

1 Genomic repetition in climate adaptation 2 across different crop species

3 Oliver Selmoni¹

4 1– Spatial Genetics of Ecosystems, Department of Geography, University of Zurich

5 oliver.selmoni@gmail.com

6

7

8 Abstract

9 Commercial crop varieties (cultivars) have a narrow genetic diversity that makes them
10 vulnerable to climate change. Yet most crops have related semi-domesticated varieties
11 (landraces) and wild relatives that thrive under climates that would be considered extreme for
12 cultivars. The genomes of such varieties are reservoirs of unknown adaptive traits, some of
13 which could be used to extend the climatic tolerance of cultivars.

14 To shed light on these traits, we built a genomic database of fourteen crop types, including
15 landraces and wild relatives of staple crops. This database included individual-level genomic
16 data for plants from over 3,000 locations worldwide. Leveraging remote sensing and climatic
17 models, we reconstructed bioclimatic conditions, soil types, and human disturbance for each
18 of these locations. Genes whose geographic variation correlates with an environmental
19 gradient are potentially implicated in local adaptation; if the same gene repeatedly shows this
20 association in different crops, its adaptive role is more likely.

21 We found 16 groups of genes showing such repeated environmental associations, mainly
22 linked to water availability and cold-season temperature. For some of these genes, the same
23 climate association was observed in 6, 7, or even up to 13 crop types. Using *A. thaliana* as a
24 reference, we found that many of these genes likely encode proteins already well known to
25 regulate cellular responses to drought and cold stress. Our findings suggest that natural
26 variation in these genes can enhance such responses, and that such variation exists in
27 landraces and wild relatives of different crops. If validated, this finding could open new
28 avenues to make cultivars more climate-tolerant via genomic-guided breeding.

29

30

31 Introduction

32 Climate change poses a major challenge to agricultural systems and global food security (1).
33 One key issue is the narrow genetic diversity of commercial crop varieties (cultivars), a
34 consequence of centuries of domestication (2). Crops with limited genetic diversity require
35 stable climate conditions for optimal growth, typically maintained through irrigation, fertilizers,
36 and pesticides (3). However, as climate instability increases, reliance on these external
37 inputs will grow, making agricultural practices more challenging and less sustainable (4). A
38 recent study projected that, by the end of the century, climate change could reduce yields by
39 up to 27% for staple crops worldwide (5). Thus, there is an urgent need to develop new crop
40 varieties capable of withstanding extreme conditions, such as recurrent drought or degraded
41 soils (3).

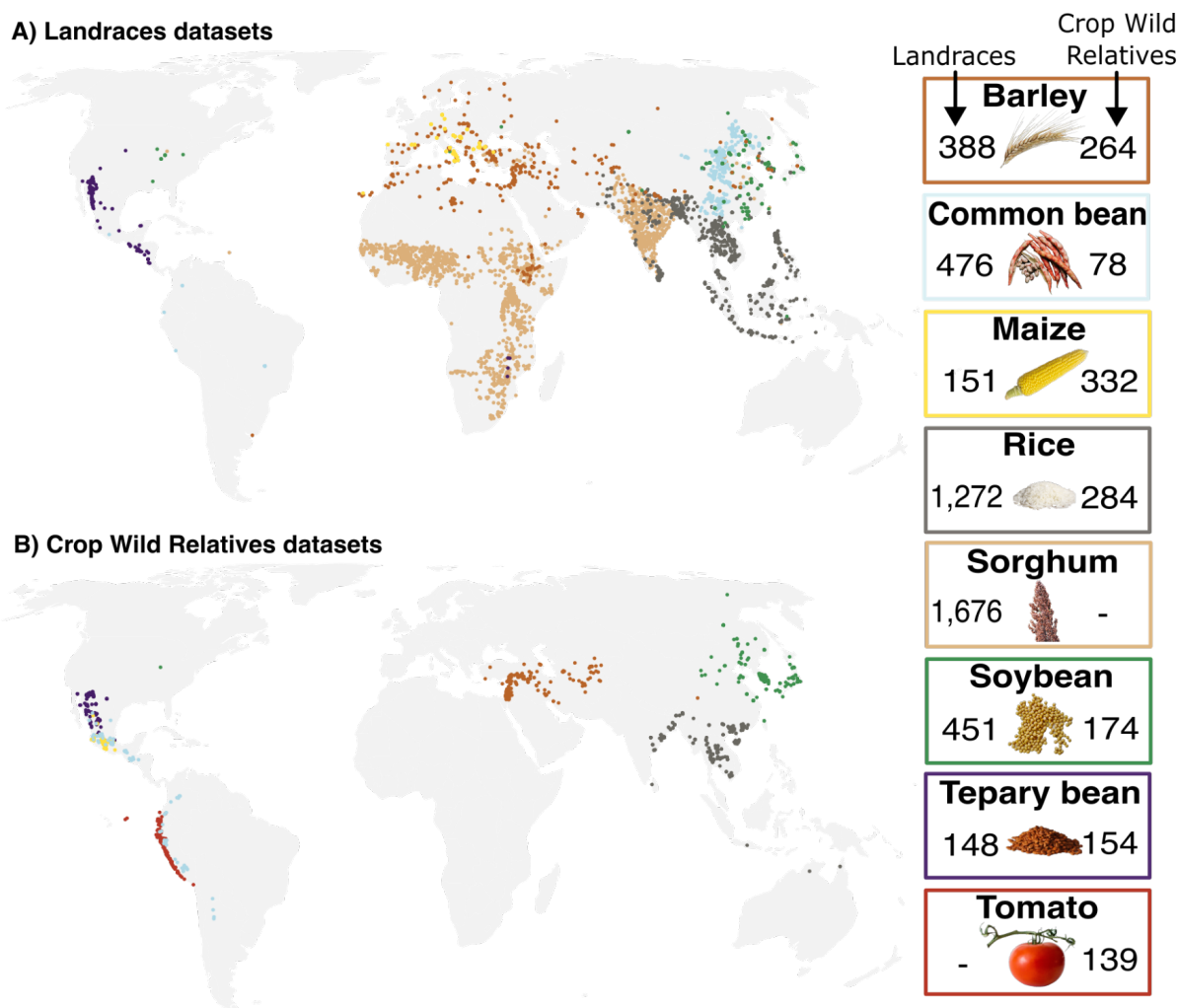
42 Crop landraces and wild relatives could provide the genetic material to make cultivars
43 more climate-resilient (6). Landraces are local varieties maintained by traditional farming
44 systems, well adapted to their local growing conditions, minimizing the need for external
45 inputs. Crop wild relatives (CWR) are varieties that have completely escaped domestication
46 and thrive in the wild. Both landraces and wild relatives retain greater genetic diversity and
47 can grow under conditions considered extreme for cultivars (2). Therefore, their genomes
48 represent unexplored reservoirs of genetic diversity that could be leveraged to adapt
49 cultivars (7), for example by assisting breeding via genomic selection (8). For these
50 strategies to succeed, the genetic variants underpinning climate adaptation in landraces and
51 CWR must first be identified (9).

52 Genome-environment association (GEA) studies are an emergent approach to
53 identify genetic variants potentially underpinning climate adaptation in the field (10). DNA
54 from hundreds of individuals is first sampled across a species' range. Conditions at sampling
55 sites are then reconstructed using climatic and environmental models, characterizing
56 variables such as bioclimatic factors (11), topography (12), soil types (13, 14), or
57 anthropogenic influences (15). The DNA of each individual is sequenced genome-wide, and
58 genetic variation is compared across environmental gradients. The goal is to uncover genetic
59 variants whose geographic distribution associates with an environmental gradient, as they
60 may underpin an adaptive role (10).

61 Despite the promises, GEA studies often suffer from false discoveries, which arise
62 when the geographic distribution of genetic variants spuriously associates with an
63 environmental gradient (16). These false associations can be due to underlying population
64 structure, to local adaptation against unmeasured environmental conditions, or to chance
65 (16). One pioneering approach to filter out spurious associations relies on comparisons
66 across species (17). In this approach, parallel GEA analyses are performed on multiple
67 species against the same environmental variable. Orthologous genes showing a repeated
68 GEA in different species are prominent candidates for an adaptive role, because it is unlikely
69 for spurious associations to hit the same gene in different species by chance. This
70 cross-species GEA approach has shed light on candidate genes that might be repeatedly
71 involved in climate adaptation in distantly related plants—as distant as lodgepole pine and
72 *Arabidopsis* (18).

73 In this study, we investigated genetic variation associated with climate adaptation in
74 landraces and wild relatives of eight staple crops: barley, common bean, maize, rice,
75 sorghum, soybean, tepary bean, and tomato. We conducted a comparative GEA study
76 across these crops to identify genes with variants repeatedly linked to climate adaptation.
77 For the most promising candidate genes, we further investigated putative molecular functions
78 and phenotypic associations in the model plant species *Arabidopsis thaliana*.

Figure 1. Genomic database of crop landraces and wild relatives. Shown is the geographic distribution of the individual samples included in the database for (A) crop landraces and (B) crop wild relatives. Point colors correspond to the different crop species shown on the right, with the respective sample size.



82

83 Results

84 A genetic diversity database for crop landraces and wild relatives

85 We built a genomic database to characterize geographic patterns of genetic variation in crop
 86 landraces and CWR worldwide. The database compiled publicly available datasets from 13
 87 published studies (19–31), covering 14 crop types: six species with both landrace and CWR
 88 datasets (barley, common bean, maize, rice, soybean, tepary bean), one with landrace-only
 89 data (sorghum), and one with wild relative-only data (tomato). The full database included
 90 10,391 precisely georeferenced plants, each with individual-level SNP data.

91 We harmonized the genomic database by applying consistent filters for missingness,
 92 allele frequencies, and linkage disequilibrium to all datasets (see Methods for details; Table
 93 S1; Figure S1–S14). After filtering, the median number of SNPs per dataset was 54,201
 94 (range: 4,346 to 583,642), and the median number of individuals was 274 (range: 78 to
 95 1,676). The median number of sampling sites (samples collected up to 10 km apart) per
 96 dataset was 112 (range: 35 to 1,307), with a median distance of 1,110 km between sites
 97 (range: 187 to 3,884 km; Figure 1). For each dataset, we visually confirmed that the
 98 observed patterns of genetic diversity aligned with those reported in the original publications
 99 (e.g., isolation by distance or subspecies presence; Figure S1–S14).

To facilitate downstream comparisons of genetic diversity across species, we identified groups of orthologous genes across datasets (*i.e.*, orthogroups), and retained only SNPs mapping within these genes. The median number of SNPs retained per dataset was 24,707 (range: 1,403 to 151,898; Table 1), corresponding to a median of 6,377 orthologous genes per dataset (range: 952 to 14,975; Table S2).

Table 1. Crop types in the genomic database. The table summarizes the 14 crop type datasets included in the database. For each crop type dataset, shown are the original dataset source, the species name, the number of individual samples (# Sam.), the number of SNPs characterized, the number of sampling sites (# Sam. sites), and their geographic extent in km (*i.e.*, median distance across sampling sites).

#	Crop types	Original dataset	Species name	# Sam.	# SNPs	# sam. sites	Geo. ext. [km]
1	Barley landraces	Chen et al. 2022 (19)	<i>Hordeum vulgare</i>	388	46,890	299	3087
2	Barley wild relatives			264	72,272	177	624
3	Common bean landraces	Wu et al. 2020 (28)	<i>Phaseolus vulgaris</i>	476	13,164	238	1370
4	Common bean wild relatives	López-Hernández and Cortés 2019 (21)	<i>Phaseolus vulgaris</i>	78	7,032	69	2159
5	Maize landraces	Aiyesa et al. 2025 (22)	<i>Zea mays</i>	332	34,784	35	845
6	Maize wild relatives	Chen et al. 2022 (20)	<i>Zea mays subsp. mexicana</i> <i>Zea mays subsp. parviglumis</i>	151	121,116	84	187
7	Rice landraces	Gutaker et al. 2020 (23) Wang et al. 2018 (24)	<i>Oryza sativa</i>	1272	14,631	503	2207
8	Rice wild relatives	Bedford et al. 2023 (25) Huang et al. 2012 (26)	<i>Oryza rufipogon</i>	284	91,715	129	1273
9	Sorghum landraces	Lasky et al. 2015 (27)	<i>Sorghum bicolor</i>	1676	8,449	1307	3884
10	Soybean landraces	Zhang et al. 2022 (29)	<i>Glycine max</i>	451	103,101	80	1141
11	Soybean wild relatives		<i>Glycine soja</i>	174	151,898	104	1080
12	Tepary bean landraces	Bornowski et al. 2023 (30)	<i>Phaseolus acutifolius</i> A. Gray	148	2,704	86	798
13	Tepary bean wild relatives			154	8,816	90	551
14	Tomato wild relatives	Gibson and Moyle 2020 (31)	<i>Solanum pimpinellifolium</i>	139	1,403	120	535

109

110 Repeated Genome-Environment Associations in different crops

To identify genes potentially involved in crop climate adaptation, we conducted 14 parallel GEA analyses, one for each crop type dataset. In each analysis, we assessed the association between genetic variation in orthogroups and 43 environmental gradients

(including bioclimatic conditions, soil types, topography, and proxies of human activity; Table S3; Figure S15–S16). Two GEA methods were employed: Latent Factor Mixed Model [LFMM (32); Figure S17, S18A] and Kendall Rank Correlation (τ ; Figures S17, S18B). Results from both methods were compared across datasets, aiming to identify orthogroups repeatedly associated with the same environmental variable in different crop types (Figure 2; Table S4).

We found 5 orthogroups with significant repetition of adaptive signals ($q < 0.05$), 2 identified under the LFMM analysis and 3 under the τ approach. When we considered marginally significant signals ($q < 0.2$), candidate adaptive orthogroups were 16 (9 from LFMM and 7 from the τ approach; Figure S19–S33). For these 16 orthogroups, the median number of crops involved in the repetition was 4 (range: 2–13). Repeated adaptive signals were most frequently observed in wild relatives of soybean (12 orthogroups), maize (9), rice (7), barley (7) and in landraces of barley (7) and soybean (9). In nearly all significant orthogroups (15/16), the repetition of adaptive signals was observed in at least one landrace–CWR pair of the same crop. Most significant orthogroups (13/16) were associated with a single environmental variable (Table S4), with the most common being Volumetric soil water (7 genes), Aridity (3 genes) and Mean temperature in the coldest quarter (2 genes).

Functional profiling of candidate adaptive orthogroups

We then wanted to investigate the potential function of the 16 orthogroups showing repeated adaptive signals in different crop types. Therefore, we focused on the well-characterized *A. thaliana* orthologs and retrieved their Gene Ontology [GO (33)] annotations (Table S5). Under an enrichment analysis focusing on these orthologs, we found 35 GO terms for biological processes as significantly overrepresented (FDR < 0.05; Table S6). These 35 GO terms spanned across six semantic groups of the biological processes ontology (Figure 3A). We also investigated whether these *A. thaliana* orthologs had been linked to phenotypic variation in previous genome-wide association studies (GWAS). Using the araGWAS database (34), we retrieved standardized GWAS results for 438 phenotypes, spanning 71 Plant Trait Ontology (PTO) terms [Table S5; (35)]. For each orthogroup, we identified the most strongly associated phenotype, which spanned across 8 PTO terms (Figure 3B).

For orthogroups associated with Volumetric Soil Water, *A. thaliana* orthologs were most frequently annotated with GO terms related to responses to water deprivation or salt stress (3/7 orthogroups). In previous GWAS, these orthologs were also repeatedly associated with root growth and morphology phenotypes (4/7 orthogroups). The top orthogroup under τ approach (OG0002438, $q = 0.003$; Figure 4) is a notable example where the *A. thaliana* orthologs feature both GO annotation for salt stress response and phenotypic association with root branching—although this association is not significant after multiple testing correction. Additional notable orthogroups associated with volumetric soil water are OG0006158 ($q = 0.109$; Figure S32) and OG0000277 ($q = 0.193$; Figure S27), as their *A. thaliana* orthologs are linked with GO terms on water deprivation response, and they also showed significant phenotypic associations with days to flowering and root morphology traits. Of note, OG0000277 is the orthogroup with the largest repetition of GEA signals across crop types—including all crops except tomato wild relatives.

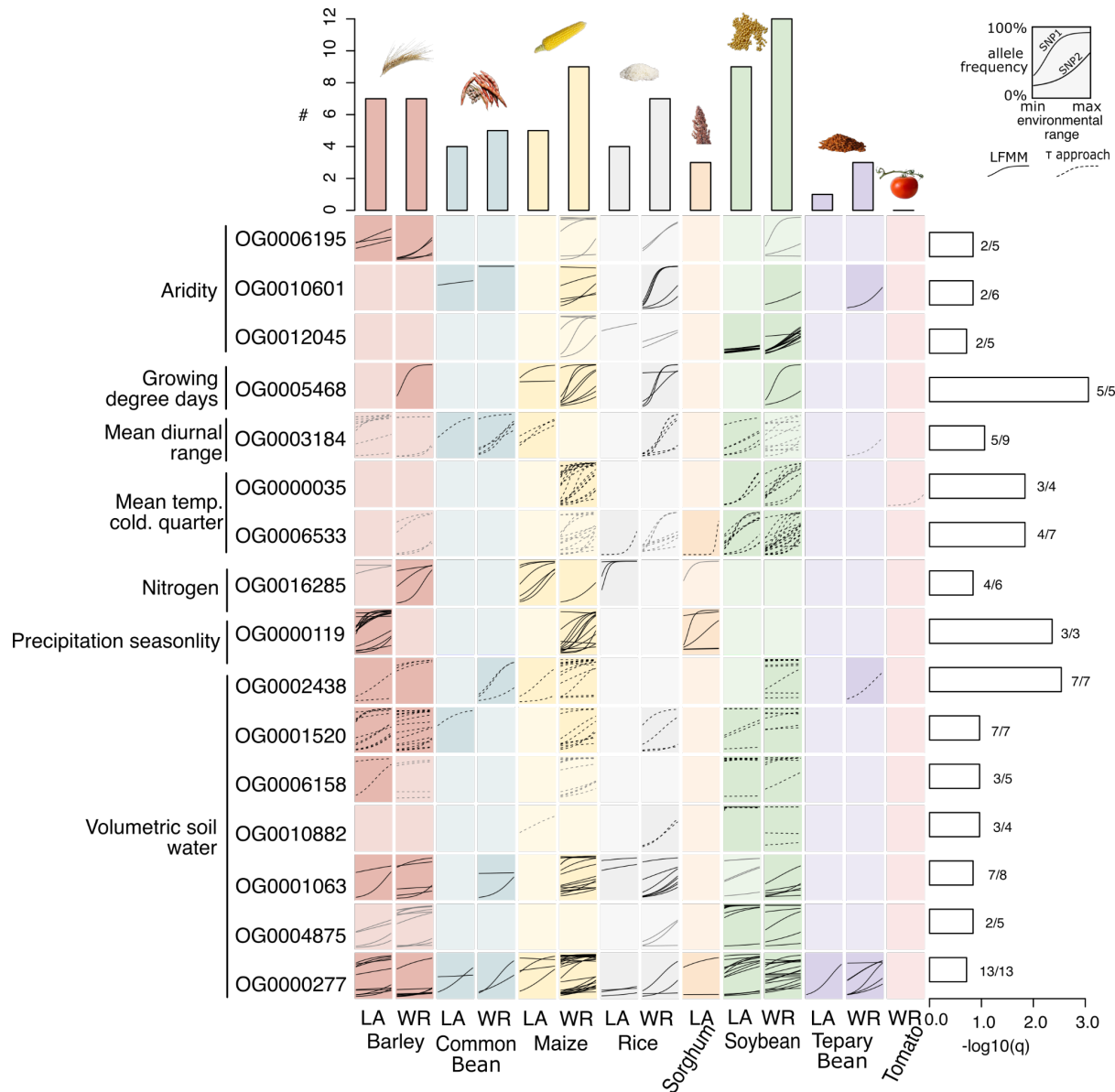
For orthogroups associated with Mean Temperature During the Coldest Quarter (OG0000035, $q = 0.014$, Figure S28; OG0006533, $q = 0.014$; Figure S29), *A. thaliana* orthologs were annotated with GO terms related to calcium homeostasis, transport, or signaling and were associated with phenotypic variation in days to flowering traits. The top orthogroup under LFMM was also associated with multiple temperature-related variables

(Table S4), with the best association being Growing Degree Days (OG0005468, $q = 0.001$; Figure S19). The orthogroup's *A. thaliana* ortholog is linked to GO terms for cold response, and was linked to days to flowering traits—although this association was not significant after multiple testing corrections.

For the remaining candidate orthogroups (associated with Aridity, Nitrogen Concentration, or Precipitation Seasonality), *A. thaliana* orthologs showed no particular GO term enrichment nor recurrent associations with specific phenotypic traits.

Figure 2. Orthogroups with repeated environmental associations across crop types. Shown are the orthogroups (rows) that showed a significant ($q < 0.05$) or marginally significant ($q < 0.2$) repetition of associations with the same environmental variable in different crop types (columns; LA = landraces, WR = wild relatives). Orthogroups are sorted by environmental variable (indicated on the left). In the central part of the figure, each plot shows the genome-environment association (GEA) between the allele frequency of each SNP in the orthogroup genes (y-axis) and the environmental variable (x-axis). Solid lines represent GEA from LFMM; dotted lines from the τ approach. Flat regression lines indicate a weak GEA, whereas lines with a steeper slope indicate a stronger GEA. Whitenized plots indicate GEA associations that were not involved in the repetition across crop types. The barplot on the right shows the significance of the GEA repetition across crop types [$-\log_{10}(q)$ from the rank analysis of Picmin]. Numbers next to the bars indicate the number of crop types involved in the repetition, out of the total number of crop types for which the orthogroup genes were genotyped. The barplot on top indicates the number of repeated signals that each crop type contributed to.

180
181



182

Figure 3. Functional annotation of top orthogroups. The figure summarizes the functional annotations of the *Arabidopsis thaliana* orthologs for each orthogroup with a repeated signal of environmental association in different crop types. Orthogroup identifiers are reported on the left, grouped by the environmental variable involved in the association. **(A)** Functional annotation based on Gene Ontology (GO) terms for Biological Processes (33). GO terms that were significantly overrepresented across the *A. thaliana* orthologs were grouped into six categories based on semantic similarity (bottom axis). Points in the grid indicate orthogroups whose *A. thaliana* orthologs are annotated with specific GO terms. The barplot on top shows the total count of a given annotation across all orthogroups. **(B)** Functional annotation based on phenotypic associations from previous genome-wide association studies (34). The phenotypes most strongly associated with each *A. thaliana* orthogroup were grouped according to the Plant Trait Ontology [bottom axis; (35)]. Points in the grid indicate which phenotype group is associated with which orthogroup. The barplot on top summarizes phenotype counts across orthogroups. Solid points and bars indicate phenotypes with a significant association after multiple testing corrections.

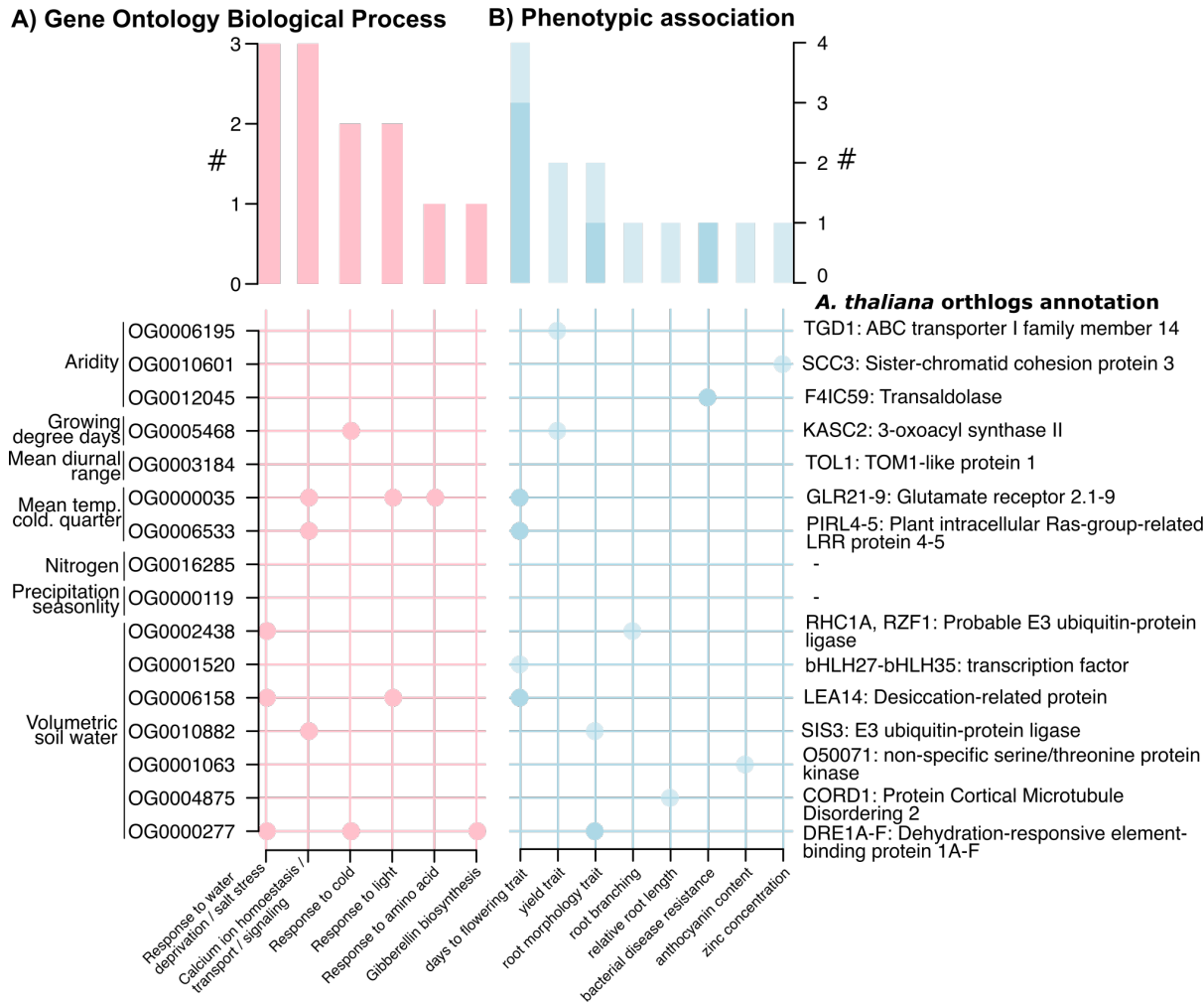
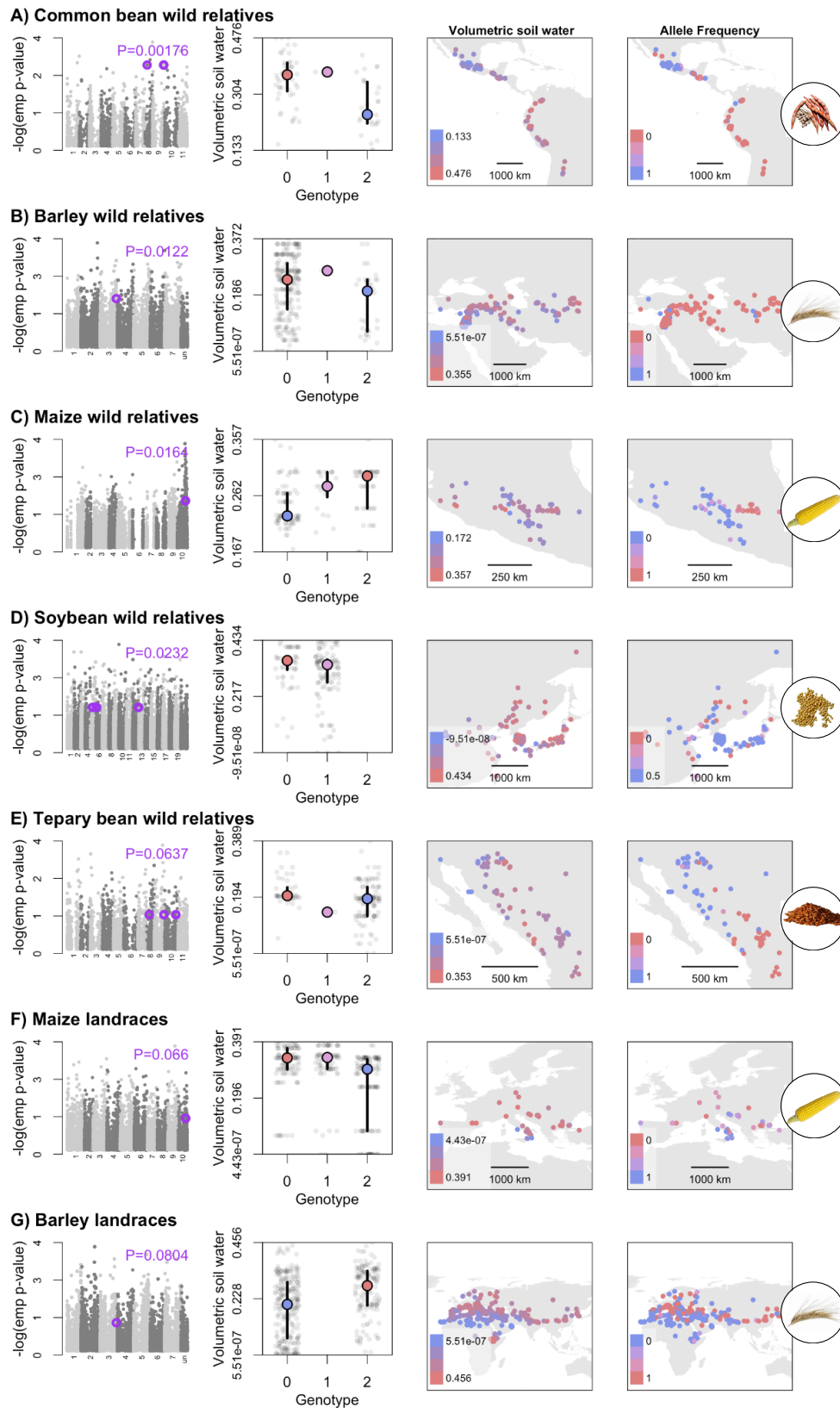


Figure 4. Repeated association with volumetric soil water in the OG0002438 orthogroup. Shown is the association with volumetric soil water in the OG0002438 orthologs across seven different crop types (A–G). For each crop type, shown are (from left to right): a Manhattan plot of the genome-wide associations with volumetric soil water, highlighting the position of the OG0002438 orthologs (in purple, with their operational p-value); a boxplot showing the genotype distribution across different levels of volumetric soil water for the top SNP underlying the orthogroup association (colored points represent median, bars interquartile range, points in the background are individual values); a map showing the geographic distribution of volumetric soil water in the study area; and a map showing the distribution of allele frequencies across the study area for the top SNP underlying the orthogroup association (for display purposes, the map points group samples located up to 50 km apart).



208 Discussion

209 Prominent candidate adaptive orthogroups

210 For the majority of the candidate adaptive orthogroups (14/16), we identified one or more
211 orthologs in the *A. thaliana* genome. In several cases, these orthologs had functional
212 annotations or phenotypic associations that aligned with the orthogroup's climate
213 association. While these functional annotations must be interpreted with caution (36), they
214 can provide additional support for the putative role of candidate adaptive orthogroups (10).
215 Below, we discuss five candidate adaptive orthogroups that appear particularly promising
216 when considering the strength of the repeated signal, the number of crop types involved, the
217 associated climate variable, the functional annotations, and previous evidence about the
218 molecular role.

- 219 - **OG0002438** was the top orthogroup under the τ approach and showed repeated
220 association with Volumetric Soil Water in seven crop types (landraces of barley and
221 maize, and wild relatives of barley, common bean, maize, soybean, and tepary bean).
222 This orthogroup has two *A. thaliana* orthologs that were associated with root
223 morphology changes under abscisic acid and salt stress treatment in previous GWAS
224 (37, 38)—although these associations were not significant after multiple testing
225 corrections. These orthologs encode E3 ubiquitin-protein ligases (RZF1 and RHC1A),
226 which are well known to be involved in responses to abiotic stress in many crops
227 (39). Indeed, a targeted mutagenesis experiment in *A. thaliana* suggested RZF1 to
228 be an important regulator of water deficit stress during the early seedling
229 development (40).
- 230 - **OG000277** showed an outstanding repetition of association with Volumetric soil
231 water, including all crop types except tomato wild relatives. Its *A. thaliana* orthologs
232 encode multiple versions of Dehydration-responsive element-binding protein 1
233 (DREB1A–F), a transcriptional activator that mediates dehydration-inducible
234 responses (41) and that had been associated with salt-induced changes in root
235 architecture in a previous GWAS (37).
- 236 - **OG0006158** was associated with Volumetric soil water in three crop types (landraces
237 of barley and soybean, and soybean wild relatives). Its *A. thaliana* orthologs encode
238 two Desiccation-related proteins (LEA14 and LEA2R), one of which was significantly
239 associated with flowering time in simulated seasons in a previous GWAS (42).
240 Proteins from this family, called LEA (Late Embryogenesis Abundant), have been
241 proposed to mitigate the intracellular damage due to water deprivation in multiple
242 crops (43–46).
- 243 - **OG0005468** is the orthogroup showing the strongest repetition of adaptive signals
244 under the LFMM approach. The orthogroup was associated with Growing Degree
245 Days in five crop types (maize landraces and wild relatives of maize, barley, rice, and
246 soybean) and was also significantly associated with five other temperature-related
247 variables (Table S4). *A. thaliana* has one ortholog in this orthogroup, encoding
248 3-oxoacyl synthase II (KASC2), a protein shown in a mutagenesis experiment to
249 confer cold resistance by maintaining chloroplast membrane integrity (47).
- 250 - **OG0000035** was associated with Mean Temperature in the Coldest Quarter in three
251 crop types (soybean landraces and wild relatives of soybean and maize). *A. thaliana*
252 has multiple orthologs in this group coding for multiple types of Glutamate receptor 2
253 (GLR2.1–9), with one of these having been significantly associated with flowering
254 time in simulated seasons in a previous GWAS (48). Other *A. thaliana* Glutamate

receptors (*GLR1.2–3*) have been suggested to promote cold tolerance by regulating jasmonate signaling in a mutagenesis experiment (49).

These five orthogroups represent prominent candidates for an adaptive role in crops. For most of these orthogroups, *A. thaliana* annotations already suggested that they are important contributors to responses to either water deprivation or cold stress. Our results suggest that genetic variation exists in these genes across diverse wild and semi-domesticated crops, and that this variation might extend the tolerance range against these climatic stressors. Future studies should harness genetic variation in these orthogroups across different crop types and rigorously test whether specific genotypes provide superior environmental tolerance in controlled trials (9, 10).

Cross-species GEA to find candidate adaptive genes in crops

Genome-environment association (GEA) analyses are a promising tool to uncover adaptive genetic variation in crops (10). However, when focusing on a single species, GEA analyses often struggle to balance sensitivity with specificity (16). Our investigation confirms this challenge: while we applied identical GEA workflows to 14 crop datasets, the number of significant genes identified varied considerably across workflows (LFMM vs. the τ approach), crop types, and environmental variables (Figure S19). The only common pattern that emerged is that GEA on crop wild relatives (CWR) tended to identify a larger fraction of climate-associated genes compared to landraces (Figure S18). This result might reflect the broader adaptive potential of wild varieties to natural environments compared to semi-domesticated ones (2).

The specificity challenge of single-species GEA can be partially addressed via cross-species comparisons (17). Our search for repeated GEA signals across different crop types allowed us to filter out spurious associations, likely reducing false discoveries. In the robust subset of retained GEA signals, *Volumetric soil water* and *Aridity* emerged as the most prominent environmental variables, together associated with 10 of the 16 candidate adaptive orthogroups. These variables are proxies for water availability in the soil, likely representing a major selective barrier in wild crops, but also in landraces as they are typically not reliant on artificial irrigation (3).

Cross-species comparisons can also increase sensitivity in single-species GEA (17). Our results showed that candidate adaptive orthogroups identified across crop types rarely included genes underlying the strongest GEA signals in individual datasets. Instead, these orthogroups often included genes that would fall below the significance threshold in single-species GEA. This finding aligns with recent simulations indicating that adaptive genotypes are not always those forming the strongest association with a climate gradient—particularly when polygenicity and pleiotropy are involved (50).

The first application of cross-species GEA analyses demonstrated the potential of this approach, enabling the identification of candidate adaptive genes across distantly related plants, from lodgepole pine to *Arabidopsis* (18). In comparison, our analysis focused on more closely related species—all annuals, and in some cases, landraces and wild relatives of the same species. The narrower taxonomic range likely explains why we obtained stronger signals compared to the first study (roughly twice as many orthogroups with $\text{FDR} < 0.2$, using about half the datasets). This comparison suggests that the power of cross-species GEA can be enhanced by focusing on closely related taxa.

303 Limitations and perspectives

304 This study is the first attempt to compare repeated GEA signals across different crops to
305 identify genes potentially underlying climate adaptation. We successfully identified several
306 such genes, despite leveraging diverse public datasets that were not optimized for GEA
307 analyses. Future studies should focus on extending this approach to additional staple crops,
308 for example by leveraging seed bank collections, for which appropriate population genomics
309 datasets are not yet available. Such new studies could substantially enhance the power of
310 this approach by addressing important methodological aspects that we could not tackle here.

311 Sampling strategy is crucial to ensure statistical power in GEA studies (16), and
312 some of our datasets were likely suboptimal in this regard. Simulation-based guidelines
313 recommend using overall sample sizes of at least 200 individuals and selecting sampling
314 sites that maximize the environmental contrasts of interest (51). In practice, this could be
315 achieved by accessing geographic metadata from seed banks and selecting seeds that best
316 cover the gradients of interest.

317 Genotyping strategy is another important element to consider. Our results clearly
318 indicated that datasets with a larger number of SNPs were more likely to be involved in
319 repeated GEA signals. Since our analysis is based on orthogroup comparisons, whole
320 exome sequencing [as used in the barley datasets (19)] would be a more cost-effective
321 approach than whole genome sequencing. However, future studies should also consider
322 whether sequence orthology can be found outside coding regions (e.g., in regulatory
323 sequences), as this would open the search for adaptive loci in unexplored genomic regions.
324 Furthermore, our analysis relied solely on SNPs, but other types of genetic variants exist
325 (e.g., indels, deletions, inversions, translocations, or duplications) and are more likely to
326 underlie major phenotypic changes (52). Emerging genotyping techniques such as k-mers
327 analysis can characterize such variants, and would likely provide a more comprehensive
328 view of adaptive genetic variation (53).

329 The final consideration concerns the methods used to run the GEA analysis. Our
330 approach treated orthogroups as independent from each other, largely overlooking the
331 polygenic architectures underlying climate adaptation. Future studies should explore new
332 avenues to test the repeatability of GEA not only by gene orthology but also by gene
333 function. Focusing on repeatability by function rather than by orthology would allow
334 researchers to consider adaptive loci that evolved in single crop types. Such loci likely
335 remain a substantial fraction of all adaptive loci in a crop genome (54), and were missed in
336 our study.

337

338

339 **Methods**

340 **Genomic datasets selection**

341 The genomic data used in this study were retrieved from published datasets. In June 2025,
342 we ran a manual search in the peer-reviewed literature (Web of Science and Google
343 Scholar) for articles matching the keywords "crop landraces", "crop wild relatives",
344 "population genomics", "genotype-environment associations", or "genome-wide association
345 study". We then manually screened these articles and retained those that:

- 346 1) Included genotypic data for at least 50 individual samples of a crop landrace and/or a
347 wild relative;
- 348 2) Provided genotypic data characterized as genome-wide single nucleotide
349 polymorphisms (SNPs; *e.g.*, from whole genome sequencing, restriction
350 site-associated DNA sequencing, genotype-by-sequencing, exome sequencing, or
351 transcriptome sequencing);
- 352 3) Made genotypic data publicly available as a genotype matrix (*e.g.*, variant calling
353 format or similar, available in supplementary materials with no access on request);
- 354 4) Included precise georeferencing (*i.e.*, coordinates) of genotyped individuals, publicly
355 available (*e.g.*, in supplementary materials with no access on request);
- 356 5) Used a genome reference that was publicly available in online repositories (such as
357 NCBI or Phytozome), including chromosome sequences (*i.e.*, nucleotide fasta format)
358 annotations of gene products (*e.g.*, gene feature or gene transfer format) and protein
359 sequences (*i.e.*, amino acids fasta format).

360 After applying these filtering criteria, 11 datasets were retained (Table 1): three included both
361 landrace and wild relative samples [barley (19), soybean (29), and tepary bean (30)], four
362 included landrace-only samples [common bean (28), corn (22), rice (23, 24), and sorghum
363 (27)], and four included wild relative-only samples (common bean (21), corn (20), rice (25,
364 26), and tomato (31)).

365

366 **Preliminary filtering of genotype matrices**

367 For each genomic dataset, we downloaded the genotype matrix and converted it to standard
368 variant calling format (VCF) using bcftools [v. 1.16; (55)], PLINK [v. 1.90; (56)], or TASSEL
369 [v. 5.2.89; (57)], depending on the original format. We then mapped the position of every
370 SNP in the genotype matrix to genes in the reference genome using a custom script in R
371 [v. 4.3.2; available in the online material; (58)]. We applied multiple filters to the genotype
372 matrix of each dataset using bcftools. Specifically, we retained only bi-allelic SNPs located
373 within coding sequences. We also applied soft filters to exclude rare variants (minor allele
374 frequency < 1%) and SNPs with excessive missing calls (present in > 50% of individuals).
375 Finally, we applied a linkage disequilibrium filter, discarding SNPs with an r^2 coefficient > 0.9
376 within a 1,000-nucleotide window. This filtering step aimed to remove nearly identical SNPs,
377 ensuring datasets with varying SNP densities (*e.g.*, whole genome sequencing vs. restriction
378 site-associated DNA sequencing) could be processed downstream using the same analytical
379 pipeline.

380

381 **Population structure analysis of crop varieties**

382 Genotype-environment association (GEA) analyses can lose statistical power when datasets
383 include highly divergent populations (*e.g.*, distinct lineages) (16). To anticipate this issue,
384 datasets including both landrace and wild relative varieties (barley, soybean, and tepary

bean) were divided into two derived datasets (Table 1). These splits resulted in 14 derived datasets (seven for landraces and seven for wild relatives, Table 1).

We processed the genotype matrix of each derived dataset independently, retaining only individuals with available georeferencing for their sampling sites. Using the SNPrelate R package [v. 1.36.1; (59)], we applied standard GEA preparatory filters for minor allele frequency ($> 5\%$) and missing rate ($< 10\%$ per SNP and per sample). We then performed a principal coordinate analysis (PCoA) on the filtered genotype matrix to verify that the genetic structure qualitatively aligned with the geographic distribution of samples (Figure S1-S14).

Finally, we imputed missing genotype calls in each matrix using the LEA R package [v. 3.12.2; (60)]. The optimal number of ancestral populations (K) was estimated via the cross-entropy criterion, running the snmf function iteratively ($K = 1$ to 20, five replicates, 1,000 random SNPs; Figure S34). The optimal K was determined when increasing to $K+1$ did not significantly reduce cross-entropy (Welch t-test, $P < 0.01$). Missing genotypes were imputed using the impute function, accounting for shared ancestry under the optimal K . Median imputation accuracy was 0.84 (IQR = 0.07), assessed on 1,000 randomly masked genotypes before imputation (Table S1).

Environmental data

The environmental characterization of the sampling sites included 43 variables, derived from four public datasets (Table S3):

- The "Global bioclimatic indicators from 1979 to 2018 derived from reanalysis" dataset [Copernicus Climate Data Store, v. 1; (61)]. This dataset provides 19 bioclimatic variables at 0.5 decimal degree (~ 50 km) resolution, based on the ERA5 climate model. We used 1979–2018 long-term averages, retaining variables that describe yearly and seasonal temperature and precipitation regimes [calculated as those in Bioclim; (11, 61)], as well as variables describing aridity, cloud cover, evaporative fraction, growing degree days, potential evaporation, volumetric soil water, annual summer days, annual dry days, maximal dry spell length, maximal dry spell intensity, and annual frost days.
- The "SoilGrids" dataset, a global predictive model of soil chemical and physical properties (13). We extracted 1 km aggregate variables (January 2023 release) for the 0–5 cm depth interval, including bulk density, cation exchange capacity (pH 7), coarse fragments, clay content, nitrogen concentration, organic carbon density, soil organic carbon stock, pH (water), and sand and silt fractions.
- The "Consensus Land Cover" product [EarthEnv, v. 1; (15)], offering 1 km global satellite-derived land cover data for 12 classes. We focused on two human activity-related classes—"Cultivated and Managed Vegetation" and "Urban/Built-up"—and calculated their average prevalence on a 1 decimal degree (~ 100 km) global grid.
- The "Global Multiresolution Topography" dataset [v. 4.4; (12)], providing worldwide topographic characterization at ~ 0.03 degree (~ 3 km) resolution.

All 43 environmental variables were prepared using the terra R package [v. 1.8-54; (62)], which was also used to extract the point values for each sampling site.

Since the ten SoilGrids variables exhibited substantial missing data (median = 0.15, IQR = 0.15; Table S7), we imputed missing values using a Random Forest model trained on longitude, latitude, and the remaining 33 environmental variables. Imputation was performed with the missForest R package [v. 1.5; (63)]. To assess imputation accuracy, we masked known SoilGrids values (replicating each dataset's missingness proportion), ran the

433 imputation, and evaluated the correlation between real and imputed values (Pearson R).
434 Imputed values showed high correlation with real values in all datasets (median R = 0.90
435 [IQR=0.07]; Figure S35–S36).

436

437 **Genotype-environment associations in orthologous genes**

438 For each of the 14 datasets, we characterized genotype-environment associations (GEA) for
439 the same environmental conditions across orthologous genes. Orthologous genes were
440 identified using OrthoFinder [v. 3.1.0; (64)] on protein sequences from each reference
441 genome. To ensure high-quality gene function annotation (see the “Functional annotation of
442 candidate adaptive genes” section), we also included protein sequences from the *A. thaliana*
443 reference genome [Ensembl Plants, TAIR10 release; (65)] in the orthogroup classification.

444 We focused solely on SNPs located in orthologous genes and evaluated their
445 association with environmental variation. GEA analyses were conducted separately for each
446 dataset using two complementary approaches:

- 447 - Latent Factor Mixed Model (LFMM): LFMM assesses the linear association between
448 each SNP's genotypic variation and environmental variation, while accounting for
449 shared ancestry (as characterized in the "Population structure analysis of crop
450 varieties" section). For each of the 43 environmental variables, we calculated the
451 association of every SNP with the focal variable and adjusted the resulting P-values
452 for the Genomic Inflation Factor using the LFMM R package [v. 1.1; (32)].
- 453 - Kendall Rank Correlation Coefficients (τ): Unlike LFMM, this method does not
454 account for population structure but can detect non-linear genotype-environment
455 associations (18). For each of the 43 environmental variables, τ values from SNP
456 associations were converted to absolute values, ranked, and transformed into
457 empirical P-values.

458 Finally, we merged GEA results by orthologous gene using the Weighted Z-Score Approach
459 [WZA; December 2023 release; (66)]. WZA combines association signals from multiple
460 SNPs within a gene by weighting P-values according to allele frequencies (*i.e.*, prioritizing
461 more heterozygous SNPs). WZA was applied separately to LFMM and τ P-values for each
462 dataset.

463

464 **Cross-datasets comparison**

465 We compared genotype-environment associations (GEA) in orthologous genes across
466 species using PicMin [v. 0.0.0.9; (17)]. PicMin ranks orthologous genes in each dataset by
467 GEA significance, and then assesses whether rank distributions across datasets deviate
468 from random expectations. For a given environmental variable, orthologous genes
469 consistently ranked highly (*i.e.*, strong GEAs) across datasets are strong candidates for an
470 adaptive role. We applied PicMin to each of the 43 environmental variables, searching for
471 adaptive signals in orthologous genes present in at least 3 of the 14 datasets. Two PicMin
472 analyses were performed: one using LFMM-derived P-values and another using τ -derived
473 P-values.

474 The significance of the repeated adaptive signals was assessed using the PicMin
475 rank analysis P-value, adjusted into a Q-value (q) that corrects for 1) false discovery rate
476 [FDR; (67)] and 2) the uneven distribution of orthologous genes across datasets. We defined
477 candidate adaptive orthogroups as those with a statistically significant repetition of adaptive
478 signals ($q < 0.05$), but also reported marginally significant signals ($q < 0.2$). Although the 0.2
479 threshold introduces more uncertainty, it allowed us to explore a broader set of candidate
480 adaptive orthogroups and compare their putative functions. This approach aligns with the

first application of PicMin to a taxonomically diverse set of plant species (from trees to annual plants), where a permissive threshold of $q = 0.5$ was used (18).

Functional annotation of candidate adaptive genes

For every group of candidate adaptive orthogroups, we further characterized the putative molecular role of their *A. thaliana* orthologs [annotation of the TAIR10 release, v 3.1; (65). We used *A. thaliana* as a reference due to its high-quality DNA and protein sequence annotations, and the extensive availability of genome-wide association studies (GWAS) linking SNP variation to diverse phenotypes. Each ortholog was annotated using three resources:

- 1) Gene ontology annotations: We retrieved manually reviewed protein annotations from UniProt [accessed July 2026; (68)] and extracted protein names and Gene Ontology (GO) terms for Biological Process, Molecular Function, and Cellular Component (33).
- 2) SNP classification: We accessed the *A. thaliana* 1001 Genome Project database (69) and retrieved annotations describing the predicted effect of every SNP in the gene sequence [TAIR10 assembly, v. 3.1; (65)].
- 3) Phenotypic associations: Using the AraGWAS Catalog [accessed in July 2026; (34)], we retrieved standardized association scores ($-\log_{10}P$) for every SNP in the ortholog. AraGWAS included standardized results from 28 GWAS studies across 438 phenotypes, storing associations with scores ≥ 4 (i.e., $P < 0.0001$) and flagging those significant after Bonferroni correction for multiple testing.

To summarize Gene Ontology (GO) annotations across candidate adaptive genes, we performed a gene set enrichment analysis using Panther [release 19.0; (70)]. This analysis identified GO terms significantly overrepresented (FDR-corrected $P < 0.05$) among candidate genes, compared to all annotated genes in the *A. thaliana* reference genome. We excluded enriched terms that referred to high level biological processes (e.g., “localization”), using the exclusion list for high level terms provided by the Gene Ontology Consortium [accessed in September 2026; (71)]. To facilitate the visualization of the enrichment analysis, we manually grouped the remaining enriched terms by considering the ontology hierarchical structure and semantic similarity (for ex., GO terms *response to cold* and *cold acclimation* were grouped into a “Response to cold” category), and calculated the average fold enrichment in every group.

515 Code and Data availability statement

516 Code and intermediary data used in the analysis are publicly available on Github
517 (<https://github.com/Oselmoni/CropAdapt>).

518

519

520 Acknowledgment

521 We are grateful to the openness of many researchers who made crop genomic data publicly
522 available and this research possible. We furthermore thank the Spatial Genetics of
523 Ecosystem laboratory for the helpful discussions.

524

525 Funding

526 This research was supported by a Spark Grant from the Swiss National Science Foundation (Grant
527 no. 229006).

528

529 Competing interests

530 The authors declare no competing financial interests.

531

532

533 References

- 534 1. M. Parry, C. Rosenzweig, M. Livermore, Climate change, global food supply and risk of hunger. *Philos. Trans. R. Soc.*
535 *Lond. B Biol. Sci.* **360**, 2125–2138 (2005).
- 536 2. C. K. Khoury, S. Brush, D. E. Costich, H. A. Curry, S. de Haan, J. M. M. Engels, L. Guarino, S. Hoban, K. L. Mercer, A. J.
537 Miller, G. P. Nabhan, H. R. Perales, C. Richards, C. Riggins, I. Thormann, Crop genetic erosion: understanding and
538 responding to loss of crop diversity. *New Phytol.* **233**, 84–118 (2022).
- 539 3. I. Elouafi, Why biodiversity matters in agriculture and food systems. *Science*, doi: 10.1126/science.ads8197 (2024).
- 540 4. Y. Yang, D. Tilman, Z. Jin, P. Smith, C. B. Barrett, Y.-G. Zhu, J. Burney, P. D'Odorico, P. Fantke, J. Fargione, J. C. Finlay,
541 M. C. Rulli, L. Sloat, K. Jan van Groenigen, P. C. West, L. Ziska, A. M. Michalak, Clim-Ag Team, D. B. Lobell, M. Clark, J.
542 Colquhoun, T. Garg, K. A. Garrett, C. Geels, R. R. Hernandez, M. Herrero, W. D. Hutchison, M. Jain, J. M. Jungers, B.
543 Liu, N. D. Mueller, A. Ortiz-Bobea, J. Schewe, J. Song, J. Verheyen, P. Vitousek, Y. Wada, L. Xia, X. Zhang, M. Zhuang,
544 Climate change exacerbates the environmental impacts of agriculture. *Science* **385**, eadn3747 (2024).
- 545 5. A. Hultgren, T. Carleton, M. Delgado, D. R. Gergel, M. Greenstone, T. Houser, S. Hsiang, A. Jina, R. E. Kopp, S. B.
546 Malevich, K. E. McCusker, T. Mayer, I. Nath, J. Rising, A. Rode, J. Yuan, Impacts of climate change on global agriculture
547 accounting for adaptation. *Nature* **642**, 644–652 (2025).
- 548 6. L. Gao, M. B. Kantar, D. Moxley, D. Ortiz-Barrientos, L. H. Rieseberg, Crop adaptation to climate change: An evolutionary
549 perspective. *Mol. Plant* **16**, 1518–1546 (2023).
- 550 7. S. Yeaman, Can wild plant adaptations help crops tolerate heat? *Science* **388**, 1148–1150 (2025).
- 551 8. A. Scheben, Y. Yuan, D. Edwards, Advances in genomics for adapting crops to climate change. *Current Plant Biology* **6**,
552 2–10 (2016).
- 553 9. Q. Campbell, J. A. Bedford, Y. Yu, A. Halpin-McCormick, N. Castaneda-Alvarez, B. Runck, J. Neyhart, P. Ewing, D.
554 Ortiz-Barrientos, L. Gao, D. Wang, M. A. Chapman, L. H. Rieseberg, M. B. Kantar, Agricultural landscape genomics to
555 increase crop resilience. *Plant Commun.* **6**, 101260 (2025).
- 556 10. J. R. Lasky, E. B. Josephs, G. P. Morris, Genotype-environment associations to reveal the molecular basis of
557 environmental adaptation. *Plant Cell* **35**, 125–138 (2023).
- 558 11. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE
559 SURFACES FOR GLOBAL LAND AREAS. *Int. J. Climatol.* **37**, 4302–4315 (2017).
- 560 12. W. B. F. Ryan, S. M. Carbotte, J. O. Coplan, S. O'Hara, A. Melkonian, R. Arko, R. A. Weissel, V. Ferrini, A. Goodwillie, F.
561 Nitsche, J. Bonczkowski, R. Zemsky, Global multi-resolution topography synthesis. *Geochem. Geophys. Geosyst.* **10**
562 (2009).
- 563 13. L. Poggio, L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro, D. Rossiter, SoilGrids 2.0: producing
564 soil information for the globe with quantified spatial uncertainty. *SOIL* **7**, 217–240 (2021).
- 565 14. M. E. Turek, L. Poggio, N. H. Batjes, R. A. Armindo, Q. de Jong van Lier, L. de Sousa, G. B. M. Heuvelink, Global
566 mapping of volumetric water retention at 100, 330 and 15 000 cm suction using the WoSIS database. *Int. Soil Water*
567 *Conserv. Res.* **11**, 225–239 (2023).

15. M.-N. Tuanmu, W. Jetz, A global 1-km consensus land-cover product for biodiversity and ecosystem modelling: Consensus land cover. *Glob. Ecol. Biogeogr.* **23**, 1031–1045 (2014).
16. C. Rellstab, F. Gugerli, A. J. Eckert, A. M. Hancock, R. Holderegger, A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* **24**, 4348–4370 (2015).
17. T. R. Booker, S. Yeaman, M. C. Whitlock, Using genome scans to identify genes used repeatedly for adaptation. *Evolution* **77**, 801–811 (2023).
18. J. R. Whiting, T. R. Booker, C. Rougeux, B. M. Lind, P. Singh, M. Lu, K. Huang, M. C. Whitlock, S. N. Aitken, R. L. Andrew, J. O. Borevitz, J. J. Bruhl, T. L. Collins, M. C. Fischer, K. A. Hodgins, J. A. Holliday, P. K. Ingvarsson, J. K. Janes, M. Khandaker, D. Koenig, J. M. Kreiner, A. Kremer, M. Lascoux, T. Leroy, P. Milesi, K. D. Murray, T. Pyhäjärvi, C. Rellstab, L. H. Rieseberg, F. Roux, J. R. Stinchcombe, I. R. H. Telford, M. Todesco, J. S. Tyrmi, B. Wang, D. Weigel, Y. Willi, S. I. Wright, L. Zhou, S. Yeaman, The genetic architecture of repeated local adaptation to climate in distantly related plants. *Nat. Ecol. Evol.*, 1–15 (2024).
19. Y.-Y. Chen, M. Schreiber, M. M. Bayer, I. K. Dawson, P. E. Hedley, L. Lei, A. Akhunova, C. Liu, K. P. Smith, J. C. Fay, G. J. Muehlbauer, B. J. Steffenson, P. L. Morrell, R. Waugh, J. R. Russell, The evolutionary patterns of barley pericentromeric chromosome regions, as shaped by linkage disequilibrium and domestication. *Plant J.* **111**, 1580–1594 (2022).
20. L. Chen, J. Luo, M. Jin, N. Yang, X. Liu, Y. Peng, W. Li, A. Phillips, B. Cameron, J. S. Bernal, R. Rellán-Álvarez, R. J. H. Sawers, Q. Liu, Y. Yin, X. Ye, J. Yan, Q. Zhang, X. Zhang, S. Wu, S. Gui, W. Wei, Y. Wang, Y. Luo, C. Jiang, M. Deng, M. Jin, L. Jian, Y. Yu, M. E. B. Zhang, X. Yang, M. B. Hufford, A. R. Fernie, M. L. Warburton, J. Ross-Ibarra, J. Yan, Genome sequencing reveals evidence of adaptive variation in the genus *Zea*. *Nat. Genet.* **54**, 1736–1745 (2022).
21. F. López-Hernández, A. J. Cortés, Last-generation genome-environment associations reveal the genetic basis of heat tolerance in common bean (*Phaseolus vulgaris* L.). *Front. Genet.* **10**, 954 (2019).
22. L. V. Aiyesa, D. Kaufmann, B. Zumbach, W. Link, S. Scholten, T. Beissinger, Individual plant genetics reveal the control of local adaptation in European maize landraces. *BMC Biol.* **23**, 138 (2025).
23. R. M. Gutaker, S. C. Groen, E. S. Bellis, J. Y. Choi, I. S. Pires, R. K. Bocinsky, E. R. Slayton, O. Wilkins, C. C. Castillo, S. Negrão, M. M. Oliveira, D. Q. Fuller, J. A. D. Guedes, J. R. Lasky, M. D. Purugganan, Genomic history and ecology of the geographic spread of rice. *Nat. Plants* **6**, 492–502 (2020).
24. W. Wang, R. Mauleon, Z. Hu, D. Chebotarov, S. Tai, Z. Wu, M. Li, T. Zheng, R. R. Fuentes, F. Zhang, L. Mansueto, D. Copetti, M. Sanciango, K. C. Palis, J. Xu, C. Sun, B. Fu, H. Zhang, Y. Gao, X. Zhao, F. Shen, X. Cui, H. Yu, Z. Li, M. Chen, J. Detras, Y. Zhou, X. Zhang, Y. Zhao, D. Kudrna, C. Wang, R. Li, B. Jia, J. Lu, X. He, Z. Dong, J. Xu, Y. Li, M. Wang, J. Shi, J. Li, D. Zhang, S. Lee, W. Hu, A. Poliakova, I. Dubchak, V. J. Ulat, F. N. Borja, J. R. Mendoza, J. Ali, J. Li, Q. Gao, Y. Niu, Z. Yue, M. E. B. Naredo, J. Talag, X. Wang, J. Li, X. Fang, Y. Yin, J.-C. Glaszmann, J. Zhang, J. Li, R. S. Hamilton, R. A. Wing, J. Ruan, G. Zhang, C. Wei, N. Alexandrov, K. L. McNally, Z. Li, H. Leung, Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49 (2018).
25. J. A. Bedford, M. Carine, M. A. Chapman, Detection of locally adapted genomic regions in wild rice (*Oryza rufipogon*) using environmental association analysis. *G3 (Bethesda)* **13**, jkad194 (2023).
26. X. Huang, N. Kurata, X. Wei, Z.-X. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, H. Lu, W. Li, Y. Guo, Y. Lu, C. Zhou, D. Fan, Q. Weng, C. Zhu, T. Huang, L. Zhang, Y. Wang, L. Feng, H. Furuumi, T. Kubo, T. Miyabayashi, X. Yuan, Q. Xu, G. Dong, Q. Zhan, C. Li, A. Fujiyama, A. Toyoda, T. Lu, Q. Feng, Q. Qian, J. Li, B. Han, A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
27. J. R. Lasky, H. D. Upadhyaya, P. Ramu, S. Deshpande, C. T. Hash, J. Bonnette, T. E. Juenger, K. Hyma, C. Acharya, S. E. Mitchell, E. S. Buckler, Z. Brenton, S. Kresovich, G. P. Morris, Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* **1**, e1400218 (2015).
28. J. Wu, L. Wang, J. Fu, J. Chen, S. Wei, S. Zhang, J. Zhang, Y. Tang, M. Chen, J. Zhu, L. Lei, Q. Geng, C. Liu, L. Wu, X. Li, X. Wang, Q. Wang, Z. Wang, S. Xing, H. Zhang, M. W. Blair, S. Wang, Resequencing of 683 common bean genotypes identifies yield component trait associations across a north-south cline. *Nat. Genet.* **52**, 118–125 (2020).
29. H. Zhang, H. Jiang, Z. Hu, Q. Song, Y.-Q. C. An, Development of a versatile resource for post-genomic research through consolidating and characterizing 1500 diverse wild and cultivated soybean genomes. *BMC Genomics* **23**, 250 (2022).
30. N. Bornowski, J. P. Hart, A. V. Palacios, B. Ogg, M. A. Brick, J. P. Hamilton, J. S. Beaver, C. R. Buell, T. Porch, Genetic variation in a tepary bean (*Phaseolus acutifolius* A. Gray) diversity panel reveals loci associated with biotic stress resistance. *Plant Genome* **16**, e20363 (2023).
31. M. J. S. Gibson, L. C. Moyle, Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Mol. Ecol.* **29**, 2204–2217 (2020).
32. K. Caye, B. Jumentier, J. Lepeule, O. François, LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Mol. Biol. Evol.* **36**, 852–860 (2019).
33. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, G. O. Consortium, Gene Ontology: Tool for The Unification of Biology. *Nat. Genet.* **25**, 25–29 (2000).
34. M. Togninalli, Ü. Seren, D. Meng, J. Fitz, M. Nordborg, D. Weigel, K. Borgwardt, A. Korte, D. G. Grimm, The AraGWAS Catalog: a curated and standardized Arabidopsis thaliana GWAS catalog. *Nucleic Acids Res* **46**, D1150–D1156 (2018).
35. L. Cooper, A. Meier, M.-A. Laporte, J. L. Elser, C. Mungall, B. T. Sinn, D. Cavaliere, S. Carbon, N. A. Dunn, B. Smith, B. Qu, J. Preece, E. Zhang, S. Todorovic, G. Gkoutos, J. H. Doonan, D. W. Stevenson, E. Arnaud, P. Jaiswal, The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res.* **46**, D1168–D1180 (2018).
36. P. Gaudet, C. Dessimoz, Gene Ontology: Pitfalls, Biases, and Remedies. *Methods Mol Biol* **1446**, 189–205 (2017).
37. M. M. Julkowska, I. T. Koevoets, S. Mol, H. Hoefsloot, R. Feron, M. A. Tester, J. J. B. Keurentjes, A. Korte, M. A. Haring, G.-J. de Boer, C. Testerink, Genetic components of root architecture remodeling in response to salt stress. *Plant Cell* **29**, 3198–3213 (2017).
38. D. Ristova, M. Giovannetti, K. Metesch, W. Busch, Natural genetic variation shapes root system responses to phytohormones in Arabidopsis. *Plant J.* **96**, 468–481 (2018).
39. Y. Su, G. L. N. Ngea, K. Wang, Y. Lu, E. A. Godana, M. Ackah, Q. Yang, H. Zhang, Deciphering the mechanism of E3 ubiquitin ligases in plant responses to abiotic and biotic stresses and perspectives on PROTACs for crop resistance. *Plant Biotechnol. J.* **22**, 2811–2843 (2024).
40. H.-W. Ju, J.-H. Min, M.-S. Chung, C. S. Kim, The *atrzf1* mutation of the novel RING-type E3 ubiquitin ligase increases proline contents and enhances drought tolerance in Arabidopsis. *Plant Sci.* **203–204**, 1–7 (2013).
41. Y. Sakuma, Q. Liu, J. G. Dubouzet, H. Abe, K. Shinozaki, K. Yamaguchi-Shinozaki, DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* **290**, 998–1009 (2002).

- 645 42. S. Atwell, Y. S. Huang, B. J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A. M. Tarone, T. T. Hu, R. Jiang,
646 N. W. Muliyati, X. Zhang, M. A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J. R. Ecker, N.
647 Faure, J. M. Kniskern, J. D. G. Jones, T. Michael, A. Nemri, F. Roux, D. E. Salt, C. Tang, M. Todesco, M. B. Traw, D.
648 Weigel, P. Marjoram, J. O. Borevitz, J. Bergelson, M. Nordborg, Genome-wide association study of 107 phenotypes in
649 *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
- 650 43. K. Goyal, L. J. Walton, A. Tunnacliffe, LEA proteins prevent protein aggregation due to water stress. *Biochem. J.* **388**,
651 151–157 (2005).
- 652 44. I. A. Arroyo-Mosso, H. N. Diaz-Ardila, A. Garcarrubio, U. G. V. S. S. Kumara, D. F. Rendón-Luna, T. B. Nava-Ramírez, T.
653 C. Boothby, J. L. Reyes, A. A. Covarrubias, A Group 6 LEA protein plays key roles in tolerance to water deficit, and in
654 maintaining the glassy state and longevity of seeds. *Plant Cell Environ.* **48**, 6874–6896 (2025).
- 655 45. M. Battaglia, A. A. Covarrubias, Late Embryogenesis Abundant (LEA) proteins in legumes. *Front. Plant Sci.* **4**, 190 (2013).
- 656 46. S. C. Hand, M. A. Menze, M. Toner, L. Boswell, D. Moore, LEA proteins during water stress: not just for plants anymore.
657 *Annu. Rev. Physiol.* **73**, 115–134 (2011).
- 658 47. A. S. Carlsson, S. T. LaBrie, A. J. Kinney, P. von Wettstein-Knowles, J. Browse, A KAS2 cDNA complements the
659 phenotypes of the *Arabidopsis fab1* mutant that differs in a single residue bordering the substrate binding pocket: A KAS2
660 cDNA Complements the *Arabidopsis fab1* Mutant. *Plant J.* **29**, 761–770 (2002).
- 661 48. Y. Li, Y. Huang, J. Bergelson, M. Nordborg, J. O. Borevitz, Association mapping of local climate-sensitive quantitative trait
662 loci in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21199–21204 (2010).
- 663 49. Y. Zheng, L. Luo, J. Wei, Q. Chen, Y. Yang, X. Hu, X. Kong, The glutamate receptors AtGLR1.2 and AtGLR1.3 increase
664 cold tolerance by regulating jasmonate signaling in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* **506**, 895–900
665 (2018).
- 666 50. K. E. Lotterhos, The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. *Proc. Natl. Acad. Sci.*
667 *U. S. A.* **120**, e2220313120 (2023).
- 668 51. O. Selmoni, E. Vajana, A. Guillaume, E. Rochat, S. Joost, Sampling strategy optimization to increase statistical power in
669 landscape genomics: A simulation-based approach. *Mol. Ecol. Resour.* **20**, 154–169 (2020).
- 670 52. A. J. Scott, C. Chiang, I. M. Hall, Structural variants are a major source of gene expression differences in humans and
671 often affect multiple nearby genes. *Genome Res.* **31**, 2249–2257 (2021).
- 672 53. Y. Voichek, D. Weigel, Identifying genetic variants underlying phenotypic variation in plants without complete genomes.
673 *Nat. Genet.* **52**, 534–540 (2020).
- 674 54. T. R. Booker, A. Brown, S. Yeaman, What fraction of the genomic basis of local adaptation are we missing?, *EcoEvoRxiv*
675 (2025). <https://ecoevovrxiv.org/repository/view/10403/>.
- 676 55. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project
677 Data Processing 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools.
678 *Bioinformatics* **25**, 2078–2079 (2009).
- 679 56. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker, M. J.
680 Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J.*
681 *Hum. Genet.* **81**, 559–575 (2007).
- 682 57. P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, E. S. Buckler, TASSEL: Software for association
683 mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
- 684 58. R Core Team, *R: A Language and Environment for Statistical Computing* (2022; <https://www.r-project.org/>).
- 685 59. X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir, A high-performance computing toolset for
686 relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
- 687 60. E. Frichot, O. François, LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**,
688 925–929 (2015).
- 689 61. H. Wouters, Global bioclimatic indicators from 1979 to 2018 derived from reanalysis, (2021);
690 <https://doi.org/10.24381/cds.bce175f0>.
- 691 62. R. J. Hijmans, *Terra: Spatial Data Analysis* (2025; <https://doi.org/10.32614/CRAN.package.terra>).
- 692 63. D. J. Stekhoven, P. Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*
693 **28**, 112–118 (2012).
- 694 64. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238
695 (2019).
- 696 65. T. Z. Berardini, L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, E. Huala, The *Arabidopsis* information resource:
697 Making and mining the “gold standard” annotated reference plant genome: Tair: Making and Mining the “Gold Standard”
698 Plant Genome. *Genesis* **53**, 474–485 (2015).
- 699 66. T. R. Booker, S. Yeaman, J. R. Whiting, M. C. Whitlock, The WZA: A window-based method for characterizing
700 genotype-environment associations. *Mol. Ecol. Resour.* **24**, e13768 (2024).
- 701 67. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R.*
702 *Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
- 703 68. A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, R. Apweiler, E. Alpi, R. Antunes, J. Arganiska, B. Bely, M. Bingley,
704 C. Bonilla, R. Britto, B. Bursteinas, G. Chavali, E. Cibrian-Uhalte, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, P. Gane,
705 L. G. Castro, P. Garmiri, E. Hatton-Ellis, R. Hieta, R. Huntley, D. Legge, W. Liu, J. Luo, A. Macdougall, P. Mutwora,
706 Nightingale, S. Orchard, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A.
707 Shypitsyna, E. Turner, V. Volynkin, T. Wardell, X. Watkins, H. Zellner, A. Cowley, L. Figueira, W. Li, H. McWilliam, R.
708 Lopez, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen,
709 P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E.
710 Coudert, B. Cuhe, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S.
711 Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemerrier, D.
712 Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nospikel, S. Paesano, I. Pedruzzi, S. Pilboud, M.
713 Pozzato, M. Pruess, C. Rivoire, B. Roecher, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M.
714 Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K.
715 Laiho, P. McGarvey, D. A. Natale, B. E. Suzek, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, M. S. Yerramalla, J. Zhang,
716 UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
- 717 69. C. Alonso-Blanco, J. Andrade, C. Becker, F. Bemm, J. Bergelson, K. M. Borgwardt, J. Cao, E. Chae, T. M. Dezwaan, W.
718 Ding, J. R. Ecker, M. Exposito-Alonso, A. Farlow, J. Fitz, X. Gan, D. G. Grimm, A. M. Hancock, S. R. Henz, S. Holm, M.
719 Horton, M. Jarsulic, R. A. Kerstetter, A. Korte, P. Korte, C. Lanz, C. R. Lee, D. Meng, T. P. Michael, R. Mott, N. W.
720 Muliyati, T. Nägele, M. Nagler, V. Nizhynska, M. Nordborg, P. Y. Novikova, F. X. Picó, A. Platzer, F. A. Rabanal, A.
721 Rodriguez, B. A. Rowan, P. A. Salomé, K. J. Schmid, R. J. Schmitz, Ü. Seren, F. G. Sperone, M. Sudkamp, H. Svoldal, M.

- 722 M. Tanzer, D. Todd, S. L. Volchenboum, C. Wang, G. Wang, X. Wang, W. Weckwerth, D. Weigel, X. Zhou, 1,135 genomes
723 reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
724 70. P. D. Thomas, D. Ebert, A. Muruganujan, T. Mushayahama, L.-P. Albou, H. Mi, PANTHER: Making genome-scale
725 phylogenetics accessible to all. *Protein Sci* **31**, 8–22 (2022).
726 71. Gene Ontology Consortium, The Gene Ontology knowledgebase in 2026. *Nucleic Acids Res.* **54**, D1779–D1792 (2026).
727