO	2016	Winter damage	1 (no damage) to 9	CANCOR

†GEA-GWAS: traits found significantly associated with GEA-GWAS outlier loci (FDR = 0.1 with GEA and GWAS). Only variables with strongest association with outlier SNPs are shown. CANCOR: Traits with norm of regression slope projection on the first phenotypic canonical plane greater than 0.90 and highly correlated ($|r| > 0.5$) to some CANCOR outlier loci (FDR = 0.1 with the CANCOR test).

‡ Trait included in the CANCOR results but with norm of regression slope smaller than 0.95.

Figures

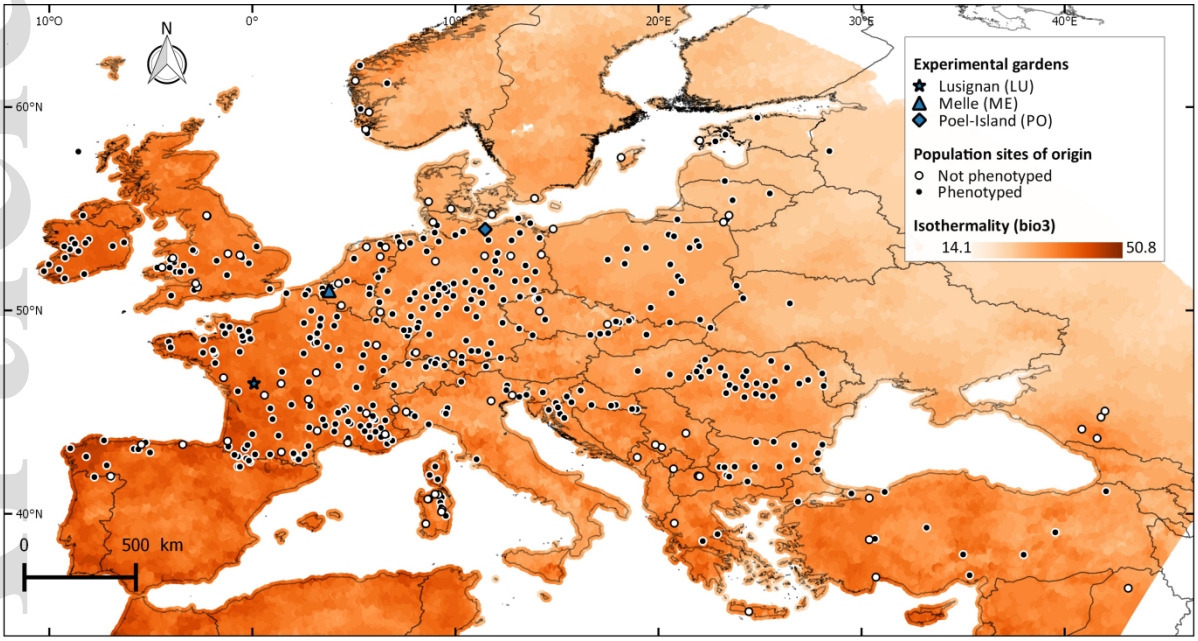

Fig. 1 – Spatial distribution of the 469 perennial ryegrass populations studied and locations of experimental gardens used for phenotyping. Isothermality values are displayed in background as an indicator of climatic variability across Europe.

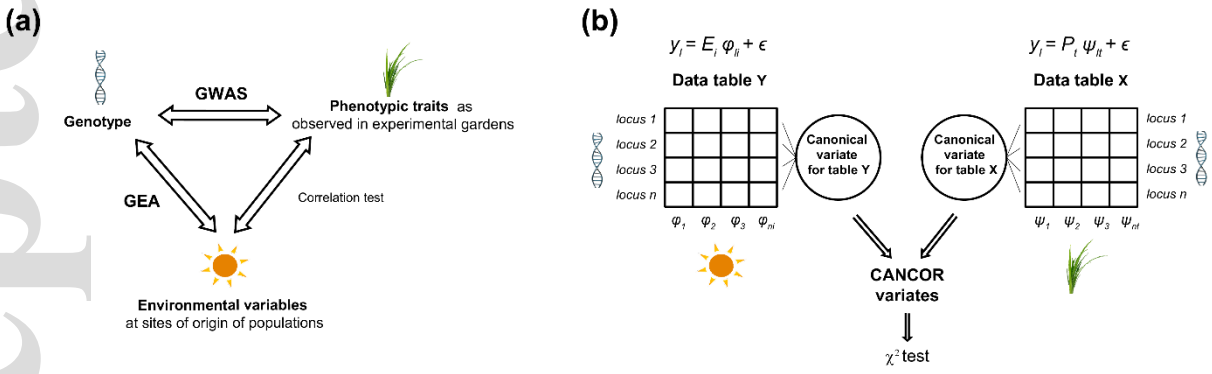

Fig. 2 – Two approaches used to detect adaptive loci. a) The GEA-GWAS approach: a locus is inferred as highly associated with both the environmental variable (GEA) and the phenotypic trait (GWAS). The environmental variable and the phenotypic trait should also be significantly correlated. b) The additive fixed effects (univariate regression slopes) of environmental variables and phenotypic traits on population alternative allele frequencies (AAFs) (y_l) of genotyped loci make up Tables Y and X, respectively. The CANCOR analysis is performed using columns of Y and X as input variables and loci as observations (see further details in Supporting Information, Methods S6). After determining the number of canonical

dimensions reporting for selection gradients (Supporting Information, Fig. S1), a χ^2 test on Mahalanobis distances is implemented to detect outlier loci (Fig. 3a and Supporting Information, Fig. S2).

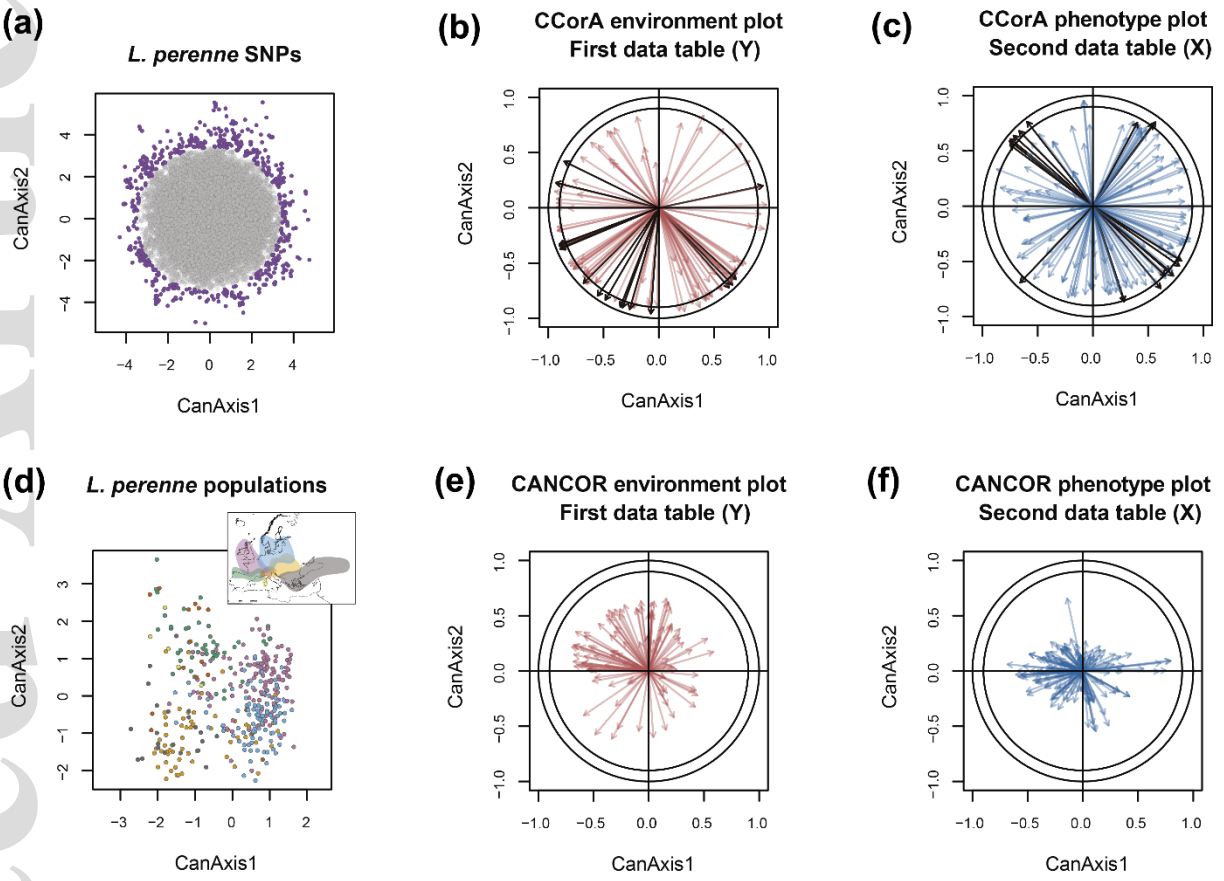

Fig. 3 – The CANCOR analyses. (a-c) Analysis using loci as observations. (d-f) Analysis using populations as observations. (b-c) Projections of slopes of univariate regressions of SNP alternative allele frequencies (AAFs) on environmental variables at sites of origin of populations (Y, see Fig. 2b) and on population mean values of phenotypic traits (X, see Fig. 2b) in the first environmental (b) and phenotypic (c) canonical planes, respectively. (e-f) Projections of environmental and phenotypic variables in the first environmental (e) and phenotypic (f) canonical planes, respectively. In (b) and (c), projections of Y and X input variables are displayed in black if their norm is greater than 0.95 and 0.9, respectively and if the correlation of the corresponding environmental or phenotypic variable with the population AAF of at least one outlier locus is such as $|r| > 0.5$. The projection of the regression slope of the environmental variable *HEA_avg* is also displayed in black although its norm equals 0.83. In (b-c) and (e-f), inner and outer circles mark 0.9 and 1 projection norm values respectively. Dots in (a) represent the coordinates of loci in the X (phenotypic) biplot of the first two canonical axes. Loci detected as significant by the CANCOR selection signal test (at FDR = 0.1) are displayed in purple. Dots in (d) represent the coordinates of populations in the X (phenotypic) biplot

of the first two canonical axes and dot colors represent neutral genetic clusters (as per Blanco-Pastor et al., 2019). Environmental and phenotypic variables whose regression slope projections are displayed in black in (b) and (c), respectively, are described in Table 1, Table 2 and Fig. 5. Detailed information about these variables is provided in Supporting Information, Methods S2 and S3.

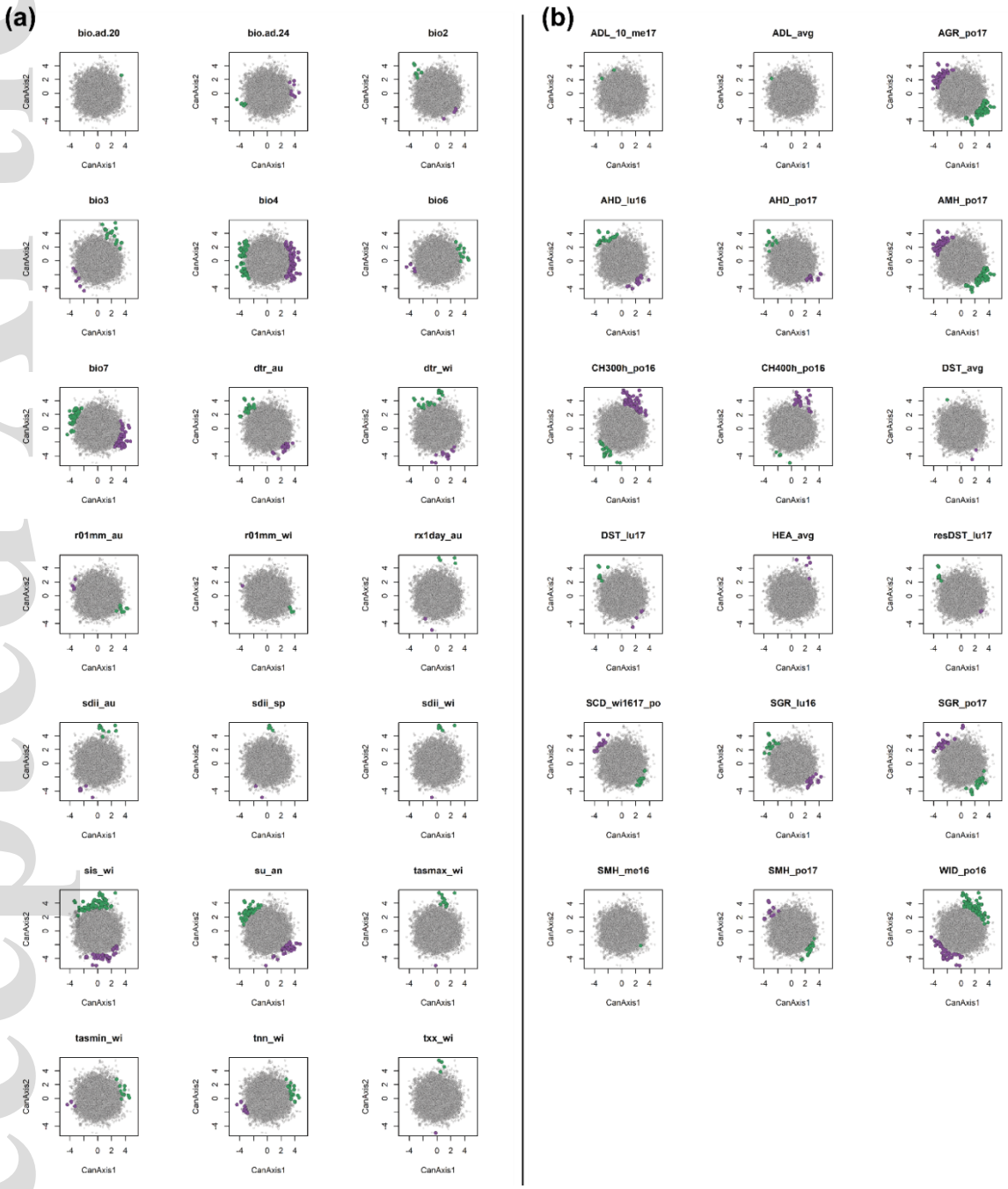

Fig. 4 – Outlier loci revealed by the CANCOR test whose alternative allele frequency is highly correlated (|r| > 0.5) with environmental variables or phenotypic traits well represented in the first environmental and phenotypic canonical planes (projection norms of corresponding input variables > 0.95 and 0.90,

respectively) or with *HEA_avg*. (a) Loci are plotted in the Y biplot representing the first two environmental canonical axes. (b) Loci are plotted in the X biplot representing the first two phenotypic canonical axes. Note that loci positions are computed on the basis of alternative allele frequencies. Purple and green colors indicate positive and negative correlations, respectively, between the locus alternative allele frequency and the environmental (a) or phenotypic (b) variable. See description of variables in Table 1 and Table 2. See detailed information on these variables in Supporting Information, Table S3, Table S5, Methods S2 and Methods S3.

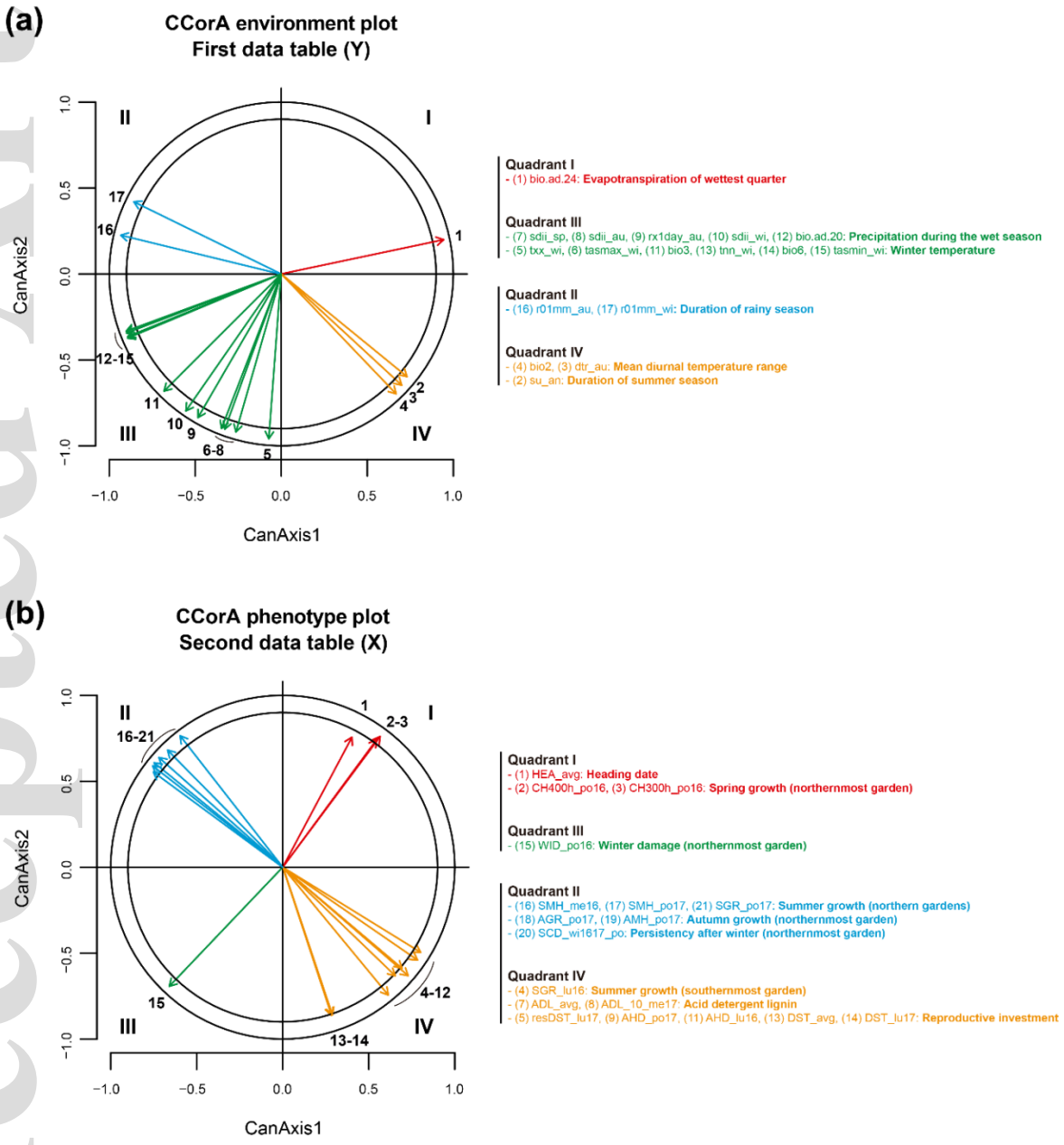

Fig. 5 – Synthetic representation of main climatic adaptations in perennial ryegrass natural populations. (a) and (b) represent the first environmental (Y) and phenotypic (X) canonical planes, respectively, of the CANCOR analysis using loci as observations. Projections of input environmental and phenotypic variables (regression slopes) are displayed if their norm is larger than 0.95 and 0.90, respectively. In addition, the corresponding environmental and phenotypic variables should be highly correlated to the population alternative allele frequency (AAF) of at least one outlier locus ($|r| > 0.5$). The projection of the regression slope of the environmental variable *HEA_avg* is also displayed although its norm equals 0.83. Colors and roman numbers I, II, III and IV indicate quadrants in the CANCOR canonical planes and groups of associated climate and phenotypic variables. Note that arrow positions are computed on the basis of the correlation between the variable and SNP alternative allele frequencies. Also note that the diagonal from quadrant I (red) to III (green) represents a cold-dry to mild-wet winter gradient whereas the diagonal from quadrant II (blue) to IV (orange) represents a long rainy season to long summer gradient.

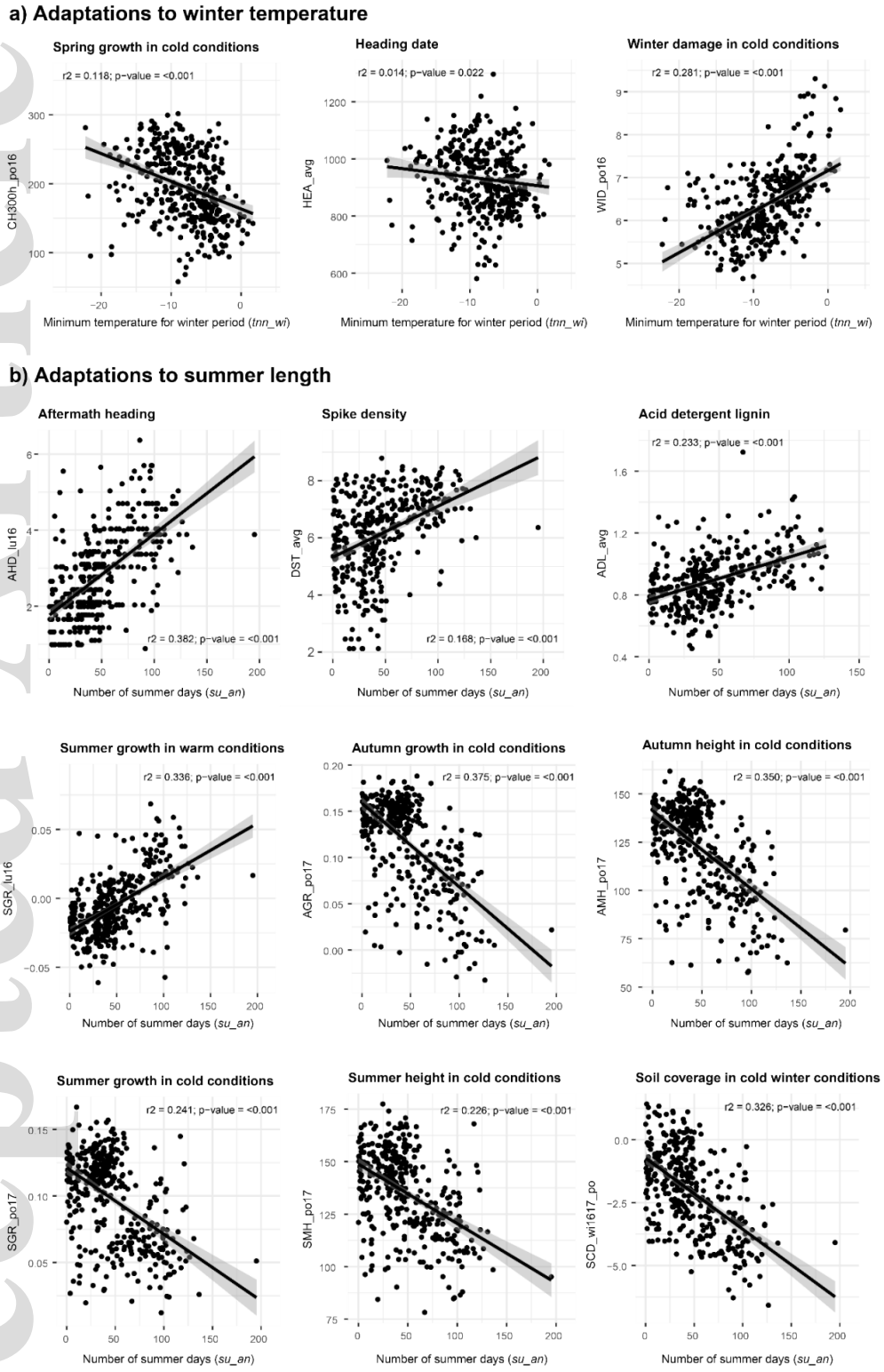

Fig. 6 – Relationships between two main selective climatic gradients represented by minimum value of daily minimum temperature in winter (*tnn_wi*) and number of summer days in the year (*su_an*) at sites of origin of populations and key phenotypic responses (mean values of populations) depicted by scatter plots.

Results of linear regressions of phenotypic traits on climatic variables are also displayed (r^2, p-values and trend lines).

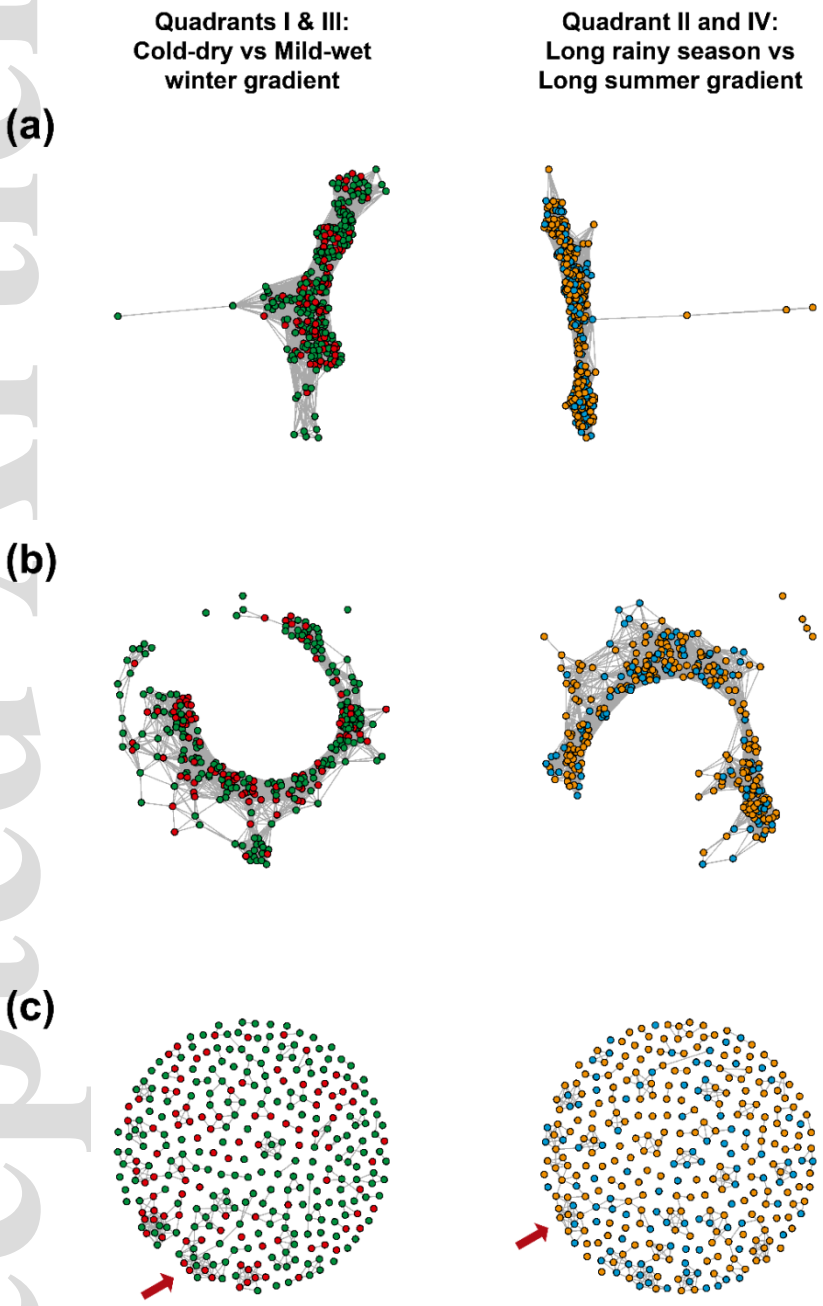

Fig. 7 – Co-association modules for the outlier SNPs identified by the CANCOR test. Each co-association network represents a distinct module. Colors schemes are according to the four quadrants of the CANCOR analysis (Fig. 5) and are displayed on the basis of alternative allele frequencies (SNPs from the same module can display different colors because one color represents the alternative allele as adaptive and the

other color represents the reference allele as adaptive). Climatic gradients corresponding to environmental variables with highest scores in each quadrant are indicated. Phenotypic traits associated with these quadrants are displayed in Fig. 5b. (a), (b) and (c) show alternative networks obtained with three different thresholds of pairwise Euclidean distances (< 1, < 0.5 and < 0.1, respectively). Red arrows point to the modules used for the Gene Ontology enrichment analyses.