Annual and perennial Medicago show signatures of parallel adaptation to climate and soil in highly conserved genes

José Luis Blanco-pastor, Isabel M. Liberal, Muhammet Sakiroglu, Yanling Wei, E. Charles Brummer, Rose L. Andrew, Bernard E. Pfeil

Received: 23 October 2020 | Revised: 24 June 2021 | Accepted: 29 June 2021

DOI: 10.1111/mec.16061

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Annual and perennial *Medicago* show signatures of parallel adaptation to climate and soil in highly conserved genes

José Luis Blanco-Pastor[1,2] | Isabel M. Liberal[1,3] | Muhammet Sakiroglu[4] | Yanling Wei[5] | E. Charles Brummer[5] | Rose L. Andrew[6] | Bernard E. Pfeil[1]

[1]Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden

[2]INRAE, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F), Lusignan, France

[3]Real Jardín Botánico de Madrid (RJB-CSIC), Madrid, Spain

[4]Department of Bioengineering, Adana Alparslan Turkes Science and Technology University, Adana, Turkey

[5]Plant Breeding Center, Department of Plant Sciences, University of California, Davis, Davis, CA, USA

[6]School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia

Correspondence
José Luis Blanco-Pastor, INRAE|Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F), Lusignan, France
Email: jlblancopastor@gmail.com

Funding information
FP7 People: Marie-Curie Actions, Grant/Award Number: 625308

Abstract

Human induced environmental change may require rapid adaptation of plant populations and crops, but the genomic basis of environmental adaptation remain poorly understood. We analysed polymorphic loci from the perennial crop *Medicago sativa* (alfalfa or lucerne) and the annual legume model species *M. truncatula* to search for a common set of candidate genes that might contribute to adaptation to abiotic stress in both annual and perennial *Medicago* species. We identified a set of candidate genes of adaptation associated with environmental gradients along the distribution of the two *Medicago* species. Candidate genes for each species were detected in homologous genomic linkage blocks using genome-environment (GEA) and genome-phenotype association analyses. Hundreds of GEA candidate genes were species-specific, of these, 13.4% (*M. sativa*) and 24% (*M. truncatula*) were also significantly associated with phenotypic traits. A set of 168 GEA candidates were shared by both species, which was 25.4% more than expected by chance. When combined, they explained a high proportion of variance for certain phenotypic traits associated with adaptation. Genes with highly conserved functions dominated among the shared candidates and were enriched in gene ontology terms that have shown to play a central role in drought avoidance and tolerance mechanisms by means of cellular shape modifications and other functions associated with cell homeostasis. Our results point to the existence of a molecular basis of adaptation to abiotic stress in *Medicago* determined by highly conserved genes and gene functions. We discuss these results in light of the recently proposed omnigenic model of complex traits.

KEYWORDS
alfalfa, climate, comparative genomics, conserved genes, *Medicago truncatula*, omnigenic model

1 | INTRODUCTION

The predicted increasing evapotranspiration and decreasing soil moisture levels in the context of climate change may result in unsuitable areas for cropping and reduced yields across temperate regions of the world (Boyer, 1982; Long & Ort, 2010; Tebaldi et al., 2006). Most crops may need to adapt to environmental changes, but their ability to do so depend on the genetic diversity at adaptive loci, which seems to be insufficient to cope with fast environmental changes (Bansal et al., 2014; Hodgkin & Bordoni, 2012; Khan et al., 2019).

Alfalfa is one of the world's most important fodder crops and the most efficient source of protein yield per hectare (Hanson, 1988), a key feature for a planet with greatly diminished arable land, but with increasing protein requirements for livestock and people. Alfalfa extracts also have tremendous potential for producing value-added goods such as biodegradable plastics, improved textiles and biofuels (Frame et al., 1998; Li et al., 2013; Samac et al., 2006), and its perennial root system also serves to reduce erosion (Small, 2011). Alfalfa is also of great interest because of the nitrogen-fixing abilities of the symbiotic bacteria housed in its roots nodules. Alfalfa crops often fix as much as 200–300 kg of nitrogen per hectare, making it a key component of sustainable agricultural systems (Peterson & Russelle, 1991). Demand for alfalfa is growing, but its yield and cultivation area is hampered by water deficits due to the need for irrigation in arid and semi-arid regions (Small, 2011; Song et al., 2019; Zhang et al., 2018). Compared to other crop species, little is known about the mechanisms by which genetic, physiological and molecular factors contribute to abiotic stress tolerance in alfalfa. Recent efforts have been made to identify potential candidate genes (Humphries et al., 2021; Shu et al., 2017; Zhang et al., 2015), but such candidate genes require further validation before their use in marker-assisted breeding, which is time consuming in perennial cultivars.

A way of indirectly validating candidate genes for abiotic stress is through the cross-association among allele frequencies, environmental variables and phenotypic trait variations in natural populations (Blanco-Pastor et al., 2021). Mining adaptive genes from natural populations is becoming an increasingly popular practice in the context of adaptation to abiotic stress (Brozynska et al., 2016; Dempewolf et al., 2014; Khan et al., 2019; Redden et al., 2015). Wild lineages of crop species have survived repeated and extreme environmental changes in the past (e.g., Blanco-Pastor et al., 2019), yet their adaptive variation remain largely unknown and untapped for crop improvement (Dempewolf et al., 2014; Ford-Lloyd et al., 2011), but see Fustier et al. (2017) or Blanco-Pastor et al. (2021).

The identification of climate-adaptive loci is inherently difficult because the architecture of adaptation to climate is polygenic (Blanco-Pastor et al., 2021; Exposito-Alonso et al., 2018; Josephs et al., 2019). Additionally, genomes of crop varieties are typically large and polyploid, often highly heterozygous, and usually composed of large amounts of repetitive elements and multicopy genes (Brozynska et al., 2016; Flagel & Wendel, 2009; Flavell et al., 1974). Nevertheless, advancements in studies of model species with diploid simpler genomes, such as *Arabidopsis thaliana* and *Medicago truncatula*, have shown that adaptation to environmental constraints is usually driven by small-effect alleles that exhibit climate specialization across the species ranges (Exposito-Alonso et al., 2019; Fournier-Level et al., 2011; Yoder et al., 2014). Generally, cultivated species may show different patterns of adaptation compared with well-known model species, which are often autogamous, short-lived and weedy (Savolainen et al., 2013). However, regardless of their apparent direct agronomic or horticultural value, it has been claimed that model species have the potential to provide building blocks for genetically enhancing adaptation of closely related modern cultivars (Bordat et al., 2011; Humphries et al., 2021; Paterson et al., 2010). Identifying adaptive functions in genes shared between model species and their crop relatives could provide useful information for enhancing adaptation to abiotic stress in crops and particularly in forage legumes (Annicchiarico et al., 2015; Araújo et al., 2015; Benny et al., 2019; Humphries et al., 2021; Joly-Lopez et al., 2017; Lenser & Theißen, 2013). Furthermore, studying adaptation genomics in a replicate context also provides an opportunity to gain statistical power over shared candidates (Balti et al., 2020; Elmer & Meyer, 2011).

In legumes, large effect genes have been identified in the model species *Medicago truncatula* or *Lotus japonicus* (Krusell et al., 2002; Lévy et al., 2004; Limpens et al., 2003; Stracke et al., 2004) that are potentially useful for legume crops. The genetic map of *M. truncatula* is largely syntenous with that of alfalfa (Choi et al., 2004; Li et al., 2014) and has been previously used to identify a gene for biotic stress resistance (i.e., to anthracnose fungal disease caused by *Colletotricum* spp.) and genes associated with freezing tolerance that could be useful for alfalfa improvement (Shu et al., 2017; Yang et al., 2007, 2008). Association mapping analyses in the model legume *M. truncatula* also identified genes involved in adaptation to climate and soil conditions, which were linked to flowering phenology and growth rate (Burgarella et al., 2016; Guerrero et al., 2018; Yoder et al., 2014). However, it remains unknown if such a set of genes useful for adaptation to the environment in an annual, self-compatible model plant such as *M. truncatula*, can be associated with abiotic stress resistance in a perennial, outcrossing crop such as alfalfa.

It is known that any type of stress in plants can trigger the same genetic responses independently of the origin of its cause (Baena-González et al., 2007; Exposito-Alonso et al., 2018). Abiotic stress can lead to similar conserved downstream responses with genetic parallelism in closely related and far-related species (Rellstab et al., 2020; Yeaman et al., 2016). Key genes involved in such downstream responses may be so conserved that they can have orthologues with similar functions in species as far-related as plants, yeast and mammalians (Baena-González & Sheen, 2008). However, as we move towards shallower evolutionary timescales it is more likely that genes associated with abiotic stress resistance in closely-related species share similar molecular pathways (Lenser & Theißen, 2013).

The recently proposed omnigenic model posits that adaptive genes can be categorized into “core” versus “peripheral” association with a given trait, as a way to distinguish between those with larger direct and interpretable mechanistic effects versus smaller indirect

and unknown effects (Boyle et al., 2017). Those core genes are presumably more likely to be the target of genetic parallelism in different species (Rellstab et al., 2020), thus the comparative approach in genome-wide association studies offers the opportunity to identify core genes of adaptation and to understand generalized molecular and phenotypic responses. Here, we used genomic, environmental and phenotypic data together with comparative association analyses in *M. sativa* and *M. truncatula* to: (i) find loci involved in adaptation to environmental gradients in wild populations of these two species; and (ii) investigate whether we could find genomic signatures of parallel adaptation to abiotic stress, as well as the gene ontologies associated with such parallel adaptive response.

2 | MATERIALS AND METHODS

2.1 | Sampling and genotyping

2.1.1 | *Medicago sativa* data set

Medicago sativa is a long-day perennial plant species native to Eurasia. It is basically an outcrossing species with little to no self-pollination. The species occurs in a variety of generally open habitats with moderately fertile soil (Small, 2011). Essentially, all *Medicago sativa* diploid individuals are wild. Polyploidization seems common (Muller et al., 2003; Small & Bauchan, 1984) and gene flow can occur from diploids to tetraploids via unreduced gametes but less commonly from tetraploid back to diploid plants (Petit et al., 1999; Small, 2011). We obtained 696 wild diploid *Medicago sativa* individuals from 150 natural populations (1–9 individuals per population), most of them conserved by the USDA National Plant Germplasm System. We selected accessions that listed "wild material" as improvement status in the Germplasm Resources Information Network (GRIN) database, and filtered out the accessions known to contain tetraploid individuals according to previous research (Brummer et al., 1999; Sakiroglu et al., 2010). Only accessions with passport information that pinpoints the collection location were used. In order to increase the number of samples in European high elevation areas, we also collected natural populations of alfalfa in the eastern Alps (Italy and Austria). To further exclude tetraploid individuals, we analysed the ploidy level on a bulk of five genotypes for each of the selected populations using flow cytometry. We performed flow cytometry following the protocol described in Brummer et al. (1999) on a FACSCalibur flow cytometry system (Becton Dickinson). Information on *M. sativa* genotypes used are available in Table S1.

We extracted genomic DNA from each individual following a CTAB-based DNA extraction protocol (Doyle & Doyle, 1987). We constructed libraries for genotyping-by-sequencing (GBS). Libraries were generated following the protocol of Elshire et al. (2011) with minor modifications as described by Li et al. (2015). We used the frequent cutting enzyme ApeKI and sets of 384 barcodes (4–8-bp) for multiplex sequencing. Sequences were obtained with an Illumina HiSeq 2500 system that generated 100-bp single-end reads. This protocol was chosen based on comparisons made among a few protocols and different enzymes, including the two-enzyme protocol by Poland et al. (2012) and the 2b-RAD protocol by Wang et al. (2012). Decision was made based on the number of sites genotyped that were shared among representative individuals (see Annicchiarico et al., 2017).

Alignments and SNP discovery described here are based on the Mt4.0v1 version of the *M. truncatula* genome sequence as a reference. We followed the TASSEL pipeline (Bradbury et al., 2007; Glaubitz et al., 2014) to obtain SNP data, assuming a sequencing error rate of 0.01. Genotypes based on fewer than 5x reads were set to missing data with SNPSIFT toolbox (Cingolani et al., 2012).

2.1.2 | *Medicago truncatula* data set

Medicago truncatula is a Mediterranean annual selfing plant that germinates and develops after the advent of the wet season, and produces seeds before summer drought becomes severe. This model species occurs in inland grasslands, meadows and shrublands (Small, 2011). We used the *M. truncatula* SNP data set from the *Medicago truncatula* Hap-Map Project (www.medicagohapmap.org) for comparative analyses with our *M. sativa* data set. Details on the sequencing technology used and SNP calling pipeline can be found in Branca et al. (2011) and Stanton-Geddes et al. (2013). Reads of 205 single-selfed plant lines (individual samples) with available GPS information were aligned to the *M. truncatula* reference genome (Young et al., 2011). We used the updated SNP calls based on the Mt4.0v1 reference genome assembly. Information on *M. truncatula* genotypes used are available in Table S2.

2.2 | Additional filters

For the *M. sativa* and *M. truncatula* data sets, rare alleles in triallelic sites were set as missing, we considered only sites for which there was sequence coverage in at least 100 individuals (Site Min Count filter in TASSEL), and for which the less-common allele was present in at least 10% of sampled individuals (minor allele frequency (MAF) filter ≥0.10). To generate a pruned subset of SNPs that were in approximate linkage equilibrium with each other we ran the --indep-pairwise command in Plink 1.9 (Chang et al., 2015) in the *M. sativa* data set. We used a window size of 50 SNPs, window step of five SNPs and an r^2 threshold of 0.5. Finally, we eliminated putative paralogous SNPs using the approach of McKinney et al. (2017) (see Supporting Information, Methods S1, Figure S1 and Figure S2). Then, we used an R script to retain those SNPs in the *M. truncatula* data set (4,375,118 SNPs after previous filters) that were within a 1 Kbp distance of the SNPs in the *M. sativa* data set. We only considered for subsequent analyses those individuals with less than 50% of missing data. After applying all filters, the *M. sativa* data set retained 14,160 SNPs in 675 individual genotypes (within 150 populations) and the *M. truncatula* data set retained 10,478 SNPs (within 145 bp mean distance to the

corresponding *M. sativa* SNP) in 202 individuals (single-selfed lines) (see Methods S1, Figure S3, Data S1 and Data S2) (Blanco-Pastor et al., 2021). For running analyses, we transformed the *M. sativa* matrix of individual genotypes to population alternative allele frequencies, where multiple individuals (independent genotypes) from the same population/USDA accession were lumped together.

2.3 | Genetic structure

We analysed the genetic structure in diploid *M. sativa* and *M. truncatula* using conStruct, a method for characterizing discrete population structure in the presence of continuous patterns of genetic differentiation (Bradburd et al., 2018). The conStruct method models genetic data as a combination of discrete layers, within each of which relatedness may decay continuously with geographic distance. We ran alternative spatial and nonspatial conStruct models with varying values of K ($K = 2$ to $K = 4$) and selected the spatial models with $K = 2$ as the best models for both *M. sativa* and *M. truncatula* based on cross-validation (see Methods S1, Figure S4). To categorize individuals, each genotype was assigned to the cluster with highest admixture proportion.

2.4 | Environmental variables

We obtained information on topoclimatic factors, including bioclimatic variables, atmospheric variables, radiation data, soil nutrients data, soil chemistry, soil granular material, soil depth, and soil water storage capacity, from the following public databases: worldclim.org (WC), The Climatic Research Unit (University of East Anglia) (CRU), The Satellite Application Facility on Climate Monitoring and The NASA Distributed Active Archive Centre for Biogeochemical Dynamics (DAAC) (see Methods S1, Figure S5 and Data S3–S4). We selected environmental factors that could potentially be linked to plant adaptation covering the complete study area and having a minimum resolution of 0.5 × 0.5°. The environmental data matrices contained 0.32% (*M. sativa*) and 2.79% (*M. truncatula*) of missing values.

2.5 | Genotype-environment associations

We first used redundancy analysis (RDA) to detect candidate genes under environmental selection in *M. sativa* and *M. truncatula*. RDA can be used as a genotype-environment association (GEA) method to detect loci under selection based on multivariate ordination (Forester et al., 2016). RDA determines how groups of loci covary in response to the multivariate environment, and can detect processes that result in weak, multilocus molecular signatures (Forester et al., 2018; Rellstab et al., 2015). Compared to other methods, RDA has shown a superior combination of low false-positive and high true-positive rates across a variety of selection scenarios (Forester et al., 2018). Another advantage of RDA is that it can be used to analyse many loci and environmental predictors simultaneously. RDA is a regression-based method, and so it is subject to problems when using highly correlated predictors. Hence, we removed correlated predictors with a correlation value of $|r| > .7$. Variable reduction was guided by an ecological interpretation of the relevance of possible predictors. We implemented a variable reduction protocol as follows: First, we performed cluster analyses of factors according to a matrix of absolute correlation values |r|. For that we used the complete linkage clustering method of the hclust function in R (R Core Team, 2021). After subsequent cluster analyses, we retained one variable in clusters with distance among variables lower than 0.3 (correlation higher than 0.7). We retained the same set of 20 variables for the *M. truncatula* and *M. sativa* data sets (see Methods S1, Figure S6). We favoured temperature variables over precipitation or radiation variables, monthly over quarterly climatic variables and topsoil over subsoil variables. After a preliminary RDA analysis, we also checked for multicollinearity using variance inflation factors (VIF) and confirmed that VIF of selected variables was <10.

For an easier interpretation of the RDA results, we performed independent RDAs using the bioclimatic data (WC and CRU) and the soil data (DAAC) separately for both *M. sativa* and *M. truncatula* (hereafter RDA-bioclim and RDA-soil). We used a test statistic based on the RDA analysis to find outlier SNPs (Capblancq et al., 2018). First, we recovered the locus loadings from the RDA analysis. A Mahalanobis distance D was then computed for each locus to identify SNPs showing extreme D values compared to the rest of the SNPs. Mahalanobis distances are distributed as chi-squared distribution with K degrees of freedom after correcting with the genomic inflation factor (Capblancq et al., 2018; Luu et al., 2017). We used the number of RDA axes (*K*) that provided the most uniform distribution of *p*-values, considering that distribution of *p*-values should be flat with enrichment only for low values (François et al., 2016). RDA tests with $K = 7$ and $K = 13$ for *M. sativa* (bioclim and soil, respectively) and $K = 7$ for *M. truncatula* (both bioclim and soil) provided the most uniform distributions of *p*-values and were therefore selected for the tests (see Methods S1, Figure S7). We ran the RDA analyses using the R package VEGAN (Oksanen et al., 2020).

In addition to RDA analyses, we also ran latent factor mixed models (LFMM), a univariate method with high power across alternative demographic scenarios that explicitly accounts for confounding effects due to population structure (Frichot et al., 2013). For running the analyses, we performed LFMM least-squares estimates with ridge penalty as implemented in the package LFMM (Caye et al., 2019). For LFMM, we ran analyses using the constrained axes of independent principal component analyses for bioclimatic and soil variables using the first $K = 3$ axes (selected after observing little additional variance explained by the fourth axis onwards, see Methods S1, Figure S8) and setting $K = 3$ latent factors for both species, which provided an appropriate lambda (genomic inflation factor) value (close to 1) (Frichot & François, 2015a, 2015b).

For RDA and LFMM analyses, significance values were calibrated using the genomic inflation factor and outlier SNPs were obtained with a calibrated *p*-value <.01 cutoff. Outlier SNPs were considered

for subsequent analyses if detected by any of the RDA and LFMM methods. We used a liberal *p*-value <.01 cutoff under the assumption that most climate and soil adaptive loci are of low effect. Only those putatively adaptive genes that were found associated with at least two different datasets, that is, that were either associated with the climate in the two species (*M. sativa* and *M. truncatula*), or associated with climate and phenotype in a single species (see below) were considered for the discussion.

Genomic location of outlier SNPs were identified using the *M. truncatula* genome as reference. We used their genomic position to identify candidate genes potentially under environmental selection located <1 Kbp upstream or downstream from any given outlier SNP, ignoring those genes at >1 Kbp from any given outlier SNP. Once these genes were identified, we used the Phytomine tool of the platform Phytozome (https://phytozome.jgi.doe.gov/phytomine/genomicRegionSearch.do) to retrieve their functional annotation on the *M. truncatula* Mt4.0v1 genome. Additionally, for the common set of GEA candidate genes (hereafter "shared GEA candidates"), we retrieved the homologues in *Arabidopsis thaliana* as determined by inParanoid (Remm et al., 2001).

To test whether we found more shared candidate genes between *M. sativa* and *M. truncatula* than expected by chance, we performed the resampling analysis of Rellstab et al. (2020). This analysis creates a random empirical distribution of expected overlap and compares it to the observed overlap in candidate genes (here the shared GEA candidates). In each of 100,000 iterations we picked a random gene subset with a size equal to the species-specific number of observed candidate genes (RDA + LFMM outliers at *p-value* < .01) from each species-specific list of annotated genes. The resulting distribution represented the random, empirical null distribution, and the proportion of observations above the real observed value denoted the empirical *p*-value.

2.6 | Genotype-phenotype associations

To identify candidate loci with a large effect on phenotypic traits, we analysed all the candidate loci of each species (SNPs identified by any of the RDA and LFMM analyses) through univariate SNP-phenotype associations. For this, we used 13 measures of traits related to growth, morphology and cell wall components in *M. sativa* from Sakiroglu et al. (2011) and 16 measures of traits related to symbiosis, growth and phenology in *M. truncatula* from Stanton-Geddes et al. (2013) and Burgarella et al. (2016). Significant univariate genotype-phenotype associations were identified with LFMM (at *p*-value <.01 threshold) using the same parameters as described above.

Additionally, we also investigated if a proportion of variance in phenotypic traits of *M. sativa* and *M. truncatula* could be explained by an optimized subset of RDA + LFMM shared GEA candidate genes. For that, we fit a Gaussian generalized linear model (GLM) via penalized maximum likelihood with LASSO penalty (Tibshirani, 1996). LASSO shrinks the coefficient estimates towards zero as the regularization parameter lambda (λ) increases, allowing to efficiently find the sparse model that involves a small subset of the parameters. We identified the best value for the lambda parameter using a 100th-fold cross-validation, which performs an automatic search across the space of lambda to identify the optimal value. We performed the GLM fit and cross-validation with the R package GLMNET (Friedman et al., 2010, ; Simon et al., 2011). Univariate LFMM and multivariate LASSO linear models were performed with data from 119 *M. sativa* populations (Sakiroglu et al., 2011) and two different sets of *M. truncatula* lines: 176 (Stanton-Geddes et al., 2013) and 154 (Burgarella et al., 2016).

2.7 | Gene ontology analysis in shared GEA candidates

We performed a gene ontology (GO) enrichment analysis of the RDA + LFMM shared GEA candidate genes with agriGO 2.0 (Tian et al., 2017). We used the singular enrichment analysis (SEA) tool with the *M. truncatula* Mt4.0v1 genome as a reference and applied the Fisher's exact test with a false discovery rate (FDR) correction (*q*-value <.01) (Benjamini & Hochberg, 1995). Relevance of GO terms for environmental adaptation was discussed in light of previous studies in other organisms.

3 | RESULTS

3.1 | Genetic structure

Genetic clusters in the *M. sativa* data set represented the infraspecific classification of *M. sativa* with the purple cluster denoting the subspecies *caerulea* (*caerulea* cluster) and the yellow cluster denoting the subspecies *falcata* (Figure 1a, c). Cluster contributions of the hybrid subspecies × hemicycla individuals ranged from 38.16% to 74.22% for the *caerulea* cluster and from 25.78% to 61.83% for the *falcata* cluster (see Table S1). *Medicago truncatula* showed a marked geographic structure with populations assigned to the blue cluster mostly located in the western side of the Mediterranean region (west cluster), whereas the populations assigned to the red cluster were located in the eastern side of Mediterranean region (east cluster) (Figure 1b, d and Table S2).

3.2 | Genotype-environment associations

RDA bioclim explained 21.7% of the genetic variance in *M. sativa* and 10.5% of the variance in *M. truncatula*. RDA soil explained 22.4% of the genetic variance in *M. sativa* and 12.7% of the variance in *M. truncatula*. For both species and analyses, the scores of the populations in the first constrained axis showed a gradient of genetic variation in response to the multivariate environment represented by RDA1, with genetic groups (*falcata*/*caerulea* for *M. sativa* and east/

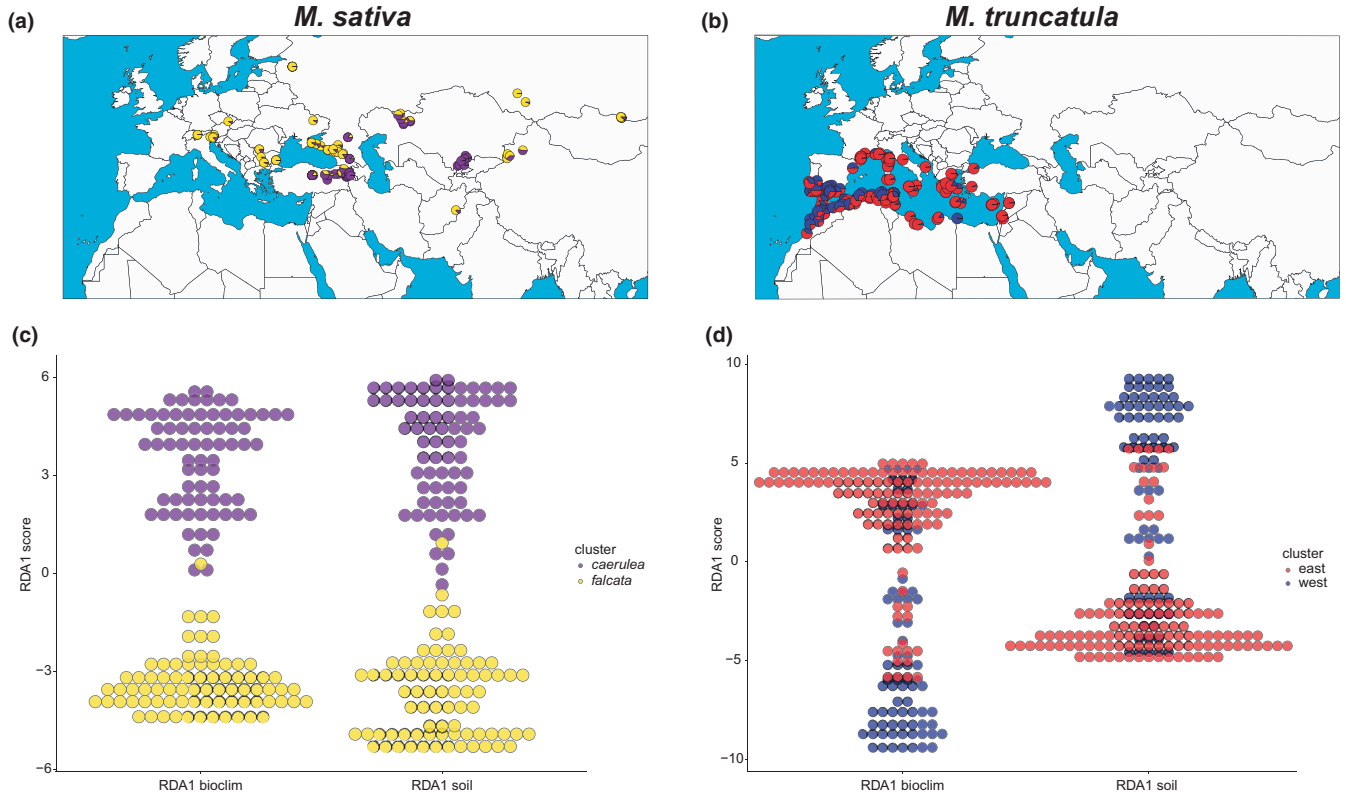

FIGURE 1 Diploid *Medicago sativa* and *M. truncatula* population structure and redundancy analyses (RDA). Spatial distribution of admixture proportions estimated for the *M. sativa* dataset (a) and the *M. truncatula* data set (b) using the spatial Bayesian model implemented in conStruct with two clusters ($K = 2$). The analysis on the *M. sativa* data set contained a matrix of 14,160 SNPs in 675 individual genotypes (within 150 populations) and the *M. truncatula* data set contained 10,478 SNPs (within 145 bp mean distance to the corresponding *M. sativa* SNP) in 202 individuals (single-selfed lines). We show the scores of the populations in the first constrained axis (which represents from 43% to 67% of the total environmentally constrained genetic variance) of all four RDAs preformed on *M. sativa* (c) and *M. truncatula* (d) data sets. Pie charts/coloured dots represent populations (*M. sativa*) and single-selfed lines (*M. truncatula*), in (c) and (d) colours represent the assignment to the cluster with highest admixture proportion

west for *M. truncatula*) representing the two extremes of the adaptive genetic variation (Figure 1c,d). The set of significant environmental variables of the RDA analyses and their scores in RDA1 axes are displayed in Table 1. We found four bioclimatic variables and two soil variables that were significant for both species (at p-value <.01): Potential evapotranspiration (CRU_PET), precipitation (CRU_PRE), precipitation of driest month (WC_BIO14), min. temperature of coldest month (WC_BIO6), soil depth with plant-extractable water (DAC_DUNNE) and subsoil pH (DAAC_S_PH_H2O). Five additional environmental variables (one bioclimatic and four soil variables) were significant for *M. sativa* and four additional environmental variables (two bioclimatic and two soil variables) were significant for *M. truncatula* (p-value <.01) (Table 1).

Using a p-value <.01 cutoff for the RDA tests, we detected 143 (bioclim) and 187 (soil) outlier SNPs for *M. sativa*, and 159 (bioclim) and 181 (soil) outlier SNPs for *M. truncatula*. Using LFMM (at p-value <.01), we detected 930 (bioclim + soil) outlier SNPs for *M. sativa*; and 884 (bioclim + soil) for *M. truncatula* (Figure 2). The joint list of outlier SNPs included 1,149 GEA outliers in *M. sativa* and 1,057 GEA outliers in *M. truncatula*. These outliers targeted 1,065 annotated genes in *M. sativa* (Supporting Information, Data S5) and 972 genes in *M. truncatula* (Supporting Information, Data S6) at 1 Kbp window size. Among these, we found 168 shared GEA candidates (Data S7), which was significantly above the expected overlap (Figure 3). None of the shared GEA candidate genes were linked to the same SNP position in both species. The resampling analysis generated a number of shared genes higher or equal to the observed number in only 70 out of the 100,000 random subsamples (empirical p-value = .0007) (Figure 3). Among the set of 168 shared GEA candidates, 118 (70.23%) had homologues in the *Arabidopsis thaliana* genome.

3.3 | Genotype-phenotype associations

From the list of 1065 candidate genes (RDA + LFMM) in *M. sativa*, we identified 143 genes (13.4%) with significant univariate phenotype-genotype associations (at p-*value* < .01) (Supporting Information, Data S5). Strongest associations were found with a set of 13 phenotypic traits (Figure 4a). From the list of 972 candidate genes in *M. truncatula*, we identified 233 genes (24%) with significant univariate phenotype-genotype associations (at p-*value* < .01) (Supporting Information, Data S6). Strongest associations were

TABLE 1 Results of ANOVA-like permutation tests for redundancy analyses (RDA) to assess the significance of individual environmental variables

		Medicago sativa				*Medicago truncatula*			
Variables (terms)	**Definition**	**Variance**	***F***	***p*-value (>*F*)**	**RDA1 score**	**Variance**	***F***	***p*-value (>*F*)**	**RDA1 score**
RDA-bioclim									
CRU_PET	Potential evapotranspiration	1303.9	16.6989	.000999**	0.760	160.1	3.3116	.000999**	−0.274
CRU_PRE	Precipitation	529.3	6.7788	.000999**	−0.009	185.8	3.8435	.000999**	−0.361
WC_BIO14	Precipitation of driest month	522.3	6.6889	.000999**	−0.597	337.2	6.9765	.000999**	0.545
WC_BIO3	Isothermality (BIO2/BIO7)x100	158.0	2.0237	.016983	0.428	119.3	2.4672	.002997*	−0.386
WC_BIO5	Max temperature of warmest month	127.5	1.6329	.044955	0.632	110.2	2.2795	.002997*	−0.164
WC_BIO6	Min temperature of coldest month	204.6	2.6197	.005994*	0.040	136.6	2.8269	.000999**	−0.098
WC_BIO8	Mean temperature of wettest quarter	226.2	2.8968	.000999**	−0.569	51.7	1.0693	.274725	0.050
RDA-soil									
DAAC_AWC_CLASS	Available water storage capacity of soil	399.6	4.9439	.000999**	0.447	102.7	2.1106	.011988	0.213
DAAC_BD	Bulk density	176.2	2.1796	.005994*	0.166	76.4	1.5703	.042957	0.010
DAAC_DUNNE	Soil depth with plant-extractable water	733.1	9.0695	.000999**	−0.703	157.9	3.2452	.000999**	0.351
DAAC_FC	Soil water content at field capacity	350.9	4.3412	.000999**	−0.368	79.9	1.6433	.032967	0.193
DAAC_S_GRAVEL	Subsoil gravel content	154.5	1.9110	.013986	−0.256	52.3	1.0755	.277722	−0.039
DAAC_S_PH_H2O	Subsoil pH (in water)	263.2	3.2557	.001998*	−0.142	177.6	3.6508	.000999**	0.386
DAAC_SMBC	Soil microbial biomass carbon	147.0	1.8182	.026973	0.004	84.4	1.7348	.024975	−0.114
DAAC_SPD	Soil phosphorus distribution	162.9	2.0148	.010989	0.527	74.3	1.5275	.053946.	−0.124
DAAC_T_CEC_CLAY	Cation exchange capacity of the clay fraction in the topsoil	130.7	1.6169	.045954	0.070	174.5	3.5873	.000999**	−0.081
DAAC_T_CLAY	Topsoil clay fraction	118.0	1.4597	.082917	0.187	112.6	2.3150	.006993*	0.463
DAAC_T_PH_H2O	Topsoil pH (in water)	197.3	2.4407	.006993*	0.611	94.4	1.9400	.013986	0.077
DAAC_T_SILT	Topsoil silt fraction	164.7	2.0377	.011988	0.010	64.8	1.3311	.092907	0.150
DAAC_TND	Total nitrogen density	169.3	2.0948	.013986	−0.191	81.4	1.6740	.033966	0.007

**p-value <.001; *p-value <.01.

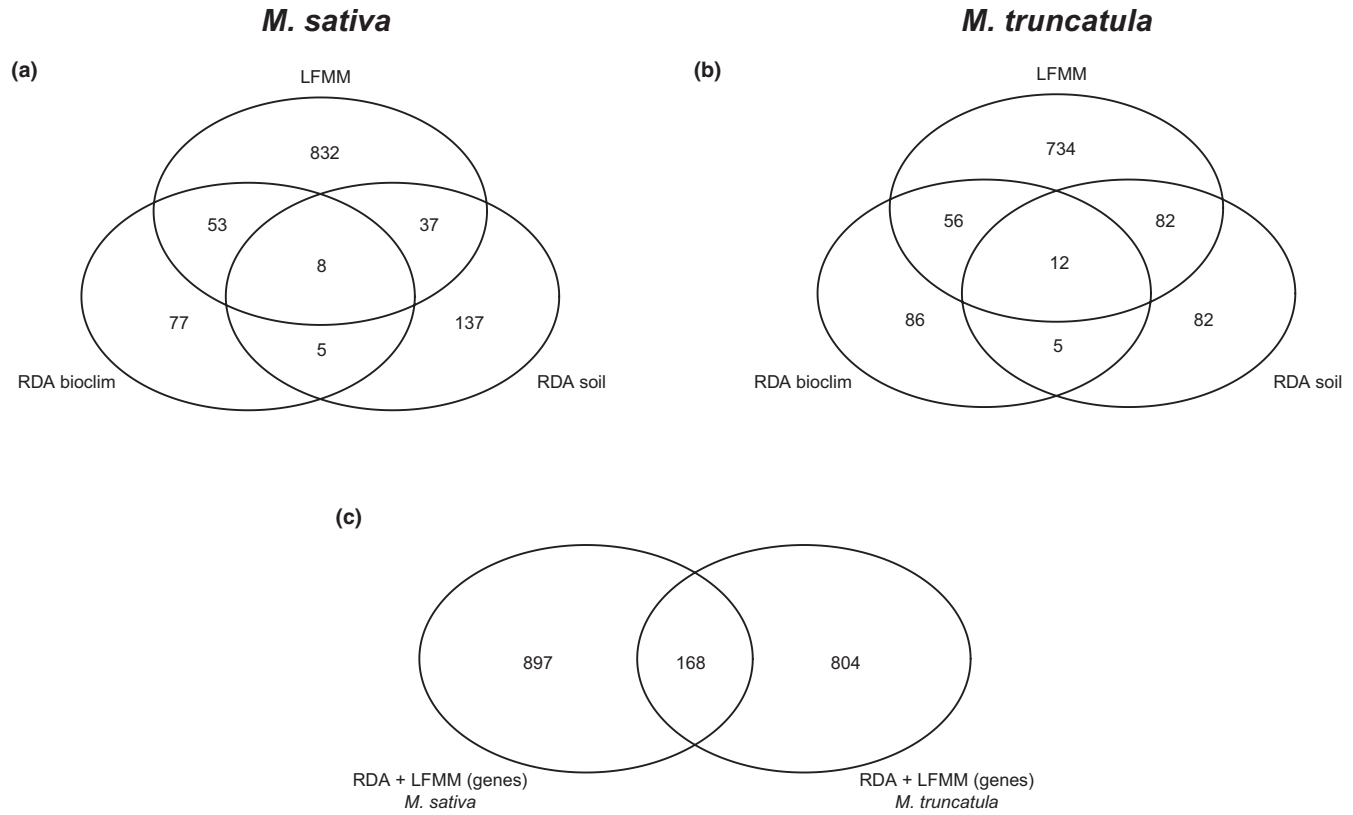

FIGURE 2 Venn diagrams displaying the number of outlier SNPs detected by the redundancy analyses (RDA) and latent factor mixed models (LFMM) in (a) *Medicago sativa* and (b) *M. truncatula*. (c) Number of candidate genes located <1 Kbp upstream or downstream from any given outlier SNP in each species and shared by both species

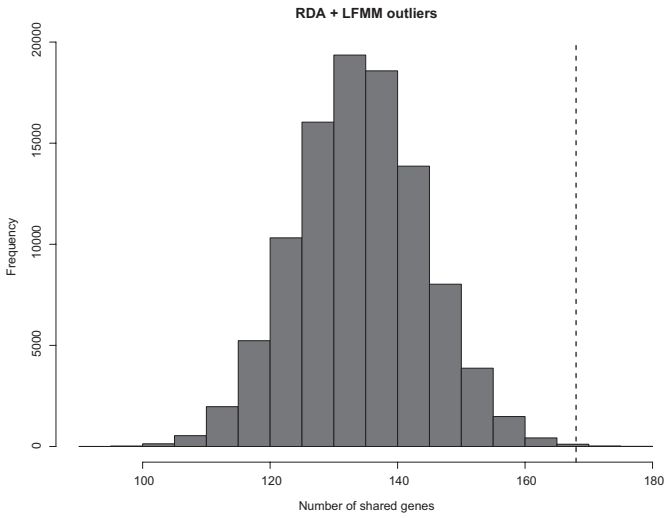

FIGURE 3 Shared signatures of adaptation in redundancy analyses and latent factor mixed model analyses (RDA + LFMM) candidate genes. We show the number of shared candidate genes (genes located <1 Kbp upstream or downstream from any given outlier SNP) between *Medicago sativa* and *M. truncatula* (dashed line, 168 genes) compared to random subsamples, iterated 100,000 times using the complete gene lists. We indicate the average expected number of shared genes, the observed number of shared genes and the empirical *p*-value

found with a set of 16 phenotypic traits (Figure 4b). From the set of 168 shared GEA candidate genes between the two species, 22 (13.1%) were significantly associated with a phenotypic trait in *M. sativa* (see Supporting Information, Data S5) and 53 (31.5%) were significantly associated with a phenotypic trait in *M. truncatula* (see Data S6). Only four of these were associated with a phenotypic trait in both species (Medtr4g048000, Medtr5g082560, Medtr5g082570 and Medtr8g106950). In *M. sativa*, these genes were associated with arabinose content and regrowth vigor after harvest, while in *M. truncatula*, they were associated with plant height (Table 2).

Optimized subsets of shared GEA candidate genes obtained with LASSO regressions explained 20%–79% of variance in phenotypic traits in *M. sativa* and 0%–48% in *M. truncatula*. Traits with highest variance explained in *M. sativa* and *M. truncatula* were also arabinose content and plant height, respectively (Table 3).

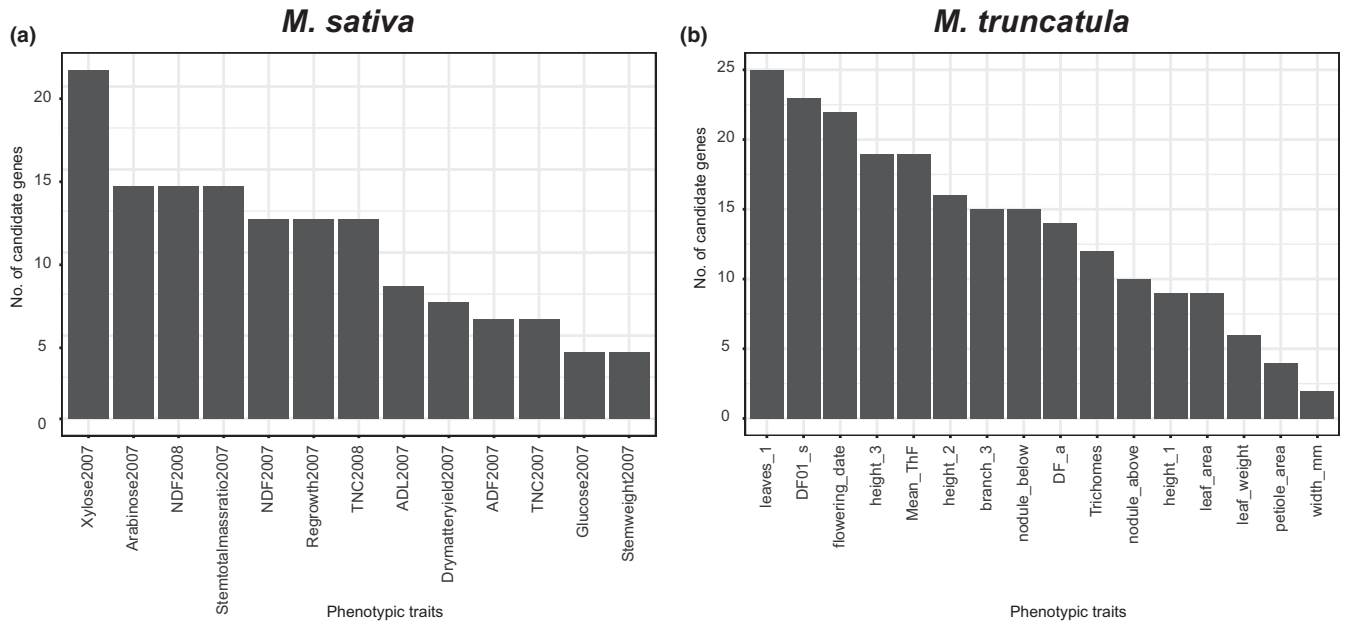

FIGURE 4 Number of candidate genes in (a) *Medicago sativa* and (b) *M. truncatula* — genes located <1 Kbp upstream or downstream from any given outlier SNP derived from redundancy analyses (RDA) or latent factor mixed models (LFMM) — associated with a phenotypic trait (phenotypic trait data from Burgarella et al., 2016; Sakiroglu et al., 2011; Stanton-Geddes et al., 2013). The bars display the number of candidates genes showing the strongest significant association with the particular trait, as found by genotype-phenotype association analyses. See Table 3 for description of phenotypic traits

3.4 | Gene ontology

The GO enrichment analysis showed 7 GO terms significantly enriched (q-value < .01) among the *M. sativa* and *M. truncatula* shared GEA candidate genes (Table 4). Shared GO terms included three biological processes: transport (GO:0006810), establishment of localization (GO:0051234) and localization (GO:0051179); two cellular components: cytoskeleton (GO:0005856) and cytoskeletal part (GO:0044430); and two molecular functions: motor activity (GO:0003774) and transporter activity (GO:0005215).

4 | DISCUSSION

Over the past few decades, evidence has been accumulating for the conservation of key genes responsible for core molecular processes even across exceptionally distant groups of species (Baena-González & Sheen, 2008). Our study shows that intraspecific variation in these highly conserved genes may play a key role for adaptation to climate in species with a wide range of variation in relatedness, habitat differences and life-history traits (see also Rellstab et al., 2020; Yeaman et al., 2016). Specifically, we show that certain genetic mechanisms of adaptation are, to some extent, genetically constrained in the genus *Medicago* and regulate adaptation in species with different life histories, mating systems and phenologies such as *M. sativa* and *M. truncatula*.

4.1 | Isolation-by-environment in annual and perennial *Medicago*

In the present study, we show that the species *M. sativa* and *M. truncatula* display a genetic structure pattern that follows the gradient of genotype-environment interactions, with two presumably neutral clusters placed at the extremes of the adaptive genetic variation (Figure 1c,d). This reflects a genome-wide pattern of isolation-by-environment (IBE) where clusters represent genetic groups with different niche breadth. This occurs for both *Medicago* species but less clearly in *M. truncatula*, where an IBE pattern cannot be clearly discerned from an isolation-by-distance (IBD) pattern (see also Burgarella et al., 2016; Sakiroglu et al., 2010; Small, 2011). One of the major difficulties in analysing adaptive genetic diversity is to remove the confounding factors of genetic structure for which alternative correction approaches have been proposed (e.g., Caye et al., 2019; Coop et al., 2010; Yu et al., 2006). However in cases with a strong IBE pattern, the typical correction for population structure used in association analyses can mask true positive associations. Using liberal thresholds and validation with two sources of data in association analyses has proven useful for finding candidate genes of local adaptation while reducing false positives in cases with strong IBE patterns (Blanco-Pastor et al., 2021; Rellstab et al., 2020; Talbot et al., 2017). Here, to maximize the chances of finding true signals of local adaptation in the presence of IBE we performed association analyses using liberal thresholds (*p-value* < .01) in complementary environmental and phenotypic data sets.

TABLE 2 Candidate genes of environmental adaptation shared by *Medicago sativa* and *M. truncatula* significantly associated with phenotypic traits in univariate genotype-phenotype associations in both species

Gene	Chr.	Start	End	SNP (*M. sativa*)	SNP (*M. truncatula*)	Length	Description	LFMM Phenotypic trait (*p*-value)	LFMM Phenotypic trait (best significant association)[a]
Medtr4g048000	chr4	16929374	16947707	chr4:16945327		18,334	(1 of 2) PTHR11206//PTHR11206:SF143 - MULTIDRUG RESISTANCE PROTEIN	.0026	Regrowth2007
Medtr4g048000	chr4	16929374	16947707		chr4:16945297	18,334	(1 of 2) PTHR11206//PTHR11206:SF143 - MULTIDRUG RESISTANCE PROTEIN	.0050	height_3
Medtr4g048000	chr4	16929374	16947707		chr4:16947025	18,334	(1 of 2) PTHR11206//PTHR11206:SF143 - MULTIDRUG RESISTANCE PROTEIN	.0055	height_3
Medtr5g082560	chr5	35495732	35496746	chr5:35497399		1015	Unknown	.0020	Arabinose2007
Medtr5g082560	chr5	35495732	35496746		chr5:35497700	1015	Unknown	.0025	height_1
Medtr5g082570	chr5	35497078	35502565	chr5:35497399		5488	(1 of 2) PTHR13711//PTHR13711:SF199 - SWI/SNF-RELATED CHROMATIN BINDING PROTEIN	.0020	Arabinose2007
Medtr5g082570	chr5	35497078	35502565		chr5:35497700	5488	(1 of 2) PTHR13711//PTHR13711:SF199 - SWI/SNF-RELATED CHROMATIN BINDING PROTEIN	.0025	height_1
Medtr8g106950	chr8	45164389	45168068	chr8:45165678		3680	(1 of 4) PF01535//PF12854//PF13041// PF14432 - PPR repeat (PPR) // PPR repeat (PPR_1) // PPR repeat family (PPR_2) // DYW family of nucleic acid deaminases (DYW_deaminase)	.0050	Arabinose2007
Medtr8g106950	chr8	45164389	45168068		chr8:45168433	3680	(1 of 4) PF01535//PF12854//PF13041// PF14432 - PPR repeat (PPR) // PPR repeat (PPR_1) // PPR repeat family (PPR_2) // DYW family of nucleic acid deaminases (DYW_deaminase)	.0017	height_3

[a]See Table 3 for description of phenotypic traits.

TABLE 3 Phenotypic variance explained by RDA + LFMM candidate genes shared by *Medicago sativa* and *M. truncatula*. Fraction of variance (deviance) explained in phenotypic traits (phenotypic trait data from Burgarella et al., 2016; Sakiroglu et al., 2011; Stanton-Geddes et al., 2013) and number of explanatory variables (number of shared candidate genes) used in the best cross-validated Gaussian generalized linear model regularized with LASSO penalty

Trait	Description	No. of shared candidate genes	Variance explained (R^2)
M. sativa			
ADF2007	Acid detergent fibre	11	.45
ADL2007	Acid detergent lignin	24	.65
Arabinose2007	Arabinose content	37	.79
Drymatteryield2007	Total dry biomass yield	24	.44
Glucose2007	Glucose content	7	.20
NDF2007	Neutral detergent fibre measured in 2007	5	.33
NDF2008	Neutral detergent fibre measured in 2008	16	.52
Regrowth2007	Spring regrowth	11	.25
Stemtotalmassratio2007	Stem proportion of the total biomass	8	.37
Stemweight2007	Stem weight	36	.64
TNC2007	Total nonstructural carbohydrates measured in 2007	24	.64
TNC2008	Total nonstructural carbohydrates measured in 2008	17	.43
Xylose2007	Xylose content	29	.58
M. truncatula			
Branch_3	Number of branches	18	.30
Flowering_date	Number of days to flowering starting from 1st of January	13	.27
Height_1	Height at t1	35	.38
Height_2	Height at t2	18	.32
Height_3	Height at t3	29	.48
Leaf_area	Leaf area	0	.00
Leaf_weight	Leaf weight	0	.00
Leaves_1	Number of leaves	10	.16
Nodule_above	Number of nodules in the top 5 cm of roots	18	.24
Nodule_below	Number of nodules below the top 5 cm of roots	23	.35
Petiole_area	Petiole area	6	.11
Trichomes	Trichome density visible at 10x magnification along a 2 mm section of the petiole	7	.13
Width_mm	Petiole width	6	.11
Mean_ThF	Mean thermal flowering time	10	.33
DF_a	Mean number of days to flowering for autumn sowing	9	.35
DF01_s	Number of days from sowing to flowering measured in 2001	11	.27

TABLE 4 Significantly enriched gene ontology (GO) terms in RDA + LFMM candidate genes shared by *Medicago sativa* and *M. truncatula*

ID	GO information	GO category	Description	FDR (q-value)	Number of genes
1	GO:0005856	CC	Cytoskeleton	0.0022	6
2	GO:0044430	CC	Cytoskeletal part	0.0022	6
3	GO:0003774	MF	Motor activity	0.0050	5
4	GO:0006810	BP	Transport	0.0066	19
5	GO:0051234	BP	Establishment of localization	0.0066	19
6	GO:0051179	BP	Localization	0.0066	19
7	GO:0005215	MF	Transporter activity	0.0080	14

Abbreviations: BP, biological process; CC, cellular component; FDR, false discovery rate; MF, molecular function.

4.2 | Species-specific signatures of adaptation

Our analyses showed that phenotypic trait variation in *M. sativa* and *M. truncatula*, provide insights into their specific strategies of adaptation to abiotic stress. From the full list of GEA candidates, 13.4% (*M. sativa*) and 24% (*M. truncatula*) were also associated with phenotypic traits (Figure 4). Traits with the highest number of strongest associations in *M. sativa* were: xylose (Xylose2007), arabinose (Arabinose2007), neutral detergent fibre (NDF2008), and stem proportion of the total biomass (Stemtotalmassratio2007), followed by others (Figure 4). High concentration of arabinose is associated with desiccation tolerance (Moore et al., 2006) and salt stress tolerance in plants (Zhao et al., 2019). Pectin-associated arabinose polymers maintain the flexibility of the cell wall during water deficit stress (Moore et al., 2008, 2013; Ying et al., 2011). Polymerization of arabinose and xylose produce arabinoxylan, which contributes to cell-wall flexibility, rehydration capacity and water use efficiency (Moore et al., 2013; Zhang et al., 2010). Previous findings in alfalfa have shown an impact of abiotic stress on cell wall remodeling (Song et al., 2019). Neutral detergent fibre and stem fibre concentration are also known to be associated with the environmental conditions in alfalfa experimental gardens (Lamb et al., 2012). Additionally, high fibre content, notably lignin, has been recently associated with adaptation to heat stress in other species (Blanco-Pastor et al., 2021). In general terms, our results suggest that the biochemistry of aerial biomass in *M. sativa* play an important role in its abiotic stress resistance.

In *M. truncatula*, traits with the highest number of strongest associations were: number of leaves (leaves_1), number of days from sowing to flowering (Df01_s), number of days to flowering starting from 1 January (flowering_date), height (height_3) and mean thermal flowering time (Mean_ThF), among others. These associations point to a clear phenotypic response to climate associated with flowering onset coupled with other mechanisms of abiotic stress adaptation associated with plant growth in *M. truncatula*, which is in line with the findings of Burgarella et al. (2016) and Yoder et al. (2014).

4.3 | Shared genomic signatures of adaptation

Adaptation to climate can be genetically constrained and genetic parallelism can play an important role for adaptation even under the highly polygenic architectures of environmental adaptation, determining the intraspecific variation in a group of genes among the full set of adaptive loci (Rellstab et al., 2020; Yeaman et al., 2016, 2018). Our study with *M. sativa* and *M. truncatula* suggests that the overall genetic mechanisms of adaptation appear to be largely dissimilar, with multiple adaptive genes that are species-specific (Figure 2 and Data S5–S6). It was already known that adaptation to soil and climate features is driven by selection on polygenic traits in *M. truncatula* (Guerrero et al., 2018; Yoder et al., 2014). On this basis, here we show that adaptation to environmental factors is also polygenic in *M. sativa* and mostly nonredundant with *M. truncatula*. These species have contrasting geographic distributions and life-history traits, but share similar environmental (climatic and edaphic) pressures as shown by the shared set of significant RDA terms (Table 1). But despite their unique signatures of adaptation, we found an above-expected number (168) of shared candidate genes between the two species (Figure 3), which was 25.4% more than expected by chance alone (134). Additionally, from the list of shared GEA candidate genes we found four shared candidates that also showed significant univariate associations with growth-related traits in both species. These were associated either with spring regrowth or arabinose in *M. sativa* and height in *M. truncatula*. The monosaccharide arabinose also participates in metabolic pathways associated with cell wall extensibility, thereby controlling plant growth and development (Kotake et al., 2016). Thus, our results suggest that this set of four shared genes may be involved in growth response to environmental signals in both species, although probably other genes are also involved in these responses.

In addition, we also found that optimized sets of shared GEA candidates explained a high proportion of phenotypic variance for certain traits, especially arabinose content (*M. sativa*) and plant height (*M. truncatula*) (Table 3), suggesting that they play an important role in the genetic determinism of polygenic phenotypic traits in both *Medicago* species. From the list of shared GEA candidates, another four genes (Medtr1g050730, Medtr2g076670, Medtr2g076680 and Medtr7g010650) were found associated with drought-related and biomass traits in a previous study using the same *M. truncatula* genomic data set (Kang et al., 2015). Among these, Medtr2g076670 was also associated with *M. sativa* total dry biomass yield in our study.

The set of shared GEA candidates showed enrichment in gene ontologies associated with cell structure and transport, including cytoskeleton, motor activity, transport or localization. These gene ontologies embrace the cellular shape modifications and other cellular functions associated with cell homeostasis, including cellular movement, cell division, endocytosis, movement of organelles or movement of substances by means of transporters, pores or motor proteins (Moustakas et al., 1998). All these components and functions have shown to play a central role in drought avoidance and tolerance mechanisms through a diversity of molecular pathways in a variety of plants (Jarzyniak & Jasiński, 2014), including alfalfa (Song et al., 2019).

4.4 | Does adaptation to climate in *Medicago* follow the omnigenic model?

The comparative approach allowed us to reveal the presence of 168 shared GEA candidate genes of adaptation to environmental conditions in *M. sativa* and *M. truncatula*. Among these, 118 (70.41%) had homologues in the *Arabidopsis thaliana* genome, which is exceptionally high given that, out of the 62,319 genes of the *M. truncatula* genome (Mt4.0v1), only 19,145 (30.72%) have their homologs identified in the *A. thaliana* (TAIR10) genome according to the Phytozome database (https://phytozome.jgi.doe.gov/).

The four shared GEA candidate genes, mentioned above, that were significantly associated with growth-related traits in both *M. sativa* and *M. truncatula* (Table 2) are Medtr4g048000, Medtr5g082560, Medtr5g082570 and Medtr8g106950. The gene Medtr4g048000 codes for a multidrug resistance-associated protein. Molecular functions of genes belonging to the multidrug resistance-associated proteins are very conserved (Klein et al., 2006). These genes were originally identified in drug-resistant human cancer cell lines and have been identified later in many different plant species, where they have been shown to play a key role in the accumulation of substances in the vacuole and in the regulation of guard cells (the pair of curved cells that surround a stoma) (Klein et al., 2006; Kreuz et al., 1996). In this regard, some wild *M. sativa*. subsp. *falcata* populations that occur in arid areas have been shown to have lower stomatal conductance (Ray et al., 2004), which is an indicator of delayed leaf senescence (Humphries et al., 2021). The SWI/SNF (SWItch/Sucrose nonfermentable) chromatin remodelling complex gene (Medtr5g082570) is also a highly conserved multiprotein complex, with homologues in humans, *Drosophila*, yeast and plants (Sarnowska et al., 2016). Stress induced transcriptional regulation relies on these SWI/SNF genes to allow for the access of the transcriptional machinery to a specific DNA sequence (Song et al., 2021). Research on SWI/SNF genes in *Arabidopsis* suggest that it is involved in numerous basic functions, including an effect on developmental control, cell differentiation, flowering time and growth, together with heat, drought and salt stress response and DNA damage response caused by excess of UV radiation (Buszewicz et al., 2016; Han et al., 2015; Jerzmanowski, 2007; Nguyen et al., 2019; Ojolo et al., 2018; Shaked et al., 2006). Its function is very conserved with a role in cellular response to UV damage also in mammalians (Gong et al., 2008). The SWI/SNF genes have been shown to play a role in the transcription of heat shock proteins (HSPs) genes, which increase cell survival under high temperatures by maintaining proper folding and preventing agglutination of denatured proteins (Qian et al., 2014), and in the transcription of genes responsible for the resource allocation decision between growth and drought tolerance (Han et al., 2012). The gene Medtr8g106950 codes for a PPR (pentatricopeptide repeat protein) deaminase with a DYW domain. This type of proteins are present in all organisms that undergo C-to-U editing of organelle RNA transcripts and play a critical role in the RNA editing process. They have been associated with growth, development and response to abiotic stresses in several plant families (Chen et al., 2018; Su et al., 2019; Wagoner et al., 2015; Xing et al., 2018), and with leaf relative water content (Zhang et al., 2015) and salt tolerance (Yu et al., 2016) in alfalfa.

The presence of a set of shared GEA candidate genes, some of them associated with growth-related traits in both species and with highly conserved key molecular functions, together with a larger set of species-specific candidate genes is in line with the omnigenic model of heritability (Boyle et al., 2017). This model poses that complex traits are ruled by both core, large-effect genes with conserved adaptive functions and multiple small-effect peripheral genes in regulatory networks. Adaptive functions in such core genes may persist longer in evolutionary time-scale than peripheral genes, the latter being responsible for unique patterns of local adaptation independent for each species (Vitti et al., 2013). While we do not know the specific role of these candidate genes in the regulatory networks of *Medicago*, *M. truncatula* gene models indicate that they are responsible of basic biological functions and, as shown here and previous studies, they also play a key role in abiotic stress tolerance.

Comparative genomics of plant and animal model organisms has revealed that a reduced set of highly conserved genes and gene networks present in distantly related organisms translate to a set of proteins that make up much of the cellular machinery (Miklos & Rubin, 1996) and are subject to adaptive evolution (Balti et al., 2020; Talbert et al., 2004). Genes encoding essential functions can evolve rapidly and can show parallel evolutionary trajectories, even when they are retained over long evolutionary periods across distantly related species (Chateigner et al., 2020; Kasinathan et al., 2020; Malik & Henikoff, 2001; Rellstab et al., 2020; Talbert et al., 2002, 2004). Although the omnigenic model is gaining relevance for understanding adaptation, it has also been criticized on the basis of its presumed oversimplification and the lack of additional evidence beyond genome-wide association studies (GWAS) (Wray et al., 2018). Identifying the functions of shared candidate genes in different species at the genome scale, as we did here, is an important complement to intraspecific effect patterns derived from GWAS. This and other recent studies (Rellstab et al., 2020; Yeaman et al., 2016, 2018) provide another perspective into the factors that constrain the diversity of viable routes to adaptation and build in favour of the omnigenic model to explain adaptation to climate.

SNP density across *M. sativa* and *M. truncatula* data sets used in the present study included one or more SNPs within a 100 Kbp window size for 88.8% and 89.1% of the *M. truncatula* genome, respectively (Methods S1, Figure S3). Nevertheless, our final data set provides a limited representation of the genomes of these two species. Our *M. sativa* genotyping led to the identification of around 14,160 SNPs, in tags of 100-bp. Given that the *M. truncatula* genome includes about 360 Mbp, only a maximum of 0.4% of the total *M. truncatula* genome size was effectively represented by the *M. sativa* SNP set, assuming no LD. Given this fairly low SNP density, most adaptive loci not in LD with this set of SNPs may have gone undetected. Whole genome resequencing of the *M. sativa* pangenome would increase the possibility of detecting more adaptive diversity with greater accuracy and facilitate the development of abiotic stress resistant alfalfa varieties in the future.

ACKNOWLEDGMENTS

We thank Jeremy Yoder for facilitating access to GPS data of *M. truncatula* collection sites and the USDA-GRIN germplasm bank for providing wild *M. sativa* populations. We also thank the three anonymous reviewers for their valuable comments that helped to improve previous versions of this manuscript. This work was supported by the European Commission through the Marie Curie Intra-European Fellowship "AlfalfaEvolution" (FP7-PEOPLE-2013-IEF, project reference 625308).

CONFLICT OF INTERESTS

There are no conflict of interests to declare.

AUTHOR CONTRIBUTIONS

José Luis Blanco-Pastor conceived the idea, José Luis Blanco-Pastor and Bernard E. Pfeil acquired funding, José Luis Blanco-Pastor, Isabel M. Liberal, Muhammet Sakiroglu, Yanling Wei and E. Charles Brummer collected data; José Luis Blanco-Pastor performed analyses, José Luis Blanco-Pastor, Isabel M. Liberal, Rose L. Andrews and Bernard E. Pfeil designed the study, José Luis Blanco-Pastor and Bernard E. Pfeil coordinated the research activity; José Luis Blanco-Pastor wrote the manuscript with input from Isabel M. Liberal, Muhammet Sakiroglu, E. Charles Brummer and Rose L. Andrews.

DATA AVAILABILITY STATEMENT

Additional Methods and Results and Information on the samples used are available as Supporting Information. *Medicago sativa* RAD-seq (GBS) data is available in the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA655898. Filtered SNPs (Data S1 and Data S2), Topoclimatic data used in this study (Data S3 and Data S4), sets of GEA candidate genes of adaptation (Data S5 and S6), and the set of GEA candidate genes shared by *M. sativa* and *M. trunatula* (Data S7) are also available from the Dryad digital repository (https://doi.org/10.5061/dryad.x95x69pgj). Code to replicate the analyses is hosted at Zenodo (https://doi.org/10.5281/zenodo.4775855).

ORCID

José Luis Blanco-Pastor https://orcid.org/0000-0002-7708-1342
Muhammet Sakiroglu https://orcid.org/0000-0002-7024-4348
Rose L. Andrew https://orcid.org/0000-0003-0099-8336

REFERENCES

Annicchiarico, P., Barrett, B., Brummer, E. C., Julier, B., & Marshall, A. H. (2015). Achievements and challenges in improving temperate perennial forage legumes. *Critical Reviews in Plant Sciences*, *34*, 327–380. https://doi.org/10.1080/07352689.2014.898462

Annicchiarico, P., Nazzicari, N., Wei, Y., Pecetti, L., & Brummer, E. C. (2017). Genotyping-by-Sequencing and its exploitation for forage and cool-season grain legume breeding. *Frontiers in Plant Science*, *8*, 679. https://doi.org/10.3389/fpls.2017.00679

Araújo, S. S., Beebe, S., Crespi, M., Delbreil, B., González, E. M., Gruber, V., Lejeune-Henaut, I., Link, W., Monteros, M. J., Prats, E., Rao, I., Vadez, V., & Vaz Patto, M. C. (2015). Abiotic stress responses in legumes: Strategies used to cope with environmental challenges. *Critical Reviews in Plant Sciences*, *34*, 237–280. https://doi.org/10.1080/07352689.2014.898450

Baena-González, E., Rolland, F., Thevelein, J. M., & Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signalling. *Nature*, *448*, 938–942. https://doi.org/10.1038/nature06069

Baena-González, E., & Sheen, J. (2008). Convergent energy and stress signaling. *Trends in Plant Science*, *13*, 474–482. https://doi.org/10.1016/j.tplants.2008.06.006

Balti, I., Benny, J., Perrone, A., Caruso, T., Abdallah, D., Salhi-Hannachi, A., & Martinelli, F. (2020). Identification of conserved genes linked to responses to abiotic stresses in leaves among different plant species. *Functional Plant Biology*, *48*(1), 54. https://doi.org/10.1071/FP20028

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benny, J., Pisciotta, A., Caruso, T., & Martinelli, F. (2019). Identification of key genes and its chromosome regions linked to drought responses in leaves across different crops through meta-analysis of RNA-Seq data. *BMC Plant Biology*, *19*, 194. https://doi.org/10.1186/s12870-019-1794-y

Blanco-Pastor, J. L., Barre, P., Keep, T., Ledauphin, T., Escobar-Gutiérrez, A., Roschanski, A. M., Willner, E., Dehmer, K. J., Hegarty, M., Muylle, H., Veeckman, E., Vandepoele, K., Ruttink, T., Roldán-Ruiz, I., Manel, S., & Sampoux, J.-P. (2021). Canonical correlations reveal adaptive loci and phenotypic responses to climate in perennial ryegrass. *Molecular Ecology Resources*, *21*, 849–870. https://doi.org/10.1111/1755-0998.13289

Blanco-Pastor, J. L., Liberal, I. M., Sakiroglu, M., Wei, Y., Brummer, E. C., Andrew, R. L., & Pfeil, B. E. (2021). Data from: Annual and perennial *Medicago* show signatures of parallel adaptation to climate and soil in highly conserved genes. https://doi.org/10.5061/dryad.x95x69pgj

Blanco-Pastor, J. L., Manel, S., Barre, P., Roschanski, A. M., Willner, E., Dehmer, K. J., Hegarty, M., Muylle, H., Ruttink, T., Roldán-Ruiz, I., Ledauphin, T., Escobar-Gutiérrez, A., & Sampoux, J.-P. (2019). Pleistocene climate changes, and not agricultural spread, accounts for range expansion and admixture in the dominant grassland species *Lolium perenne* L. *Journal of Biogeography*, *46*, 1451–1465. https://doi.org/10.1111/jbi.13587

Bordat, A., Savois, V., Nicolas, M., Salse, J., Chauveau, A., Bourgeois, M., Potier, J., Houtin, H., Rond, C., Murat, F., Marget, P., Aubert, G., & Burstin, J. (2011). Translational genomics in legumes allowed placing in silico 5460 unigenes on the Pea functional map and identified candidate genes in *Pisum sativum* L. *G3; Genes|Genomes|Genetics*, *1*, 93–103. https://doi.org/10.1534/g3.111.000349

Boyer, J. S. (1982). Plant productivity and environment. *Science*, *218*, 443–448. https://doi.org/10.1126/science.218.4571.443

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, *210*, 33–52. https://doi.org/10.1534/genetics.118.301333

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., Bharti, A. K., Woodward, J. E., May, G. D., Gentzbittel, L., Ben, C., Denny, R., Sadowsky, M. J., Ronfort, J., Bataillon, T., Young, N. D., & Tiffin, P. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, E864–E870. https://doi.org/10.1073/pnas.1104032108

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

Brummer, E. C., Cazcarro, P. M., & Luth, D. (1999). Ploidy determination of alfalfa germplasm accessions using flow cytometry. *Crop Science*,

39, 1202–1207. https://doi.org/10.2135/cropsci1999.0011183X003900040041x

Burgarella, C., Chantret, N., Gay, L., Prosperi, J.-M., Bonhomme, M., Tiffin, P., Young, N. D., & Ronfort, J. (2016). Adaptation to climate through flowering phenology: A case study in *Medicago truncatula*. *Molecular Ecology*, *25*, 3397–3415. https://doi.org/10.1111/mec.13683

Buszewicz, D., Archacki, R., Palusiński, A., Kotliński, M., Fogtman, A., Iwanicka-Nowicka, R., Sosnowska, K., Kuciński, J., Pupel, P., Olędzki, J., Dadlez, M., Misicka, A., Jerzmanowski, A., & Koblowska, M. K. (2016). HD2C histone deacetylase and a SWI/SNF chromatin remodelling complex interact and both are involved in mediating the heat stress response in *Arabidopsis*. *Plant, Cell & Environment*, *39*, 2108–2122. https://doi.org/10.1111/pce.12756

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*, 1223–1233. https://doi.org/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*, 852–860. https://doi.org/10.1093/molbev/msz008

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Chateigner, A., Lesage-Descauses, M.-C., Rogier, O., Jorge, V., Leplé, J.-C., Brunaud, V., Roux, C.-L., Soubigou-Taconnat, L., Martin-Magniette, M.-L., Sanchez, L., & Segura, V. (2020). Gene expression predictions and networks in natural populations supports the omnigenic theory. *BMC Genomics*, *21*. 416. https://doi.org/10.1186/s12864-020-06809-2

Chen, G., Zou, Y., Hu, J., & Ding, Y. (2018). Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genomics*, *19*, 720. https://doi.org/10.1186/s12864-018-5088-9

Choi, H.-K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J.-H., Kalo, P., Penmetsa, R. V., Seres, A., Kulikova, O., Roe, B. A., Bisseling, T., Kiss, G. B., & Cook, D. R. (2004). A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics*, *166*, 1463–1502. https://doi.org/10.1534/genetics.166.3.1463

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SNPSIFT. In: Toxicogenomics in Non-Mammalian Species. *Frontiers in Genetics*, *3*, 35. https://doi.org/10.3389/fgene.2012.00035

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

Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V., & Toll, J. (2014). Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecology and Sustainable Food Systems*, *38*, 369–377. https://doi.org/10.1080/21683565.2013.870629

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology & Evolution*, *26*, 298–306. https://doi.org/10.1016/j.tree.2011.02.008

Elshire, R. J., Glaubitz, J. C., Sun, Q. I., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*, 1–10. https://doi.org/10.1371/journal.pone.0019379

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

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

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, *183*, 557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Flavell, R. B., Bennett, M. D., Smith, J. B., & Smith, D. B. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochemical Genetics*, *12*, 257–269. https://doi.org/10.1007/BF00485947

Ford-Lloyd, B. V., Schmidt, M., Armstrong, S. J., Barazani, O. Z., Engels, J., Hadas, R., Hammer, K., Kell, S. P., Kang, D., Khoshbakht, K., Li, Y., Long, C., Lu, B.-R., Ma, K., Nguyen, V. T., Qiu, L., Ge, S., Wei, W., Zhang, Z., & Maxted, N. (2011). Crop wild relatives—undervalued, underutilized and under threat? *BioScience*, *61*, 559–565. https://doi.org/10.1525/bio.2011.61.7.10

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*, 104–120. https://doi.org/10.1111/mec.13476

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*, 2215–2233. https://doi.org/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*, 86–89. https://doi.org/10.1126/science.1209271

Frame, J., Charlton, J. F. L., & Laidlaw, A. S. (1998). *Temperate forage legumes*. Cab International.

François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*, 454–469. https://doi.org/10.1111/mec.13513

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

Frichot, E., & François, O. (2015b). *A short manual for LFMM version 1.4* [WWW Document]. Retrieved from http://membres-timc.imag.fr/Olivier.Francois/lfmm/files/note.pdf

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*, 1687–1699. https://doi.org/10.1093/molbev/mst063

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1. https://doi.org/10.18637/jss.v033.i01

Fustier, M.-A., Brandenburg, J.-T., Boitard, S., Lapeyronnie, J., Eguiarte, L. E., Vigouroux, Y., Manicacci, D., & Tenaillon, M. I. (2017). Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Molecular Ecology*, *26*, 2738–2756. https://doi.org/10.1111/mec.14082

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. I., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity Genotyping By Sequencing analysis pipeline. *PLoS One*, *9*, e90346. https://doi.org/10.1371/journal.pone.0090346

Gong, F., Fahy, D., Liu, H., Wang, W., & Smerdon, M. J. (2008). Role of the mammalian SWI/SNF chromatin remodeling complex in the

cellular response to UV damage. *Cell Cycle*, *7*, 1067–1074. https://doi.org/10.4161/cc.7.8.5647

Guerrero, J., Andrello, M., Burgarella, C., & Manel, S. (2018). Soil environment is a key driver of adaptation in *Medicago truncatula*: New insights from landscape genomics. *New Phytologist*, *219*, 378–390. https://doi.org/10.1111/nph.15171

Han, S.-K., Sang, Y., Rodrigues, A., BIOL425 F2010, Wu, M.-F., Rodriguez, P. L., & Wagner, D. (2012). The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in *Arabidopsis*. *The Plant Cell*, *24*, 4892–4906. https://doi.org/10.1105/tpc.112.105114

Han, S.-K., Wu, M.-F., Cui, S., & Wagner, D. (2015). Roles and activities of chromatin remodeling ATPases in plants. *The Plant Journal*, *83*, 62–77. https://doi.org/10.1111/tpj.12877.

Hanson, A. A. (1988). *Alfalfa and alfalfa improvement*. American society of agronomy.

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

Humphries, A. W., Ovalle, C., Hughes, S., del Pozo, A., Inostroza, L., Barahona, V., Yu, L., Yerzhanova, S., Rowe, T., Hill, J., Meiirman, G., Abayev, S., Brummer, E. C., Peck, D. M., Toktarbekova, S., Kalibayev, B., Espinoza, S., Ivelic-Saez, J., Bingham, E., … Kilian, B. (2021). Characterization and pre-breeding of diverse alfalfa wild relatives originating from drought-stressed environments. *Crop Science*, *61*, 69–88. https://doi.org/10.1002/csc2.20274

Jarzyniak, K. M., & Jasiński, M. (2014). Membrane transporters and drought resistance – A complex issue. *Frontiers in Plant Science*, *5*, https://doi.org/10.3389/fpls.2014.00687

Jerzmanowski, A. (2007). SWI/SNF chromatin remodeling and linker histones in plants. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. *Plant Chromatin: Structure and Expression*, *1769*, 330–345. https://doi.org/10.1016/j.bbaexp.2006.12.003

Joly-Lopez, Z., Forczek, E., Vello, E., Hoen, D. R., Tomita, A., & Bureau, T. E. (2017). Abiotic stress phenotypes are associated with conserved genes derived from transposable elements. *Frontiers in Plant Science*, *8*, 2027. https://doi.org/10.3389/fpls.2017.02027

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*, 989–1004. https://doi.org/10.1534/genetics.118.301786

Kang, Y., Sakiroglu, M., Krom, N., Stanton-Geddes, J., Wang, M., Lee, Y.-C., Young, N. D., & Udvardi, M. (2015). Genome-wide association of drought-related and biomass traits with HapMap SNPs in *Medicago truncatula*. *Plant, Cell & Environment*, *38*, 1997–2011. https://doi.org/10.1111/pce.12520

Kasinathan, B., Colmenares, S. U., McConnell, H., Young, J. M., Karpen, G. H., & Malik, H. S. (2020). Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in Drosophila. *eLife*, *9*, e63368. https://doi.org/10.7554/eLife.63368

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., & Varshney, R. K. (2019). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends in Plant Science*, *25*, 148–158. https://doi.org/10.1016/j.tplants.2019.10.012

Klein, M., Burla, B., & Martinoia, E. (2006). The multidrug resistance-associated protein (MRP/ABCC) subfamily of ATP-binding cassette transporters in plants. *FEBS Letters*, ABC Transporters, *580*, 1112–1122. https://doi.org/10.1016/j.febslet.2005.11.056

Kotake, T., Yamanashi, Y., Imaizumi, C., & Tsumuraya, Y. (2016). Metabolism of l-arabinose in plants. *Journal of Plant Research*, *129*, 781–792. https://doi.org/10.1007/s10265-016-0834-z

Kreuz, K., Tommasini, R., & Martinoia, E. (1996). Old enzymes for a new job (herbicide detoxification in plants). *Plant Physiology*, *111*, 349–353. https://doi.org/10.1104/pp.111.2.349

Krusell, L., Madsen, L. H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., Duc, G., Kaneko, T., Tabata, S., de Bruijn, F., Pajuelo, E., Sandal, N., & Stougaard, J. (2002). Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature*, *420*, 422–426. https://doi.org/10.1038/nature01207

Lamb, J. F. S., Hans-Joachim, G. J., & Riday, H. (2012). Harvest impacts on alfalfa stem neutral detergent fiber concentration and digestibility and cell wall concentration and composition. *Crop Science*, *52*, 2402–2412. https://doi.org/10.2135/cropsci2012.01.0027

Lenser, T., & Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Science*, *18*, 704–714. https://doi.org/10.1016/j.tplants.2013.08.007

Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.-P., Ané, J.-M., Lauber, E., Bisseling, T., Dénarié, J., Rosenberg, C., & Debellé, F. (2004). A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science*, *303*, 1361–1364. https://doi.org/10.1126/science.1093038

Li, R., di Virgilio, N., Guan, Q., Feng, S., & Richter, G. M. (2013). Reviewing models of land availability and dynamics for biofuel crops in the United States and the European Union. *Biofuels, Bioproducts and Biorefining*, *7*, 666–684. https://doi.org/10.1002/bbb.1419

Li, X., Wei, Y., Acharya, A., Hansen, J. L., Crawford, J. L., Viands, D. R., Michaud, R., Claessens, A., & Brummer, E. C. (2015). Genomic prediction of biomass yield in two selection cycles of a tetraploid alfalfa breeding population. *The Plant Genome*, *8*, 1–10. https://doi.org/10.3835/plantgenome2014.12.0090

Li, X., Wei, Y., Acharya, A., Jiang, Q., Kang, J., & Brummer, E. C. (2014). A saturated genetic linkage map of autotetraploid alfalfa (*Medicago sativa* L.) developed using Genotyping-by-Sequencing is highly syntenous with the *Medicago truncatula* genome. *G3 Genes|Genomes|Genetics*, *4*, 1971–1979. https://doi.org/10.1534/g3.114.012245

Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., & Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science*, *302*, 630–633. https://doi.org/10.1126/science.1090074

Long, S. P., & Ort, D. R. (2010). More than taking the heat: Crops and global change. *Current Opinion in Plant Biology*, *13*, 240–247. https://doi.org/10.1016/j.pbi.2010.04.008

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

Malik, H. S., & Henikoff, S. (2001). Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics*, *157*, 1293–1298. https://doi.org/10.1093/genetics/157.3.1293

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*, 656–669. https://doi.org/10.1111/1755-0998.12613

Miklos, G. L. G., & Rubin, G. M. (1996). The role of the genome project in determining gene function: Insights from model organisms. *Cell*, *86*, 521–529. https://doi.org/10.1016/S0092-8674(00)80126-9

Moore, J. P., Farrant, J. M., & Driouich, A. (2008). A role for pectin-associated arabinans in maintaining the flexibility of the plant cell wall during water deficit stress. *Plant Signaling & Behavior*, *3*, 102–104. https://doi.org/10.4161/psb.3.2.4959

Moore, J. P., Nguema-Ona, E., Chevalier, L., Lindsey, G. G., Brandt, W. F., Lerouge, P., Farrant, J. M., & Driouich, A. (2006). Response of the leaf cell wall to desiccation in the resurrection plant *Myrothamnus flabellifolius*. *Plant Physiology*, *141*, 651–662. https://doi.org/10.1104/pp.106.077701

Moore, J. P., Nguema-Ona, E. E., Vicré-Gibouin, M., Sørensen, I., Willats, W. G. T., Driouich, A., & Farrant, J. M. (2013). Arabinose-rich polymers as an evolutionary strategy to plasticize resurrection plant cell walls against desiccation. *Planta*, *237*, 739–754. https://doi.org/10.1007/s00425-012-1785-9

Moustakas, A., Theodoropoulos, P. A., Gravanis, A., Häussinger, D., & Stournaras, C. (1998). The Cytoskeleton in cell volume regulation. *Cell Volume Regulation*, *123*, 121–134. https://doi.org/10.1159/000059925

Muller, M. H., Prosperi, J. M., Santoni, S., & Ronfort, J. (2003). Inferences from mitochondrial DNA patterns on the domestication history of alfalfa (*Medicago sativa*). *Molecular Ecology*, *12*, 2187–2199. https://doi.org/10.1046/j.1365-294X.2003.01897.x

Nguyen, N. H., Jung, C., & Cheong, J.-J. (2019). Chromatin remodeling for the transcription of type 2C protein phosphatase genes in response to salt stress. *Plant Physiology and Biochemistry*, *141*, 325–331. https://doi.org/10.1016/j.plaphy.2019.06.012

Ojolo, S. P., Cao, S., Priyadarshani, S. V. G. N., Li, W., Yan, M., Aslam, M., Zhao, H., & Qin, Y. (2018). Regulation of plant growth and development: A review from a chromatin remodeling perspective. *Frontiers in Plant Science*, *9*, 1232. https://doi.org/10.3389/fpls.2018.01232

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. & Wagner, H.. (2013). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Paterson, A. H., Freeling, M., Tang, H., & Wang, X. (2010). Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*, *61*, 349–372. https://doi.org/10.1146/annurev-arplant-042809-112235

Peterson, T. A., & Russelle, M. P. (1991). Alfalfa and the nitrogen cycle in the Corn Belt. *Journal of Soil and Water Conservation*, *46*, 229–235.

Petit, C., Bretagnolle, F., & Felber, F. (1999). Evolutionary consequences of diploid-polyploid hybrid zones in wild species. *Trends in Ecology & Evolution*, *14*, 306–311. https://doi.org/10.1016/S0169-5347(99)01608-0

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, *7*, e32253. https://doi.org/10.1371/journal.pone.0032253

Qian, J., Chen, J., Liu, Y. F., Yang, L. L., Li, W. P., & Zhang, L. M. (2014). Overexpression of *Arabidopsis* HsfA1a enhances diverse stress tolerance by promoting stress-induced Hsp expression. *Genetics and Molecular Research: GMR*, *13*, 1233–1243. https://doi.org/10.4238/2014.February.27.8

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ray, I. M., Segovia-Lerma, A., & Murray, L. W. (2004). Diallel analysis of carbon isotope discrimination and its association with forage yield among nine historically recognized alfalfa germplasms. *Crop Science*, *44*, 1970–1975. https://doi.org/10.2135/cropsci2004.1970

Redden, R., Yadav, S. S., Maxted, N., Dulloo, M. E., Guarino, L., & Smith, P. (2015). *Crop wild relatives and climate change*. 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*, 4348–4370. https://doi.org/10.1111/mec.13322

Rellstab, C., Zoller, S., Sailer, C., Tedder, A., Gugerli, F., Shimizu, K. K., Holderegger, R., Widmer, A., & Fischer, M. C. (2020). Genomic signatures of convergent adaptation to Alpine environments in three Brassicaceae species. *Molecular Ecology*, *29*, 4350–4365. https://doi.org/10.1111/mec.15648

Remm, M., Storm, C. E. V., & Sonnhammer, E. L. L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology*, *314*, 1041–1052. https://doi.org/10.1006/jmbi.2000.5197

Sakiroglu, M., Doyle, J. J., & Brummer, E. C. (2010). Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theoretical and Applied Genetics*, *121*, 403–415. https://doi.org/10.1007/s00122-010-1319-4

Sakiroglu, M., Moore, K. J., & Brummer, E. C. (2011). Variation in biomass yield, cell wall components, and agronomic traits in a broad range of diploid alfalfa accessions. *Crop Science*, *51*, 1956–1964. https://doi.org/10.2135/cropsci2010.11.0658

Samac, D. A., Jung, H., & Lamb, J. F. S. (2006). Development of alfalfa (*Medicago sativa* L.) as a feedstock for production of ethanol and other bioproducts. *Chemical Industries-New York-Marcel Dekker-*, *112*, 79.

Sarnowska, E., Gratkowska, D. M., Sacharowski, S. P., Cwiek, P., Tohge, T., Fernie, A. R., Siedlecki, J. A., Koncz, C., & Sarnowski, T. J. (2016). The Role of SWI/SNF chromatin remodeling complexes in hormone crosstalk. *Trends in Plant Science*, *21*, 594–608. https://doi.org/10.1016/j.tplants.2016.01.017

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

Shaked, H., Avivi-Ragolsky, N., & Levy, A. A. (2006). Involvement of the *Arabidopsis* SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. *Genetics*, *173*, 985–994. https://doi.org/10.1534/genetics.105.051664

Shu, Y., Li, W., Zhao, J., Zhang, S., Xu, H., Liu, Y., & Guo, C. (2017). Transcriptome sequencing analysis of alfalfa reveals CBF genes potentially playing important roles in response to freezing stress. *Genetics and Molecular Biology*, *40*, 824–833. https://doi.org/10.1590/1678-4685-gmb-2017-0053

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, *39*, 1. https://doi.org/10.18637/jss.v039.i05

Small, E. (2011). *Alfalfa and relatives: Evolution and classification of* Medicago. NCR Research Press.

Small, E., & Bauchan, G. R. (1984). Chromosome numbers of the *Medicago sativa* complex in Turkey. *Canadian Journal of Botany*, *62*, 749–752. https://doi.org/10.1139/b84-110

Song, Y., Lv, J., Ma, Z., & Dong, W. (2019). The mechanism of alfalfa (*Medicago sativa* L.) response to abiotic stress. *Plant Growth Regulation*, *89*, 239–249. https://doi.org/10.1007/s10725-019-00530-1

Song, Z.-T., Liu, J.-X., & Han, J.-J. (2021). Chromatin remodeling factors regulate environmental stress responses in plants. *Journal of Integrative Plant Biology*, *63*, 438–450. https://doi.org/10.1111/jipb.13064

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., Bharti, A. K., Farmer, A. D., Zhou, P., Denny, R., May, G. D., Erlandson, S., Yakub, M., Sugawara, M., Sadowsky, M. J., Young, N. D., & Tiffin, P. (2013). Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS One*, *8*, e65688. https://doi.org/10.1371/journal.pone.0065688

Stracke, S., Sato, S., Sandal, N., Koyama, M., Kaneko, T., Tabata, S., & Parniske, M. (2004). Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana*, for positional cloning of a legume symbiosis gene. *Theoretical and Applied Genetics*, *108*, 442–449. https://doi.org/10.1007/s00122-003-1438-2

Su, H.-G., Li, B. O., Song, X.-Y., Ma, J., Chen, J., Zhou, Y.-B., Chen, M., Min, D.-H., Xu, Z.-S., & Ma, Y.-Z. (2019). Genome-wide analysis of the DYW subgroup PPR gene family and identification of GmPPR4 responses to drought stress. *International Journal of Molecular Sciences*, *20*, 5667. https://doi.org/10.3390/ijms20225667

Talbert, P. B., Bryson, T. D., & Henikoff, S. (2004). Adaptive evolution of centromere proteins in plants and animals. *Journal of Biology*, *3*, 18. https://doi.org/10.1186/jbiol11

Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., & Henikoff, S. (2002). Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *The Plant Cell*, *14*, 1053–1066. https://doi.org/10.1105/tpc.010425

Talbot, B., Chen, T.-W., Zimmerman, S., Joost, S., Eckert, A. J., Crow, T. M., Semizer-Cuming, D., Seshadri, C., & Manel, S. (2017). Combining genotype, phenotype, and environment to infer potential candidate genes. *Journal of Heredity*, *108*, 207–216. https://doi.org/10.1093/jhered/esw077

Tebaldi, C., Hayhoe, K., Arblaster, J. M., & Meehl, G. A. (2006). Going to the extremes. *Climatic Change*, *79*, 185–211. https://doi.org/10.1007/s10584-006-9051-4

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120. https://doi.org/10.1146/annurev-genet-111212-133526

Wagoner, J. A., Sun, T., Lin, L., & Hanson, M. R. (2015). Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. *Journal of Biological Chemistry*, *290*, 2957–2968. https://doi.org/10.1074/jbc.M114.622084

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*, 808–810. https://doi.org/10.1038/nmeth.2023

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., & Visscher, P. M. (2018). Common disease is more complex than implied by the core gene omnigenic model. *Cell*, *173*, 1573–1580. https://doi.org/10.1016/j.cell.2018.05.051

Xing, H., Fu, X., Yang, C., Tang, X., Guo, L. I., Li, C., Xu, C., & Luo, K. (2018). Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. *Scientific Reports*, *8*, 2817. https://doi.org/10.1038/s41598-018-21269-1

Yang, S., Gao, M., Deshpande, S., Lin, S., Roe, B. A., & Zhu, H. (2007). Genetic and physical localization of an anthracnose resistance gene in *Medicago truncatula*. *Theoretical and Applied Genetics*, *116*, 45–52. https://doi.org/10.1007/s00122-007-0645-7

Yang, S., Gao, M., Xu, C., Gao, J., Deshpande, S., Lin, S., Roe, B. A., & Zhu, H. (2008). Alfalfa benefits from *Medicago truncatula*: The RCT1 gene from *M. truncatula* confers broad-spectrum resistance to anthracnose in alfalfa. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 12164–12169. https://doi.org/10.1073/pnas.0802518105

Yeaman, S., Gerstein, A. C., Hodgins, K. A., & Whitlock, M. C. (2018). Quantifying how constraints limit the diversity of viable routes to adaptation. *Plos Genetics*, *14*, e1007717. https://doi.org/10.1371/journal.pgen.1007717

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., Nurkowski, K. A., Smets, P., Wang, T., Gray, L. K., Liepe, K. J., Hamann, A., Holliday, J. A., Whitlock, M. C., Rieseberg, L. H., & Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, *353*, 1431–1433. https://doi.org/10.1126/science.aaf7812

Ying, R., Barron, C., Saulnier, L., & Rondeau-Mouro, C. (2011). Water mobility within arabinoxylan and β-glucan films studied by NMR and dynamic vapour sorption. *Journal of the Science of Food and Agriculture*, *91*, 2601–2605. https://doi.org/10.1002/jsfa.4498

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

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F. X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D. R., Meyers, B. C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., … Roe, B. A. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, *480*, 520–524. https://doi.org/10.1038/nature10625

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*, 203–208. https://doi.org/10.1038/ng1702

Yu, L.-X., Liu, X., Boge, W., & Liu, X.-P. (2016). Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid Alfalfa (*Medicago sativa* L.) Using Genotyping-by-Sequencing. *Frontiers in Plant Science*, *7*, 956. https://doi.org/10.3389/fpls.2016.00956

Zhang, B., Liu, W., Chang, S. X., & Anyia, A. O. (2010). Water-deficit and high temperature affected water use efficiency and arabinoxylan concentration in spring wheat. *Journal of Cereal Science*, *52*, 263–269. https://doi.org/10.1016/j.jcs.2010.05.014

Zhang, T., Kesoju, S., Greene, S. L., Fransen, S., Hu, J., & Yu, L.-X. (2018). Genetic diversity and phenotypic variation for drought resistance in alfalfa (*Medicago sativa* L.) germplasm collected for drought tolerance. *Genetic Resources and Crop Evolution*, *65*, 471–484. https://doi.org/10.1007/s10722-017-0546-9

Zhang, T., Yu, L.-X., Zheng, P., Li, Y., Rivera, M., Main, D., & Greene, S. L. (2015). Identification of loci associated with drought resistance traits in heterozygous autotetraploid alfalfa (*Medicago sativa* L.) using genome-wide association studies with Genotyping by Sequencing. *PLoS One*, *10*, e0138931. https://doi.org/10.1371/journal.pone.0138931

Zhao, C., Zayed, O., Zeng, F., Liu, C., Zhang, L., Zhu, P., Hsu, C.-C., Tuncil, Y. E., Tao, W. A., Carpita, N. C., & Zhu, J.-K. (2019). Arabinose biosynthesis is critical for salt stress tolerance in Arabidopsis. *New Phytologist*, *224*, 274–290. https://doi.org/10.1111/nph.15867

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Blanco-Pastor, J. L., Liberal, I. M., Sakiroglu, M., Wei, Y., Brummer, E. C., Andrew, R. L., & Pfeil, B. E. (2021). Annual and perennial *Medicago* show signatures of parallel adaptation to climate and soil in highly conserved genes. *Molecular Ecology*, 00, 1–18. https://doi.org/10.1111/mec.16061