Genome-wide scans of selection highlight the impact of biotic and abiotic constraints in natural populations of the model grass *Brachypodium distachyon*.

Yann Bourgeois[1], Christoph Stritt[2], Jean-Claude Walser[3], Sean P. Gordon[4], John P. Vogel[4], Anne C. Roulin[2*]

[1] New York University Abu Dhabi, PO Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates

[2] Institute of Plant and Microbial Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

[3] Genetic Diversity Centre, ETH Zürich, Universitätstrasse 16, Zurich, Switzerland

[4]DOE Joint Genome Institute, Walnut Creek, CA 94598, USA.

* Corresponding author: anne.roulin@botinst.uzh.ch

Key words: local adaptation, genome scans of selection, grasses, *Brachypodium distachyon*, host-pathogen interaction.

Summary

Grasses are essential plants for ecosystem functioning. Thus, quantifying the selection pressures that act on natural variation in grass species is essential regarding biodiversity maintenance. In this study, we investigated the selection pressures that act on natural populations of the grass model *Brachypodium distachyon* without prior knowledge about the traits under selection. To do so, we took advantage of whole-genome sequencing data produced for two natural populations of *B. distachyon* and used complementary genome-wide scans of selection (GWSS) methods to detect genomic regions under balancing and positive selection. We show that selection is shaping genetic diversity at multiple temporal and spatial scales in this species and affects different genomic regions across the two populations. Gene Ontology annotation of candidate genes reveals that pathogens may constitute important factors of selection in *Brachypodium distachyon*. We eventually cross-validated our results with QTL data available for leaf-rust resistance in this species and demonstrated that, when paired with classical trait mapping, GWSS can help pinpointing candidate genes for further molecular validation. Our study revealed widespread signatures of natural selection on genes involved in adaptation in *B. distachyon* and suggests that pathogens may constitute an important driving force of genetic diversity and evolution in this system. Thanks to a near-base perfect reference genome and the large collection of freely available natural accessions collected across its natural range, *B. distachyon* appears as a prime system for studies in ecology, population genomics and evolutionary biology.

Introduction

Grasses cover more than 40% of the world land area (Gibson, 2009) and dominate a wide variety of ecosystems, from tropical to temperate regions (Clayton, 1981; Gibson, 2009). Grasses also play a key role in eco- and agrosystem functioning as they provide habitats for many animal species (Groves, 2000) and represent the main source of grain and forage (Stromberg, 2011). Increasing crop production to meet the food and energy requirements of the world's growing population is however putting great pressure on natural grasslands (Wallace, 1997; Helm *et al.*, 2009; Ceballos *et al.*, 2010). Faced with constant deterioration and fragmentation due to anthropic activities (Kiviniemi, 2002), these ecosystems are highly endangered (Ceballos *et al.*, 2010), but little is known about their evolutionary resilience. Assessing the genetic basis of adaptation and quantifying the selection pressures that act on natural variation in grass species is therefore crucial with respect to biodiversity maintenance and food security.

To date, reciprocal transplant experiments have been extensively used to test the effect of selection on adaptive differentiation across populations (for review see Savolainen *et al.*, 2013). Based on a "home vs. foreign" effect on fitness, reciprocal transplants are indeed powerful to unravel overall genotype by environment (GxE) interactions and demonstrated the prevalence of local adaptation in grasses and plants in general (for review see Bischoff *et al.*, 2006; Wadgymar *et al.*, 2017). However, reciprocal transplant experiments use proxy such as survival, vegetative growth or seed production to measure the effect of the habitat on fitness (Bischoff *et al.*, 2006). Hence, they provide little information about the functional and genetic bases of adaptation, unless combined with trait mapping such as quantitative trait locus (QTL) analyses and genome-wide association studies (GWAS) (Latta, 2009). QTL analyses and GWAS, on the other hand, are largely constrained by the effort and time necessary for high-resolution mapping. In grasses, while these trait-by-trait approaches have been valuable to decipher the genetic architecture of important characters with regard to crop genetic improvement (Huang *et al.*, 2002; Barbieri *et al.*, 2012; Morris *et al.*, 2013; Slavov *et al.*, 2014), they remain of limited value to grasp the overall selective forces that act on natural populations.

An efficient alternative to provide insights about evolutionary forces in natural populations consists in identifying genes under various types of selection at a whole

genome scale, then describing their function and the type of selection acting on them (Mitchell-olds *et al.*, 2007). For instance, new mutations that are beneficial in some populations will be positively selected and are more likely to quickly increase in frequency. Such so-called selective sweeps tend to reduce genetic diversity, increase differentiation among populations, and lead to extended haplotypes in the vicinity of the locus under selection due to genetic hitchhiking (Nielsen, 2005; Hermisson, 2009). Various genome-wide selection scans (GWSS) methods have been developed to detect such footprints of positive selection in genomes while taking into account demographic history (Tang *et al.*, 2007; Gautier *et al.*, 2012; Stamatakis *et al.*, 2013; Messer, 2015), and thanks to the remarkable progress of sequencing technologies, GWSS are now emerging as complementary approaches to classical trait mapping.

While local adaptation is commonly associated to positive selection on new advantageous polymorphisms, recent studies have demonstrated that balancing selection is also playing an important role in this evolutionary process (Mitchell-Olds *et al.*, 2007; Rasmussen *et al.*, 2014; Wu *et al.*, 2017). The term balancing selection is an "umbrella" concept (Fijarczyk & Babik, 2015) which describes the maintenance of genetic diversity over longer periods of time through adaptation to spatial heterogeneity, heterozygote advantage and negative frequency-dependent selection (Mitchell-Olds *et al.*, 2007; Rasmussen *et al.*, 2014). Leading to the recycling of polymorphisms and to selection on standing variation (Richman, 2000; Turchin *et al.*, 2012), this process is more difficult to detect than positive selection (Fijarczyk & Babik, 2015) since older alleles had more time to recombine and may lead to narrow signatures around selected sites. As a consequence, the effect of balancing selection is still largely overlooked in genome scans, which remain strongly biased towards the detection of recent positive selection (Hassl & Payseur, 2016).

In this study, we capitalize on the near base-perfect quality of the reference genome of the Mediterranean grass *Brachypodium distachyon* (https://phytozome.jgi.doe.gov) to investigate how both positive and balancing selection are shaping diversity in this species. In the last decade, *B. distachyon* has been developed as a powerful model for research on temperate grass species as it is closely related to major crop cereals and to some of the grasses used for biofuel production (The international brachypodium Consortium, 2010). Entirely sequenced, its small diploid genome (272Mb) is fully assembled into five chromosomes and has been exhaustively annotated (The

international brachypodium Consortium, 2010). In addition, *B. distachyon* is broadly distributed around the Mediterranean rime (Dell'Acqua *et al.*, 2014; Gordon *et al.*, 2014; Tyler *et al.*, 2016), providing access to natural populations from contrasting habitats for which a large collection has been collected. It constitutes therefore a unique and prime system to investigate the genetic basis of local adaptation in natural grass populations, opening the way to further fundamental and applied research.

Here, we took advantage of whole-genome sequencing data produced for 44 *B. distachyon* natural accessions originating mainly from Spain and Turkey (Gordon *et al.*, 2017). We identified over 6 million SNPs and used four complementary GWSS methods to detect genomic regions under different regimes of selection (Figure 1). Namely, we asked i) at what time and geographical scale is selection acting in *B. distachyon* populations? ii) what are the selective constrains that shape diversity and adaptation in these populations? iii) whether positive selection is acting on the same genomic regions in the two populations or, on the opposite, on distinct loci?

Results

Population structure

In this study, we used whole-genome sequencing data (paired-end; Illumina technology) with a 86-fold median coverage of 44 *B. distachyion* accessions originating from Turkey, Iraq, Spain and France (Figure 2A, Table S1, (Gordon *et al.*, 2017). After filtering, we identified 6,204,029 SNPs. An ADMIXTURE analysis, where K=2 was identified as the best model, highlighted two distinct genetic clusters, an eastern and a western one, with extremely little admixture between the two (Figure 2B). For the rest of the study, accessions from Turkey and Iraq will be referred to as the eastern population while accessions from Spain and France will be referred to as the western population. The western population showed a lower level of nucleotide diversity (Wilcoxon test; P-value < 2.2e-16, Figure 2C) and haplotype diversity (P-value < 2.2e-16, data not shown) than the eastern one. Excluding the reference accession Bd21, which has been artificially inbred before sequencing, the average level of heterozygosity in these accession is of 8% and ranges from 4 to 17.4% (Table S1).

Functional clustering of the genome of B. distachyon

GWSS outputs provide information about the likelihood for a given locus to be under selection. A classical approach applied to analyze the results of GWSS consists in selecting genomic regions containing top 1% outliers for signals of selection and then assessing whether some biological functions or processes are significantly over-represented in the gene sets under selection through a Gene Ontology (GO) annotation (Kelley *et al.*, 2006; Hancock *et al.*, 2011; Nelson *et al.*, 2017). While recombination rate is relatively high in *B. distachyon* (Huo *et al.*, 2011), signals of selection around focal loci may decrease slowly due to locally stronger linkage disequilibrium and subsequent genetic hitchhiking. 1% outlier regions may thus contain several adjacent genes. Because genes having the same function or being involved in the same biological process tend to be physically clustered (Hammond-kosack & Jones, 1996; Michelmore & Meyers, 1998; Takos *et al.*, 2011; Nutzmann & Osbourn, 2014; Singh *et al.*, 2015), we anticipated that this non-random organization of genomes could lead to an over-representation of some biological pathways or functions in small genomic regions and to an artificial enrichment for some GO terms in GWSS 1% outlier regions (Pavlidis *et al.*, 2012).

To assess whether the genome of *B. distachyon* harbors such functional clusters of genes, we first performed a GO annotation for the 32,712 genes annotated in the reference genome. We then controlled for potential gene clustering by following the procedure described in (Al-Shahrour *et al.*, 2010). Briefly, we split the genome into overlapping windows of 50 consecutive genes and performed enrichment analyses on each window. We identified 272 windows significantly enriched for at least one biological process (Table S2). Several windows were enriched for processes that may be associated to adaptation to local environmental conditions such response to stress and defense response (Table S2). This prompted us to narrow down top 1% outlier regions by keeping only the genes located at and in the immediate vicinity (-10% of the peak value) of each of the peaks of selection. With the exception of the coalescence analysis, which is a window-based approach (see methods), all analyses subsequent to GWSS reported in the following sections were performed on these filtered outputs.

Genes under balancing selection due to environmental heterogeneity

Variation in abiotic conditions can drive local adaptation at a fine-grained spatial scale and lead to correlations between genotypes and environment. We first used an environmental association analysis approach to detect loci that may have been repeatedly selected by convergent climatic conditions across the two populations. If alleles at these loci have been recycled in front of environmental conditions common to Spain and Turkey, they should display a detectable signal of association at the scale encompassed by our study. Note that geographically varying selection is sometimes considered as being distinct from balancing selection. While it can also be referred to as local adaptation in the literature (Mitchell-olds *et al.*, 2007), we kept it here under the term of balancing selection.

To detect such loci and to highlight alleles shared between the two populations, we selected seven bioclimatic variables showing variation within both the eastern and western populations but little variation between them (Figure 2D). As all these variables were either associated to temperature or to precipitation and are likely to be correlated, we summarized them with a PCA. The first axis of the PCA explained 62.6% of the variance but did not discriminate populations (Figure 2D) and was used to perform an environmental association analysis with all the SNPs identified across the 44 accessions. We identified 26 genomic regions associated with the first PCA axis. These regions harbored 71 genes (Figure 3A). No significant enrichment for any specific biological process was observed (Table 1).

Genes harboring extremely long coalescence time

Balancing selection can be detected with coalescence approaches, as ancient alleles are associated with older coalescence times (Charlesworth, 2006). To detect additional candidate regions for balancing selection in *B. distachyon*, we used the software ARGWeaver (Rasmussen *et al.*, 2014). Briefly, ARGWeaver models the coalescent process along chromosomes and across non-recombining blocks of sequences to address their evolutionary history. It allows recovering several statistics that describe local genealogies and recombination, such as times since the most recent common ancestor (TMRCA), which should be increased near ancient alleles such as those under ancient balancing selection. By doing so, we identified 72 regions harboring 115 genes under what will be referred to as long-term balancing selection in

the following. We observed a significant enrichment for genes involved in phosphorylation (Table 1).

Genes under disruptive selection between western and eastern populations

As new mutations providing a selective advantage rise in frequency through positive selection, neutral mutations that are physically close tend to remain strongly linked to them. Recent positive selection should therefore lead to a signature of long haplotypes near selected mutations. We used the Rsb test (Tang *et al.*, 2007) to detect such signatures of recently or almost completed selective sweeps. This test detects haplotypes that are positively selected in one population by estimating the length of haplotypes around each allele at a core SNP, then comparing these lengths between populations. In contrast to the two approaches presented above, the Rsb test should detect regions that are genetically differentiated across the two populations (Figure 1). We identified 312 regions harboring 824 genes and 319 regions harboring 1212 genes in the eastern and western populations respectively (Figure 3A). The selected regions contained more genes in the western than in the eastern population (Wilcoxon test, P-value=0.001; Figure 3B). We observed a significant enrichment in genes involved in response to stress, particularly for defense response and response to oxidative stress in the eastern population (Table 1, Table S3). The gene set associated to the process of defense response contained well-known types of resistance genes (R-genes), i.e genes with NBS-LRR domains (Table S3). Eventually, we observed a significant enrichment for genes involved in nitrogen transport in the western population (Table 1, Table S3).

Genes under ongoing selection within population

In the case of an ongoing and partial selective sweep (Figure 1), not all individuals within the population will display haplotype extension in the region under selection, which can lead to more subtle patterns that are not detected by Rsb test. We used the program H-scan to detect such incomplete patterns and genes under ongoing positive selection. We identified 142 regions harboring 487 genes and 79 regions harboring 463 genes in the eastern and western population respectively (Figure 3A). The gene set under selection in the eastern population was significantly enriched for genes involved in stress response, including the processes of defense response and response

to cadmium ion (Table 1). The gene set associated to the process of defense response also contained many R-genes (Table S3). The gene set under selection in the western population showed a significant enrichment for genes involved in pyruvate metabolic process (Table 1, Table S3). As in the previous Rsb analysis, the selected windows contained more genes in the western than in the eastern population (Wilcoxon test, P-value=0.003; Figure 3B).

GWSS outliers display allele frequency spectra and coalescence patterns consistent with expectations

To test whether candidate regions displayed an allele frequency spectrum consistent with the type of selection they were supposed to detect, we computed different statistics and compared candidate regions to the rest of the genome. We first computed Tajima's D (Tajima, 1989), which is a measure of genetic diversity influenced by both selection and demographic variation. Positive values are associated with balancing selection and bottlenecks, while negative values suggest recent expansion or positive/purifying selection. We then computed nucleotide diversity, genome-wide relative (F_{ST}) and absolute (d_{XY}) measures of population differentiation (Cruickshank & Hahn, 2014). Note that d_{XY} can be interpreted as $F_{ST,}$ but is correlated to the time to coalescence of alleles from all populations, making it independent of diversity within populations. In addition to TMRCA values, we finally extracted relative TMRCA halftime (RTH) values from the output of ARGWeaver. RHT captures coalescence events are skewed toward the recent past but is independent of the overall coalescence rate (Rasmussen *et al.*, 2014). Assuming similar selection strength acting on Rsb and H-scan outliers, RTH is thus expected to be smaller in regions under ongoing positive selection (H-scan) than in regions under disruptive selection where one allele already reached near-fixation in one of the two populations (Rsb).

Regions associated with bioclimatic variables or under long-term balancing selection displayed significantly higher Tajima's D than the genomic background (Wilcoxon test, all P-values$<1.0.10^{-10}$, Figure 3C) except for associated loci in Spain for which Tajima's D values were only marginally higher than in the rest of the genome (P-value=0.055). These regions also display a significantly higher level of nucleotide diversity within regional groups (P-values$<1.0.10^{-10}$) than the rest of the genome.

Coalescence times (TMRCA) within regional groups were higher for windows covering loci associated to environmental variables than for genomic background (P-values=1.10^{-14} and 0.02 for eastern and western populations respectively). As expected, regions under long-term balancing selection also displayed higher TMRCAs as they were specifically selected as belonging to the top 1% outlier for this statistic. Tajima's D, nucleotide diversity and TMRCA statistics are thus consistent with our expectations and confirmed that both the environmental association and coalescence analyses detected older polymorphisms shared across the two populations.

On the other hand, regions identified with the H-scan and Rsb approaches harbored negative and significantly lower Tajima's D than the rest of the genome, which is consistent under positive selection (all P-values < 1.10^{-5}, Figure 3C). Outlier windows for Rsb also displayed higher relative (F_{ST}, P-values<$2.2.10^{-16}$) and absolute levels of differentiation (d_{XY}; P-values<$2.2.10^{-16}$) than the rest of the genome. These latter patterns are coherent with the high frequencies of divergent haplotypes between the western and eastern population for Rsb outliers, resulting in an increased differentiation between them (Figure 1). H-scan and Rsb outliers also displayed a lower RTH compared to genomic background (all P-values<$2.2.10^{-16}$ except for Turkish Rsb outliers for which P-value = 0.002), which is also consistent with selective sweeps where selection skews coalescence times towards the recent past for most (but not necessarily all) lineages (Rasmussen *et al.*, 2014). Eventually, H-scan outliers displayed lower RTH than Rsb outliers (all P-values<$2.2.10^{-16}$), which confirms that this test tented to detect more recent sweeps than the Rsb.

Overlap between outputs from the four different approaches

As expected given the specificity of each test, we found little overlap between the outputs of the four different tests (Figure 3A). Out of a total of 1,262 and 1,617 genes under positive selection, only 49 and 58 genes were common to the H-scan and Rsb approaches in the eastern and the western population respectively. Similarly, little overlap (10 genes in total) was observed between the environmental association analysis and the Rsb or H-scan tests (Figure 3A). No overlap was observed between the gene sets detected to be under selection with the coalescence approach and with any the three other tests.

Genomic regions affected by positive selection

We also tested whether distinct loci/genomic regions were affected by recent positive selection (Rsb and H-scan outliers combined) in the two populations using linear models. We found no association between the density of genes under selection in the eastern and western populations on chromosomes 2 and 3 (p-value = 0.4 and 0.3 respectively), which indicates that, as observed on Figure 3D, distinct regions are affected by positive selection on these two chromosomes. We found a significant association between these variables on chromosomes 1, 4 and 5 (p-value = 0.004, 1.07e-11 and 0.02 respectively). In these latter cases, however, R^2 were small (7.366e-05, 0.0004 and 4.42e-05) suggesting that many regions affected by selection remain specific to each population on these chromosomes as well, as depicted on Figure 3D.

Identification of candidate genes in known QTL regions

Combining association mapping and analyses of selection constitutes a powerful approach to identify candidate genes and to address their selective regime. We identified many candidate regions harboring resistance genes. As a proof of concept, we aimed at assessing whether regions identified as resistance loci against known pathogens were also highlighted in our scans of selection. In *B. distachyon*, the genetic basis of resistance to the rust fungus *Puccinia brachypodii* has been deciphered through a QTL mapping (Barbieri *et al.*, 2012) which showed that leaf-rust resistance is controlled by three main QTL located on chromosome 2 (from nucleotide 37,949,269 to 40,903,216), 3 (from nucleotide 13,943,000 to 14,512,222) and 4 (from nucleotide 9,649,152 to 10,679,750). For the rest of the study, these regions will be referred to as $QTL_{rust\text{-}2}$, $QTL_{rust\text{-}3}$ and $QTL_{rust\text{-}4}$.

We screened these three regions for evidence of selection and found strong Rsb signals in $QTL_{rust\text{-}3}$ and $QTL_{rust\text{-}4}$ (Figure 4A). These regions were not detected as outliers with any other test but belonged to the top 0.05% p-values of Rsb outliers. The Rsb signal in $QTL_{rust\text{-}3}$ reaches its highest point in a serine/threonine phosphatase (Bradi3g16320: 14,486,831- 14,488,838; Figure 4A, left panel) and 60 kb upstream the QTL peak identified in this region (Barbieri *et al.*, 2012). The two adjacent Rsb signals in $QTL_{rust\text{-}4}$ reach their highest point into two NBS-LRR resistance genes (Bradi4g10153: 9,807,879-9,812,927; and Bradi4g10171: 9,828,236-9,835,003; Figure 3A, right panel). These two genes are respectively located 6 kb and 22 kb

upstream the QTL peak. We observed extended haplotypes in the eastern population in both QTL_{rust-3} and QTL_{rust-4} (Figure 4B). Congruent with the signal detected by the Rsb test, the large majority of the eastern accessions displayed the extended haplotype in these regions, indicating a nearly completed selective sweep, especially in QTL_{rust-4} (Figure 4B).

QTL_{rust-4} showed a striking enrichment for genes involved in defense response (displayed by grey boxes in Figure 4A, P-value = 8E-09) and in immune signaling process such as phosphorylation (P-value = 5.6E-06). Among 113 genes covered by QTL rust-$_4$, 40 correspond either to a gene with a NBS-LRR, a receptor-like protein kinase (RLK) or a F-box domains, three types of genes that can confer resistance in plants. The presence of such a large gene clusters, as found in other regions of the genome (Table S2), further demonstrates the importance of narrowing down regions under selection to top outlier genes for unbiased GO annotation.

Discussion

Assessing the time and spatial scales at which selection acts is a key to understand how genetic diversity is maintained or lost through adaptation (Stinchcombe & Hoekstra, 2008; Fuller *et al.*, 2015). In plants, and especially in grasses, this question has been largely restricted to crops, which biases our understanding of evolutionary processes that have shaped genomes in their natural ancestors and extant relatives. In this study, we investigate the selective forces influencing adaptation in two populations consisting of 44 natural accessions of the wild Mediterranean grass *B. distachyon.* We found that ancient balancing and recent positive selection left distinct signatures on specific gene categories, and that positive selection affects distinct loci across the two populations. Importantly, our results support a role for pathogens in driving population differentiation and confirm that GWSS constitute effective approaches to pinpoint candidate genes as a complement to classical trait mapping.

Time- and space-varying selection is shaping diversity in B. distachyon

The ending of the last glaciation period 10,000 years ago led to drastic and recent changes of plant communities in Eurasia (Svenning *et al.*, 2008; Binney *et al.*, 2017).

At that time, climate warmed and species distribution expanded over Europe (Hewitt, 1999). Pollen-based studies show that vegetation expansion was fast, reaching up to 2km per year for some species (Hewitt, 1999). To our knowledge, no fossil pollen records are available for *B. distachyon*, which prevents reconstructing the geographical distribution of this species before and during the last ice age. Yet, a previous study showed that the two populations analyzed here experienced a severe population size reduction during the last glaciation followed by a rapid expansion within the last 10,000 years (Stritt *et al*., in press). Even though unraveling the history of populations in southern peninsulas is more complex than in northern regions (Hewitt, 2000; Feliner, 2011), these results are congruent with the recent global postglacial recolonization of Europe by plants (Hewitt, 1999; Svenning *et al*., 2008; Binney *et al.*, 2017) and imply that *B. distachyon* populations had to adapt to newly colonized habitats in the recent past.

Balancing selection associated to spatial heterogeneity (Richman, 2000; Charlesworth, 2006; Fijarczyk & Babik, 2015; Wu *et al.* 2017) may have maintained ancestral polymorphisms over long periods of time in natural populations of *B. distachyon.* Our environmental association analysis indeed shows that loci associated to bioclimatic variables display higher diversity, older alleles and more shared variation between regional groups when compared to genomic background. We exclusively focused the analysis on bioclimatic variables that displayed variation across localities and not between (Figure 2D). This approach, together with our relatively small sample size, could explain why this analysis highlighted fewer genes than the other methods used in the rest of the study. Our observations nonetheless suggest that adaptation to climate after recolonization does not necessarily involve *de novo* mutations, even after strong bottlenecks. As a further support of balancing selection, we identified an even older set of shared polymorphisms with a coalescence approach. The environmental heterogeneity encountered by *B. distachyon* populations (Lopez-Alvarez *et al.*, 2015) thus seems to have provided selective pressure strong enough to maintain polymorphisms over long periods of time within each population. This result is congruent with a previous analysis of natural populations of *B. distachyon* originating exclusively from Turkey (Dell'Acqua *et al.*, 2014) which showed, at a smaller geographical scale than the one investigated here, that populations are adapted to local habitats (Dell'Acqua *et al.*, 2014).

On the other hand, we also found evidence for positive selection acting on younger polymorphisms with both the Rsb and H-scan tests. These tests also revealed that positive selection is targeting different loci in the two populations. Interestingly, these loci appear to be non-randomly distributed along chromosomes (Figure 3D). As recombination rate is relatively high in *B. distachyon* (Huo *et al.*, 2011), we do not believe that this pattern is due to extended linkage disequilibrium and the subsequent process of linked selection along such large genomic regions (Cutter & Payseur, 2013; Slotte, 2014). Rather, the peaks of selection we identified were narrow and allowed to pinpoint genes (Figure 4), indicating that while *B. distachyon* is primarily inbreeding, outcrossing events must be frequent enough to limit extended linkage disequilibrium. This is also congruent with the higher level of heterozygosity we observed here in *B. distachyon* compared to other selfing plants such as *A. thaliana* (Platt *et al.*, 2010). Whether these regions form islands of divergence and result from a complex interaction between recombination rate variation, gene flow and selection (Renaut *et al.*, 2013; Samuk *et al.*, 2017) remains to be investigated. As we obtained enrichments for different GO terms in the eastern and western populations, our study nonetheless shows that contrasting abiotic and biotic factors are shaping population diversity at different regions of the genome of *B. distachyon* through positive selection.

B. distachyon occurs exclusively in Mediterranean habitats which may appear at a first glance to be homogeneous and unlikely to promote local adaptation. Our results defeat this prediction and revealed that natural selection affected different genes and genomic regions across populations. Even though we identified more genes under positive than under balancing selection, it would be daring to conclude that the former selection regime, less challenging to detect (Delph & Kelly, 2014), is a predominant process shaping diversity in this species. Rather, we believe that we provide here genomic evidence that large-scale balancing selection also leads to the adaptation of *B. distachyon* populations to local environmental conditions.

Pathogens as a potential driving force of population evolution

Host-pathogen interactions lead to a strong coevolutionary dynamics and are considered as a major factor shaping diversity (Karasov *et al.*; Fumagalli *et al.*, 2011; Krattinger & Keller, 2016a). Two main types of interaction have been proposed.

Under an arms race model, repeated innovation from both sides results in repeated fixation of advantageous alleles (Brown & Tellier, 2011). This interaction can therefore lead to positive selection that can be detected by tests focusing on extended haplotypes. The other type of interaction is often referred to as Red Queen dynamics or trench warfare, where alleles involved in the interaction are recycled by negative-frequency dependence and can therefore subsist for long periods of time in populations through balancing selection (Brown & Tellier, 2011).

Plant immune system machinery is complex. On one hand, it is composed of two tiers of extracellular and intracellular receptors (Krattinger & Keller, 2016a,b) that efficiently detect the presence of pathogens and constitute a first level of defense (for review Greeff *et al.*, 2012; Couto & Zipfel, 2016; Eckardt, 2017). Further mechanisms, such as oxidative bursts which are produced at an early stage in case of pathogen invasion, can act as additional levels of defense and prevent pathogen proliferation (Wojtaszek, 1997; Torres *et al.*, 2006; Fones & Preston, 2011; Sewelam *et al.*, 2016). In this study, we found a significant enrichment of signals of selection at genes involved in these two levels of defense in the eastern population. More specifically, we found many of well-characterized R-genes, i.e. genes displaying NBS-LRR domains (Mchale *et al.*, 2006; Jacob *et al.*, 2013; Liu *et al.*, 2014; Couto & Zipfel, 2016; Eckardt, 2017; Ooijen *et al.*, 2017), and genes involved into oxidative stress response to be under ongoing and/or disruptive selection in the eastern population (Table 1). On the other hand, we also found additional R-genes and genes involved in phosphorylation, a process especially important for immune signaling response in plants (for review Park *et al.*, 2012), to be under balancing selection. Our results are thus consistent with the two classical models of host-pathogens coevolution (Mondragón-palomino *et al.*, 2002; Mchale *et al.*, 2006; Gos *et al.*, 2012; Mace *et al.*, 2014; Zhong *et al.*, 2015; Wu *et al.*, 2017). Overall, and as shown in other organisms (Karasov *et al.* 2014; Fumagalli *et al.*, 2011; Krattinger & Keller, 2016a,b; Bourgeois *et al.*, 2017), our genome-wide approach suggests that pathogens may constitute an important driving force of population and genome evolution in *B. distachyon.*

Surprisingly, the significant enrichments for resistance genes in the Rsb and H-scan outlier gene sets were only observed in the eastern population. The geographical origin of the populations could play a role in this pattern. The Middle East is indeed

the center of origin of many grasses, including *B. distachyon*, and of their associated pathogens (Wyand & Brown, 2003; Stukenbrock *et al.*, 2005; Opanowicz *et al.*, 2008; Hovmøller *et al.*, 2011). Many studies found that both resistance genes in plants and effector genes in pathogens can be organized in clusters evolving through an arm race resulting in a gene birth-and-death process (Michelmore & Meyers, 1998; Dong *et al.*, 2015; Singh *et al.*, 2015). Because centers of origin are usually associated to higher diversity, it is therefore possible that a higher level of pathogen diversity drove selection at a larger number of resistance genes in the eastern population.

We can, however, not rule out technical biases inherent to the sampling design and to the history of the studied populations. The western population indeed experienced a stronger bottleneck than the eastern one (Stritt *et al.*, in press), which, together with the smaller geographical area sampled, may be responsible for the reduced nucleotide and haplotype diversity we observed in this population. In addition, the bottleneck may have led to longer haplotypes in the western population, which could explain why windows under positive selection were on average larger and contained more genes in this population than in the eastern one. Discriminating the effect of selection from the one of demography remains difficult in such bottlenecked populations (Long *et al.*, 2013). As a consequence, confounding effects may have resulted in the detection of more false positives and blurred the GO annotation in the western population. As we applied stringent filtering criteria and used approaches which are expected to be robust to demographic history (Tang *et al.*, 2007), we nonetheless believe that we provide the community with a reliable set of candidate loci, even for the western population. It is worth noting that, despite no significant enrichment for defense response, footprints of positive selection at R-genes were also found in the western population. This indicates that while milder, pathogens may constitute a selection pressure in this population as well.

GWSS as complementary approach to QTL and GWAS

Disentangling the mechanisms that promote or prevent adaptation requires more integrated studies, using both experimentations in controlled conditions and methods to characterize genetic diversity in natural populations (Feder & Mitchell-olds, 2003; Stinchcombe & Hoekstra, 2008; Flood & Hancock, 2017). Several studies used GWSS to validate genes functionally characterized or previously identified through

GWAS and to propose stronger hypotheses on the mode of selection operating on traits relevant for adaptation (Roulin *et al.* 2016; Tang *et al.*, 2007; Fumagalli *et al.*, 2011; Bourgeois *et al.* 2017). Following this idea, we inspected three QTL regions responsible for the resistance of *B. distachyon* to the leaf-rust fungus *Puccinia brachypodii*, a natural pathogen of *B. distachyon* expected to exert strong selection on natural accessions.

For two of these QTL regions, we found strong Rsb outliers and reduced haplotype diversity in the eastern cluster in the vicinity of the QTL peaks, but no sign of balancing selection. These results strongly indicate that large-scale positive selection is shaping rust resistance in *B. distachyon* natural populations, as suggested in other species (Dodds & Thrall, 2009; Chavan *et al.*, 2015). Interestingly, the QTL region identified on chromosome 3 displays a strong signal of selection in a serine/threonine phosphatase, a class of genes known for their role in defense response and stress signaling (País *et al.*, 2009; Durian *et al.*, 2016). The region identified on chromosome 4 is more complex and consists of a cluster of resistance and stress signaling genes. Nevertheless, and while other genes display evidence of positive selection, a strong peak of selection co-localizes with the peak of the QTL and points at two R-genes. Such genes have been shown to confer resistance to rust in other species (Bettgenhaeuser *et al.*, 2014) and constitute prime candidates for further functional characterization.

B. distachyon is closely related to major crop cereals as well as to grass species used for biofuel production. Translating research from *B. distachyon* to plants of agronomical and economical interest will require a deeper understanding of the genetic architecture of traits involved in the response to environmental stresses. The molecular basis of tolerance to various abiotic stresses such as drought, salt and cold has been investigated in this species (Luo *et al.*, 2011; Manzaneda, 2013; Carmo & Charron, 2014; Gordon *et al.*, 2014; Marais & Juenger, 2015; Sun *et al.*, 2015; Mur & Bosch, 2016). Here, we also highlighted cadmium pollution as a potential factor of selection in the eastern population. As pollution with heavy metals including cadmium has been reported in Turkey in regions where accessions were collected for this study (Bakirdere & Yaman, 2008; Mor & Ceylan, 2017), our results suggest that *B. distachyon* could be used to investigate the tolerance to this stress. As genetic transformation is highly efficient in this species relative to other grasses, we anticipate

that combining classical trait mapping analyses with GWSS will assist allele mining for additional eco-responsive traits.

Conclusion

Our results revealed widespread signatures of natural selection at genes involved in adaptation in *B. distachyon*. We also found that pathogens may constitute an important driving force of genetic diversity and evolution in this system. While we limited our analysis to classical point mutations, recent studies showed that copy number variants (CNVs) and transposable element polymorphisms are abundant across *B. distachyon* populations (Gordon *et al.*, 2017; Stritt *et al.*, in press). Hence, the important genomic resources currently developed in this species open new avenues of research to further investigate the role of structural variation in natural population evolution and adaptation. To date, *B. distachyon* remains a classical model for research on grass genomics with a strong orientation towards applied research. Thanks to the high quality of its reference genome and the existence of large collections of freely available natural accessions collected from the species native range, it also constitutes a prime system for studies in ecology, population genomics and evolutionary biology.

Experimental procedures

SNPs calling, population structure and genetic diversity

We used paired-end Illumina sequencing data generated for 44 accessions of *B. distachyon* (Gordon *et al*., 2017) originating from Spain (N=16), France (N=1), Turkey (N=23) and Iraq (N=4, Table S1 for information about the origin of the accessions and sequencing effort). Reads were aligned to the reference genome v2.0 with BWA-MEM (standard settings; Li, 2013). After removing duplicates with Sambamba (Tarasov *et al.*, 2017), single nucleotide polymorphisms (SNPs) were called with Freebayes (Garrison & Marth, 2016). The output was then filtered by removing SNPs with more than 10 missing genotypes or more than 2 alleles, a quality lower than 20, a minor allele frequency of 0.05 and a mean depth lower than 20 or

higher than 200. Data were phased using the software BEAGLE V4 (Browning & Browning, 2007) using default settings.

We then used the program Admixture (Alexander & Novembre, 2009) to identify the genetic structure of the two populations. The analysis was run for K values from 1 to 5, and the best model was determined as the model with the lowest cross-validation error. Summary statistics such as within-population nucleotide or haplotype diversity were computed with the R package PopGenome (Pfeifer *et al.*, 2014). Nucleotide or haplotype diversity values were square-root transformed to fit a normal distribution and a t-test was used to compare the two within-population distributions. Levels of heterozygosity were calculated with VCFtools (Danecek *et al.*, 2011) .

Detecting balancing selection with an environmental association analysis

Bioclimatic variables were downloaded at a 30 arc-seconds resolution from http://www.worldclim.org/ and extracted for each locality using the R libraries gdal and raster. We then picked seven bioclimatic variables (bio6, bio8, bio11, bio12, bio13, bio16, bio18) that displayed substantial variation within geographical groups, but little between them. Bio6, bio8, bio11, bio12, bio13, bio16 and bio18 correspond respectively to minimum temperature of coldest month, mean temperature of wettest quarter, mean temperature of coldest quarter, annual precipitation, precipitation of wettest month, precipitation of wettest quarter and precipitation of warmest quarter. As likely to be correlated, these variables were summarized in a principal components analysis (PCA) in R. The coordinates of the first axis were used to test for correlation between allele frequencies and environment in the R package GENABEL (Aulchenko *et al.*, 2007). We accounted for relatedness between samples using a PCA correction and used corrected P-values of association.

Detecting balancing selection with ancestral recombination graphs (ARG)

We used the software ARGWeaver to detect additional candidate regions for balancing selection with a coalescence approach (Rasmussen *et al.*, 2014). We included in the analysis a subset of 12 accessions with high sequencing depth and covering the largest geographical range (6 accessions from each population) to limit computation time. We used a mutation rate of 1.4x10-9/bp/generation, a recombination rate of 5.9x10-8/bp/generation. Mutation rate was estimated by

aligning the orthologs of 100 genes and using rice as an out-group (divergence estimated at 40My (The international brachypodium Consortium, 2010)). Note that we subsequently used an outlier approach to identify the oldest polymorphisms present in the two populations (top 1% outliers). Therefore, potential biases inherent to the use of a molecular clock do not affect our analysis. ARGWeaver is also flexible with regard to recombination rate as it reconstructs ancestral recombination graphs and accommodates variable recombination rates and genealogies along the genome. The algorithm was run for 1000 iterations, using 20 discretized time steps, a maximum coalescence time of 3 million generations and a prior effective population size of 100,000 individuals.

Detecting disruptive positive selection

We used the Rsb test (Tang *et al.*, 2007) to detect signatures of recent or almost completed hard sweeps. This test detects haplotypes that are positively selected in one population by using a second population as a contrast. While the output of the test provides P-value of significance, it also indicates in which population a given allele is under selection. Rsb statistics were computed for each SNP with the R package rehh2.0 (Gautier *et al.*, 2012) with default settings. We further visualized the extension of haplotypes at candidate regions using the bifurcation.diagram() function of the rehh.

Detecting ongoing positive selection within populations

We eventually used the software H-scan (Schlamp *et al.*, 2016) to detect incomplete ongoing positive selection. To do so, we calculated average pairwise haplotypes lengths using the number of segregating sites spanned by each tract within each population. The statistics is expected to be larger as the number of extended haplotypes increases in a population. This method is specifically dedicated to the detection of ongoing sweeps where one (hard sweep) or several (soft sweep) haplotypes are under positive selection and provides statistics for each SNP. We ran the method on Turkish and Spanish accessions independently to detect selective sweeps within each geographical group.

Testing for functional clustering

We first performed a GO annotation for the 32,712 genes annotated in the reference genome (version 2.1) with Blast2GO (Conesa *et al.*, 2005). We then controlled for potential gene clustering by following the procedure described in (Al-Shahrour *et al.*, 2010). The entire gene set of the reference genome was split into windows of 50 consecutive genes. Windows were moved along chromosomes in steps of 25 genes to allow for half-window overlaps. Enrichment analyses of biological processes were then performed for all the generated windows with the R package GOstats (Falcon & Gentleman, 2017) using Fisher's exact test. P-values were subsequently adjusted for multiple testing with a Benjamin-Hochberg correction. Regions were considered significantly enriched for a biological process when they displayed a corrected P-value ≤ 0.01 and also harbored at least five genes associated to the given process.

GWWS subsequent filtering

Both the H-scan and the Rsb tests compute statistics at each SNP. To limit false positives, we first selected 10 kb windows displaying at least four significant SNPs within the top 1% outliers. Overlapping significant windows were merged. We narrowed down the selected windows by keeping only the genes located at and around (-10% of the peak value) each of the top 1% peaks of selection. For the association test, we also selected 10 kb windows displaying at least four significant SNPs, i.e. with a corrected P-value ≤ 0.001 ($-\log_{10}$(P-value) ≥ 3) and narrowed them down in the same manner. These filtering criteria, however, were not applied to the output of ARGweaver, which is a window-based approach and for which we only kept the top 1% outlier windows.

Overlap between the different approaches

Venn diagrams were drawn with the R package Vennerable to visualize potential overlap between the different gene sets under selection. To compare the distribution along chromosomes of candidate genes for recent positive selection (H-scan and Rsb candidate genes combined) in each population, we used linear models where the density of selected genes along each chromosome identified in the eastern and western population (100,000 bins per chromosome) were entered as variables. We eventually used the function plotBed of the R package Sushi (Phanstiel, 2015) to

vizualize the density of genes under positive selection as a heat map along each chromosome. The R package ggplot2 (Wickham, 2009) was used to display the density of all the annotated genes in the genome along each chromosome as a line.

Summary statistics and coalescence characterization at candidate loci

Tajima's D, F_{ST} and d_{XY} were computed in the R package PopGenome (v2.2.3) over 5kb windows across the genome. We then compared values between windows overlapping with candidate regions and windows outside these regions. We extracted TMRCA and relative TMRCA halftime (RTH) from the output of ARGWeaver. Because the coalescence approach is window-based, summary statistics were averaged across windows including 300 non-recombining blocks (around 10kb windows). Significant windows were merged. We then compared TMRCA and RTH values between windows overlapping with candidate regions and those outside.

GO annotation of genes under selection

For each test, we extracted the genes located in the filtered regions with bedtools (Quinlan & Hall, 2010). We then examined potential enrichment for biological processes for each of the selected gene sets with the R package GOstats (Falcon & Gentleman, 2017) using the Fisher's exact test. P-values were subsequently adjusted for multiple testing with a Benjamin-Hochberg correction. Gene sets were considered significantly enriched for a biological process when they displayed a P-value ≤ 0.01 and harbored at least five genes associated to the given process. The ancestor and child terms of each significant process were determined using QuickGO (http://www.ebi.ac.uk/QuickGO) and used to simplify Fisher test outputs and keep non-redundant terms.

QTL for leaf-rust resistance validation

A QTL analysis performed in *B. distachyon* revealed three genomic regions involved in the resistance to *P. brachypodii* (Barbieri *et al.*, 2012). The coordinates of these three QTL were extracted from (Barbieri *et al.*, 2012) from v.2.0 of the *B. distachyon* reference genome. We then assessed weather those three regions were identified as outliers in at least one of the tests of selection.

Acknowledgements

We thank the Genetic Diversity Center-ETH Zurich for providing access the Euler high-performance cluster. We also would like to thank Beat Keller's group as well as Simon Krattinger and Mahendra Mariadassou for their advises during the elaboration of the study. This work is supported by the Swiss National Science Foundation (PZ00P3_154724). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.

Availability of data and materials

The raw outputs of each GWWS will be archive upon acceptance of the manuscript. All whole-genome sequences data are available at the NCBI Sequence Read archive (SRA available in Gordon *et al*., 2017).

Conflict of interest

There is no conflict of interest issue related to this work.

Supporting Information

Table S1: Geographical coordinates of the 44 accessions and sequencing effort.

Table S2: Significant GO terms in each 50 gene-window of the reference genome

Table S3: Genes under selection with functional annotation

References

Al-Shahrour, Minguez P, Marques-Bonet T, Gazave E, Navarro A, Dopazo J. **2010**. Selection upon Genome Architecture□: Conservation of Functional Neighborhoods with Changing Genes. *PLoS computational biology* **6**: e1000953.

Alexander DH, Novembre J. **2009**. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome research* **19**: 1655–1664.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. **2007**. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**: 1294–6.

Bakirdere S, Yaman M. **2008**. Determination of lead , cadmium and copper in roadside soil and plants in Elazig , Turkey. *environmental monitoring Assessment* **136**: 401–410.

Barbieri M, Marcel TC, Niks RE, Francia E, Pasquariello M, Mazzamurro V, Garvin DF. **2012**. QTLs for resistance to the false brome rust Puccinia brachypodii in the model grass. *Genome* **163**: 152–163.

Bettgenhaeuser J, Gilbert B, Ayliffe M, Moscou MJ. **2014**. Nonhost resistance to rust pathogens – a continuation of continua. *frontiers in plant science* **5**: 1–15.

Binney H, Edwards M, Macias-fauria M, Lozhkin A, Anderson P, Kaplan JO, Andreev A, Bezrukova E, Blyakharchuk T, Jankovska V, *et al.* 2017. Vegetation of Eurasia from the last glacial maximum to present□: Key biogeographic patterns. *Quaternary Science Reviews* **157**: 80–97.

Bischoff A, Crémieux L, Smilauerova M, Lawson CS, Mortimer SR, Dolezal J, Edwards AR, Brook AJ, Macel M, Leps JAN, *et al.* 2006. Detecting local adaptation in widespread grassland species – the importance of scale and local plant community. *journal of ecology* **94**: 1130–1142.

Bourgeois Y, Roulin AC, Müller K, Ebert D. **2017**. Parasitism drives host genome evolution: Insights from the *Pasteuria ramosa* - *Daphnia magna* system. *Evolution* **71**:1106–13.

Brown JKM, Tellier A. **2011**. Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology* **49**: 345–67.

Browning SR, Browning BL. **2007**. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *the american journal of human genetics* **81**: 1084–1097.

Carmo S Do, Charron J. **2014**. Comparative analysis of the cold acclimation and

freezing tolerance capacities of seven diploid Brachypodium distachyon accessions. *Annals of botany* **113**: 681–693.

Ceballos, G., Davidson, A., List, R., Pacheco, J., Manzano-Fisher, P., Santos-Barrera, G., Cruzado J. **2010**. Rapid Decline of a Grassland System and Its Ecological and Conservation Implications. *PloS one* **5**: e8562.

Charlesworth D. **2006**. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* **2**: e64.

Chavan S, Gray J, Smith SM. **2015**. Diversity and evolution of Rp1 rust resistance genes in four maize lines. *Theoretical and Applied Genetics* **128**: 985–998.

Clayton WD. 1981. Evolution and Distribution of Grasses. *Annals of the Missouri Botanical Garden* **68**: 5–14.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. **2005**. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–6.

Couto D, Zipfel C. **2016**. Regulation of pattern recognition receptor signalling in plants. *nature reviews immunology* **16**: 537–552.

Cruickshank TE, Hahn MW. **2014**. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**: 3133–3157.

Cutter AD, Payseur BA. **2013**. Genomic signatures of selection at linked sites□: unifying the disparity among species. *Nature reviews Genetics* **14**: 262–274.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dell'Acqua MD, Zuccolo A, Tuna M, Gianfranceschi L, Pè ME. **2014**. Targeting environmental adaptation in the monocot model Brachypodium distachyon□: a multi-faceted approach. *BMC Genomics* **15**: 801.

Delph LF, Kelly JK. **2014**. On the importance of balancing selection in plants Lynda. *New phytologist* **201**: 1–22.

Des Marais DL, Juenger TE. **2015**. Brachypodium and the Abiotic Environment. In: Vogel J, editor. *Genetics and Genomics of Brachypodium*, 291–311.

Dodds P, Thrall P. **2009**. Recognition events and host–pathogen co-evolution in gene-for- gene resistance to flax rust. *Functional plant biology* **36**: 395–408.

Dong S, Raffaele S, Kamoun S. **2015**. ScienceDirect The two-speed genomes of

filamentous pathogens□: waltz with plants. *Current Opinion in Genetics & Development* **35**: 57–65.

Durian G, Rahikainen M, Alegre S, Brosché M. **2016**. Protein Phosphatase 2A in the Regulatory Network Underlying Biotic Stress Resistance in Plants. *frontiers in immunology* **7**: 1–17.

Eckardt NA. **2017**. The Plant Cell Reviews Plant Immunity□: Receptor-Like Kinases, ROS-RLK Crosstalk, Quantitative Resistance, and the Growth / Defense Trade-Off. *Plant cell* **29**: 601–602.

Falcon SÃ, Gentleman R. **2017**. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**: 257–258.

Feder ME, Mitchell-olds T. **2003**. Evolutionary and ecological functional genomics. *Nature reviews Genetics* **4**: 649–655.

Feliner GN. **2011**. Southern European glacial refugia□: A tale of tales. *taxon* **60**: 365–372.

Fijarczyk A, Babik A. **2015**. Detecting balancing selection in genomes□: limits and prospects. *Molecular ecology* **24**: 3529–3545.

Flood J, Hancock AM. **2017**. ScienceDirect The genomic basis of adaptation in plants. *Current opinion in plant biology* **36**: 88–94.

Fones H, Preston GM. **2011**. Reactive oxygen and oxidative stress tolerance in plant pathogenic Pseudomonas. *FEMS Microbiology letter* **1**: 1–8.

Fuller ZL, Niño EL, Patch HM, Bedoya-Reina OC, Baumgarten T, Muli E, Mumoki F, Ratan A, McGraw J, Frazier M, *et al.* 2015. Genome-wide analysis of signatures of selection in populations of African honey bees (Apis mellifera) using new web-based tools. *BMC genomics* **16**: 518.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-admettla A, Pattini L. **2011**. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLoS genetics* **7**: e1002355.

Garrison E, Marth G. **2016**. Haplotype-based variant detection from short-read sequencing. Available at:https://arxiv.org/abs/1207.3907

Gautier M, Vitalis R, Baillarguet D, Cedex F-M. **2012**. rehh□: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**: 1176–1177.

Gibson DJ. **2009**. *Grasses and grassland ecology* (Oxford university press, Ed.).

Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S,

Stritt C, Roulin AC, Schackwitz W, Tyler et al. Extensive gene content hetlvariation in the Brachypodium distachyon pan-genome correlates with phenotypic variation. *Nature Communications* **8**: 2184

Gordon SP, Priest H, Marais DL Des, Schackwitz W, Figueroa M, Martin J, Jennifer N, Tyler L, Lee C, Bryant D, *et al.* 2014. Genome diversity in Brachypodium distachyon□: deep sequencing of highly diverse inbred lines. Plant Journal **79**: 361–374.

Gos G, Slotte T, Wright SI. **2012**. Signatures of balancing selection are maintained at disease resistance loci following mating system evolution and a population bottleneck in the genus Capsella. *BMC Evolutionary Biology* **12**: 152.

Greeff C, Roux M, Mundy J, Petersen M. **2012**. Receptor-like kinase complexes in plant innate immunity. *frontiers in plant scienceontiers in plant science* **3**: 1–7.

Groves RH. **2000**. Temperate grasslands of the southern hemisphere. In: Jacobs SWL, Everett J, editors. *Grasses Systematics and Evolution*, 3781–3791.

Hammond-kosack KE, Jones JDG. **1996**. Resistance Gene-Dependent Plant Defense Responses. *The Plant cell* **8**: 1773–1791.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. **2011**. Adaptation to climate across the Arabidopsis thaliana genome. *Science* **334**: 83–6.

Hassl RJ, Payseur BA. **2016**. Detecting selection in natural populations: Making sense of genome scans and towards alternative solutions. *Molecular Ecology* **25**: 5–23.

Helm A, Oja T, Saar L, Takkis K, Talve T, Pa M. **2009**. Human influence lowers plant genetic diversity in communities with extinction debt. *journal of ecology* **97**: 1329–1336.

Hermisson J. **2009**. Who believes in whole-genome scans for selection□? *Heredity* **103**: 283–284.

Hewitt GM. **1999**. Post-glacial re-colonization of European biota. *Biological journal of the Linnean society* **68**: 87–112.

Hewitt G. **2000**. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Hovmøller MS, Sørensen CK, Walter S, Justesen AF. **2011**. Diversity of Puccinia striiformis on Cereals and Grasses. *Annual Review of Phytopathology* **49**: 197–220.

Huang J, Pray C, Rozelle S. **2002**. Enhancing the crops to feed the poor. *Nature* **418**: 678–684.

Huo N, Garvin DF, You FM, Luo SMM, Gu YQ, Lazo GR, Philip J. **2011**. Comparison of a high-density genetic linkage map to genome features in the model grass Brachypodium distachyon. *Theoretical and Applied Genetics* **123**: 455–464.

Jacob F, Vernaldi S, Maekawa T. **2013**. Evolution and conservation of plant NLR functions. *frontiers in immunology* **4**: 1–16.

Karasov TL, Horton MW, Bergelson J. ScienceDirect Genomic variability as a driver of plant – pathogen coevolution? *Current Opinion in Plant Biology* **18**: 24–30.

Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. **2006**. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome research*: **206**: 980–989.

Kiviniemi K, Eriksson O. 2002. Size-related deterioration of semi-natural grassland fragments in Sweden. *diversity and distribution* **8**: 21–29.

Krattinger SG, Keller B. **2016a**. Tansley review Molecular genetics and evolution of disease resistance in cereals. *New phytologist*: 320–332.

Krattinger SG, Keller B. **2016b**. Trapping the intruder — immune receptor domain fusions provide new molecular leads for improving disease resistance in plants. *Genome Biology* **17**:23.

Latta RG. **2009**. Testing for local adaptation in Avena barbata: a classic example of ecotypic divergence. *Molecular Ecology* **18**: 3781–3791.

Li H. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. preprint arXiv:1303.3997.

Liu W, Frick M, Huel R, Nykiforuk CL, Wang X, Gaudet DA, Eudes F, Conner RL, Kuzyk A, Chen Q, *et al.* 2014. The Stripe Rust Resistance Gene Yr10 Encodes an Evolutionary-Conserved and Unique CC – NBS – LRR Sequence in Wheat. *Molecular Plant* **7**: 1740–1755.

Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, Zhang Q, Vilhjálmsson BJ, Korte A, Nizhynska V, *et al.* 2013. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature genetics* **45**: 884–890.

Lopez-Alvarez D, Manzaneda AJ, Rey P, Giraldo P, Benavente E, Allainguillaume J *et al.* 2015. Environmental niche variation and evolutionary diversifi cation of the Brachypodium distachyon grass complex species in their native. *American journal of botany* **102**: 1073–1088.

Luo N, Liu J, Yu X, Jiang Y. **2011**. Natural variation of drought response in Brachypodium distachyon. *physiologia plantarum* **141**: 19–29.

Mace E, Tai S, Innes D, Godwin I, Hu W, Campbell B, Gilding E, Cruickshank A, Prentis P, Wang J, *et al.* 2014. The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC plant biology* **14**: 253.

Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. **2013**. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in Brachypodium distachyon (Poaceae) Antonio. *New phytologist* **193**: 797–805.

Mchale L, Tan X, Koehl P, Michelmore RW. **2006**. Plant NBS-LRR proteins□: adaptable guards. *Genome Biology* **7**: 212.

Messer PW. **2015**. H-scan□: Detecting hard and soft sweeps in population genomic data. **3**: 1–4.

Michelmore RW, Meyers BC. **1998**. Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process. *Genome research* **530**: 1113–1130.

Mitchell-Olds T, Willis JH, Goldstein DB. **2007**. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature reviews Genetics* **8**: 845–56.

Mondragón-palomino M, Meyers BC, Michelmore RW, Gaut BS. **2002**. Patterns of Positive Selection in the Complete NBS-LRR Gene Family of Arabidopsis thaliana. *Genome research* **12**: 1305–1315.

Mor F, Ceylan S. **2017**. Cadmium and lead contamination in vegetables collected from industrial , traffic and rural areas in Bursa Province, Turkey. *Food Additives & Contaminants*□**25**: 611–615.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD. **2013**. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences* **110**: 453–458.

Mur LAJ, Bosch M. **2016**. Linking Dynamic Phenotyping with Metabolite Analysis to Study Natural Variation in Drought Responses of Brachypodium distachyon. *frontiers in plant science* **7**: 1–15.

Nelson RM, Wallberg A, Sim P, Lawson DJ, Webster MT. **2017**. Genomewide analysis of admixture and adaptation in the Africanized honeybee. *Molecular Ecology* **26**: 3603–3617.

Nielsen R. **2005**. Molecular Signatures of Natural Selection. Annual Review of Genetics. **39**: 197–218.

Nutzmann H, Osbourn A. **2014**. Gene clustering in plant specialized metabolism. *current opinion in biotechnology* **26**: 91–99.

Ooijen G Van, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC, Takken FLW. **2017**. Structure – function analysis of the NB-ARC domain of plant disease resistance proteins. *jouranl of experimental botany* **59**: 1383–1397.

Opanowicz M, Vain P, Draper J, Parker D, Doonan JH. **2008**. Brachypodium distachyon□: making hay with a wild grass. *Trends in Plant Science* **13**: 172–177.

París SM, Téllez-iñón MT, Capiati DA. **2009**. Serine / threonine protein phosphatases type 2A and their roles in stress signaling. *Plant signaling and behaviour* **4**: 1013–1015.

Park C, Caddell DF, Ronald PC. **2012**. Protein phosphorylation in plant immunity□: insights into the regulation of pattern recognition receptor-mediated signaling. *frontiers in plant science* **3**: 1–9.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A. **2012**. A Critical Assessment of Storytelling□: Gene Ontology Categories and the Importance of Validating Genomic Scans Research article. *Molecular biology and evolution* **29**: 3237–3248.

Phanstiel DH. **2015**. Sushi: Tools for visualizing genomics data. *R package version 1.14.0*.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. **2014**. PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution* **31**: 1929–1936.

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Bossdorf O, Byers D, Donohue K, Dunning M, *et al.* 2010. The Scale of Population Structure in Arabidopsis thaliana. *plos genetics* **6**: e1000843.

Quinlan AR, Hall IM. **2010**. BEDTools□: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. **2014**. Genome-Wide Inference of Ancestral Recombination Graphs. *PLoS genetics* **10**: e1004342.

Renaut S, Grassa CJ,Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. **2013**. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* **4**: 1–8.

Richman A. **2000**. Evolution of balanced genetic polymorphism. *Molecular Ecology*

9: 1953–1963.

Roulin A, Bourgeois Y, Stiefel U, Walser J-C, Ebert D. A photoreceptor underlies natural variation in Daphnia magna diapause induction. *Molecular Biology and Evolution*. **12**:3194–3204.

Samuk K, Rennison DJ, Owens GL, Delmore KE, Miller SE, Schluter D. **2017**. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology* **26**: 4378–4390.

Savolainen O, Lascoux M, Merilä J. **2013**. Ecological genomics of local adaptation. *Nature reviews Genetics* **14**: 807–820.

Schlamp F, Van Der Made J, Stambler R. **2016**. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Molecular Ecology* **25**: 342–356.

Sewelam N, Kazan K, Schenk PM. **2016**. Global Plant Stress Signaling□: Reactive Oxygen Species at the. **7**: 1–21.

Singh S, Chand S, Singh NK, Sharma TR. **2015**. Genome-Wide Distribution , Organisation and Functional Characterization of Disease Resistance and Defence Response Genes across Rice Species. *Plos one* **10**:e0125964

Slavov GT, Nipper R, Robson P, Farrar K, Allison GG, Bosch M, Clifton-brown JC, Donnison IS, Jensen E. **2014**. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass Miscanthus sinensis. *New phytologist* **201**: 1227–1239.

Slotte T. **2014**. The impact of linked selection on plant genomic variation. *briefings in functional genomics* **13**: 268–275.

Stamatakis A, Alachiotis N, Pavlidis P, Daniel Z. **2013**. SweeD□: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Molecular biology and evolution* **30**: 2224–2234.

Stinchcombe JR, Hoekstra HE. **2008**. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.

Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. Transposable elements as a major source of genetic diversity: insights from genome-wide analyses of Brachypodium distachyon natural accessions. *Genome biology and evolution*. In press. Available at https://doi.org/10.1093/gbe/evx276.

Stromberg C. **2011**. Evolution of Grasses and Grassland Ecosystems. *Annual review*

of earth and planetary science **39**: 517–544.

Stukenbrock EH, Banke S, Javan-nikkhah M, Mcdonald BA. **2005**. Origin and Domestication of the Fungal Wheat Pathogen Mycosphaerella graminicola via Sympatric Speciation. *Molecular biology and evolution* **24**: 398–411.

Sun J, Hu W, Zhou R, Wang L. **2015**. The Brachypodium distachyon BdWRKY36 gene confers tolerance to drought stress in transgenic tobacco plants. *Plant cell reports* **34**: 23–35.

Svenning J, Normand S, Kageyama M. **2008**. Glacial refugia of temperate trees in Europe□: insights from species distribution modelling. *journal of ecology* **96**: 1117–1127.

Tajima F. **1989**. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–95.

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K, *et al.* 2011. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. *Plant Journal* **68**: 273–286.

Tang K, Thornton KR, Stoneking M. **2007**. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS biology* **5**:e171

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. **2017**. Genome analysis Sambamba□: fast processing of NGS alignment formats. *Bioinformatics* **31**: 2032–2034.

The international brachypodium Consortium. **2010**. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **463**: 763–8.

Torres MA, Jones JDG, Dangl JL. **2006**. Reactive Oxygen Species Signaling in Response to Pathogens. **141**: 373–378.

Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Investigation G, Giant T, Hirschhorn JN. **2012**. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* **44**: 1015–1019.

Tyler AL, Lee SJ, Young ND, Deiulio GA, Benavente E, Reagon M, Sysopha J, Baldini RM, Troìa A, Hazen SP, *et al.* 2016. Population structure in the model grass Brachypodium distachyon is highly correlated with flowering differences

across broad geographic areas. *The Plant Genome* **9**:1–55.

Wadgymar SM, Lowry DB, Gould BA, Byron CN, Mactavish RM, Anderson JT. **2017**. Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods in Ecology and Evolution* **8**: 738–749.

Wallace JS, Batchelor CH. **1997**. Managing water resources for crop production. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **352**: 937–947.

Wickham H. **2009**. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wojtaszek P. **1997**. Oxidative burst□: an early plant response to pathogen infection. *biochemical journal* **692**: 681–692.

Wu Q, Han T, Chen X, Chen J, Zou Y, Li Z, Xu Y. **2017**. Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. *Genome biology* **18**: 217.

Wyand RA, Brown K. 2003. Genetic and forma specialis diversity in Blumeria graminis of cereals and its implications for host-pathogen co-evolution. *Molecular plant pathology* **4**: 187–198.

Zhong Y, Li Y, Huang K, Cheng Z. **2015**. Species-specific duplications of NBS-encoding genes in Chinese chestnut (Castanea mollissima). *Scientific reports* **5:** 16638.

Figures

Figure 1: Schematic representation of the regimes of selection analyzed in this study

In the case of association between SNPs and environmental variables displaying variation within population, shared polymorphisms will be maintained across the two populations. The same outcome is expected under long-term balancing selection. The Rsb test, on the other hand, detects mutations that rose in frequency and reached near fixation in one of the two populations. The H-scan test targets mutations currently under selection and rising in frequency in one of the two populations.

Figure 2: Geographical origin and genetic diversity of the 44 accessions

A) Geographical origin of the 44 accessions used in the study. Blue and orange dots represent accessions from the western and eastern population respectively. The asterisk indicates the origin of the reference accession Bd21 B) Structure plot. Each bar represents the genetic data of one accession. An accession presents some admixture when the bar displays different colors. The high of the bar is proportional to the admixture level C) Nucleotide diversity (in 5 kb windows) with the eastern and western population D) PCA summarizing bioclimatic variables varying within the eastern and western population.

Figure 3: Venn diagram and descriptive statistics for each test of selection

A) Overlap between environmental association, H-scan and Rsb results in the eastern and western populations, in number of genes B) Number of genes per window after keeping the genes located at and in the vicinity of each peak of selection C) Comparison between genome-wide (GW) Tajima's D and Tajima's D in regions under the different selection regimes tested D) Overall gene density (line) and heat maps displaying the density of genes under positive selection (Rsb + H-scan) along the five chromosomes in the eastern (E) and western (W) populations.

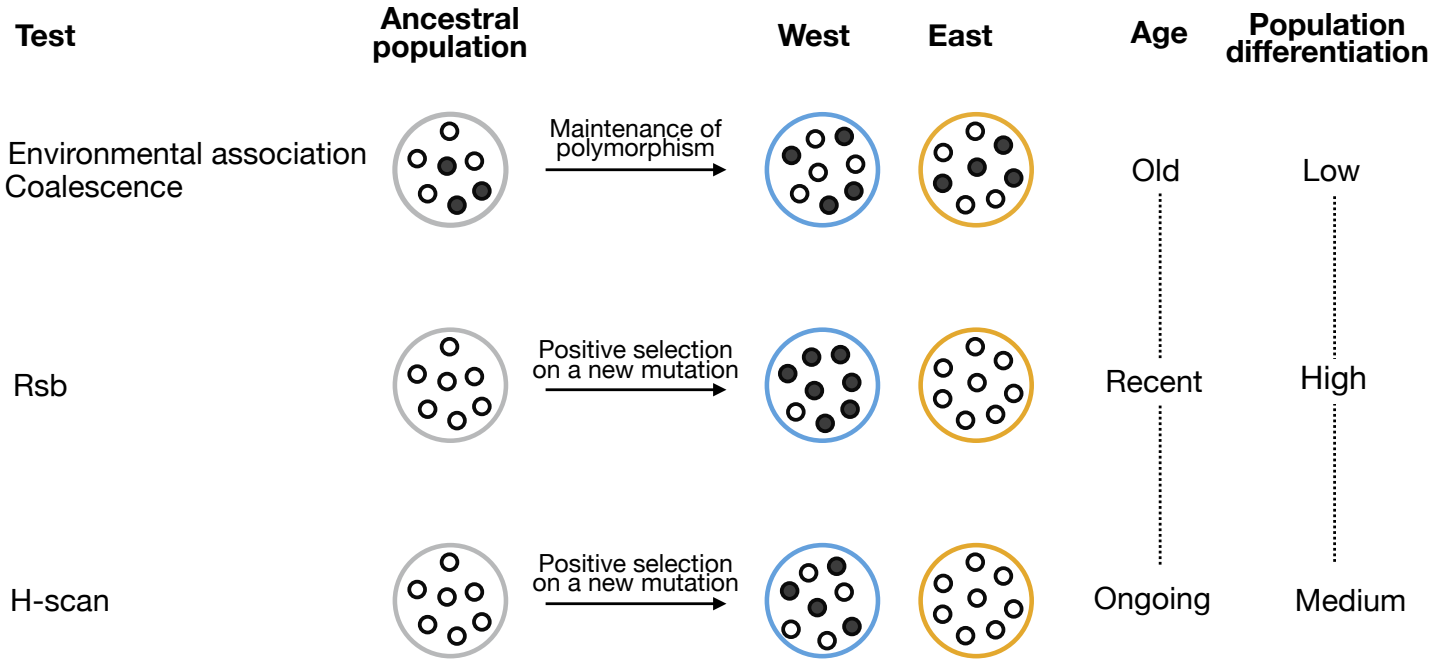
Figure 4: QTL involved in resistance to rust

A) Rsb signals observed in $QTL_{rust\text{-}3}$ (left panel) and $QTL_{rust\text{-}4}$ (right panel). The dashed line indicates the 0.1% $-\log_{10}$(Rsb P-values) outlier threshold. Note that in order to reduce the size of the figure, points with $-\log_{10}$(Rsb P-values) values <1 are not displayed. Grey boxes display the position of gene potentially involved in defense

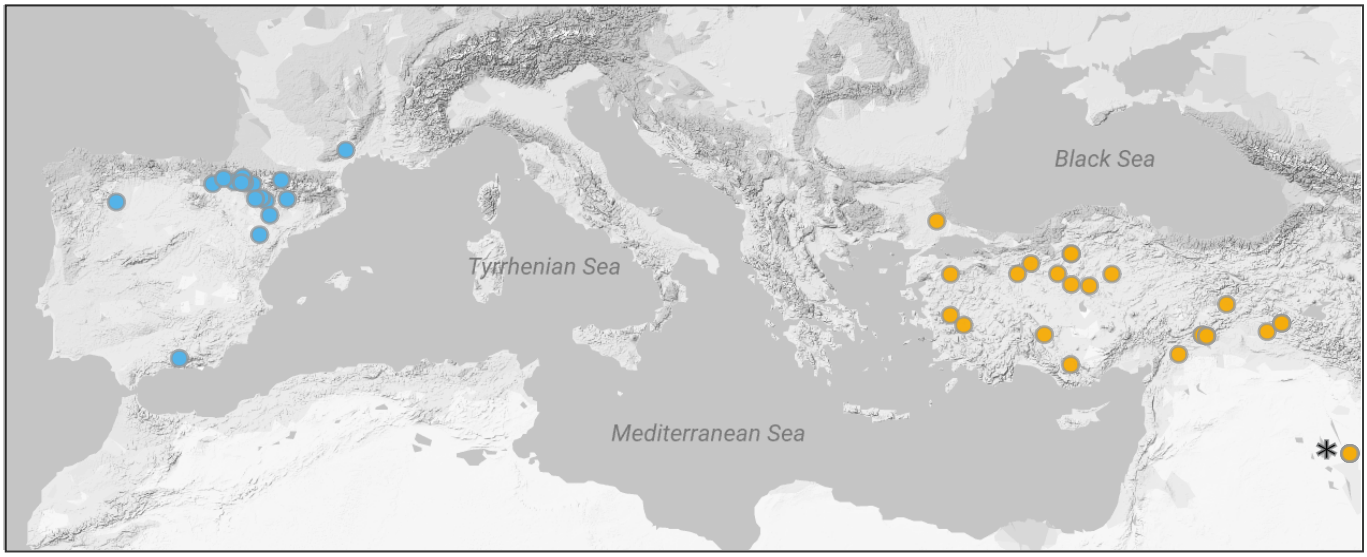
response (genes with NBS-LRR, F-box or RLK domains) if present in the region B) Haplotype bifurcation diagram in sub-regions displaying strong Rsb outliers. Left panel: zoom in on chromosome 3; right panel: zoom in on chromosome 4. The diagrams visualize the breakdown of LD at increasing distances from the selected focal SNP, displayed by a vertical black line. Each horizontal blue line represents a haplotype. Each accession is represented by two lines, one for each haplotype. Horizontal blue lines are merged when two accessions share the same haplotype. The thickness of the line is therefore correlated to the number of accessions sharing the same haplotype.

Tables

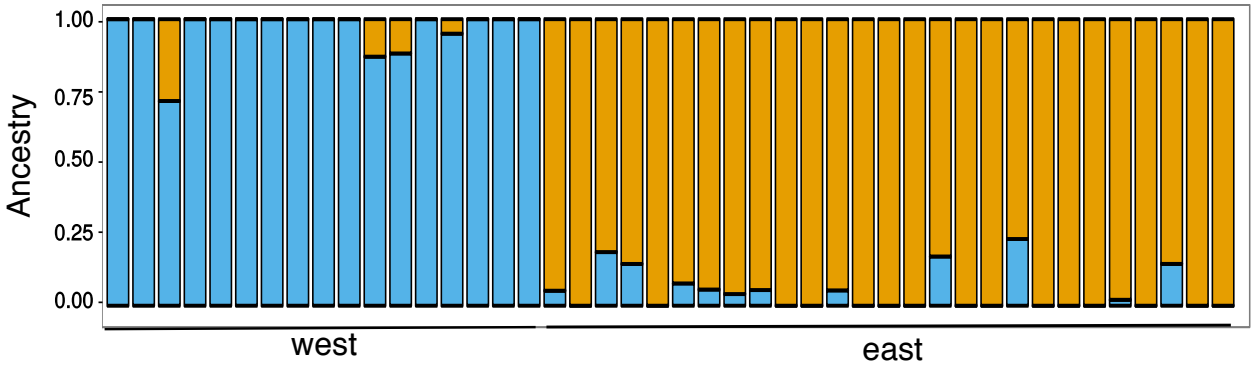
Table 1: Significant GO term associated to each test of selection


Test
Ancestral population
West
East
Age
Population differentiation
Environmental association
Coalescence
Maintenance of polymorphism
Old
Low
Rsb
Positive selection on a new mutation
Recent
High
H-scan
Positive selection on a new mutation
Ongoing
Medium

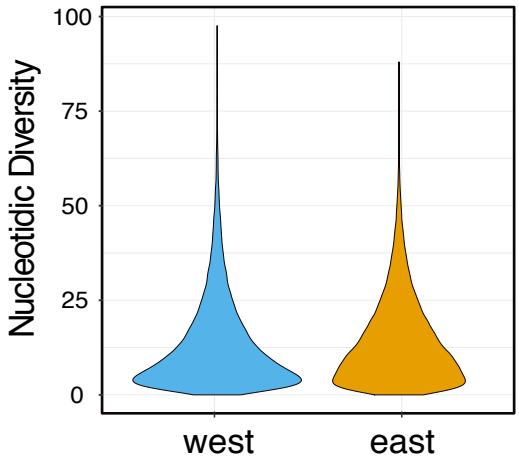
A


Black Sea
Tyrrhenian Sea
Mediterranean Sea

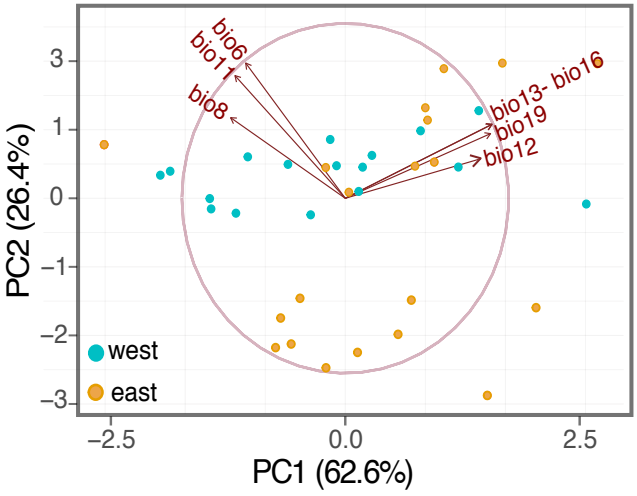
B


Ancestry
1.00
0.75
0.50
0.25
0.00
west
east

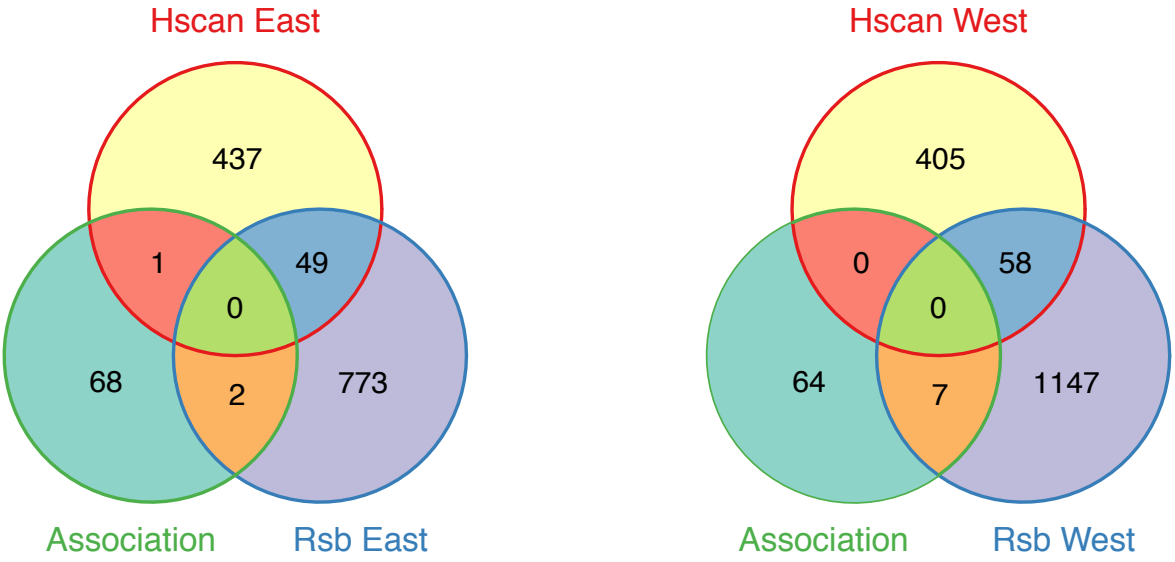
C


Nucleotidic Diversity
100
75
50
25
0
west
east

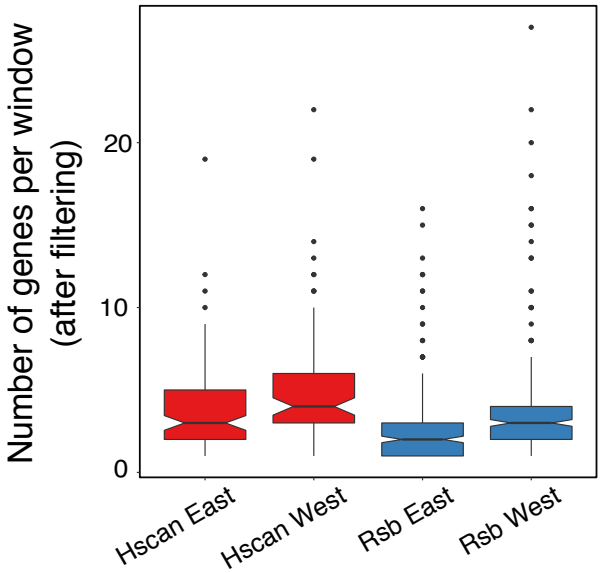
D


bio6
bio11
bio8
bio13- bio16
bio19
bio12
PC2 (26.4%)
PC1 (62.6%)
3
1
0
−1
−2
−3
−2.5
0.0
2.5
west
east

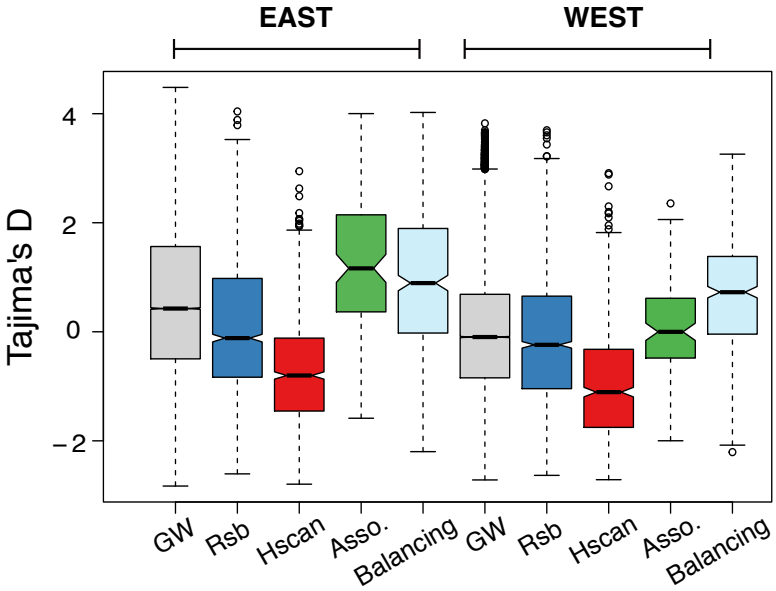
A


Hscan East
437
1
49
0
68
2
773
Association
Rsb East
Hscan West
405
0
58
0
64
7
1147
Association
Rsb West

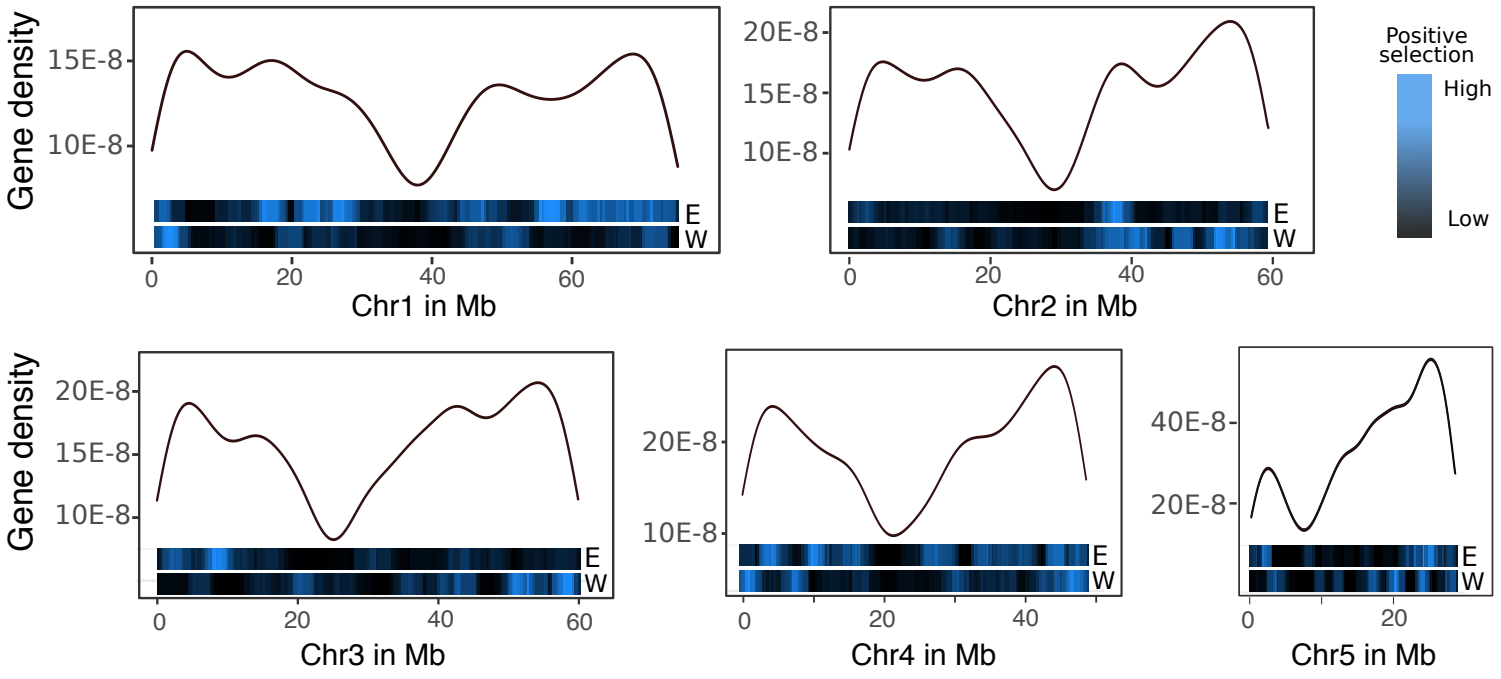
B


Number of genes per window
(after filtering)
0
10
20
Hscan East
Hscan West
Rsb East
Rsb West

C


EAST
WEST
Tajima's D
4
2
0
−2
GW
Rsb
Hscan
Asso.
Balancing
GW
Rsb
Hscan
Asso.
Balancing

D


Gene density
15E-8
10E-8
E
W
0
20
40
60
Chr1 in Mb
20E-8
15E-8
10E-8
Chr2 in Mb
Positive selection
High
Low
Gene density
20E-8
15E-8
10E-8
Chr3 in Mb
20E-8
10E-8
Chr4 in Mb
40E-8
20E-8
Chr5 in Mb

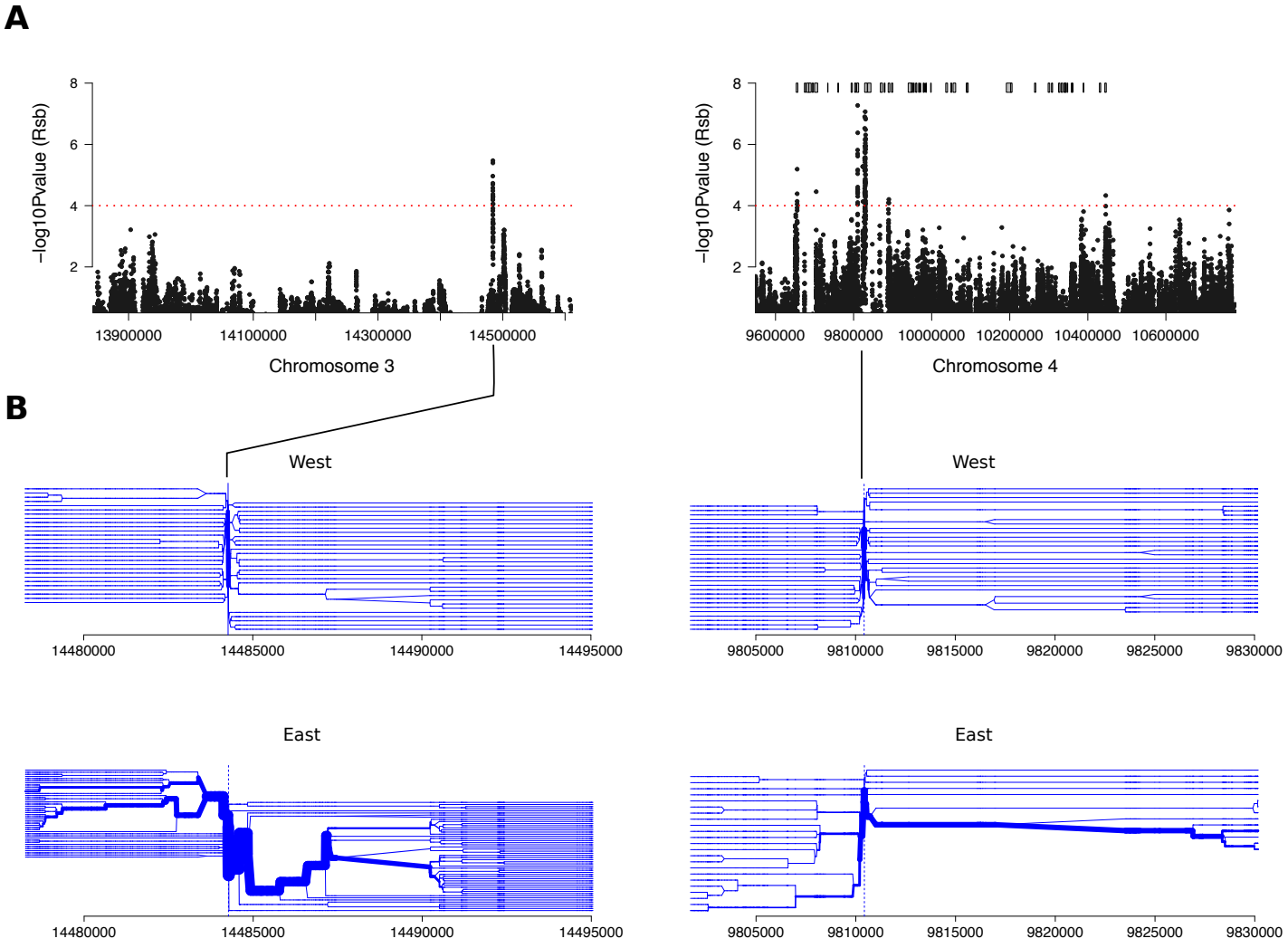

A
−log10Pvalue (Rsb)
8
6
4
2
13900000
14100000
14300000
14500000
Chromosome 3
9600000
9800000
10000000
10200000
10400000
10600000
Chromosome 4
B
West
14480000
14485000
14490000
14495000
9805000
9810000
9815000
9820000
9825000
9830000
East

Table 1: significant GO term associated to each test of selection

Test	Significant GO process	Pvalue
Association	None	NA
TMRCA	Phosphorylation	3.13E-08
Rsb East	Response to stress including:	3.00E-05
	- defense response	2.98E-03
	- response to oxidative stress	2.98E-03
RsbWest	Nitrogen compound transport	4.70E-03
H-scan East	Response to stress including:	2.66E-04
	- defense response	7.82E-05
	- response to cadmium ion	2.66E-04
H-scan West	Pyruvate metabolic process	4.00E-04