

Genome architecture and diverged selection shaping pattern of genomic differentiation in wild barley

Wenyang Zhang^{1,+} , Cong Tan^{2,+} , Haifei Hu^{2,†}, Rui Pan^{1,+} , Yuhui Xiao³, Kai Ouyang³, Gaofeng Zhou², Yong Jia² , Xiao-Qi Zhang⁴, Camilla Beate Hill², Penghao Wang⁴, Brett Chapman², Yong Han^{2,5} , Le Xu¹, Yanhao Xu¹, Tefera Angessa², Hao Luo², Sharon Westcott⁵, Darshan Sharma⁵, Eviatar Nevo⁶, Roberto A. Barrero⁷, Matthew I. Bellgard⁷, Tianhua He^{2,4,*}, Xiaohai Tian^{1,*} and Chengdao Li^{2,4,5,*} 

¹Hubei Collaborative Innovation Centre for Grain Industry, Yangtze University, Jingzhou, China

²Western Crop Genetics Alliance, Future Food Institute, Western Australian State Agricultural Biotechnology Centre, College of Science, Health, Engineering and Education, Murdoch University, Murdoch, Western Australia, Australia

³Grandomics Biotechnology Co., Ltd, Wuhan, China

⁴College of Science, Health, Engineering and Education, Murdoch University, Murdoch, Western Australia, Australia

⁵Department of Primary Industries and Regional Development, South Perth, Western Australia, Australia

⁶Institute of Evolution, University of Haifa, Haifa, Israel

⁷eResearch Office, Queensland University of Technology, Brisbane, Queensland, Australia

Received 20 May 2022;

revised 9 August 2022;

accepted 19 August 2022.

*Correspondence (Tel/Fax +61 8 93607519; email c.li@murdoch.edu.au (C.L.), Tel/Fax +86 716 8066541; email xiaohait@sina.com (X.T.), Tel +61 8 93606816; fax +61 8 93606303; email tianhua.he@murdoch.edu.au (T.H.))

†These authors contributed equally to this work.

Summary

Divergent selection of populations in contrasting environments leads to functional genomic divergence. However, the genomic architecture underlying heterogeneous genomic differentiation remains poorly understood. Here, we *de novo* assembled two high-quality wild barley (*Hordeum spontaneum* K. Koch) genomes and examined genomic differentiation and gene expression patterns under abiotic stress in two populations. These two populations had a shared ancestry and originated in close geographic proximity but experienced different selective pressures due to their contrasting micro-environments. We identified structural variants that may have played significant roles in affecting genes potentially associated with well-differentiated phenotypes such as flowering time and drought response between two wild barley genomes. Among them, a 29-bp insertion into the promoter region formed a cis-regulatory element in the *HvWRKY45* gene, which may contribute to enhanced tolerance to drought. A single SNP mutation in the promoter region may influence *HvCO5* expression and be putatively linked to local flowering time adaptation. We also revealed significant genomic differentiation between the two populations with ongoing gene flow. Our results indicate that SNPs and small SVs link to genetic differentiation at the gene level through local adaptation and are maintained through divergent selection. In contrast, large chromosome inversions may have shaped the heterogeneous pattern of genomic differentiation along the chromosomes by suppressing chromosome recombination and gene flow. Our research offers novel insights into the genomic basis underlying local adaptation and provides valuable resources for the genetic improvement of cultivated barley.

Keywords: chromosome inversion, cis-regulatory mutation, drought response, evolution canyon, flowering time, local principal component analysis.

Introduction

Genomic changes accumulate non-randomly across the genome with locally elevated genomic differentiation as populations diverge (Renaut *et al.*, 2013; Vijay *et al.*, 2016). Adapting to different environments may drive changes in genomic regions through divergent selection (Shah *et al.*, 2020; Zong *et al.*, 2021). Genomic regions with locally diverged sequences can be revealed in comparisons of closely related populations and species, and such genomic regions are often assumed to contain functional genetic variants conferring fitness in a local environment (Seehausen *et al.*, 2014). However, genomic regions with distinctive differentiation may arise from processes that differ from local adaptation (Van Doren *et al.*, 2017), and genetic differentiation can evolve through mechanisms that reduce local genetic diversity within the genomic regions (Burri, 2017; Charlesworth, 1998). Therefore, understanding the pattern of

genomic differentiation requires exploring multiple yet piecemeal processes underlying genomic divergence. Recent studies suggest that SV have a significant role in ecological adaptation and population divergence (Göktay *et al.*, 2021; Hu *et al.*, 2022; Huang *et al.*, 2020; Todesco *et al.*, 2020). But, despite its relevance in adaptation and evolution, the contribution of different types of SVs to local adaptation and genomic divergence remains to be explored. We also know little about the contribution of different types of SVs to local adaptation and genomic divergence in heterogeneous environments or how adaptive and neutral processes shape the evolution of different types of SVs.

The 'Evolution Canyon' (EC) in Mount Carmel, Israel, has been proposed as an optimal model to understand the effect of environmental heterogeneity on population differentiation and adaptation (Nevo, 1995, 2012). The EC has two geologically similar slopes separated by c. 100 m at the bottom of the canyon. These slopes present dramatic abiotic contrasts and biotic

divergence due to the significantly higher solar radiation (200%–800% more) on the south-facing slope (hereafter S_FS), compared to the north-facing slope (hereafter N_FS; Figure S1). The S_FS features higher daily temperatures and drought, whereas a cooler climate characterizes the N_FS, with higher relative humidity (1%–7%) than the S_FS despite geographic proximity (Nevo, 2012; Pavlíček *et al.*, 2003). The sharp microclimatic divergence between the two slopes has resulted in contrasting populations of animals, fungi, and plants with well-differentiated phenotypes, e.g., flowering time, drought tolerance (Nevo, 2012, 2014), and disease resistance (Wang *et al.*, 2020; Yin *et al.*, 2015) despite their close geographic proximity. The EC model offers an opportunity for direct investigation of the genomic divergence associated with local adaptation in the presence of gene flow, without the complications resulting from isolation-by-distance. Studies are emerging to explore the genomic underpinnings of local adaptation in the EC systems (Richardson *et al.*, 2014). However, the origin and build-up of heterogeneous genomic differentiation and the mechanisms driving and maintaining the differentiation remain to be elucidated.

The barley genomic study has focused mainly on cultivated barleys that usually have a narrow genetic diversity due to genetic bottlenecks from domestication (Jayakodi *et al.*, 2020; Mascher *et al.*, 2017, 2021). Wild barley (*Hordeum spontaneum* K. Koch), the ancestor of cultivated barley (*Hordeum vulgare* L.), has a wide eco-geographic distribution across highly diverse environments throughout the Near East and is common in the EC (Zohary *et al.*, 2012). Wild barley has been accessed in a wide range of agronomic traits, especially resistance and tolerance to biotic and abiotic stresses. It possesses tremendous potential for genetic improvement of cultivated barley (Nevo and Chen, 2010). Structural variants have been known to play a decisive role in barley breeding (Jayakodi *et al.*, 2020). Knowledge of wild barley genomes, particular the dynamics and architecture of their genome organization, can provide viable solutions to the future challenge in barley production.

Here we assemble two chromosome-scale wild barley genomes from contrasting environments in the EC system. We then sampled and re-sequenced multiple individuals from each environment to analyse the genomic differentiation at the population level. We finally investigated the pattern-differentiated gene expression of wild barley from contrasting environments under water stress. We deciphered the relative roles of adaptive and alternative processes shaping the landscape of genomic differentiation and shed light on how local selection interacts with genomic architecture in heterogeneous environments to drive genomic evolution.

Results

Long-read *de novo* assemblies and gene annotations of two wild barley accessions

We first sequenced and *de novo* assembled the genomes of two wild barley accessions, EC-S1 from the S_FS and EC-N1 from the N_FS, using Oxford Nanopore long-read sequencing technology. The contigs were anchored to the chromosome using Hi-C sequencing and Bionano optical mapping with conflicts resolved. With a total of 295.6 Gb (EC-S1) and 285.4 Gb (EC-N1), Nanopore GridION/PromethION raw reads, we assembled the primarily polished sequences for the two accessions with 5.03 Gb (EC-S1) and 5.05 Gb (EC-N1; Table 1, Data S1: Table 1). The final

polished EC-S1 and EC-N1 genomes have a contig N50 of 3.52 and 3.45 Mb, respectively, showing significant improvement in continuity compared to previously published wild barley genomes (Table 1). BioNano optical mapping and Hi-C sequencing anchored c. 90%, or 4.5 Gb, of assembled sequence into seven chromosomes for EC-S1 and EC-N1 (Table 1).

Using RNA-seq/PacBio IsoSeq transcript mapping combined with *ab initio* prediction and homologous protein searches, we recovered 39 179 and 38 737 high-confidence protein-coding genes in the EC-S1 and EC-N1 assemblies, respectively (Table 1; Data S1: Tables 2 and 3). Over 90% of the annotated genes have a functional annotation in both assemblies (Data S1: Table 4). Repeat sequences (including those centromere-specific) and transposable elements in the two assemblies have been characterized (Data S1: Tables 1, 5, 6). Benchmarking Universal Single-Copy Orthologs (BUSCO) results (Table 1) suggested high-quality assemblies. EC-S1 shared more genes with EC-N1 than domesticated barley (var. Morex). The wild barley specific orthologous gene clusters were significantly enriched in functions associated with the oxidation–reduction process, photosynthesis, and response to light stimulus (Data S1: Tables 7–9), which may reflect the genomic evolution in response to, for example, changed light environments of seed germination and seedling development in tilled soil (Civán *et al.*, 2021), or the practice of maximizing plant productivity under stress-free situations in cultivation (Considine and Foyer, 2014).

We sequenced an additional 22 wild barleys collected from the opposing slopes at the Evolution Canyon, with 11 accessions from each of the S_FS and N_FS, with an average of 17.5× genome coverage of the wild barley assembly, though most samples are around 12× and others are closer to 40× coverage (Data S2: Table 1). A total of 71.9 million SNPs and small InDels was identified for the 22 barley accessions (Data S2: Table 2) using the EC-S1 assembly as the reference genome. The 22 wild barley accessions and two assembled genomes clustered genetically into two clearly defined groups based on their sampling location from the two opposite slopes at the EC (Figure S2). The whole-genome nucleotide diversity (π) of both populations was generally low (Figure 1a) being 1.4×10^{-3} in the S_FS, and 1.8×10^{-3} in the N_FS population, with a significant difference between the two populations ($P < 0.001$). Low nucleotide diversity at large pericentromeric and centromeric genomic

Table 1 Genome assembly and annotation statistics for the two wild barley assemblies compared to domesticated barley (cultivar Morex, Mascher *et al.*, 2021) and two wild barleys (accession B1K-04-12, Jayakodi *et al.*, 2020; WB-01, Liu *et al.*, 2020)

	EC-S1	EC-N1	Morex V3	B1K-04-12	WB-01
Assembly total length (Mb)	5025	5052	4200	4214	4280
Total contig number	2593	2628	439	105 572	177 289
Max contig length (kb)	19 859	23 442	–	663	4913
Contig N50 (kb)	3525	3451	69 600	82	724
Anchored length (Mb)	4524	4471	4200	4182	–
Gene number	39 179	38 737	38 352	36 366	36 395
BUSCO completeness	96.2%	96.3%	97.4%	97.0%	95.3%

regions was observed in both populations, like the distribution of the density of SNPs and SVs along the chromosomes (Figure 1a). The analysis of genomic differentiation using Wright's F -statistic uncovered remarkable genomic divergences between the S_FS and N_FS wild barley population, with 10.4% of SNPs putatively under differentiated selection ($P < 0.05$), which involved 4220 (11%) high confidence genes, including genes involved in drought response, flowering time, and plant disease resistance (Data S2: Table 3).

Extensive structural variations in wild barley

The comparison of the two wild barley assemblies revealed extensive SV (Figure 1b). A total of 1.21 Gb (or 27% of the genome) contained SVs that were either deletions, insertions, duplications, translocations, or inversions (Data S3: Table 1). The 16 890 deletions and insertions were generally small in size, with 50% smaller than 420 bp, while the largest deletion spanned 1.9 Mb (Figure S3, Data S3: Table 2). A total of 225 chromosome inversions were detected. These inversions ranged from 620 bp to 37.12 Mb, with 12 inversions spanning more than 10 Mb. For example, the large inversions in Chr3H and Chr4H spanned 36.6 and 37.2 Mb, respectively. Large chromosome inversions were validated through reciprocal mapping of Hi-C data for EC-S1 and EC-N1 (Figure S4). Similar structural variations have been observed between assemblies and domesticated barley (var. Morex; Figure S5). Meanwhile, over 77 (63%) of inversions revealed by comparing EC-S1 and EC-N1 assemblies were genotyped (Data S3: Table 3), and 14 inversions showed population-level differentiation (Data S3: Table 4). Both the Illumina data (Chr3H inversion presence frequency: NFP: 11 vs SFP: 3 and Chr4H inversion presence frequency: NFP: 8 vs SFP: 0) and PCR results (Figure S6) suggest the large inversions (>20 Mb) in Chr3H and Chr4H are differentiated at the population level.

SVs overlapped 9055 genes (gene CDS located within the SVs regions) and potentially affected their gene expression in wild barley (Data S3: Table 5). GO enrichment analysis showed that genes affected by SV were functionally associated with photosynthesis and response to abiotic and biotic stress ($P < 0.05$, Fisher's exact test; Data S3: Table 6). Notably, 65 of the 225 annotated genes putatively related to plant disease resistance and 21 drought response-related genes (out of the 199 genes curated in the plant drought stress genes database DroughtDB, Alter *et al.*, 2015) were associated with one of the SVs (Data S3: Table 3).

Structural variations and differentially expressed genes between the wild barley populations from opposing slopes

The pattern of gene expression was strikingly different between the wild barley populations from the two opposing slopes (Data S4: Table 1). RNA-seq experiments revealed that a suite of genes showed differentiated expression between wild barley from the S_FS and N_FS under water stress (Data S4: Tables 2 and 3). Among the 3045 genes with differentiated expression, 785 (25.8%) genes are overlapped with SVs (Figure 2a). GO term enrichment for genes highly expressed in the S_FS wild barley suggested that many of these genes are involved in biological processes that are associated with drought response, such as the three most significant processes: photosynthesis, response to water, and dephosphorylation (Figure 2b, Data S4: Table 4). Meanwhile, SVs potentially impacted the level of gene expression, whether in plants under water stress or not, in both S_FS and N_FS (Figure S7). The expression of those genes in wild barley from the S_FS was increased more than two-fold compared to the N_FS (fold change >2; Figure 2c). Eight of the 17 genes

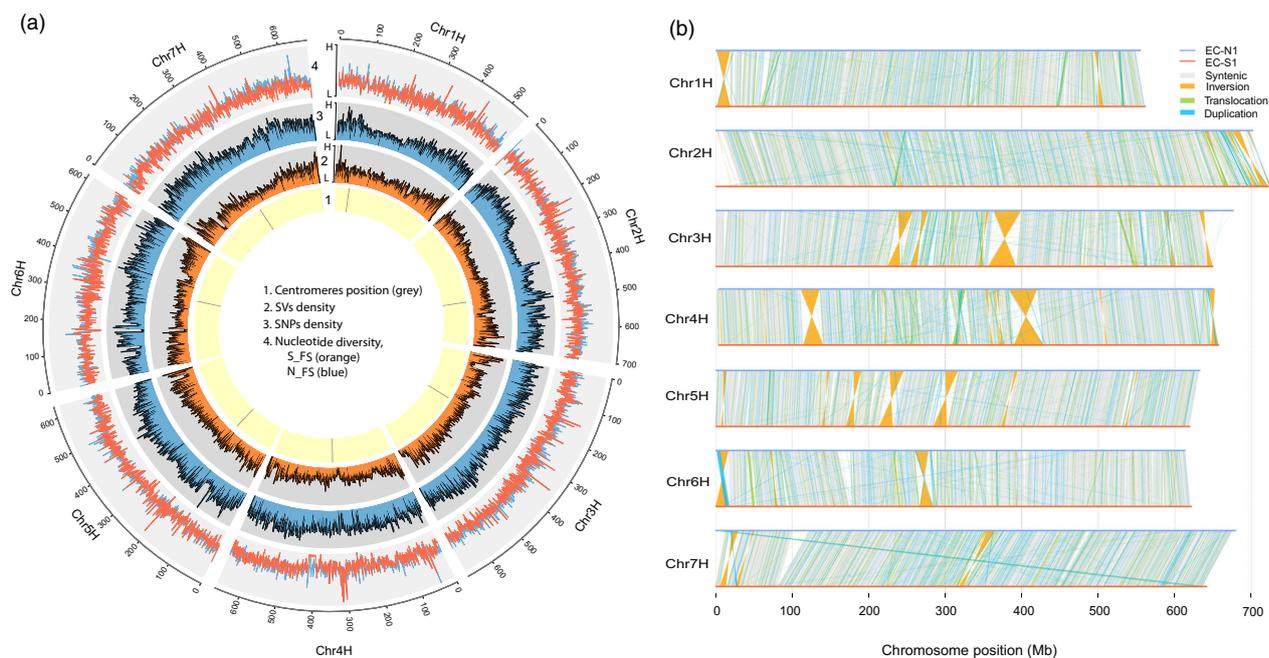


Figure 1 Genome-wide characteristics of the two wild barley assemblies. (a) Pattern of distribution of SVs and SNPs density, and centromere in EC-S1, and nucleotide diversity along the seven chromosomes in EC-S1 and EC-N1. (b) Pattern of specific types of SVs in the seven chromosomes.

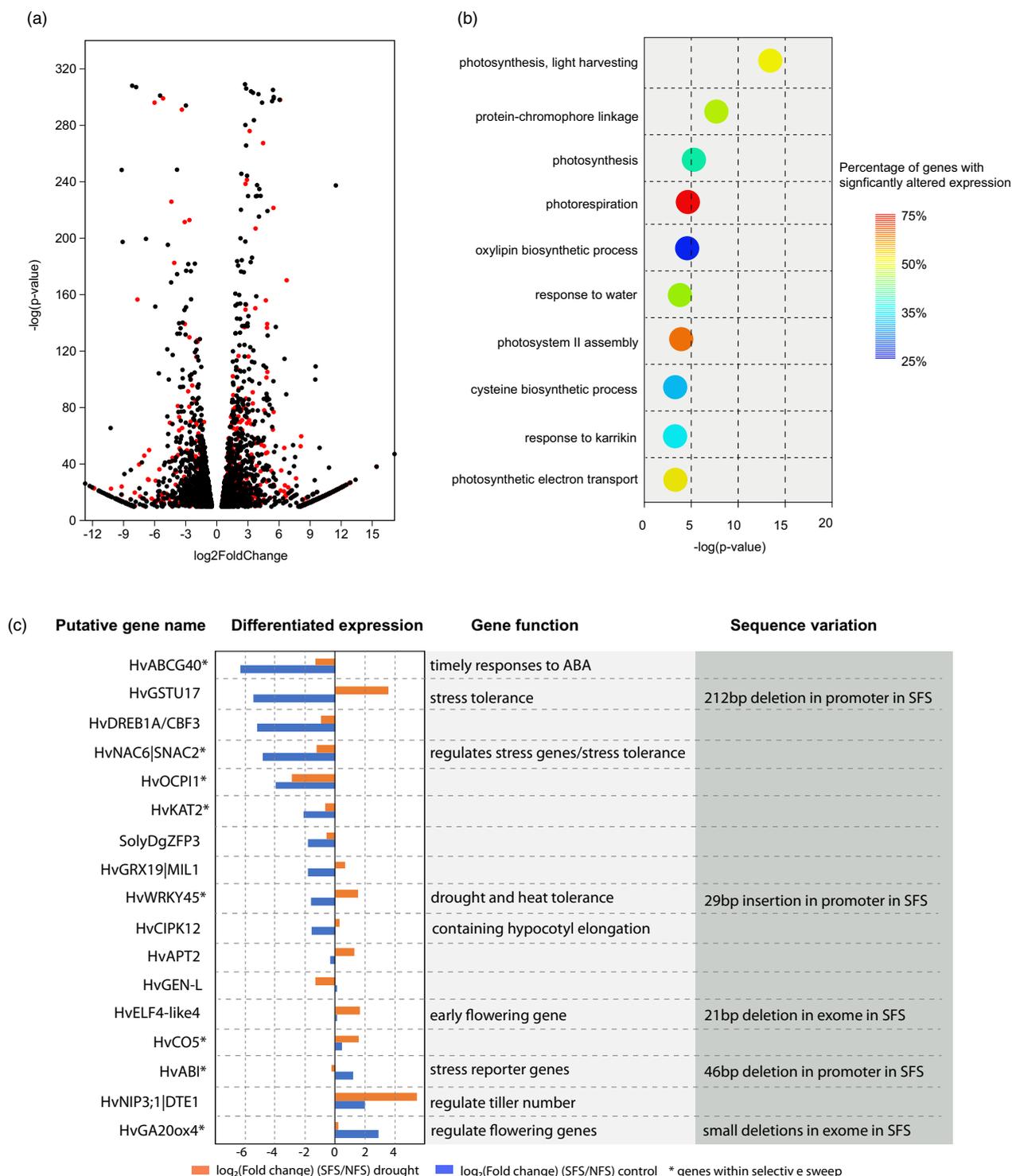


Figure 2 Differentiated gene expression in adaptive divergence and the connections to structural variations. (a) Genes with an altered expression between the S_{FS} and N_{FS} wild barley under water stress. Red dots indicate genes associated with an SV. (b) GO term enrichment of genes with an altered expression between the S_{FS} and N_{FS} wild barley under water stress. (c) Genes with differentiated expression (fold change >2) in both drought-stressed and unstressed conditions, their putative function, and the confirmed SVs in those genes; blank cells indicate that no information is available, asterisks indicate that those genes are putatively under differentiated selection. Gene expression data in panel c were obtained from Dai *et al.* (2014).

were located within genomic regions with significant F_{ST} ($P < 0.05$), and the genes were related to drought response (e.g. *HvABVG40* and *HvWRKY45*) or development and phenology

(*HvELF4-like4* and *HvGA20ox4*). Five genes were found to harbour a small SV either in the promoter region or in their exons (Figure 2c).

Large chromosome inversions and the heterogeneous pattern of genomic differentiation between the two adjacent populations

To decipher the underlying mechanisms of genomic divergence between the wild barley accessions from the two slopes, we examined genomic window-based population genetic structure (local PCA; Li and Ralph, 2019), gene flow (local N_m , effective number of migrants per generation) between the two populations, the deficiency of heterozygotes (F_{IS}), and recombination rates (Rho). Both the population relatedness in the genomic window, measured by PCA (MDS1, the first principal component), and the rate of effective gene flow (N_m) were

heterogeneous across the chromosome and genome (Figure 3; Figure S8), revealing that multiple biological processes drive this variation. Although the overall gene flow was low with $N_m = 0.9$, the rate of gene flow was heterogeneous across the chromosome and the genome. We observed that the gene flow of a few large chromosome regions was uniformly low ($N_m < 0.01$; Figure 3). These genomic regions ranged from 50 Mb to as much as 334 Mb and were mainly located in the pericentromeric regions of the chromosome (Figure 3; Figure S8).

Despite the relative higher gene flow ($N_m = 0.2-0.9$) towards both ends of the chromosomes, these genome regions harboured numerous small regions with elevated F_{ST} containing functional genes related to stress response and phenology (Figure 3). However, neither genetic relatedness (MDS1) nor genomic

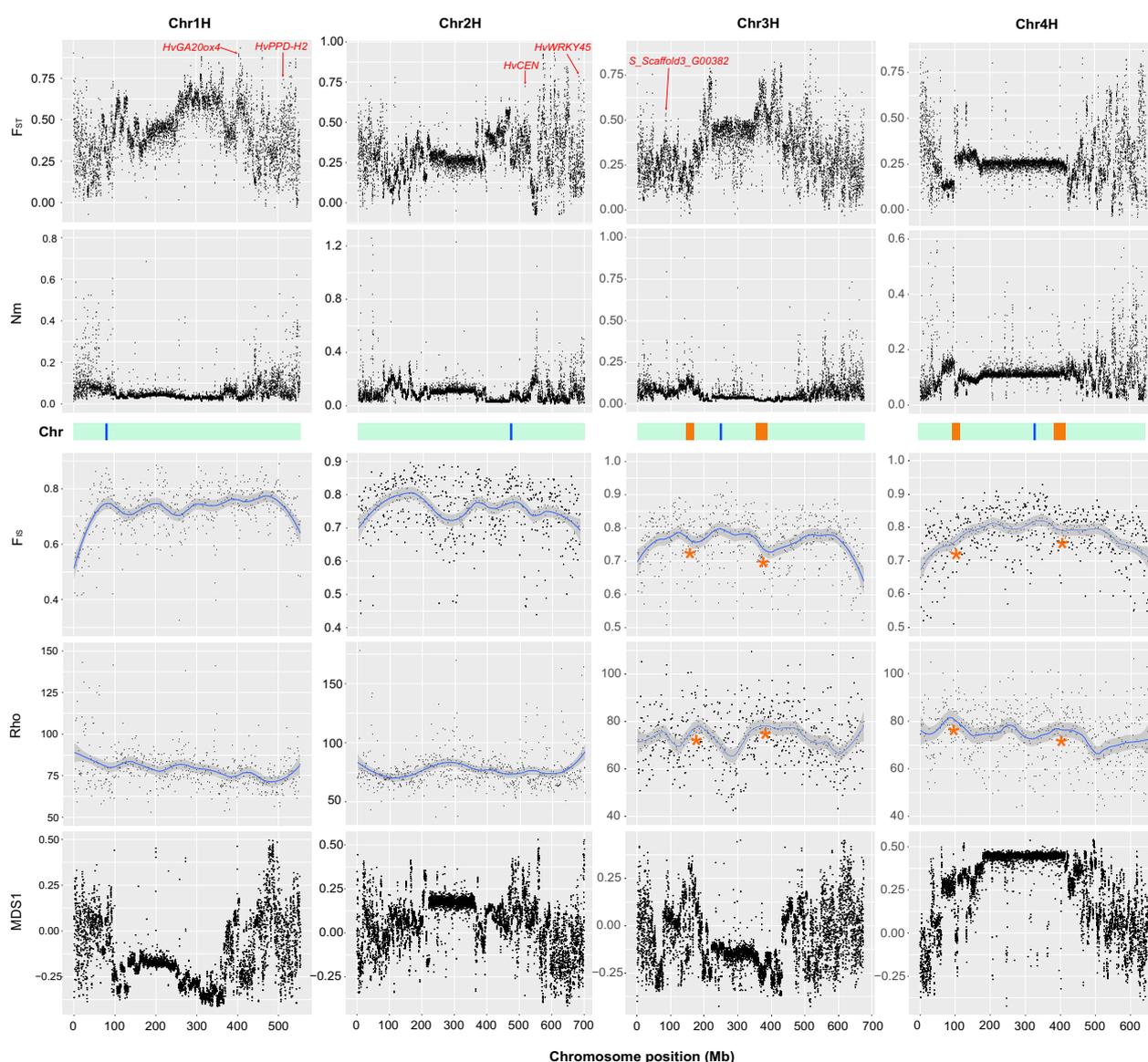


Figure 3 The impact of large chromosome inversions on the heterogeneous pattern of genomic differentiation (F_{ST}), gene flow (N_m), deficiency of heterozygotes (F_{IS}), recombination rate (Rho), and window-based genetic relatedness (MDS1). Examples of genes with elevated F_{ST} ($P < 0.05$) and a function in drought response and phenology adaptation are highlighted in red font. Blue bars on the light green stripes represent the possible location of centromeres in the chromosome, orange bars, and asterisks indicate locations of large inversions. Lines for F_{IS} and Rho (average within 1-Mbp window) were fitted by loess smoothing with 95% confidence intervals. Additional figures for Chr5H, Chr6H, and Chr7H are shown in Figure S8.

differentiation (F_{ST}) was uniformly correlated with the rate of gene flow (Nm) across the genomic regions. For example, gene flow was low (average Nm = 0.03) across a large region of chromosome Chr1H (from c. 106 to 440 Mb, the chromosome positions are in relation to EC-S1, and throughout unless specified; Figure 2). At the same time, genetic relatedness was uncorrelated with Nm and fluctuated from 106–160 Mb (Figure S9). F_{ST} maintained relatively low (average 0.381) with Nm varying between 0.04 and 0.5 in 106–258 Mb, while rapidly elevated to an average of 0.612 with a small reduction in Nm (from 0.04 to 0.02; Figure S9). Suppressed Nm (<0.02) corresponded with elevated F_{ST} between 398 and 468 Mb in chromosome Chr2H (Figure 3) and between 362.8 and 399.4 Mb in Chr3H (Figure 3).

Striking variations in the heterogeneous patterns of genetic relatedness, gene flow, and genomic differentiation along the chromosome may be due to large inversions (Figure 3). The two large inversions found on Chr4H, spanning 25.2 Mb (108.9–134.0 Mb) and 37.2 Mb (399.0–436.5 Mb), respectively, defined the entire genomic region between them with a uniform and relative low Nm (<0.1), and a homogenous genetic structure and low genetic differentiation ($F_{ST} = 0.24$). In comparison, genomic regions outside the two inversions showed fluctuating Nm, genetic relatedness, and elevated F_{ST} . The pattern of genetic relatedness and genomic differentiation showed a clear covariation at 362.8 and 399.4 Mb, defined by a 36.6 Mb inversion in Chr3H (Figure 3), and at 227.5 and 246.2 Mb that defined by an 18.7 Mb inversion in Chr4H (Figure 3). These large inversions may have led to an increased deficiency of heterozygotes by suppressing the recombination of chromosomes. In both Chr3H and Chr4H, the average level of deficiency of heterozygotes (F_{IS}) of the 50 Mb genomic region from the inversion towards the end of the chromosome was significantly higher within the 50 Mb genomic region from the inversion inwards ($P < 0.001$), while the recombination rate was significantly lower ($P < 0.001$). Noticeably, centromere may have impacted the recombination rate and level of deficiency of heterozygotes similarly, e.g., in Chr1H (Figure 3). However, not all inversions and centromeres had a noticeable impact on genomic differentiation. Uniform patterns of F_{ST} , Nm, F_{IS} , Rho, and MSD1 spanned across the inversions at 266.6–284.0 Mb in Chr6H and 355.0–366.1 Mb in Chr7H (Figure S8).

The role of structural variation and functional genes in local adaptation and phenotypic differentiation

Structural variations form a significant part of adaptive genetic variation in wild barley and may contribute to local adaptation and phenotypic differentiation. In the glasshouse experiment, wild barley from the S_FS and N_FS showed different growth responses under both control and drought-stressed conditions. Three independent accessions from the S_FS grew significantly longer roots than those from the N_FS under both growing conditions (Figure 4a). Wild barleys from the S_FS were less impacted by drought and showed continued root growth under drought-stressed conditions compared to N_FS wild barley (Figure 4b–d). We detected 15 genes located in the genomic regions with significant F_{ST} with a function in drought response (Data S2: Table 3). Further, four genes (28.6%, higher than the 27.4% of the genome associated with SVs) were found to be impacted by SVs. *Hordeum_vulgae_N_Scaffold86_G00274* was

annotated as an F-box domain protein gene, associated with drought tolerance in wheat (An *et al.*, 2019). This gene has five exons in Chr3H in the N_FS population (Figure 4e). Two small inversions in the S_FS population created a new F-box domain protein gene by inverting two exons of *Hordeum_vulgae_N_Scaffold86_G00274* by one 1133 bp inversion and coupling two additional exons through another 954 bp inversions (Figure 4e). The two inversions were 165 402 bp apart in the EC-N1 assembly, while only 302 bp in EC-S1.

The WRKY family genes have a significant role in stress response in plants (Pandey and Somssich, 2009; Jiang *et al.*, 2017). *HvWRKY45* showed differentiated expression in the S_FS wild barley both in stress and unstressed conditions (Data S4: Table 1) consistent with the over-expression of rice *WRKY45* enhancing drought tolerance in rice (Tao *et al.*, 2011). By examining the reads being mapped to the samples from both slopes, we observed a 29-bp insertion in the promoter region of the *HvWRKY45* gene consistently present in the S_FS population while absent in all samples from N_FS (Figure 4f). This insertion created a sequence of 'AGCCACC', which is a Box S cis-regulatory element that significantly regulates the expression of genes in stress response and signalling pathway (Rushton *et al.*, 2002). Noticeably, both the F-box domain gene (gene ID: *Hordeum_vulgare_S_5HG000962*) and *HvWRKY45* (gene ID: *Hordeum_vulgare_S_2HG004343*) were significantly up-regulated in the S_FS population when under drought-stress treatment (Data S2: Table 3).

The genomic divergence between the two wild barley populations also translated into a difference in flowering time. In both *in situ* and glasshouse conditions, wild barley from the S_FS flowered 10–14 days earlier than those from the N_FS (Figure 5a), with the flowering time of cross-fertilized offspring falling between the S_FS and N_FS parental lines (Parnas, 2006). Among the 174 genes associated with phenology in domesticated barley (He *et al.*, 2019; Hill *et al.*, 2019), 15 were located within genomic regions with significant F_{ST} in wild barley (Data S2: Table 3). Further examination revealed three genes (including the 2 kb promoter region) contained polymorphic variants that could be linked to divergent selection. The first gene, *HvCO5*, is one of the genes associated with flowering time in domesticated barley. A single SNP mutation G->C in the promoter region in the S_FS wild barley formed an 'AGCCACC' sequence, a Box S cis-regulatory element, which may have influenced *HvCO5* expression and putatively linked to the earlier flowering time of wild barley from the S_FS (Figure 5b). The second gene, *HvCK2α* (gene ID: *Hordeum_vulgare_S_5HG004177*), shows three SNPs that were fixed for each slope. These SNPs result in two missense mutations coding different amino acids (Figure 5c), although it is not clear how the missense mutations impacted protein function. The *CK2α* gene is evolutionarily conserved in plants and functions as a core component of the circadian clock system playing a critical role in regulating flowering (Ogiso *et al.*, 2010). The overexpression of the *CK2α* gene leads to changes in the period of the circadian clock and early flowering in *Arabidopsis thaliana* (Sugano *et al.*, 1999). The third gene, *HvGA20ox4* (gene ID: *Hordeum_vulgare_S_Contig520G000001*) contains several small deletions in the exons unique to the S_FS wild barley. The deletions shifted the open reading frame and altered the amino acid sequence and secondary protein structure (Figure 5d), which may have resulted in differential expression of *HvGA20ox4* between the two slopes (Figure 3c).

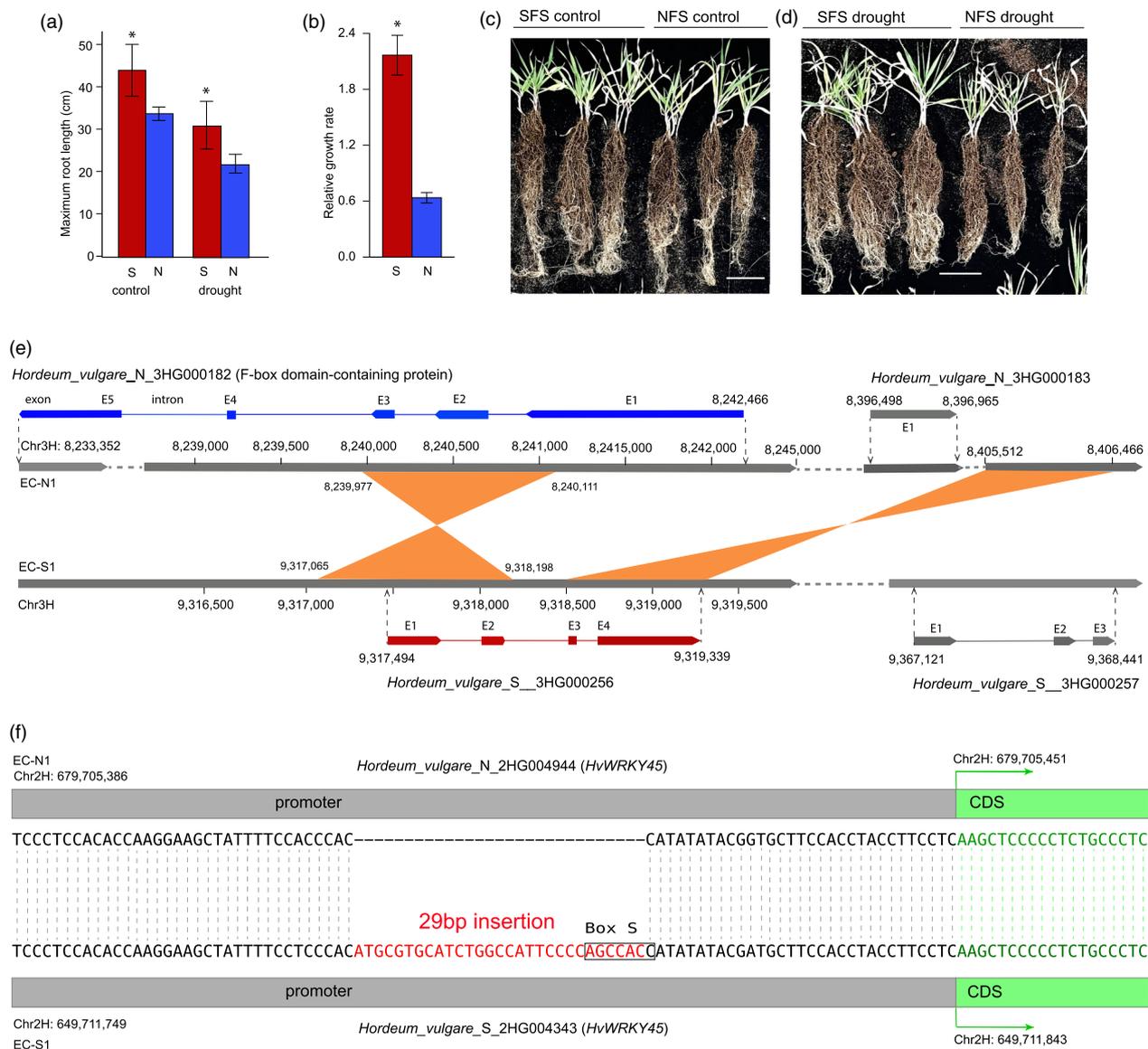


Figure 4 Examples of significant genes affected by SVs and associated with differential drought responses that vary between the S_FS and N_FS wild barley populations. (a) Measurement of root length of the S_FS and N_FS wild barley in control and the drought-stressed conditions. (b) Growth rate of wild barley from the S_FS and N_FS in the drought-stressed condition in relation to the control condition. (c) Root morphology of S_FS and N_FS wild barley growing in control conditions. (d) Root morphology of S_FS and N_FS wild barley growing in drought-stressed conditions. (e) Two chromosome inversions (shown in orange) generated a modified F box domain protein gene. Note that exon E4 in *Hordeum_vulgare_N_3HG000182* (shown in blue) shares the same sequence as E3 in *Hordeum_vulgare_S_3HG000256* (shown in dark red). The downstream gene in both S_FS and N_FS was also shown (in grey). (f) A 29-bp insertion into promoter regions of the *HvWRKY45* gene created a sequence of 'AGCCACC', a Box S cis-regulatory element.

Discussion

In this study, we used Nanopore long reads to generate two high-quality wild barley genomes (*Hordeum spontaneum*) sampled from S_FS and N_FS in an Evolution Canyon system in Israel. The two wild barley assemblies have high contiguity with a contig N50 of 3.5 Mb, which is superior to other published wild barley genomes (Jayakodi *et al.*, 2020; Liu *et al.*, 2020). Our assemblies with the assembled length of 5.02–5.05 Gb are similar to these previous reports (Doležel *et al.*, 1998; Mayer *et al.*, 2012; Mascher *et al.*, 2017), with a corresponding genome size of 4.95–5.15 Gb of *Hordeum vulgare*. The two high-quality

chromosome level wild barley assemblies reveal a massive and full spectrum of genomic structural variations (SVs). SVs can affect the order and proximity of genetic elements, disrupt the functionality of extant genes, generate new functional genes, interfere with chromosome recombination, and cause differentiated survival of gametes (Hoffmann and Rieseberg, 2008; Faria and Navarro, 2010; Lye and Purugganan, 2019; Wellenreuther *et al.*, 2019; Mérot *et al.*, 2020; Jayakodi *et al.*, 2020). Our results suggest SVs may have played an important role in the phenotypic differentiation in wild barley through divergent adaptation. Genetic differentiation at several significant drought-response and phenology-related genes linked to

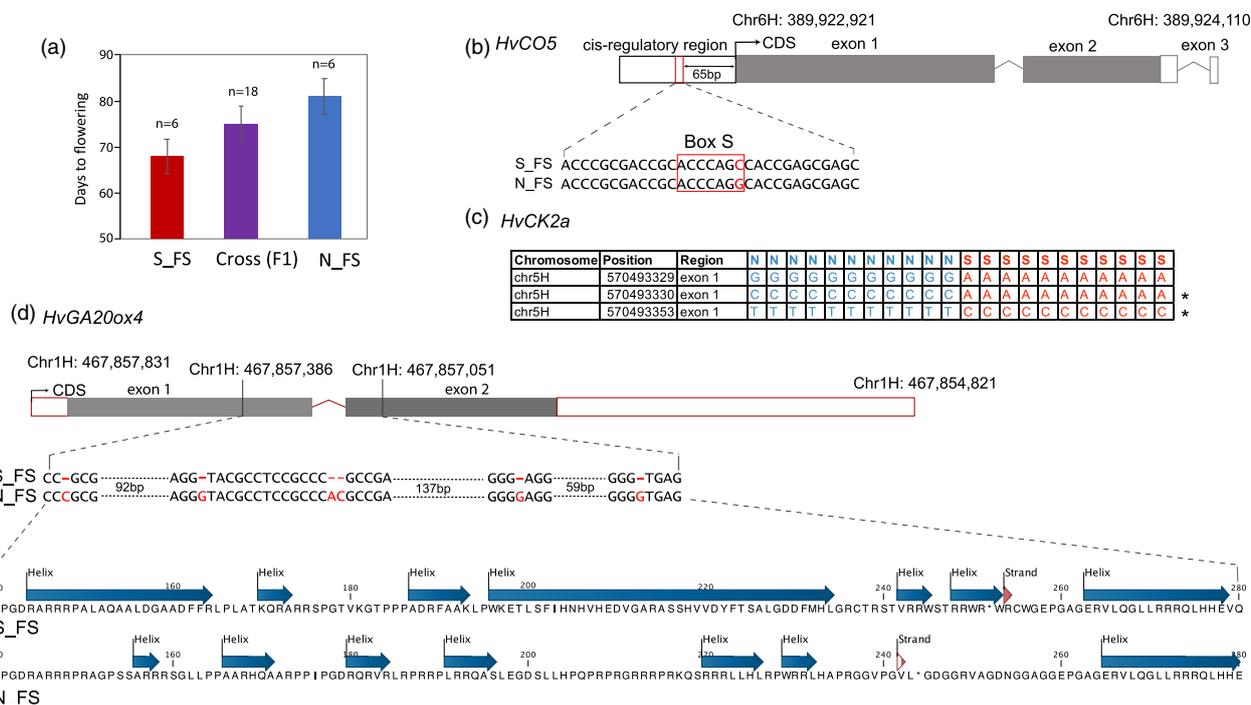


Figure 5 Significant genes are affected by SVs and associated with flowering time. (a) Different flowering times of wild barley from the S_{FS} and N_{FS} in glasshouse conditions (based on data from Parnas, 2006). (b) Flowering gene *HvCO5* and the significant SNP mutation in the cis-regulatory element. (c) Fixed SNPs in the *HvCK2a* gene in the S_{FS} and N_{FS} wild barley. Asterisks indicating missense mutations. (d) Small deletions in the flowering gene *HvGA20ox4* in the S_{FS} wild barley and the impact of deletion on the secondary structure of the consequent protein. Protein secondary structure was estimated with CLC Genomic Workbench 9.0 (CLC Bio, Qiagen, Aarhus, Denmark).

population-specific SVs may have resulted in the different expression patterns of those genes or altered protein structure, affecting phenotypic differentiation. The local microclimatic environments may have selected and maintained these differentiated SVs, evident by the fixation of slope (population)-specific SVs and alleles. For example, the *HvWRKY45* gene in all the samples from the S_{FS} population had a 29-bp insertion in the promoter region, while this insertion was missing from the N_{FS} population. The insertion may have altered the cis-regulatory element that significantly regulates the expression of genes in stress response and signalling pathway. Similar population-specific SV was observed in the *HvCO5* gene that was associated with flowering time in domesticated barley (He *et al.*, 2019; Hill *et al.*, 2019). As such, we suggest that structural genomic variants provide significant genetic material to drive adaptive evolution in wild barley. Conversely, heterogeneous environments select and maintain adaptive SVs through the effect of adaptive selection.

Although geographical proximity facilitates gene flow between the S_{FS} and N_{FS} wild barley populations, natural selection and adaptation to the contrasting environments of the EC over a long period of time have led to a clear genomic signal of evolutionary divergence. The sharp microclimatic divergence between the two slopes in the EC has resulted in differentiated drought response and flowering phenology in wild barley (Nevo, 2012, 2014). Gene flow, albeit low, was evident in the form of seed dispersal between the two slopes (Nevo, 2012). Our results show that adaptive phenotypic divergence and genomic differentiation could prevail at fine spatial scales with sharply divergent ecologies in the presence of gene flow. Those genomic regions directly

related to specific abiotic and biotic selection pressures (e.g., heat, drought, illuminance stresses, pathogen, and edaphic conditions) would be differentiated. The remainder of the genome would be subjected to alternative processes such as chromosome rearrangement (inversions; Crown *et al.*, 2018) or gene flow that homogenizes genetic elements (Hübner *et al.*, 2009).

Genomic differentiation is often heterogeneous across the genomes of diverging populations (Seehausen *et al.*, 2014; Wolf and Ellegren, 2017). Linking observed patterns of genomic differentiation to underlying processes reveals the mechanistic basis of local adaptation. Diverse types of SVs have increasingly been found to differentiate locally adapted populations (Mérot *et al.*, 2020; Todesco *et al.*, 2020). However, these results were based on materials with large geographic isolations and thus may be complicated by other factors such as isolation by distance and genetic drift. Using the geographic close wild species with a single ancestor has overcome these limitations. Here, we show that small SVs best explain the genetic differentiation observed at the gene level, influenced through divergent selection. Functional SVs are more likely to be small, taking the form of deletions or insertions in genetic elements essential to gene expression. They can also be small inversions creating novel functional genes by rearranging genetic elements. Pan-genome analyses of several other species have also revealed that cis-regulatory variation is often associated with small SVs and is linked to modified gene expression and phenotypic variation (Alonge *et al.*, 2020; Liu *et al.*, 2020; Wang *et al.*, 2021).

In contrast, large SVs may have shaped genomic divergence as part of a different process. Large chromosome inversions may

have significantly shaped the landscape of genomic differentiation through incompatible genomic architecture (Wellenreuther and Bernatchez, 2018). We observed the presence of several genomic regions spanning up to 100–250 Mb, where genetic differentiation and gene flow were uniform but seems decoupled, likely resulting from the presence of large chromosome inversions. Large chromosome inversions reduce gene flow through suppression of genetic recombination that may have occurred through the displacement of crossing-overs away from those large inversions (Crown et al., 2018). With two or multiple large inversions, chromosome recombination in the genomic region situated between those inversions is significantly reduced, leaving a long stretch of the genome with suppressed recombination and low gene flow, as we observed in Chr3H and Chr4H. Contrasting to the often observed pattern that ongoing gene flow prevents genetic differentiation (Pinho and Hey, 2010), our results suggest that reduced genomic differentiation in genomic regions could result from suppressed gene flow due to large chromosome inversions affecting chromosome recombination. On the other hand, these large SVs, which may be neutral without phenotypic effect (Jayakodi et al., 2020), could be maintained within the population boundary as the result of suppressed gene flow.

The small population size and proximity of sampling locations could limit the inference of selective loci. Without a larger diversity panel allowing analysis of selective sweeps and genetic mappings such as a QTL mapping or genome-wide association studies to delineate genomic regions associated with the locally adaptive variation, other significant genes may have been missed from our analysis. Other caveats may stem from our analysis of genomic differentiation using Wright's *F*-statistic to uncover genomic divergences between the S_FS and N_FS populations. We only included two populations in the fsthet analysis, which may cause inflation of the type I error rates (false positives). Future research requires analysis on a larger diversity panel and genomic association analysis to identify candidate genes involved in phenotypic differentiation in wild barley, and further functional validation through molecular target analysis. The way forward requires functional characterizations of these significant genes, such as *HvWRKY45*, *HvCO5*, and *HvCK2α*, through molecular target analysis for the functional impact of revealed structural and SNP variations and their roles in local adaptation.

Cultivated barley, the world's fourth largest cereal crop, was domesticated about 10 000 years ago in the mid-east. The capacity of the S_FS wild barley to persist in a drought and heat stress environment has important implications in barley breeding. For thousands of years, agronomic traits in barley have been selected for easy harvest, high yield, and high grain quality, while genetic variations that render the survival during environmental stresses have been weakened or even lost (Palmgren et al., 2015). Therefore, the wild barley population studies here hold great potential to provide a rich gene pool of drought-tolerant and disease-resistant genes that might be introduced into domestic barley, as wild barley readily crosses with domesticated barley. Our results also showed that gene-associated SVs in cis-regulatory regions could have significant phenotypic effects. Resolving these functional impacts of SV will facilitate the exploitation of natural and engineered SVs in the genetic improvement of barley and other crops (Alonge et al., 2020; Wang et al., 2021). Finally, the extensive genomic structural variations and a significant number of functional genes associated with SVs, as observed in the two wild barley genomes studied

here, suggest that single reference genomes do not capture the full diversity within a species. Therefore, future crop genomics and improvement should embrace the framework of the pan-genome (Bayer et al., 2020; Della Coletta et al., 2021; Jayakodi et al., 2020; Tao et al., 2021), especially for exploring genetic variations in the wild species.

Methods

Sample preparation and sequencing

High-molecular-weight DNA preparation and long-read sequencing with Oxford nanopore technologies

Seeds of two samples, EC-S1 and EC-N1, were collected from the S_FS and N_FS, respectively, at the 'Evolution Canyon' in Mount Carmel, Israel. Seeds were germinated and grown in a glasshouse at Yangtze University (Jingzhou, Hubei Province, China), and mature leaves were collected from plants of EC-S1 and EC-N1 for DNA preparation and sequencing. High-molecular-weight genomic DNA was prepared by the CTAB method and purified with QIAGEN® Genomic kit (Cat#13343, QIAGEN, Denmark) following the manufacturer's standard operating procedure. DNA degradation and contamination of the extracted DNA were monitored on 1% agarose gels, and DNA purity was examined using a NanoDrop™ One UV–Vis spectrophotometer (Thermo Fisher Scientific, USA). DNA concentration was further measured with a Qubit® 4.0 Fluorometer (Invitrogen, USA).

A total amount of 3–4 µg DNA per sample was used as input material for the ONT library preparations. After the sample was qualified, size-selection of long DNA fragments was performed using the PippinHT system (Sage Science, USA). Next, the ends of DNA fragments were repaired, and an A-ligation reaction was conducted with NEBNext Ultra II End Repair/dA-tailing Kit (Cat# E7546). The adapter in the SQK-LSK109 (Oxford Nanopore Technologies, UK) was used for further ligation reaction, and the DNA library was measured by Qubit® 4.0 Fluorometer. About 700 ng DNA was used to construct the library and sequenced on a Nanopore PromethION sequencer instrument (Oxford Nanopore Technologies, UK) at the Genome Centre of Grandomics (Wuhan, China).

Hi-C sequencing and BioNano optical mapping

Hi-C sequencing and BioNano optical mapping were used to anchor hybrid scaffolds onto the chromosome. Genomic DNA was extracted from EC-S1 and EC-N1 following the manufacturer's recommended protocols for library preparation (Plant DNA Isolation Kit, QIAGEN 80003, Denmark). We constructed the Hi-C library and obtained sequencing data via the Illumina Novaseq/MGI-2000 platform. In brief, freshly harvested leaves were cut into 2 cm pieces and vacuum infiltrated in nuclei isolation buffer supplemented with 2% formaldehyde. Crosslinking was stopped by adding glycine and additional vacuum infiltration. Fixed tissue was then grounded to powder before re-suspending in nuclei isolation buffer to obtain a nuclei suspension. The purified nuclei were digested with 100 units of DpnII and marked by incubating with biotin-14-dATP. Biotin-14-dATP from non-ligated DNA ends was removed owing to the exonuclease activity of T4 DNA polymerase. The ligated DNA was sheared into 300–600 bp fragments and then was blunt-end repaired and A-tailed, followed by purification through biotin-streptavidin-mediated pull-down. Finally, the Hi-C libraries were quantified and sequenced using the Illumina Novaseq/MGI-2000 platform.

For BioNano physical mapping, DNA extracted from EC-S and EC-N were subject to the manufacturer-recommended protocols for library preparation (Plant DNA Isolation Kit, 80 003) and optical scanning provided by BioNano Genomics (<https://bionanogenomics.com>), with the labelling enzyme Direct Label Enzyme (DLE; Bionano PrepDLS Labeling DNA Kit, 80 005). Labelled DNA samples were loaded and run on the Saphyr system (BioNano Genomics).

RNA preparation for Pacbio ISO-seq sequencing

Wild barley was germinated and grown in glasshouse. Leaves, roots, and inflorescences, and developing seeds were collected for RNA extraction. Total RNA was extracted by grinding tissue in TRIzol reagent (TIANGEN, China) on dry ice and processed following the manufacturer's protocol. The integrity of the RNA was monitored with agarose gel electrophoresis and an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The purity and concentration of the RNA were measured with the Nanodrop and Qubit. Only a high-quality RNA sample (OD260/280 = 1.8–2.2, OD260/230 \geq 2.0, RIN \geq 7, >1 μ g) was used to construct the sequencing library. The mRNA was enriched by Oligo (dT) magnetic beads. Then the enriched mRNA was reverse transcribed into cDNA using Clontech SMARTer PCR cDNA Synthesis Kit (TAKARA, Japan). PCR cycle optimization was used to determine the optimal amplification cycle number for the downstream large-scale PCR reactions (PrimeSTAR® GXL DNA polymerase). The optimized cycle number was then used to generate double-stranded cDNA. The large-scale PCR was performed for the next SMRTbell library construction. cDNAs were prepared with DNA damage repair, end-repaired, and sequencing adapters ligation using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, USA). The SMRTbell template was annealed to a sequencing primer and bound to polymerase and sequenced on the PacBio Sequel platform using Sequel Binding Kit 3.0 (Pacific Biosciences) with 20 h movies.

Genome assembly

Data quality control and sequencing data filtering

To ensure reads are reliable, Illumina paired-ended sequenced raw reads of EC-S1 and EC-N1 for the genomic survey were first filtered using the fastp (v.0.20.0; Chen *et al.*, 2018) preprocessor (set to default parameters) to remove low-quality reads, adapters, and reads containing poly-N. The low-quality reads were filtered under the following conditions: (1) reads with \geq 10% unidentified nucleotides (N); (2) reads with >10 nucleotides aligned to the adapter, allowing \leq 10% mismatch; (3) reads with >50% bases having Phred quality <5; (4) removing putative PCR duplicates generated by PCR amplification in the library construction process. To examine contamination in sequencing reads, 100 000 reads were selected randomly and matched against sequences in the NT (NCBI Resource Coordinators, 2013) library by BLAST+ v 2.2.3 (Camacho *et al.*, 2009). Nanopore sequencers output FAST5 files containing signal data, and base-calling were first performed to convert the FAST5 files to FASTQ format with Guppy (Wick *et al.*, 2019). The raw reads of fastq format with mean_qscore_template <7 were then filtered, resulting in pass reads.

Genome de novo assembly

For de novo genome assembly, an ONT-only assembly was constructed by using an OLC (overlap layout-consensus)/string graph method with NextDenovo. Considering the high error rate

of ONT raw reads, the original subreads were first self-corrected using NextCorrect and generated consistent sequences (CNS reads). Comparing CNS was then performed with the NextGraph module to capture correlations of CNS. Based on the correlation of CNS, the preliminary genome was assembled. To improve the assembly accuracy, we refined the contigs with Racon using ONT long reads and Nextpolish using Illumina short reads with default parameters. Sequence similarity searches were performed to generate final assemblies with the parameters 'identity 0.8–overlap 0.8' to remove redundant contigs.

The completeness of genome assemblies was assessed using BUSCO v3.0.2 (Benchmarking Universal Single-Copy Orthologs) with default settings using the 'eukaryota_odb9' dataset (Simao *et al.*, 2015). BUSCO was rerun for the wild barley genomes B1K-04-12 and WB-01, and the Morex V3 genome to generate comparable results. All the Illumina paired-end reads were mapped to the assembled genome using BWA-MEM v 0.7.17 to evaluate the accuracy of the assembly. The mapping rate and genome coverage of sequencing reads were assessed using samtools v0.1.1855 (Li *et al.*, 2009). Base accuracy of the assembly was calculated with bcftools v1.8.0 (Li *et al.*, 2009). The coverage of expressed genes of the assembly was examined by aligning all the RNA-seq reads against the assembly using HISAT2 v2.1.0 with default parameters (Kim *et al.*, 2015). To avoid including mitochondrial sequences in the assembly, the draft genome assemblies were searched against the NCBI NT library (NCBI Resource Coordinators, 2013) using BLAST+ v 2.2.3 (Camacho *et al.*, 2009), and aligned sequences were eliminated.

Chromosome anchoring with BioNano mapping and Hi-C sequence data

Raw BioNano data were cleaned by removing molecules matching any of the following rules: length <150 kb, molecule signal-to-noise ratio <2.75 and label signal-to-noise ratio <2.75, or label intensity greater 0.8. De novo assembly of BioNano molecules into genome maps was performed using the script pipelineCL.py in the BioNano Solve package v3.3 (BioNano Genomics) with parameters '-d -U -N 6 -y -i 3 -F 1 -a optArguments_nonhaplotype_noES_noCut_saphyr.xml'. Hybrid scaffolds were assembled from ONT assembly and BioNano genome maps using the script hybridScaffold.pl in the Solve package with parameters '-c hybridScaffold_DLE1_config.xml -u CTTAAG -B 2 -N 2 -f'.

For the HiC data, low-quality sequences (quality scores < 20), adaptor sequences, and sequences shorter than 30 bp were firstly filtered out using fastp with default settings (Chen *et al.*, 2018), and then the clean paired-end reads were applied to the draft assembled sequence using bowtie2 v2.3.2 with setting '-end-to-end --very-sensitive -L 30' (Langmead and Salzberg, 2012) to get the unique mapped paired-end reads. Valid interaction paired reads were identified and retained by HiC-Pro 13 v2.8.1 (Servant *et al.*, 2015) from unique mapped paired-end reads for further analysis. The scaffolds were further clustered, ordered, and oriented scaffolds onto chromosomes by LACHESIS (Burton *et al.*, 2013) with the setting 'CLUSTER_MIN_RE_SITES=100, CLUSTER_MAX_LINK_DENSITY=2.5, CLUSTER_NONINFORMATIVE_RATIO=1.4, ORDER_MIN_N_RES_IN_TRUNK=60, ORDER_MIN_N_RES_IN_SHREDS=60'. Finally, placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

To confirm the accuracy of HiC data, the EC-S1 and EC-N1 HiC data were also mapped to the Morex V3 genome using Juicer tools v.1.6 (Durand *et al.*, 2016; Data S1: Table 10). The large

inversions identified using SyRI (Goel *et al.*, 2019) between Morex and EC-S1 and EC-N1 were validated using the HiC contact alignment (Figure S10).

Genome annotation

Repeat annotation

We first annotated the tandem repeats using the software GMATA v2.3 (Wang and Wang, 2016) and Tandem Repeats Finder (TRF; Benson, 1999), where GMATA identifies the simple repeat sequences (SSRs) and TRF recognizes all tandem repeat elements in the whole genome. Transposable elements (TE) in the EC-S1 and EC-N1 genomes were then identified using a combination of *ab initio* and homology-based methods. Briefly, *ab initio* repeat libraries for EC-S1 and EC-N1 were first predicted using MITE-hunter (Han and Wessler, 2010) and RepeatModeler 2 (Flynn *et al.*, 2020) with default parameters, in which LTR_FINDER v1.2 (Xu and Wang, 2007), ltr_harvester, and LTR_retriever were also included for a plant genome. The obtained library was then aligned to TEclass Repbase (<http://www.girinst.org/repbase>) to classify the type of each repeat family. For further identification of the repeats throughout the genome, RepeatMasker was applied to search for known and novel TEs by mapping sequences against the *de novo* repeat library and Repbase TE library. Overlapping transposable elements belonging to the same repeat class were collated and combined. Telomeric repeats were identified using BLAST+ (version 2.2.3, *E*-value < 1e-5) search Arabidopsis-type telomere sequences [CCCTAA/TTTAGGG] against the EC-S1 and EC-N1 genome sequences. Centromeric repeats were identified using BLAST+ (version 2.2.3, *E*-value < 1e-5) search barley centromere-specific [AGGGAG]5 repeats against the genome sequence.

Gene prediction

Three independent approaches, including *ab initio* prediction, homology search, and reference guided transcriptome assembly, were used for gene prediction in a repeat-masked genome. In detail, GeMoMa (Keilwagen *et al.*, 2019) was used to align the homologous peptides from related species to the assembly and then provided gene structure information, which was homologue prediction. For RNA-seq-based gene prediction, filtered mRNA-seq reads were aligned to the reference genome using STAR (Dobin *et al.*, 2013). The transcripts were then assembled using stringtie v1.34 (Pertea *et al.*, 2015) and open reading frames (ORFs) were predicted using PASA (Haas *et al.*, 2008). For the *de novo* prediction, RNA-seq reads were *de novo* assembled using stringtie and analysed with PASA to produce a training set. Augustus v3.1 (Stanke and Morgenstern, 2005) with default parameters were then utilized for *ab initio* gene prediction with the training set. Finally, EVIDENCEModeler (EVM; Haas *et al.*, 2008) was used to produce an integrated gene set of which genes with TE were removed using TransposonPSI package (<http://transposonpsi.sourceforge.net/>) and the miscoded genes were further filtered. Untranslated regions (UTRs) and alternative splicing regions were determined using PASA based on RNA-seq assemblies. We retained the longest transcripts for each locus, and regions outside of the ORFs were designated UTRs.

Functional annotation of gene models

We selected the best-matching reference protein for each gene as a template sequence and defined the transcript sequence with maximum coverage of the template sequence as a gene

representative. Genes were defined as high confidence (HC) genes using the similar method of Mascher *et al.* (2017), which if they had a significant BLAST hit to reference proteins and if their representative protein had a similarity to the respective template sequence above a threshold which we determined on the basis of the origin of template sequences (>60% for *Arabidopsis thaliana*, sorghum and rice, >65% for *Brachypodium distachyon*, and >85% for barley). Gene function information, motifs, and domains of their proteins were assigned by comparing with public databases including SwissProt, KEGG, KOG, and Gene Ontology. The high-confidence gene set was functionally annotated using EggNOG v4.5 (Huerta-Cepas *et al.*, 2016) mappings with DIAMOND (Buchfink *et al.*, 2015). Emapper annotations were filtered at an *e*-value threshold of 1e⁻¹⁰. For the annotation of non-coding RNAs (ncRNAs), two strategies were used to obtain the non-coding RNA (ncRNA): searching against the database and prediction with the model. Transfer RNAs (tRNAs) were predicted using tRNAscan-SE with eukaryote parameters. MicroRNA, rRNA, small nuclear RNA, and small nucleolar RNA were detected using Infernal cmscan to search the Rfam database. The rRNAs and their subunits were predicted using RNAmmer (Lagesen *et al.*, 2007).

The BLAST reciprocal best hits of any two proteins are defined as potential ortholog genes. Orthologous genes among the EC-S1, EC-N1, and Morex genomes were identified by Orthovenn2 (Xu *et al.*, 2019), a web tool used to identify orthologous and paralogous genes with a pairwise sequence similarity cut-off of 10–5 and inflation of 1.5 to define orthologous cluster structure. Orthologous clusters were analysed by UniProt search and TopGO (Alexa and Rahnenführer, 2009) for functional annotation.

BARE transposable elements analysis

The *de novo* transposable element (TE) libraries were constructed for EC-S1 and EC-N1 using EDTA v1 (Ou *et al.*, 2019). The available barley BARE-1 TE sequence (Accession: U73173.1) were downloaded from National Center for Biotechnology Information (NCBI), and cd-hit v4.6.8 (-c 0.9; Li and Godzik, 2006) was used to remove the highly similar sequences to generate the non-redundant BARE sequences. Using BLAST+ v 2.2.3 (Camacho *et al.*, 2009), the BARE sequences were used to search against (with >85% sequence identity and *e*-value < 10⁻⁵) the EC-S1 and EC-N1 TE libraries to identify the BARE TEs. Genes overlapped with the BARE TEs in EC-S1 and EC-N1 genomes were extracted and InterProScan v5.27 (Jones *et al.*, 2014) was used for functional annotation of these overlapped genes.

Structural variations identification and validation

We used the EC-S1 genome as the reference for SV identification. The EC-N1 genome was aligned to the EC-S1 genome using Mummer v4.0 (Marçais *et al.*, 2018) with the parameters (-l 50 -c 100 -maxmatch). The raw alignments results were further filtered using delta filter with parameters (-m -i 90 -l 100). The resulting filtered delta files were used to detect structural variations using the SyRI pipeline with default parameters (Goel *et al.*, 2019). The detected variations from SyRI comprise a two-hierarchy structure: genome rearrangements and sequential variations (which can occur in both rearrangement regions and syntenic regions). The rearrangements and sequential variations in syntenic regions were used for the SV convert. The sequential variants which are embedded in the rearrangements were not included for further analysis. According to the definitions of sequence variation in SyRI outputs, we converted these variations into five types of SVs:

Insertion, Deletion, Duplication, Inversion, and Translocation. The identical type SVs with continuous (or overlapped) coordinates on a reference or query assembly were merged as a single SV which covered all sequences of continuous SVs. We calculated the density of SV breakpoints for each 10 Mb window along each chromosome. Then, all 10 Mb windows were ranked in descending order according to the numbers of SVs within the window. If the CDS of a gene is within an SV, this gene is defined as overlapping with an SV.

To confirm the identified inversions in EC-S1 and EC-N1 were not artificial results caused by, for example, assembly errors, we aligned the clean Hi-C data of EC-N1 against that of EC-S1, and EC-S1 against EC-N1, using Juicer V1.6 Hi-C data analysing pipeline (Durand *et al.*, 2016). The aligned results were visualized by JuiceBox V2.1.10 (Robinson *et al.*, 2018).

Analysis of structural variations in the S_FS and N_FS populations

We implemented three independent procedures to investigate and validate the distribution of the identified SVs in the S_FS and N_FS populations. We first employed Paragraph (Chen *et al.*, 2019) to genotype the SVs (Deletion, Duplication, Inversion, and Insertion) that have been confirmed in comparisons of the two assemblies (EC-S1 and EC-N1) in the 22 samples. Additionally, only the homozygous genotypes were used for SV frequency count. A genotyped SV with a higher presence frequency of 6 (half of the sample size) in one population over another is defined as a differentiated SV at the population level. SURVIVOR (Jeffares *et al.*, 2017) was used to merge the individual genotyped VCF files.

To confirm the presence of large inversions in the Chr3H and Chr4H in the wild barley populations, we designed polymerase chain reaction (PCR) primers to amplify the short sequence spanning the detected breakpoints of the inversions (Figure S11). The absence of the breakpoint would allow successful amplification of c. 200 bp sequence, while the presence of breakpoint would fail the amplification with no PCR product detectable. Primers were designed with BarleyVarDB (Tan *et al.*, 2020). Polymerase chain reactions were carried out in 10- μ L volume containing 50 μ M each of the primers, 200 μ M each dNTPs in BIO-TAQ system (Bioline Australia), and 1.5 mM MgCl₂, with a Veriti Thermal cycler. The PCR products were visualized on 2% agarose gels. Besides the 22 wild barley samples from the EC, we also included 8 wild barley samples collected from other parts of Israel (provided by Prof Nevo Eviatar) in the assay. Genomic DNA for PCRs was isolated from leaves of 3-week-old seedlings using the standard CTAB method.

For several significant SVs that were assumed to be the candidate functional variants linking to phenotypic divergence, we manually inspected the pattern of reads mapping at the corresponding genomic positions for all 22 samples using the Integrative Genomic Viewer (Robinson *et al.*, 2011). The presence and absence of the SVs in each sample were then manually identified (Figure S12).

Differentiated growth response under water stress

A deep, wide-spreading, and much-branched root system is one of the essential characteristics of drought tolerance (Ye *et al.*, 2018). To measure the differentiated capacity of drought tolerance of wild barley in the S_FS and N_FS, we compared the root growth of wild barley collected from the S_FS and N_FS

under southern hemisphere growing conditions in the glasshouse at Murdoch University (Perth, Western Australia). Seeds from three wild barley accessions collected from each slope were cleaned and germinated in Petri-dishes with wet filter papers at 25 °C and 60% relative humidity. The germinated seeds were then transferred to pots (25 cm in diameter) filled with potting mix soils with one plant per pot and no additional fertilizer. The pots were placed randomly on a bench. Our experiment simulated the natural drought environment with a long but gradual drought effect. At the experiment's beginning, all the pots, including in the control and drought treatments, were irrigated routinely with sufficient water until the seedlings were established with two leaves. We then only watered the wild barley in the control treatment and left the plants under drought treatment un-watered during the experiment period. The drought treatment lasted for 2 months before the plants were harvested. The fresh root weight and maximum length were measured.

Differentiation gene expression under water stress

Plant materials preparation

The experiment was conducted at Yangtze University (Jingzhou, China). Seeds of wild barley from the S_FS and N_FS were cleaned and germinated in Petri dishes with wet filter papers at 25 °C and 60% relative humidity. The germinated seeds were sown into a growth pot (9 cm top diameter, 5.5 cm bottom diameter, 8 cm height) with 120 g well-mixed substrate and vermiculite (3:1), in a growth chamber with a temperature of 20/18 °C (day/night) and light of 12/12 h (day/night). During this stage, 200 mL Hoagland nutrient solution was provided every 3 days until the two-leaf stage. Before drought stress treatment, all potted plants were calibrated for water by weighing. In the control group, the plants were watered every other day. The water supply was eliminated for drought treatment to form arid soil gradually, and the relative air humidity was controlled at 40%. The drought treatment lasted for a total of 18 days.

Differentiated growth response to drought

At the conclusion of drought treatment, five plants each from the S_FS and N_FS were harvested from both the control group and group under drought treatment, and their dry whole-plant biomass (DW) was measured after the plants were dried with an oven (105 °C for 30 min and then 75 °C for 2 weeks). Following the same procedure, dry whole-plant biomass was measured for S_FS and N_FS wild barley in control and under drought treatment after the 18-days drought treatment. For S_FS and N_FS wild barley, the relative growth rate was calculated as:

$$(DW_{\text{after_drought}} - DW_{\text{before_drought}}) / DW_{\text{before_drought}}$$

DNA preparation for differentiated gene expression

Fresh leaves from five plants of the S_FS and N_FS wild barley from the above drought treatment, including those in control with adequate water supply, were harvested at the beginning of the drought treatment and then 14 days into the drought treatment. These materials were separate from the plants harvested for biomass measurement. Total RNA was extracted with TRNzol Universal total RNA extraction Kit (TIANGEN)/TRIzol® Reagent (TIANGsEN) and treated with DNase I (GenStar). The integrity of the total RNA was examined by 2100 Bioanalyzer (Agilent Technologies) and quantified using the NanoDrop (Thermo Scientific, USA). Only a high-quality RNA sample

(OD260/280 = 1.8–2.2, OD260/230 \geq 2.0, RIN \geq 8, $>1 \mu\text{g}$) was used to construct the sequencing library. RNA-seq libraries were prepared with Illumina TruSeq RNA Library Prep Kits v2. PolyA mRNA was purified from total RNA using oligo-dT-attached magnetic beads and then fragmented by fragmentation buffer. The first-strand cDNA was synthesized using these short fragments as templates, with reverse transcriptase and random primers, followed by second stranded cDNA synthesis. The synthesized cDNA was then subjected to end-repair, phosphorylation, and 'A' base addition according to the library construction protocol. Sequencing adapters were then added to both strands of the cDNA fragments. After PCR amplification for cDNA fragments, the target fragments were clean. Paired-end sequencing (read length 150 bp) was then carried out on an Illumina platform.

Analysis of differentiated gene expression

Low-quality reads, adapters, and reads containing poly-N in the raw RNA-Seq data were removed using the fastp v 0.20.0 with default parameters (Chen *et al.*, 2018). Reads were mapped to EC-S1 reference genome using Hisat2 v 2.1.0 with default parameters (Kim *et al.*, 2015), and the number of mapped reads was counted using HTSeq version 0.6.1 (Anders *et al.*, 2015). RNA-Seq reads were normalized to TPM (Transcripts Per Kilobase Million) and low expression reads (TPM < 1) were removed. Differential expression analysis was performed using the DESeq2 R package 1.18.0 (Love *et al.*, 2014) between EC-S1 and EC-N1 samples and between S_FS and N_FS populations under control and drought treatment conditions, respectively. GO annotation of differentially expressed genes detected using the RNA-seq data was performed by TopGO (Alexa and Rahnenführer, 2009). We also re-analysed the transcriptome sequences of the S_FS and N_FS barley population previously reported (Dai *et al.*, 2014), following the pipelines as described above.

DNA sample preparation and short-read sequencing with Illumina technologies for population genetics analysis

Twenty-two additional samples (11 from the S_FS, 11 from the N_FS) were sampled from the 'Evolution Canyon'. Samples were collected in the middle of the slopes (sampling stations 6 and 7 for the N_FS, sampling stations 1 and 2 for the S_FS wild barley, Figure S1) to avoid possible mixture at the bottom of the canyon. Samples were collected at least 30 meters apart to minimize collecting genetically related samples from the same site. Mature and healthy leaves were harvested, dried, and stored in activated silica gel for processing. Genomic DNA was extracted from the leaves of the wild barley samples using Qiagen DNeasy kits (Qiagen). The DNA was evaluated for quality and concentration using agarose gels and an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies).

Approximately 1.5 μg of genomic DNA was first fragmented and then used to prepare Illumina paired-end libraries, following the manufacturer's instructions (Paired-End Sample Preparation Guide, Illumina, 1005063). Insert fragment lengths of 500 bp, 700 bp, and 1 kb were generated. Libraries for Illumina paired-end genome sequencing were constructed using Truseq Nano DNA HT Sample Preparation Kit (Illumina USA) following the standard manufacturer's protocol (Illumina). Libraries were sequenced on the HiSeq 2000 platform with a paired-end sequencing strategy at BGI-Shenzhen (Beijing Genomics Institute-Shenzhen, Shenzhen, China).

SNPs discovery

Clean reads were mapped to the EC-S1 assembly as reference genome using BWA MEM v0.7.17 with default parameters (Li, 2013). Default settings were used and duplicates were removed by Picard tools (<http://broadinstitute.github.io/picard/>). Reads were realigned by GATK v3.8-1-0 RealignerTargetCreator and IndelRealigner (McKenna *et al.*, 2010), followed by variant calling using GATK HaplotypeCaller. The resulting SNPs were filtered (QD < 2.0 || MQ < 40.0 || FS > 60.0 || QUAL < 60.0 || MQrankSum < -12.5 || Read-PosRankSum < -8.0) so that low-quality SNPs were removed. High-confidence SNPs were obtained by further filtering out the SNPs with minor allele frequency < 0.05 and missing genotype rate $> 10\%$ using VCFtools (Danecek *et al.*, 2011). Nucleotide diversity values (π) were calculated with PiXY (Korunes and Samuk, 2021) and F_{ST} with VCFtools (Danecek *et al.*, 2011) using a 100 kb window. Neighbour-joining phylogenetic trees were constructed based on SNPs with 1000 bootstraps using MEGA5 (Tamura *et al.*, 2011). Based on the SNP data, population structure was determined using ADMIXTURE (v1.3; Alexa and Rahnenführer, 2009) with a block-relaxation algorithm. Principal component analysis was performed with GCTA v1.9.2 (Yang *et al.*, 2011).

Analysis of genomic differentiation (F_{ST}) and identification of putative genes under selection

Genomic differentiation (F_{ST}) analysis was conducted by measuring the patterns of per SNP allele frequencies (Fixation index F_{ST}) using the R package 'fsthet' (Flanagan and Jones, 2017). To reduce the false positive elevated F_{ST} caused by other factors such as population demography, we apply the method to simulate the relationship between F_{ST} and heterozygosity to identify loci putatively under selection by comparing the empirical dataset to datasets simulated under a null model to identify outliers (Beaumont and Nichols, 1996; Flanagan and Jones, 2017). Genes with SNPs with excessively high F_{ST} value ($P < 0.05$) than neutral expectations were defined as genes putatively under differentiation selection.

We first focused on the functional divergence of two groups of genes related to well-differentiated flowering time and response to drought in the S_FS and N_FS populations, namely phenology related genes and drought response genes. Previously, 174 barley genes were described as putatively involved in phenology (He *et al.*, 2019; Hill *et al.*, 2019). Meanwhile, the drought stress gene database (DroughtDB) listed 199 genes as associated with drought response in plants (Alter *et al.*, 2015). These drought stress genes were involved in drought stress response in nine species (Alter *et al.*, 2015). Homology of these genes in the wild barley genomes was identified based on protein sequence similarity using BLASTP (e-value $< 1e^{-5}$ and identity $> 60\%$), and those genes containing significantly differentiated SNPs [with excessively high F_{ST} value ($P < 0.05$)] were identified.

Recent research suggested biotic stresses, such as pathogens, may also have asserted differentiated selective pressure shaping the genomic differentiation and consequent phenotypic divergence, such as differentiated resistance to plant disease in the EC system (Wang *et al.*, 2020; Yin *et al.*, 2015). We therefore examined the pattern of differentiation of disease-resistance related genes. To do so, we employed a comprehensive genome-wide gene discovery pipeline to identify the NBS-LRR disease resistance genes. First, we used all the NBS-LRR disease resistance

genes compiled by Habachi-Houimli *et al.* (2018) as queries to search against the EC-S1 genome sequences by BLASTP (Camacho *et al.*, 2009) with an *e* value cut-off of $1e^{-10}$. HMMER searches (version 3.2) were conducted on the sequences using NB-ARC HMM profile with an *e*-value of 0.01, as suggested by the HMMER user manual (Finn *et al.*, 2011). The hidden Markov model (HMM) profile of the NB-ARC domain (PF00931) was downloaded from Pfam (<http://pfam.xfam.org/>; Finn *et al.*, 2014). The gene sequences identified by both BLASTP and HMMsearch were retrieved as putative disease resistance genes, and further those genes containing significantly differentiated SNPs [with excessively high F_{ST} value ($P < 0.05$)] were identified.

Local principal component analysis

The program *lostruct* (local PCA/population structure, v.0.0.0.9) was used to detect genomic regions with abnormal population structures (Li and Ralph, 2019). 'lostruct' was run with 100 kb windows for each chromosome. 'lostruct' divides the genome into non-overlapping windows and calculates a PCA for each window. The results of PCAs derived from each window are used to calculate a similarity score. The matrix of similarity scores is then visualized using multidimensional scaling (MDS) transformation. The first MDS axis was then visualized by plotting the MDS score against the position of each window in the chromosome.

Local gene flow analysis

The effective number of migrants per generation (N_m) between the two populations was estimated using the private allele method with R package 'genepop' (Rousset, 2008). The private allele method implemented a multilocus estimate of N_m , and a sample-sized corrected estimate of N_m was given using the values from the closest regression curve (Barton and Slatkin, 1986). The SNP profiles of the 22 wild barley samples were used for the analysis. 'genepop' was run with 100 kb windows for each chromosome. To estimate the overall gene flow between S_{FS} and N_{FS} populations, we also ran 'genepop' with the whole genome SNPs after filtering with the LD method using Plink v1.9 (Chang *et al.*, 2015).

Local genomic recombination rate and deficiency of heterozygotes calculation

Genomic recombination rate (ρ) was calculated using FastEPRR v2 (Gao *et al.*, 2016) with 100 kb windows for each chromosome. The deficiency of heterozygotes was measured using inbreeding coefficient (F_{IS}) and was estimated across 100 kb windows using Plink v1.9 (--hardy; Chang *et al.*, 2015) and R package 'windowscanr' (<https://github.com/tavareshugo/WindowScanR>). Trend lines for F_{IS} and ρ (average within 1 Mbp window) were fitted by loess smoothing with 95% confidence intervals with R package ggplot2 (Wickham, 2016).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31471496), Grain Research and Development Corporation (9176507), and Yangtze University Scientific and Technological Innovation Team Foundation. We thank Claire Mérot and Peter Civián for constructive comments on an earlier version of this paper.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author contributions

CL, XT, WZ, EN, and TH perceived the project concept; WZ, RP, YX, KO, HH, LX, YX, XT, and CL conducted genome sequencing, assembly and annotation; CT, HH, TH, PH, BC, XZ, CBH, YH, HL, MB, RB, and CL conducted genome resequencing; EN, GZ, RP, XZ, TA, SW, DS, and WZ collected materials and conducted phenotyping; HH, CT, RP, and TH conducted data analysis; TH, CT, HH, YJ, WZ, and CL wrote and revised the paper with inputs from other authors.

Code availability

All custom codes are available from https://github.com/lakeseafly/Wild_barley_genomes_scripts.

Data availability statement

The EC_S1 and EC_N1 assemblies are available from China National GeneBank Database (CNP0003286); RNA-seq, NGS, Hi_C, and Nanopore data sets are available at NCBI under Bioproject accession PRJNA748178; Bionano data sets at NCBI Supplementary Files under accession SUPPF_0000004010 (EC_S1) and SUPPF_0000004011 (EC_N1). Gene annotation, variant results, and other results are accessible from: <https://doi.org/10.5281/zenodo.6859785>.

References

- Alexa, A. and Rahnenführer, J. (2009) Gene set enrichment analysis with topGO. *Bioconductor Improv.* **27**, 1–26.
- Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H. *et al.* (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, **182**, 145–161.
- Alter, S., Bader, K.C., Spannagl, M., Wang, Y., Bauer, E., Schön, C.C. and Mayer, K.F. (2015) DroughtDB: an expert-curated compilation of plant drought stress genes and their homologs in nine species. *Database*, **1**, 1.
- An, J., Li, Q., Yang, J., Zhang, G., Zhao, Z., Wu, Y., Wang, Y. *et al.* (2019) Wheat F-box protein TaFBA1 positively regulates plant drought tolerance but negatively regulates stomatal closure. *Front. Plant Sci.* **10**, 1242.
- Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
- Barton, N.H. and Slatkin, M.A. (1986) Quasi-Equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity*, **56**, 409–415.
- Bayer, P.E., Golicz, A.A., Scheben, A., Batley, J. and Edwards, D. (2020) Plant pan-genomes are the new reference. *Nat. Plants*, **6**, 914–920.
- Beaumont, M.A. and Nichols, R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **263**, 1619–1626.
- Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580.
- Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, **12**, 59–60.
- Burri, R. (2017) Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. *Mol. Ecol.* **26**, 3853–3856.
- Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O. and Shendure, J. (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.
- Charlesworth, B. (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* **15**, 538–543.
- Chen, S., Krusche, P., Dolzhenko, E., Sherman, R.M., Petrovski, R., Schlesinger, F., Kirsche, M. et al. (2019) Paragraph: a graph-based structural variant genotyper for short-read sequence data. *Genome Biol.* **20**, 1–3.
- Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.
- Civián, P., Drosou, K., Armisen-Gimenez, D., Duchemin, W., Salse, J. and Brown, T.A. (2021) Episodes of gene flow and selection during the evolutionary history of domesticated barley. *BMC Genomics*, **22**, 1–7.
- Considine, M.J. and Foyer, C.H. (2014) Redox regulation of plant development. *Antioxid Redox Signal*, **21**, 1305–1326.
- Crown, K.N., Miller, D.E., Sekelsky, J. and Hawley, R.S. (2018) Local inversion heterozygosity alters recombination throughout the genome. *Curr. Biol.* **28**, 2984–2990.
- Dai, F., Chen, Z.H., Wang, X., Li, Z., Jin, G., Wu, D., Cai, S. et al. (2014) Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. *Proc. Natl. Acad. Sci. USA*, **111**, 13403–13408.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Della Coletta, R., Qiu, Y., Ou, S., Hufford, M.B. and Hirsch, C.N. (2021) How the pan-genome is changing crop genomics and improvement. *Genome Biol.* **22**, 1–19.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
- Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M.A., Nardi, L. and Obermayer, R. (1998) Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann. Bot.* **82**, 17–26.
- Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S. and Aiden, E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98.
- Faria, R. and Navarro, A. (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A. et al. (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**(D1), D222–D230.
- Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**(suppl_2), W29–W37.
- Flanagan, S.P. and Jones, A.G. (2017) Constraints on the F_{ST} -heterozygosity outlier approach. *J. Hered.* **108**, 561–573.
- Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C. and Smit, A.F. (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA*, **117**, 9451–9457.
- Gao, F., Ming, C., Hu, W. and Li, H. (2016) New software for the fast estimation of population recombination rates (FastPRR) in the genomic era. *G3: Genes, Genomes, Genet.* **6**, 1563–1571.
- Goel, M., Sun, H., Jiao, W.B. and Schneeberger, K. (2019) SyRl: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277.
- Göktay, M., Fulgione, A. and Hancock, A.M. (2021) A new catalog of structural variants in 1,301 *A. thaliana* lines from africa, eurasia, and north america reveals a signature of balancing selection at defense response genes. *Mol. Biol. Evol.* **38**, 1498–1511.
- Haas, B.J., Salzberg, S.L., Zhu, W., Perte, M., Allen, J.E., Orvis, J., White, O. et al. (2008) Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, 1–22.
- Habachi-Houimli, Y., Khalfallah, Y., Mezghani-Khemakhem, M., Makni, H., Makni, M. and Bouktila, D. (2018) Genome-wide identification, characterization, and evolutionary analysis of NBS-encoding resistance genes in barley. *3 Biotech*, **8**, 1–6.
- Han, Y. and Wessler, S.R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199.
- He, T., Hill, C.B., Angessa, T.T., Zhang, X.Q., Chen, K., Moody, D., Telfer, P. et al. (2019) Gene-set association and epistatic analyses reveal complex gene interaction networks affecting flowering time in a worldwide barley collection. *J. Exp. Bot.* **70**, 5603–5616.
- Hill, C.B., Angessa, T.T., McFawn, L.A., Wong, D., Tibbits, J., Zhang, X.Q., Forrest, K. et al. (2019) Hybridisation-based target enrichment of phenology genes to dissect the genetic basis of yield and adaptation in barley. *Plant Biotechnol. J.* **17**, 932–944.
- Hoffmann, A.A. and Rieseberg, L.H. (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Ann. Rev. Ecol. Evol. Sys.* **39**, 21–42.
- Hu, H.F., Scheben, A., Verpaalen, B., Tirnaz, S., Bayer, P.E., Hodel, R.G.J., Batley, J. et al. (2022) Amborella gene presence/absence variation is associated with abiotic stress responses that may contribute to environmental adaptation. *New Phytol.* **233**, 1548–1555.
- Huang, K., Andrew, R.L., Owens, G.L., Ostevik, K.L. and Rieseberg, L.H. (2020) Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol. Ecol.* **29**, 2535–2549.
- Hübner, S., Höffken, M., Oren, E., Haseneyer, G., Stein, N., Graner, A., Schmid, K. et al. (2009) Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Mol. Ecol.* **18**, 1523–1536.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T. et al. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293.
- Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V.S., Gundlach, H., Monat, C., Lux, T. et al. (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, **588**, 284–289.
- Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F. et al. (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Comm.* **8**, 14061.
- Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J. and Zhang, J. (2017) WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* **59**, 86–101.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H. et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1340.
- Keilwagen, J., Hartung, F. and Grau, J. (2019) GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* **1962**, 161–177.
- Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
- Korunes, K.L. and Samuk, K. (2021) pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1568.
- Lagesen, K., Hallin, P.F., Rødland, E., Stærfeldt, H.H., Rognes, T. and Ussery, D.W. (2007) Rnammer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Res.* **35**, 3100–3108.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Li, H. (2013) *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *arXiv preprint arXiv:1303.3997*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Li, H. and Ralph, P. (2019) Local PCA shows how the effect of population structure differs along the genome. *Genetics*, **211**, 289–304.
- Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
- Liu, M., Li, Y., Ma, Y., Zhao, Q., Stiller, J., Feng, Q., Tian, Q. et al. (2020) The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. *Plant Biotech. J.* **18**, 443–456.

- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21.
- Lye, Z.N. and Purugganan, M.D. (2019) Copy number variation in domestication. *Trends Plant Sci.* **24**, 352–365.
- Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L. and Zimin, A. (2018) MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944.
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V. et al. (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature*, **544**, 427–433.
- Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C.S., Ens, J. et al. (2021) Long-read sequence assembly: a technical evaluation in barley. *Plant Cell*, **33**, 1888–1906.
- Mayer, K.F., Waugh, R., Langridge, P., Close, T.J., Wise, R.P., Graner, A., Matsumoto, T. et al. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K. et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Mérot, C., Oomen, R.A., Tigano, A. and Wellenreuther, M. (2020) A roadmap for understanding the evolutionary significance of structural genomic variation. *Trend Ecol. Evol.* **35**, 561–572.
- NCBI Resource Coordinators (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **41**, D8–D20.
- Nevo, E. (1995) Asian, African and European biota meet at 'Evolution Canyon' Israel: local tests of global biodiversity and genetic diversity patterns. *Proc. R. Soc. Biol. Sci.* **262**, 149–155.
- Nevo, E. (2012) "Evolution Canyon," a potential microscale monitor of global warming across life. *Proc. Natl. Acad. Sci. USA*, **109**, 2960–2965.
- Nevo, E. (2014) Evolution of wild barley at "Evolution Canyon": adaptation, pre-agricultural collection, and barley improvement. *Isr. J. Plant Sci.* **62**, 22–32.
- Nevo, E. and Chen, G. (2010) Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant Cell Environ.* **33**, 670–685.
- Ogiso, E., Takahashi, Y., Sasaki, T., Yano, M. and Izawa, T. (2010) The role of casein kinase II in flowering time regulation has diversified during evolution. *Plant Physiol.* **152**, 808–820.
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R., Hellinga, A.J., Lugo, C.S. et al. (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 1–8.
- Palmgren, M.G., Edenbrandt, A.K., Vedel, S.E., Andersen, M.M., Landes, X., Osterberg, J.T., Falhof, J. et al. (2015) Are we ready for back-to-nature crop breeding? *Trends Plant Sci.* **20**, 155–164.
- Pandey, S.P. and Somssich, I.E. (2009) The role of WRKY transcription factors in plant immunity. *Plant Physiol.* **150**, 1648–1655.
- Parnas, T. (2006) *Evidence for Incipient Sympatric Speciation in Wild Barley Hordeum Spontaneum*, at "Evolution Canyon", Mount Carmel, Israel, Based on Hybridization and Physiological and Genetic Diversity Estimates. Master degree thesis. The University of Haifa, Faculty of Science and Science Education, Department of Evolutionary and Environmental Biology.
- Pavlíček, T., Sharon, D., Kravchenko, V., Saaroni, H. and Nevo, E. (2003) Microclimatic interslope differences underlying biodiversity contrasts in "Evolution Canyon," Mt. Carmel, Israel. *Isr. J. Earth Sci.* **52**, 19.
- Perteau, M., Perteau, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295.
- Pinho, C. and Hey, J. (2010) Divergence with gene flow: models and data. *Ann. Rev. Ecol. Evol. Syst.* **41**, 215–230.
- Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., Bowers, J.E. et al. (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun.* **4**, 1827.
- Richardson, J.L., Urban, M.C., Bolnick, D.I. and Skelly, D.K. (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol. Evol.* **29**, 165–176.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotech.* **29**, 24–26.
- Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdóttir, H., Mesirov, J.P. and Aiden, E.L. (2018) Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256–258.
- Rousset, F. (2008) genePop'007: a complete re-implementation of the genePop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106.
- Rushton, P.J., Reinstadler, A., Lipka, V., Lippok, B. and Somssich, I.E. (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen-and-wound-induced signalling. *Plant Cell*, **14**, 749–762.
- Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L. et al. (2014) Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E. et al. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259.
- Shah, N., Wakabayashi, T., Kawamura, Y., Skovbjerg, C.K., Wang, M.Z., Mustamin, Y., Isomura, Y. et al. (2020) Extreme genetic signatures of local adaptation during *Lotus japonicus* colonization of Japan. *Nat. Commun.* **11**, 1–15.
- Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.
- Stanke, M. and Morgenstern, B. (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467.
- Sugano, S., Andronis, C., Ong, M.S., Green, R.M. and Tobin, E.M. (1999) The protein kinase CK2 is involved in regulation of circadian rhythms in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **96**, 12362–12366.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tan, C., Chapman, B., Wang, P., Zhang, Q., Zhou, G., Zhang, X.Q., Barrero, R.A. et al. (2020) BarleyVarDB: a database of barley genomic variation. *Database*, **2020**, baaa091.
- Tao, Y., Luo, H., Xu, J., Cruickshank, A., Zhao, X., Teng, F., Hathorn, A. et al. (2021) Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat. Plants*, **7**, 766–773.
- Tao, Z., Kou, Y., Liu, H., Li, X., Xiao, J. and Wang, S. (2011) OsWRKY45 alleles play different roles in abscisic acid signalling and salt stress tolerance but similar roles in drought and cold tolerance in rice. *J. Expt. Bot.* **62**, 4863–4874.
- Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.S., Soudi, S., Burge, D.O., Huang, K. et al. (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, **584**, 602–607.
- Van Doren, B.M., Campagna, L., Helm, B., Illera, J.C., Lovette, I.J. and Liedvogel, M. (2017) Correlated patterns of genetic diversity and differentiation across an avian family. *Mol. Ecol.* **26**, 3982–3997.
- Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P. and Wolf, J.B. (2016) Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Comm.* **7**, 13195.
- Wang, H., Yin, H., Jiao, C., Fang, X., Wang, G., Li, G., Ni, F. et al. (2020) Sympatric speciation of wild emmer wheat driven by ecology and chromosomal rearrangements. *Proc. Natl. Acad. Sci. USA*, **117**, 5955–5963.
- Wang, X., Aguirre, L., Rodríguez-Leal, D., Hendelman, A., Benoit, M. and Lippman, Z.B. (2021) Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat. Plants*, **7**, 419–427.
- Wang, X. and Wang, L. (2016) GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Front. Plant Sci.* **7**, 1350.
- Wellenreuther, M. and Bernatchez, L. (2018) Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* **33**, 427–440.
- Wellenreuther, M., Mérot, C., Berdan, E. and Bernatchez, L. (2019) Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Mol. Ecol.* **28**, 1203–1209.
- Wick, R.R., Judd, L.M. and Holt, K.E. (2019) Performance of neural network base calling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129.

- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wolf, J.B.W. and Ellegren, H. (2016) Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* **18**, 87–100.
- Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G. et al. (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **47**, W52–W58.
- Xu, Z. and Wang, H. (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268.
- Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82.
- Ye, H., Rookiwal, M., Valliyodan, B., Zhou, L., Chen, P., Varshney, R.K. and Nguyen, H.T. (2018) Genetic diversity of root system architecture in response to drought stress in grain legumes. *J. Expt. Bot.* **69**, 3267–3277.
- Yin, H., Ben-Abu, Y., Wang, H., Li, A., Nevo, E. and Kong, L. (2015) Natural selection causes adaptive genetic resistance in wild emmer wheat against powdery mildew at “evolution canyon” microsite, mt. Carmel, Israel. *PLoS One*, **10**, e0122344.
- Zohary, D., Hopf, M. and Weiss, E. (2012) *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford: Oxford University Press.
- Zong, S.B., Li, Y.L. and Liu, J.X. (2021) Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. *Mol. Biol. Evol.* **38**, 1317–1329.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 The S_{FS} and N_{FS} wild barley populations in the Evolution Canyon, Israel, and the contrasting microclimate of the two slopes.

Figure S2 A. Orthologous clustering analysis of annotated high-confidence genes of the two wild barley genome assemblies and the reference genome of cultivar Morex.

Figure S3 Size and probability density of major SVs between EC-S1 and EC-N1.

Figure S4 Identification and validation of large inversions in chromosome Chr3H (A) and Chr4H (B) by mapping Hi-C data.

Figure S5 Pattern of specific types of SVs in the seven chromosomes.

Figure S6 Polymerase chain reactions show the population-level differentiation of large inversion in Chr3H and Chr4H.

Figure S7 The genes impacted by SVs showed significantly different gene expression compared with genes not under SVs impact in the EC-S1 and EC-N1 wild barley under drought-stressed and un-stressed growing conditions ($P < 0.05$).

Figure S8 Heterogeneous pattern of genomic differentiation (F_{ST}), gene flow (Nm), deficiency of heterozygotes (F_{IS}), recombination rate (Rho), and window-based genetic relatedness (MDS1) for Chr5H, Chr6H, and Chr7H.

Figure S9 The level of gene flow Nm was not uniformly correlated with genetic relatedness in the genomic window

MDS1 (A) and genetic differentiation F_{ST} (B) in the genomic region.

Figure S10 Identification and validation of large inversions by mapping EC-S1 and EC-N1 HiC data to Morex reference genomes to confirm the accuracy of wild barley HiC data.

Figure S11 Primers and sequences for confirmation of population-level differentiation of large inversions in Chr3H and Chr4H.

Figure S12 Manually inspecting population-level differentiation of SVs with Integrative Genomics Viewer.

Data S1 Table 1 Summary of genome assembly and gene and TE annotation of EC-S1 and EC-N1.

Table 2 Functional annotation of high-confidence genes in the EC-S1 assembly using EggNOG-mapper.

Table 3 Functional annotation of high-confidence genes in the EC-N1 assembly using EggNOG-mapper.

Table 4 Number of genes with functional annotation and known Pfam domains.

Table 5 Detection of telomeric repeats in EC-S1 and EC-N1 genomes.

Table 6 The genome-wide BARE TE detection in EC-S1, EC-N1, and Morex.

Table 7 GO enrichment of gene clusters shared between wild barley EC-S1 and EC-N1 assemblies.

Table 8 GO enrichment of gene clusters shared between Morex and EC-S1.

Table 9 GO enrichment of gene clusters shared between Morex and EC-N1.

Table 10 HiC mapping rate among different genomes.

Data S2 Table 1 Geographical location of 22 samples used in this study and the summary statistics of Illumina re-sequencing data per sample.

Table 2 The summary statistics of SNP and InDels variants calling using the WGS sequencing data.

Table 3 List of genes under putative selection and associated with flowering time, drought response, and disease resistance.

Data S3 Table 1 List of structural variations detected by Syri.

Table 2 Summary of identified SVs.

Table 3 Genotyping the SVs (INV, INS, DUP, and DEL) identified by Syri at the population level. 1: this genotype is present 0: this genotype is absent.

Table 4 Summary of the genotyped SVs (INV, INS, DUP, and DEL) identified by Syri at the population level.

Table 5 List of genes that are overlapped and affected by SVs.

Table 6 GO enrichment of genes overlapped with structural variations.

Data S4 Table 1 DEGs identified in both EC-S1 and EC-N1 samples by comparing the control and water stress condition.

Table 2 Average expression level (TPM) between non-SV genes and SV genes in four different treatments.

Table 3 Differentially expressed Non-SV genes and SV genes between EC-S1 and EC-N1 under control and drought treatment conditions.

Table 4 GO annotation of DEGs identified in both EC-S1 and EC-N1 samples by comparing the control and water stress conditions.