



HAL
open science

Adaptive diversification through structural variation in barley

Murukarthick Jayakodi, Qiongxian Lu, Helene Pidon, M. Timothy Rabanus-Wallace, Micha Bayer, Thomas Lux, Yu Guo, Benjamin Jaegle, Ana Badea, Wubishet Bekele, et al.

► **To cite this version:**

Murukarthick Jayakodi, Qiongxian Lu, Helene Pidon, M. Timothy Rabanus-Wallace, Micha Bayer, et al.. Adaptive diversification through structural variation in barley. 2024. hal-04493576

HAL Id: hal-04493576

<https://hal.inrae.fr/hal-04493576v1>

Preprint submitted on 7 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Adaptive diversification through structural variation in barley**

2 **Murukarthick Jayakodi^{1,31}, Qiongxin Lu^{2,31}, H el ene Pidon^{1,3,31}, M. Timothy Rabanus-**
3 **Wallace^{1,31}, Micha Bayer⁴, Thomas Lux⁵, Yu Guo¹, Benjamin Jaegle⁶, Ana Badea⁷, Wubishet**
4 **Bekele⁸, Gurcharn S. Brar⁹, Katarzyna Braune², Boyke Bunk¹⁰, Kenneth J. Chalmers¹¹, Brett**
5 **Chapman¹², Morten Egevang J orgensen², Jia-Wu Feng¹, Manuel Feser¹, Anne Fiebig¹,**
6 **Heidrun Gundlach⁵, Wenbin Guo⁴, Georg Haberer⁵, Mats Hansson¹³, Axel Himmelbach¹,**
7 **Iris Hoffie¹, Robert E. Hoffie¹, Haifei Hu^{12,14}, Sachiko Isobe¹⁵, Patrick K onig¹, Sandip M.**
8 **Kale², Nadia Kamal⁵, Gabriel Keeble-Gagn ere¹⁶, Beat Keller⁶, Manuela Knauft¹, Ravi**
9 **Koppolu¹, Simon G. Krattinger¹⁷, Jochen Kumlehn¹, Peter Langridge¹¹, Chengdao Li^{12,18},**
10 **Marina P. Marone¹, Andreas Maurer¹⁹, Klaus F.X. Mayer^{5,20}, Michael Melzer¹, Gary J.**
11 **Muehlbauer²¹, Emiko Murozuka², Sudharsan Padmarasu¹, Dragan Perovic²², Klaus Pillen¹⁹,**
12 **Pierre A. Pin²³, Curtis J. Pozniak²⁴, Luke Ramsay⁴, Pai Rosager Pedas², Twan Rutten¹, Shun**
13 **Sakuma²⁵, Kazuhiro Sato^{26,15}, Danuta Sch uler¹, Thomas Schmutzer¹⁹, Uwe Scholz¹, Miriam**
14 **Schreiber⁴, Kenta Shirasawa¹⁵, Craig Simpson⁴, Birgitte Skadhauge², Manuel Spannagl⁵,**
15 **Brian J. Steffenson²⁷, Hanne C. Thomsen², Josquin F. Tibbits¹⁶, Martin Toft Simmelsgaard**
16 **Nielsen², Corinna Trautewig¹, Dominique Vequaud²³, Cynthia Voss², Penghao Wang¹²,**
17 **Robbie Waugh^{4,28}, Sharon Westcott¹², Magnus Wohlfahrt Rasmussen², Runxuan Zhang⁴,**
18 **Xiao-Qi Zhang¹², Thomas Wicker⁶ , Christoph Dockter² , Martin Mascher^{1,29} & Nils**
19 **Stein^{1,30}**

20 ¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland,
21 Germany, ²Carlsberg Research Laboratory, Copenhagen, Denmark, ³IPSiM, Univ Montpellier,
22 CNRS, INRAE, Institut Agro, Montpellier, France, ⁴The James Hutton Institute, Dundee,
23 UK, ⁵PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich - German
24 Research Center for Environmental Health, Neuherberg, Germany, ⁶Department of Plant
25 and Microbial Biology, University of Zurich, Zurich, Switzerland, ⁷Brandon Research and
26 Development Centre, Agriculture et Agri-Food Canada, Brandon, Canada, ⁸Ottawa Research
27 and Development Centre, Agriculture et Agri-Food Canada, Ottawa, Canada, ⁹Faculty of
28 Land and Food Systems, The University of British Columbia, Vancouver, Canada, ¹⁰DSMZ-
29 German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig,
30 Germany, ¹¹School of Agriculture, Food and Wine, University of Adelaide, Urrbrae,
31 Australia, ¹²Western Crop Genetics Alliance, Food Futures Institute/School of Agriculture,
32 Murdoch University, Murdoch, Australia, ¹³Department of Biology, Lund University, Lund,
33 Sweden, ¹⁴Rice Research Institute, Guangdong Academy of Agricultural Sciences,
34 Guangzhou, China, ¹⁵Kazusa DNA Research Institute, Kisarazu, Japan, ¹⁶Agriculture Victoria,
35 Department of Jobs, Precincts and Regions, Agribio, La Trobe University, Victoria, Bundoora,
36 Australia, ¹⁷Plant Science Program, Biological and Environmental Science and Engineering
37 Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi
38 Arabia, ¹⁸Department of Primary Industry and Regional Development, Government of
39 Western Australia, Perth, Australia, ¹⁹Institute of Agricultural and Nutritional Sciences,
40 Martin Luther University Halle-Wittenberg, Halle, Germany, ²⁰School of Life Sciences
41 Weihenstephan, Technical University Munich, Freising, Germany, ²¹Department of
42 Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota, USA, ²²Institute
43 for Resistance Research and Stress Tolerance, Julius Kuehn-Institute (JKI), Federal Research
44 Centre for Cultivated Plants, Quedlinburg, Germany, ²³SECOBRA Recherches, Maule,
45 France, ²⁴Department of Plant Sciences and Crop Development Centre, University of

46 Saskatchewan, Saskatoon, SK, Canada, ²⁵Faculty of Agriculture, Tottori University, Tottori,
47 Japan, ²⁶Institute of Plant Science and Resources, Okayama University, Kurashiki,
48 Japan, ²⁷Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota,
49 USA, ²⁸School of Life Sciences, University of Dundee, Dundee, UK, ²⁹German Centre for
50 Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, ³⁰Center for
51 Integrated Breeding Research, Georg-August-University, Göttingen, Germany. ³¹These
52 authors contributed equally to this work.

53 ✉ email: wicker@botinst.uzh.ch; christoph.dockter@carlsberg.com; [mascher@ipk-](mailto:mascher@ipk-gatersleben.de)
54 gatersleben.de; stein@ipk-gatersleben.de.

55
56 **Pangenomes are collections of annotated genome sequences of multiple individuals of a**
57 **species. The structural variants uncovered by these datasets are a major asset to genetic**
58 **analysis in crop plants. Here, we report a pangenome of barley comprising long-read**
59 **sequence assemblies of 76 wild and domesticated genomes and short-read sequence data**
60 **of 1,315 genotypes. An expanded catalogue of sequence variation in the crop includes**
61 **structurally complex loci that have become hot spots of gene copy number variation in**
62 **evolutionarily recent times. To demonstrate the utility of the pangenome, we focus on four**
63 **loci involved in disease resistance, plant architecture, nutrient release, and trichome**
64 **development. Novel allelic variation at a powdery mildew resistance locus and population-**
65 **specific copy number gains in a regulator of vegetative branching were found. Expansion of**
66 **a family of starch-cleaving enzymes in elite malting barleys was linked to shifts in enzymatic**
67 **activity in micro-malting trials. Deletion of an enhancer motif is likely to change the**
68 **developmental trajectory of the hairy appendages on barley grains. Our findings indicate**
69 **that rapid evolution at structurally complex loci may have helped crop plants adapt to new**
70 **selective regimes in agricultural ecosystems.**

71
72 Reliable crop yields fueled the rise of human civilizations. As people embraced a new way of
73 life, cultivated plants, too, had to adapt to the needs of their domesticators. There are
74 different adaptive requirements in a wild compared to an arable habitat. Crop plants and their
75 wild progenitors differ in how many vegetative branches they initiate or how many seeds or
76 fruits they produce and when. For example, barley (*Hordeum vulgare*) in six-rowed forms of
77 the crops, thrice as many grains set as in the ancestral two-rowed forms. This change was
78 brought about by knock-out mutations¹ of a recently evolved regulator² of inflorescence
79 development. Consequently, six-rowed barleys came to predominate in most barley-growing
80 regions³. Taking a broader view of the environment as a set of exogeneous factors that drive
81 natural selection, barley provides another fascinating, and economically important example.
82 The process of malting involves the sprouting of moist barley grains, driving the release of
83 enzymes that break down starch into fermentable sugars. In the wild, various environmental
84 cues can trigger germination to improve the odds of the emerging seedling encountering
85 favorable weather conditions for subsequent growth⁴. In the malt house, by contrast,
86 germination of modern varieties has to be fast and uniform to satisfy the desired specifications
87 of the industry. In addition to these examples, traits such as disease resistance, plant
88 architecture and nutrient use have been both a focus for plant breeders and studied
89 intensively barley geneticists⁵. While barley genetic analyses flourished during a “classical”
90 period⁶ in the first half of the 20th century, it started to lag behind small-genome models due

91 to difficulties in adapting molecular biology techniques to a large genome rich in repeats⁷.
92 However, interest in barley as diploid model for temperate cereals has surged again as DNA
93 sequencing became more powerful. High-quality sequences of several barley genomes have
94 been recently assembled⁸. New sequencing technologies have shifted the focus of barley
95 genomics: from the modest ambition of a physical map of all genes to a “pangenome”, i.e.
96 near-complete sequence assemblies⁹ of many genomes. Here, we report a pangenome
97 comprising 76 chromosome-scale sequences assembled from long-reads as well as short-read
98 sequences of 1,315 barley genomes. These data in conjunction with genetic and genomic
99 analyses provide insights into the effects of structural variation at loci related to crop evolution
100 and adaptation.

101

102 *An expanded annotated pangenome of barley*

103

104 As in previous diversity studies^{8,10}, we aimed for a judicious mix of representativeness,
105 diversity and integration with community resources (**Fig. 1a**, **Extended Data Fig. 1a-c**,
106 **Supplementary Table 1**). We selected (i) diverse domesticated germplasm with a focus on
107 genebank accessions from barley’s center of diversity in the Middle East; (ii) 23 accessions of
108 barley’s conspecific wild progenitor *H. vulgare* subsp. *spontaneum* from across that taxon’s
109 geographic range (**Extended Data Fig. 1d**); and (iii) cultivars of agronomic or scientific
110 relevance. Examples of the last category are Bonus, Foma and Bowman, three parents of
111 classical mutants¹¹. Genome sequences of each accession were assembled to contig-level
112 from PacBio HiFi accurate long reads¹² and scaffolded with conformation capture sequencing
113 (Hi-C) data¹³ to chromosome-scale pseudomolecules (**Extended Data Fig. 2a**, **Supplementary**
114 **Table 1**). Gene models were annotated with the help of transcriptional evidence and
115 homology. Illumina RNA sequencing and PacBio isoform sequencing of five different tissues
116 (**Supplementary Table 2**) were generated for 20 accessions. Gene models predicted in these
117 genomes were projected onto the remaining 56 sequence assemblies (**Supplementary Table**
118 **3**). Out of 4,896 single-copy genes conserved across the Poales, on average fewer than 92
119 (1.9%) were absent in the pangenome annotations (**Supplementary Table 3**). Our assemblies
120 also met the other quality metrics proposed by the EarthBiogenome project¹⁴
121 (**Supplementary Table 1**).

122

123 *An atlas of structural variation*

124

125 Gene content variation was abundant in the barley pangenome. The gene models in the 76
126 genomes were clustered into 95,735 orthologous groups (**Extended Data Fig. 3**), of which only
127 16,672 (17.4%) were present in all 76 genotypes. Of these groups, 14,736 had a single
128 representative in each of the genomes. At the level of individual gene models, a third were
129 considered conserved because they belong to an orthologous group with representatives
130 from each accession (**Extended Data Fig. 3b**). As expected for conspecific populations
131 connected by gene flow, wild and domesticated barleys were not strongly differentiated in
132 their gene content: of 78,565 orthologous groups subject to presence/absence variation, only
133 863 and 397 were private to wild and domesticated barleys, respectively. The functional
134 annotations of clusters restricted to specific gene pools (wild forms, landraces, cultivars and
135 combinations of these groups) pointed to an involvement in biotic and abiotic stress responses
136 (**Supplementary Table 4**).

137 To expand the catalogue of presence/absence variants (PAV), insertion and deletions (indels)
138 and polymorphic inversions, we aligned the genome sequences and detected structural
139 variants (**Fig. 1b**, **Extended Data Fig. 2b-d**, **Extended Data Fig. 3c**). Noteworthy were two
140 reciprocal interchromosomal translocations, the first in HOR 14273, an Iranian landrace, and
141 the second in HID055, a wild barley from Turkey (**Fig. 1b**). The latter event joins the short arm
142 of chromosome 2H with the long arm of chromosome 4H (and vice versa) and manifests itself
143 in a biparental population between HID055 and Barke¹⁵ in interchromosomal linkage (**Fig. 1c**)
144 and incomplete seed set in the offspring. This illustrates that inadvertent selection of
145 germplasm with structural variants can create obstacles for the use of plant genetic resources.
146 The presence of both wild and domesticated barleys in our panel made it possible to compare
147 the levels of structural diversity in the two taxa. Graph structures tabulating the presence and
148 absence of single-copy loci in individual genomes⁸ grew faster in wild than in cultivated forms
149 (**Fig. 1d**): a larger amount of single-copy sequence was present in 23 wild barley genomes than
150 in 53 genomes of the domesticate. This pattern was also seen in a whole-genome graph
151 constructed with minigraph¹⁶ (**Extended Data Fig. 4e**). The pangenome graph improves the
152 accuracy of read alignment and variant calling: more reads were aligned as proper pairs, and
153 with fewer mismatches, to the graph than to a single reference genome (**Extended Data Fig.**
154 **4b**). The genome-wide distribution of structural variants encapsulated in the graph matched
155 that inferred from pairwise alignments (**Extended Data Fig. 4c-d**). However, owing to high
156 computational requirements¹⁷, pangenome graph construction with packages supporting
157 small variants (< 50 bp) is still computationally prohibitive in barley.
158 Despite domestication bottlenecks, genetic diversity is high in cultivated barley⁵. To quantify
159 the completeness of the haplotype inventory of our pangenome, we compared our
160 assemblies against short-read data of a global diversity panel (**Supplementary Table 5**). A core
161 set of 1,000 genotypes selected from a collection of 22,626 barleys³ was sequenced to three-
162 fold haploid genome coverage. Nested therein, 200 genomes⁸ were sequenced to 10-fold
163 depth and the gene space of 46 accessions was represented in the contigs assembled from
164 50-fold short-read data (**Extended Data Fig. 5a**, **Supplementary Table 6**). A total of 315 elite
165 cultivars of European ancestry were sequenced to 3-fold coverage (**Extended Data Fig. 5a**,
166 **Supplementary Table 5**). More than 164.5 million single-nucleotide polymorphisms (SNPs)
167 and indels were detected across all panels (**Extended Data Fig. 5b**). Overlaying these with the
168 pangenome showed that the 76 chromosome-scale assemblies captured almost all pericentric
169 haplotypes of cultivated barley (**Extended Data Fig. 2d-f**). Coverage decreased to as low as
170 50% in distal regions, where haplotypes of plant genetic resources lacked a close relative in
171 the pangenome more often than those of elite cultivars (**Extended Data Fig. 2e-f**). This
172 suggests that, thanks to broad taxon sampling, short-read sequencing will remain
173 indispensable for the time being, but in the future population-scale long-read sequencing¹⁸
174 will be a desirable in agricultural genetics as it is in medical genetics.

175 176 *An inventory of complex loci*

177
178 Long-read sequencing has the power to resolve structurally complex genomic regions, where
179 repeated cycles of tandem duplication, mutation of duplicated genes and elimination by
180 deletion or recombination have created a panoply of diverged copies of one or multiple genes
181 in varied arrangements (**Extended Data Fig. 6a**). Many complex loci are intimately linked to
182 the evolution of resistance genes¹⁹. An illustrative example is barley's *Mildew resistance locus*
183 *a* (*Mla*)^{20,21}, which contains three families of resistance gene homologs, each with multiple

184 members at the locus. A 40 kb region containing members of two families is repeated four
185 times head-to-tail in RGT Planet, but is not present in even a single complete copy in 62
186 accessions of our pangenome (**Extended Data Fig. 6b-c**). *Mla* genes *sensu strictu*, i.e. those
187 that have been experimentally proven to provide functional powdery mildew resistance, are
188 among members of a subfamily that resides outside of this duplication but close to its distal
189 border (**Fig. 2a-b, Extended Data Fig. 6b-c**). Twenty-nine *Mla* alleles in the narrow sense have
190 been defined to date²². Gene models identical to seven were identified in our pangenome
191 (**Fig. 2a**). However, the sequence variation went beyond this observation: 149 unique gene
192 models were different from, but highly similar to known *Mla* alleles, with nucleotide
193 sequences at least 98% identical. Some of these genes were present in multiple copies. HOR
194 8117, a landrace from Nepal, contained 11 different close homologs of *Mla*, two of which were
195 present in five copies each (**Supplementary Fig. 1**). Genome sequences alone cannot inform
196 us how this sequence diversity relates to resistance to powdery mildew or other diseases²³.
197 Until the advent of long-read sequencing, it was virtually impossible to resolve the structure
198 of the *Mla* locus in multiple genomes at once, but now it is a corollary of pangenomics.
199 We employed a gene-agnostic method²⁴ to scan the genome sequence of Morex for
200 structurally complex loci harbouring genes, focusing on examples that had evidently caused
201 gene copy number variation across the pangenome via the expansion or collapse of long
202 tandem repeats. A total of 173 loci ranging in size from 20 kb to 2.2 Mb (median: 125 kb)
203 matched our criteria (**Fig. 2c, Supplementary Table 7**). Their copy numbers were variable in
204 the pangenome. The most extreme case was a cluster of genes annotated by homology as
205 thionin genes, which are possibly involved in resistance to herbivory²⁵. The locus had as few
206 as three thionin gene copies in the wild barley WBDC103 and up to 78 copies in WBDC199,
207 another wild barley (**Extended Data Fig. 6d**). Genes associated with such complex loci
208 possessed functional annotations suggesting involvement in various biological processes (**Fig.**
209 **2c, Supplementary Table 7**). Complex loci were enriched in distal chromosomal regions (**Fig.**
210 **2d**). In this regard, they follow the same distal-to-proximal gradient as genetic diversity and
211 recombination frequency. The latter process might play a role in their amplification and
212 contraction owing to unequal homologous recombination between neighboring repeat units²⁶
213 (**Extended Data Fig. 6a**). Molecular dating of the tandem duplications in Morex is consistent
214 with rapid evolution (**Extended Data Fig. 7**): loci with many gene copies appear to have gained
215 them within the last three million years (**Extended Data Fig. 7c**), after the *H. vulgare* lineage
216 split from that of its closest relative *H. bulbosum*²⁷. In addition, 63 loci (36.4%) underwent at
217 least one duplication in the last 10,000 years, that is, after domestication (**Extended Data Fig.**
218 **7d**). Forty-five loci expanded so recently that the genes they harboured were identical
219 duplicates of each other.
220 One interesting case of such recent diversification was a duplication at the *HvTB1* locus (also
221 known as *INTERMEDIUM-C* [*INT-C*] or *SIX ROWED SPIKE 5*). *HvTB1* is a TEOSINTE BRANCHED
222 1, CYCLOIDEA, PCF1 (TCP) transcription factor involved in basal branching (tillering) and other
223 aspects of plant architecture in cereal grasses²⁸⁻³⁰. In barley, both tillering and the fertility of
224 lateral spikelets is increased in knock-out mutants^{30,31}. Just two alleles, *Int-c.a* and *int-c.b*.
225 dominate in six-rowed and two-rowed forms³⁰, respectively, and *HvTB1* is not genetically
226 linked to the *SIX ROWED SPIKE 1* gene. Both alleles of *HvTB1* are thought to be functional and
227 occur also in wild barley^{30,32}. These patterns have defied easy explanation. Expression
228 differences owing to regulatory variation have been postulated but not proven³⁰. The
229 pangenome adds another twist. *HvTB1* is a single-copy gene in all 22 *H. spontaneum*
230 accessions and 23 two-rowed domesticates except HOR 7385 (**Supplementary Table 8**). Six-

231 rowed forms, however, have up to four copies of a 21 kb segment that contains *HvTB1* and ~5
232 kb of its upstream sequence (**Fig. 2b**). The reference cultivar Morex has three copies, although
233 these were falsely collapsed in previous short-read assemblies of that variety³³. On top of
234 variable copy numbers, the pangenome revealed six hitherto unknown HvTB1 protein variants
235 (**Extended Data Fig. 6d, Supplementary Table 8**). Reduced tillering in maize has been
236 attributed to overexpression of *TB1*. The barley pangenome will help developmental
237 geneticists reveal if copy number gains had analogous effects in six-rowed forms.

238

239 *Amplification of α -amylases in malting barley*

240

241 Among the complex loci we examined, the *amy1_1* locus of α -amylases is arguably the one of
242 greatest economic importance. These enzymes cleave the polysaccharide starch into short-
243 chain forms, which are then digested further into sugars³⁴. In both wild and cultivated forms,
244 the speed and efficiency of that process determines the energy supply to and hence the vigor
245 of the young seedling³⁵. In grains of domesticated barley, the enzymatic conversion of starch
246 into fermentable sugars by α -amylases initiates the malting process. Barley α -amylases are
247 subdivided into four families, which occupy distinct genomic loci (**Extended Data Fig. 8a,**
248 **Supplementary Tables 9 and 10**). Earlier genome sequences assembled from short reads
249 hinted at the presence of structural variation at the *amy1_1* locus on chromosome 6H,
250 respectively, but failed to resolve copy numbers³⁶. By contrast, each of our long-read
251 assemblies covered *amy1_1* in a single contig (**Extended Data Fig. 9a**). Copy numbers of
252 *amy1_1* in 76 complete genomes varied between two and eight, with on average more copies
253 in domesticated than in wild forms (**Fig. 3a, b**). Individual copies were addressable by 21-mers
254 that overlap sequence variants. We counted these 21-mers in the short-reads of 1,315
255 genotypes and also determined SNP haplotypes around the *amy1_1* locus in these data
256 (**Extended Data Figs. 8e-f, 9b, Supplementary Tables 11 and 12**). Eight clusters were
257 discernible and could be related to population structure. Three-quarters of hulless barleys
258 were in cluster #7. Six-rowed barleys belonged mostly to clusters #1 and #6. Among 315
259 European varieties, clusters #5 and #6 were most common. Clusters #3 and #8 with fewer
260 *amy1_1* copies were exclusive to plant genetic resources. Barleys from eastern and central
261 Asian countries tend to have high copy numbers. *Amy1_1* copy numbers were higher on
262 average in elite varieties than in other barleys (**Fig. 3b**). Structural diversity was accompanied
263 by differences in gene sequence owing to SNPs and indels in open reading frames and
264 promoters. The 76 genome assemblies had 94 distinct *amy1_1* haplotypes (**Fig. 3c, Extended**
265 **Data Fig. 8b, Supplementary Tables 13-16**). Twelve had insertions of transposable elements
266 (**Supplementary Table 17**). At the protein level, there were 38 unique AMY1_1 isoforms
267 (**Supplementary Tables 18 and 19**), some of which were predicted to affect protein³⁷ stability
268 and thereby influence α -amylase activity (**Fig. 3d, e**).

269 We investigated in more detail the elite malting barleys Morex, Barke and RGT Planet (**Fig. 3,**
270 **Supplementary Tables 20 and 21**). Prior to its use as a genome reference cultivar, Morex was
271 a successful variety in North America. It had six nearly identical (> 99 % similarity) *amy1_1*
272 copies. (**Fig. 3a**). The fifth copy was disrupted by the insertion of a transposable element. Full-
273 length copies were verified by PacBio amplicon sequencing. Barke, a European cultivar, had
274 six full-length copies, albeit of a different haplotype. RGT Planet, currently a successful cultivar
275 in many barley-growing regions around the world, had five copies, one of which was likely to
276 be inactivated by a 32 bp deletion in a pyr-box (CTTT(A/T) core) promoter binding site that is
277 essential for α -amylase transcription³⁸. We tested overall α -amylase activity in micro-malting

278 trials with RGT Planet and near-isogenic lines (NILs) that carried Morex and Barke *amy1_1*
279 haplotypes in the genomic background of RGT Planet. It was observed that α -amylase activity
280 was highest in *amy1_1*-Barke NILs (**Fig. 3e**). The Barke haplotype is common not only in
281 cultivars favored by European maltsters, but also among those from other regions of the
282 world, where barley α -amylases need to be abundant enough to cleave starch from adjuncts
283 such as maize and rice (**Supplementary Table 22**). The patterns of sequence variation at
284 *amy1_1* uncovered by the barley pangenome pave the way for the targeted deployment,
285 possibly even design, of *amy1_1* haplotypes in breeding.

286

287 *A regulatory variation controls trichome development*

288

289 Our last example sits at the intersection of developmental genetics, breeding and
290 domestication. Hairy appendages to grains and awns are conducive to seed dispersal in wild
291 plants, but have lost this function in domesticates³⁹. A pertinent example are the hairs on the
292 rachillae of barley grains. In barley, the rachilla is the rudimentary secondary axis of the
293 inflorescence, where multiple grains are set in wheat⁴⁰. In the single-grained spikelets of
294 barley, the rachilla is a thin and hairy thread-like structure nested in the ventral crease of the
295 grains. The long hairs of the rachillae of wild barleys and most cultivated forms are unicellular,
296 while the short hairs of some domesticated types are multicellular and branched (**Fig. 4a**,
297 **Extended Data Fig. 10a**). This seemingly minor difference in a vestigial organ belies its
298 importance in variety registration trials⁴¹, where breeders would like to predict the trait with
299 a diagnostic marker. *Short rachilla hair 1* (*srh1*) is also a classical locus in barley genetics⁴². It
300 has been mapped genetically^{8,43} (**Fig. 4b**) and both long- and short-haired genotypes are
301 included in our pangenome. Fine-mapping in a population of 2,398 recombinant inbred lines
302 derived from a cross^{36,44} between cultivars Morex (short, *srh1*) and Barke (long, *Srh1*)
303 delimited the causal variant to a 113 kb interval on the long arm of chromosome 5H (**Fig. 4c**,
304 **Supplementary Table 23**). Outside of this interval (which is itself devoid of annotated gene
305 models), but within 11 kb of the distal flanking marker is a homolog of a *SIAMESE-RELATED*
306 (*SMR*) gene of the model plant *Arabidopsis thaliana*^{45,46}. Members of this family of cyclin-
307 dependent kinase inhibitors control endoreduplication in trichomes of that species. In barley,
308 hair cell development is likewise accompanied by endopolyploidy-dependent cell size
309 increases (**Extended Data Fig. 10b**). The *SMR*-homolog was expressed in the rachilla's
310 developing trichomes (**Extended Data Fig. 10e**), but there were no differences between Morex
311 and Barke in the sequence of this otherwise plausible candidate gene. Despite this conflicting
312 evidence, we proceeded with mutational analysis and obtained several mutants using FIND-
313 IT⁴⁷ (**Extended Data Fig. 10c,d**) and Cas9-mediated targeted mutagenesis (**Fig. 4d**,
314 **Supplementary Fig. 2, Supplementary Tables 24 and 25**). Mutants of long-haired genotypes
315 with knock-out variants or a nonsynonymous change in a Pro phosphorylation motif (Thr62-
316 Pro63) had short, multicellular rachillae, supporting the idea that the gene in question,
317 *HORVU.MOREX.r3.5HG0492730*, is indeed *HvSRH1*. Sequence variants in *HvSRH1* identified in
318 the pangenome did not lend itself to easy explanation: 18 protein haplotypes caused by 23
319 non-synonymous variants bore no obvious relation to the phenotype (**Supplementary Table**
320 **26**). Thus, we then examined regulatory variation. All 14 short-haired genotypes in the
321 pangenome lacked a 4,273 bp sequence (**Fig. 4c**), which was exceptionally well conserved in
322 long-haired types, with 95% overall identity to Barke. Within this sequence, we found the
323 motif CATCGGATCCTT, matching the sequence [ATC]T[ATC]GGATNC[CT][ATC], which is
324 recognized by regulators of *SMR* expression in *A. thaliana*⁴⁸. That sequence was repeated five

325 times in Barke. The closest unit in long-haired types was no further than 13.6 kb from the
326 gene, while the minimum distance between the gene and its putative enhancer motif in short-
327 haired types was 22.3 kb, owing to the 4.3 kb deletion (**Fig. 4c**). *HvSRH1* expression during
328 rachilla hair elongation is higher in long-haired than in short-haired genotypes (**Extended Data**
329 **Fig. 10f**). Gene edits of the enhancer region, guided by the pangenome sequences, will further
330 elucidate the transcriptional regulation of *HvSRH1*.

331

332 *Discussion*

333

334 The recently published human draft pangenome demonstrated how contiguous long-read
335 sequences help make sense of reams of sequence data⁴⁹. Our study on barley pangenome
336 sheds light on crop evolution and breeding. The shortcomings of previous short-read
337 assemblies made it all but impossible to see patterns that now emerge from their long-read
338 counterparts. We were able for the first time to study the evolution of structurally complex
339 loci of nearly identical tandem repeats. Our developmental insights are admittedly still
340 cursory: true to the hypothesis-generating remit of genomics, and at least as many questions
341 were raised as answered. We studied four loci – *Mla*, *HvTB1*, *amy1_1*, *HvSRH1* – and the traits
342 they control: disease resistance, plant architecture, starch mobilization and the hairiness of a
343 rudimentary appendage to the grain. In two of these examples, phenotypic diversity has
344 visibly increased in domesticated forms: there are no six-rowed or short-haired wild barleys.
345 Malting created new selective pressures that only cultivated forms experienced. Novel allelic
346 variation at disease resistance loci is both illustrative of the power of pangenomics and in line
347 with our understanding of how disease resistance genes evolve. Structural variation at
348 *amy1_1* has been known for some time, but previous attempts at resolving the structure of
349 the locus had been thwarted by incomplete genome sequences. Tandem duplications and
350 deletions of regulatory elements, respectively, at *HvTB1* and *HvSRH1* was surprising since for
351 many years barley geneticists considered the loci as monofactorial recessive. Much of the
352 variation seems to have arisen after domestication, either because mutations that appear with
353 clock-like regularity were absent or copy numbers were lower in the wild progenitor than in
354 the domesticated forms. A common concern among crop conservationists is dangerously
355 reduced genetic diversity in cultivated plants⁵⁰. But crop evolution need not be a
356 unidirectional loss of diversity. This study has shown that valuable diversity can arise after
357 domestication. Rapid evolution at structurally complex loci may endow domesticated plants
358 with a means of adaptive diversification that aptly fulfills the needs of farmers and breeders.
359 More diverse crop pangenomes will help us understand how the counteracting forces of past
360 domestication bottlenecks and newly arisen structural variants influence future crop
361 improvement in changing climates.

362

363 **Acknowledgments**

364 We are grateful for the technical assistance of Susanne König, Ines Walde, Sabine Sommerfeld,
365 John Fuller, Nicola McCallum and Malcolm Macaulay and thank Thomas Münch, Jens
366 Bauernfeind and Heiko Miede for IT administration. Andreas Börner supported the
367 development of the core1000 diversity panel. Sequencing of Chikurin Ibaraki 1 was performed
368 at the Institute for Clinical Molecular Biology, Competence Centre for Genomic Analysis
369 (CCGA), Kiel University, Kiel, Germany under the supervision of Dr. Janina Fuß. Sequencing of
370 HOR 4224 was performed at Genomics & Transcriptomics Labor (BMFZ) Heinrich-Heine-
371 Universität, Universitätsklinikum Düsseldorf, Germany under the supervision of Prof. Dr. Karl

372 Köhrer. RGT Planet and Maximus were sequenced at Genomics WA under the supervision of
373 Dr. Alka Saxena and at Novogene, respectively. N.S., M.M., K.F.X.M., M.Spannagl and U.S. were
374 supported by grants from the German Ministry of Research and Education (BMBF, 031B0190
375 and 031B0884). D.P. was supported by BMBF grant 031B0199B and the German Federal
376 Ministry of Food and Agriculture (grant 2818BIJP01). U.S.'s research is supported by the
377 German Research Foundation (DFG, grant 442032008). C.D., Q.L., P.R.P. and B.S. gratefully
378 acknowledge support from the Carlsberg Foundation to B.S. (grants CF15-0236, CF15-0476,
379 CF15-0672) and thank F. Lok, S. Knudsen, G. B. Fincher and K. G. Jørgensen for providing
380 valuable scientific thoughts and discussions on barley α -amylases and malt quality. R.W., M.B.,
381 M.Schreiber, W.G., R.Z. and C.S. received funding from the Rural and Environment Science and
382 Analytical Services Division (RESAS, grant KJHI-B1-2). W.G. and R.Z. were supported by grant
383 BB/S020160/1 from the Biotechnology and Biological Sciences Research Council (BBSRC). C.P.
384 acknowledges support from Genome Canada and Genome Prairie. B.K. acknowledges a grant
385 from the Swiss National Science Foundation (310030_204165). C.L., K.C. and P.L. were
386 supported by a grant from the Grain Research and Development Corporation (UMU1806-
387 002RTX), by the Department of Primary Industry and Regional Development Western Australia
388 and by Pawsey Australia (for computational resources). K.P. and T.S. received a grant from DFG
389 (HAPPAN, 433162815). G.S.B. has received support from the Saskatchewan Ministry of
390 Agriculture (grant ADF20200165) and from the Saskatchewan Barley Development
391 Commission, Western Grains Research Foundation, Alberta Barley Commission and the
392 Manitoba Crop Alliance (grant ADF20210677). A.B. and W.B. are recipients of the TUGBOAT
393 grant from the Agriculture and Agri-Food Canada – Genomics Research and Development
394 Initiative and the grant “Unlocking barley endophyte microbiome to enhance plant health and
395 grain quality” from Agriculture and Agri-Food Canada – A-Base – Foundational Science. M.H.
396 is supported by the Swedish Research Council (VR 2022-03858), the Swedish Research Council
397 for Environment, Agricultural Sciences, and Spatial Planning (FORMAS 2018-01026), the Erik
398 Philip-Sörensen Foundation and the Royal Physiographic Society in Lund. K.Sato's research is
399 funded by the Japan Science and Technology Agency (grant no. 18076896) and the Japan
400 Society for the Promotion of Science (JSPS, grant no. 23H00333). K.Shirasawa is supported by
401 the JSPS grants nos. 22H05172 and 22H05181. S.S. acknowledges the Research Support
402 Project for the Next Generation at Tottori University. B.S. is supported by the Lieberman-
403 Okinow Endowment at the University of Minnesota and S.G.K. by baseline funding at KAUST.
404 The authors acknowledge the Research/Scientific Computing teams at The James Hutton
405 Institute and NIAB for providing computational resources and technical support for the "UK's
406 Crop Diversity Bioinformatics HPC" (BBSRC grant BB/S019669/1), use of which has contributed
407 to the results reported in this paper.

408

409 **Author contributions**

410 N.S. and M.Mascher designed the study. N.S. coordinated experiments and sequencing.
411 M.Mascher and M.J. supervised sequence assembly. M.Spannagl and K.F.X.M. supervised
412 annotation. U.S. supervised data management and submission. Selection of genotypes: A.B.,
413 W.B., G.S.B., K.J.C., Y.G., M.H., B.K., S.G.K., P.L., C.L., M.Mascher, A.M., G.J.M., D.P., K.P., C.J.P.,
414 S.S., K.Sato, T.S., B.S., N.S., R.W. Genome sequencing: B.B., A.H., S.I., M.K., C.L., S.P., S.S., K.Sato,
415 T.S., M.Schreiber, K.Shirasawa, N.S., S.W., X.Z. Sequence assembly: B.C., H.H., M.J., G.K.-G.,
416 M.Mascher, S.P., K.Sato, T.S., J.F.T. Transcriptome sequencing and analysis: W.G., A.H., S.P., C.S.,
417 N.S., R.Z. Annotation: H.G., G.H., N.K., T.L., K.F.X.M., M.Spannagl. Analysis and interpretation
418 of structural variants: M.B., B.C., J.-W.F., Y.G., M.J., C.L., M.P.M., A.M., S.P., H.P., K.P., T.S.,

419 M.Schreiber, P.W. Analysis of complex loci: B.J., B.K., M.T.R.-W., N.S., T.W. Analysis of *amy1_1*
420 locus: K.B., C.D., M.E.J., B.J., M.J., S.M.K., Q.L., M.P.M., E.M., P.A.P., P.R.P., B.S., H.C.T., M.T.S.N.,
421 D.V., C.V., M.W.R. *srh1* analysis: C.D., M.E.J., I.H., R.E.H., M.J., R.K., J.K., Q.L., M.Melzer, H.P.,
422 L.R., P.R.P., T.R., B.S., N.S., H.C.T., C.T., C.V., M.W.R. Data management and submission: M.B.,
423 M.F., A.F., M.J., P.K., M.Mascher, D.S., U.S. Writing: M.B., C.D., M.E.J., M.J., Q.L., M.Mascher,
424 N.S., T.W. Coordination: C.D., M.Mascher, N.S. All authors read and commented on the
425 manuscript.

426

427 **Competing interests**

428 K.B., C.D., M.E.J., S.M.K., Q.L., E.M., P.R.P., B.S., H.C.T., M.T.S.N., C.V., M.W.R. are current or
429 previous Carlsberg A/S employees. P.A.P. and D.V. are SECOBRA Recherches employees. All
430 other authors declare no competing interests.

431

432 **List of supplementary items**

433

434 **Supplementary Figure 1:** Structure and copy number variation at *Mla* at different thresholds
435 for alignment similarity.

436 **Supplementary Figure 2:** Targeted mutagenesis at *HvSRH1*.

437 **Supplementary Table 1:** Passport data of 76 genotypes and statistics and accession codes of
438 their long-read assemblies

439 **Supplementary Table 2:** Accession codes of transcriptome data

440 **Supplementary Table 3:** Gene annotation statistics

441 **Supplementary Table 4:** Gene ontology enrichment in genepool-specific orthologous groups

442 **Supplementary Table 5:** Passport data of 1,315 genotypes sequenced with short reads,
443 accession codes and mapping stats

444 **Supplementary Table 6:** Statistics and accession codes of 46 gene-space assemblies

445 **Supplementary Table 7:** List of 173 structurally complex loci.

446 **Supplementary Table 8:** Allelic profiles of 76 barley accessions at the 4H_015772 locus (*Int-*
447 *c*) and at *Vrs1*.

448 **Supplementary Table 9:** α -amylase gene IDs and chromosomal locations in Morex.

449 **Supplementary Table 10:** Sequence identity matrix of germination-related *amy1*, *amy2* and
450 *amy3* genes in the Morex genome. *amy4* genes involved in general starch metabolism were
451 excluded due to low sequence identity with other α -amylases.

452 **Supplementary Table 11:** SNP haplotype clustering analyses and k-mer based *amy1_1* copy
453 number estimation in 1,000 plant genetic resources

454 **Supplementary Table 12:** SNP haplotype clustering analyses and k-mer based *amy1_1* copy
455 number estimation in 315 European elite cultivars

456 **Supplementary Table 13:** Overview of *amy1_1* unique ORFs (start to stop codon including
457 intron). HORVU.MOREX.PROJ.6HG00545380 was used as the reference.

458 **Supplementary Table 14:** Overview of *amy1_1* ORF haplotypes (ORFHap#).

459 **Supplementary Table 15:** Overview of *amy1_1* unique CDS.

460 HORVU.MOREX.PROJ.6HG00545380 was used as the reference.

461 **Supplementary Table 16:** Overview of *amy1_1* CDS haplotypes (CDSHap#).

462 **Supplementary Table 17:** *amy1_1* genes with insertions of transposable elements in the
463 genome assemblies.

464 **Supplementary Table 18:** Overview of *amy1_1* unique proteins.

465 HORVU.MOREX.PROJ.6HG00545380 was used as the reference.

- 466 **Supplementary Table 19:** Overview of *amy1_1* protein haplotypes (ProtHap#).
467 **Supplementary Table 20:** Amino acid variation in three *amy1_1* haplotypes commonly found
468 in elite varieties (Morex, Barke and RGT Planet).
469 **Supplementary Table 21:** DynaMut2 prediction of protein stability changes by amino acid
470 variants found in three BPGv2 representatives of widely used *amy1_1* haplotypes found in
471 elite breeding material (Morex, Barke and RGT Planet).
472 **Supplementary Table 22:** *amy1_1*-Barke haplotype genotyping of AMBA(American Malting
473 Barley Association)-recommended two-row spring malting barley varieties accredited for
474 adjunct brewing.
475 **Supplementary Table 23:** PACE markers designed in this study.
476 **Supplementary Table 24:** Oligonucleotides used for gRNA cloning and for PCR amplification
477 of the target region.
478 **Supplementary Table 25:** Summary of lesions induced in *HvSRH1* by Cas9-mediated targeted
479 mutagenesis.
480 **Supplementary Table 26:** *srh1* phenotypes and *HvSRH1* gene coordinates in 76 pangenome
481 accessions.

482 483 **References**

- 484
485 1 Komatsuda, T. *et al.* Six-rowed barley originated from a mutation in a homeodomain-
486 leucine zipper I-class homeobox gene. *Proceedings of the National Academy of*
487 *Sciences* **104**, 1424-1429 (2007). <https://doi.org/10.1073/pnas.0608580104>
488 2 Sakuma, S. *et al.* Divergence of expression pattern contributed to
489 neofunctionalization of duplicated HD-Zip I transcription factor in barley. *New*
490 *Phytologist* **197**, 939-948 (2013). <https://doi.org/10.1111/nph.12068>
491 3 Milner, S. G. *et al.* Genebank genomics highlights the diversity of a global barley
492 collection. *Nat Genet* **51**, 319-326 (2019). [https://doi.org/10.1038/s41588-018-0266-](https://doi.org/10.1038/s41588-018-0266-x)
493 [x](https://doi.org/10.1038/s41588-018-0266-x)
494 4 Abbo, S. *et al.* Plant domestication versus crop evolution: a conceptual framework for
495 cereals and grain legumes. *Trends in Plant Science* **19**, 351-360 (2014).
496 <https://doi.org/10.1016/j.tplants.2013.12.002>
497 5 Dawson, I. K. *et al.* Barley: a translational model for adaptation to climate change.
498 *New Phytol* **206**, 913-931 (2015). <https://doi.org/10.1111/nph.13266>
499 6 Lundqvist, U. Scandinavian mutation research in barley – a historical review.
500 *Hereditas* **151**, 123-131 (2014). <https://doi.org/10.1111/hrd2.00077>
501 7 Schulte, D. *et al.* The international barley sequencing consortium--at the threshold of
502 efficient access to the barley genome. *Plant Physiol* **149**, 142-147 (2009).
503 <https://doi.org/10.1104/pp.108.128967>
504 8 Jayakodi, M. *et al.* The barley pan-genome reveals the hidden legacy of mutation
505 breeding. *Nature* **588**, 284-289 (2020). <https://doi.org/10.1038/s41586-020-2947-8>
506 9 Mascher, M. *et al.* Long-read sequence assembly: a technical evaluation in barley.
507 *Plant Cell* (2021). <https://doi.org/10.1093/plcell/koab077>
508 10 Russell, J. *et al.* Exome sequencing of geographically diverse barley landraces and
509 wild relatives gives insights into environmental adaptation. *Nat Genet* **48**, 1024-1030
510 (2016). <https://doi.org/10.1038/ng.3612>
511 11 Druka, A. *et al.* Genetic Dissection of Barley Morphology and Development. *Plant*
512 *Physiology* **155**, 617-627 (2011). <https://doi.org/10.1104/pp.110.166249>

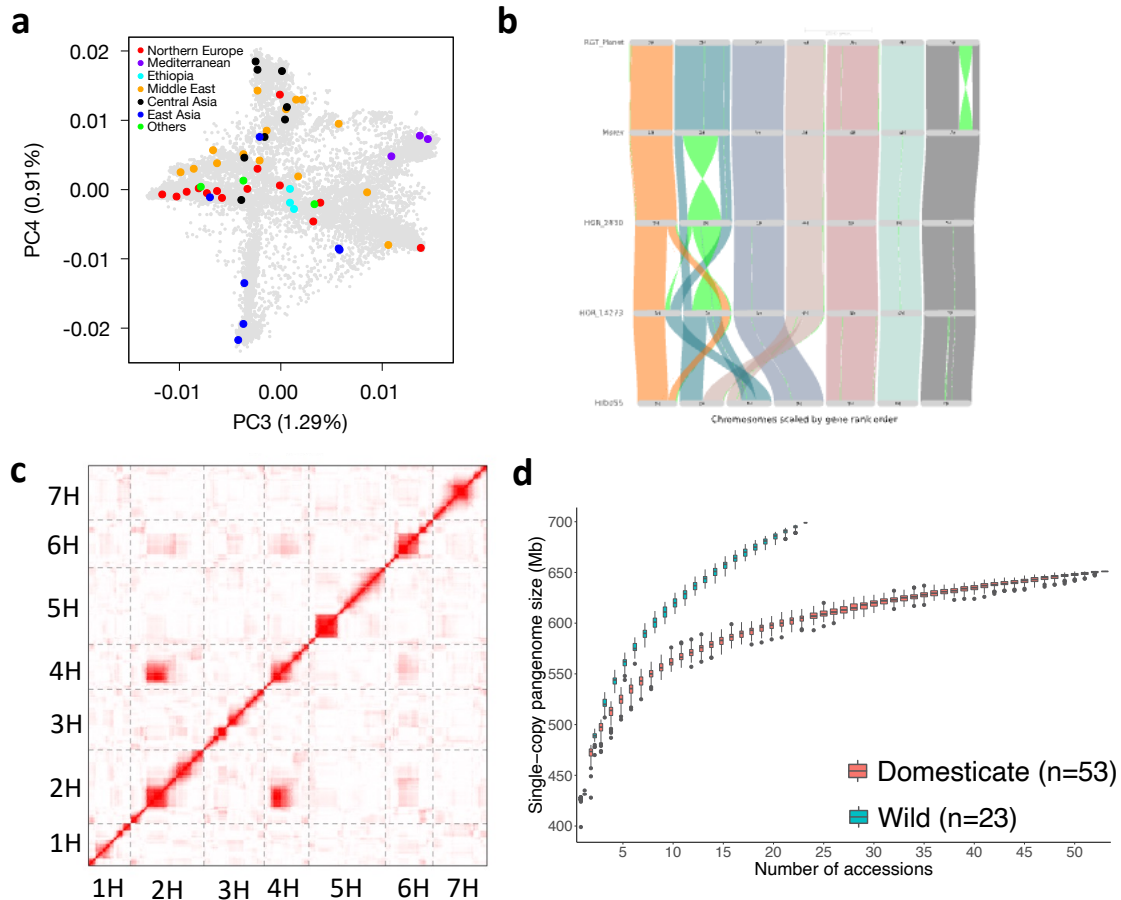
- 513 12 Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves
514 variant detection and assembly of a human genome. *Nature Biotechnology* **37**, 1155-
515 1162 (2019). <https://doi.org/10.1038/s41587-019-0217-9>
- 516 13 Padmarasu, S., Himmelbach, A., Mascher, M. & Stein, N. In Situ Hi-C for Plants: An
517 Improved Method to Detect Long-Range Chromatin Interactions. *Methods Mol Biol*
518 **1933**, 441-472 (2019). https://doi.org/10.1007/978-1-4939-9045-0_28
- 519 14 Lawniczak, M. K. N. *et al.* Standards recommendations for the Earth BioGenome
520 Project. *Proceedings of the National Academy of Sciences* **119**, e2115639118 (2022).
521 <https://doi.org/doi:10.1073/pnas.2115639118>
- 522 15 Maurer, A. *et al.* Modelling the genetic architecture of flowering time control in
523 barley through nested association mapping. *BMC Genomics* **16**, 290 (2015).
524 <https://doi.org/10.1186/s12864-015-1459-7>
- 525 16 Li, H., Feng, X. & Chu, C. The design and construction of reference pangenome graphs
526 with minigraph. *Genome Biology* **21**, 265 (2020). <https://doi.org/10.1186/s13059-020-02168-z>
- 528 17 Andreade, F., Lechat, P., Dufresne, Y. & Chikhi, R. Construction and representation of
529 human pangenome graphs. *bioRxiv*, 2023.2006.2002.542089 (2023).
530 <https://doi.org/10.1101/2023.06.02.542089>
- 531 18 De Coster, W., Weissensteiner, M. H. & Sedlazeck, F. J. Towards population-scale long-
532 read sequencing. *Nature Reviews Genetics* **22**, 572-587 (2021).
533 <https://doi.org/10.1038/s41576-021-00367-3>
- 534 19 Michelmore, R. W. & Meyers, B. C. Clusters of resistance genes in plants evolve by
535 divergent selection and a birth-and-death process. *Genome Res* **8**, 1113-1130 (1998).
536 <https://doi.org/10.1101/gr.8.11.1113>
- 537 20 Wei, F., Wing, R. A. & Wise, R. P. Genome Dynamics and Evolution of the
538 *Mla* (Powdery Mildew) Resistance Locus in Barley. *The Plant Cell* **14**,
539 1903-1917 (2002). <https://doi.org/10.1105/tpc.002238>
- 540 21 Bettgenhaeuser, J. *et al.* The barley immune receptor *Mla* recognizes multiple
541 pathogens and contributes to host range dynamics. *Nature Communications* **12**, 6915
542 (2021). <https://doi.org/10.1038/s41467-021-27288-3>
- 543 22 Seeholzer, S. *et al.* Diversity at the *Mla* powdery mildew resistance locus from
544 cultivated barley reveals sites of positive selection. *Mol Plant Microbe Interact* **23**,
545 497-509 (2010). <https://doi.org/10.1094/mpmi-23-4-0497>
- 546 23 Brabham, H. J. *et al.* Barley *MLA3* recognizes the host-specificity determinant *PWL2*
547 from rice blast (*M. oryzae*). *bioRxiv*, 2022.2010.2021.512921 (2022).
548 <https://doi.org/10.1101/2022.10.21.512921>
- 549 24 Rabanus-Wallace, M. T., Wicker, T. & Stein, N. Replicators, genes, and the C-value
550 enigma: High-quality genome assembly of barley provides direct evidence that self-
551 replicating DNA forms 'cooperative' associations with genes in arms races. *bioRxiv*,
552 2023.2010.2001.560391 (2023). <https://doi.org/10.1101/2023.10.01.560391>
- 553 25 Escudero-Martinez, C. M., Morris, J. A., Hedley, P. E. & Bos, J. I. B. Barley
554 transcriptome analyses upon interaction with different aphid species identify
555 thionins contributing to resistance. *Plant Cell Environ* **40**, 2628-2643 (2017).
556 <https://doi.org/10.1111/pce.12979>
- 557 26 Wicker, T., Yahiaoui, N. & Keller, B. Illegitimate recombination is a major evolutionary
558 mechanism for initiating size variation in plant resistance genes. *The Plant Journal* **51**,
559 631-641 (2007). [https://doi.org:https://doi.org/10.1111/j.1365-313X.2007.03164.x](https://doi.org/https://doi.org/10.1111/j.1365-313X.2007.03164.x)

- 560 27 Brassac, J. & Blattner, F. R. Species-Level Phylogeny and Polyploid Relationships in
561 Hordeum (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of
562 Multiple Nuclear Loci. *Syst Biol* **64**, 792-808 (2015).
563 [https://doi.org:10.1093/sysbio/syv035](https://doi.org/10.1093/sysbio/syv035)
- 564 28 Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize.
565 *Nature* **386**, 485-488 (1997). [https://doi.org:10.1038/386485a0](https://doi.org/10.1038/386485a0)
- 566 29 Dixon, L. E. *et al.* TEOSINTE BRANCHED1 Regulates Inflorescence Architecture and
567 Development in Bread Wheat (*Triticum aestivum*). *The Plant Cell* **30**, 563-581 (2018).
568 [https://doi.org:10.1105/tpc.17.00961](https://doi.org/10.1105/tpc.17.00961)
- 569 30 Ramsay, L. *et al.* INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is
570 an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nature*
571 *Genetics* **43**, 169-172 (2011). [https://doi.org:10.1038/ng.745](https://doi.org/10.1038/ng.745)
- 572 31 Lundqvist, U., ABEBE, B. & LUNDQVIST, A. Gene interaction of induced intermedium
573 mutations of two-row barley. *Hereditas* **111**, 37-47 (1989).
574 [https://doi.org:10.1111/j.1601-5223.1989.tb00374.x](https://doi.org/10.1111/j.1601-5223.1989.tb00374.x)
- 575 32 Youssef, H. M. *et al.* Natural diversity of inflorescence architecture traces cryptic
576 domestication genes in barley (*Hordeum vulgare* L.). *Genetic Resources and Crop*
577 *Evolution* **64**, 843-853 (2017). [https://doi.org:10.1007/s10722-017-0504-6](https://doi.org/10.1007/s10722-017-0504-6)
- 578 33 Monat, C. *et al.* TRITEX: chromosome-scale sequence assembly of Triticeae genomes
579 with open-source tools. *Genome Biol* **20**, 284 (2019).
580 [https://doi.org:10.1186/s13059-019-1899-5](https://doi.org/10.1186/s13059-019-1899-5)
- 581 34 Janeček, Š., Svensson, B. & MacGregor, E. A. α -Amylase: an enzyme specificity found
582 in various families of glycoside hydrolases. *Cell Mol Life Sci* **71**, 1149-1170 (2014).
583 [https://doi.org:10.1007/s00018-013-1388-z](https://doi.org/10.1007/s00018-013-1388-z)
- 584 35 Karrer, E. E., Chandler, J. M., Foolad, M. R. & Rodriguez, R. L. Correlation between α -
585 amylase gene expression and seedling vigor in rice. *Euphytica* **66**, 163-169 (1992).
586 [https://doi.org:10.1007/BF00025299](https://doi.org/10.1007/BF00025299)
- 587 36 Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the
588 barley genome. *Nature* **544**, 427-433 (2017). [https://doi.org:10.1038/nature22043](https://doi.org/10.1038/nature22043)
- 589 37 Kadziola, A., Sogaard, M., Svensson, B. & Haser, R. Molecular structure of a barley α -
590 amylase-inhibitor complex: implications for starch binding and catalysis
591 Edited by R. Huber. *Journal of Molecular Biology* **278**, 205-217 (1998).
592 [https://doi.org:https://doi.org/10.1006/jmbi.1998.1683](https://doi.org/https://doi.org/10.1006/jmbi.1998.1683)
- 593 38 Zou, X., Neuman, D. & Shen, Q. J. Interactions of two transcriptional repressors and
594 two transcriptional activators in modulating gibberellin signaling in aleurone cells.
595 *Plant Physiol* **148**, 176-186 (2008). [https://doi.org:10.1104/pp.108.123653](https://doi.org/10.1104/pp.108.123653)
- 596 39 Fuller, D. Q. Contrasting Patterns in Crop Domestication and Domestication Rates:
597 Recent Archaeobotanical Insights from the Old World. *Annals of Botany* **100**, 903-924
598 (2007). [https://doi.org:10.1093/aob/mcm048](https://doi.org/10.1093/aob/mcm048)
- 599 40 Sakuma, S. & Koppolu, R. Form follows function in Triticeae inflorescences. *Breed Sci*
600 **73**, 46-56 (2023). [https://doi.org:10.1270/jsbbs.22085](https://doi.org/10.1270/jsbbs.22085)
- 601 41 Yu, J. K. & Chung, Y. S. Plant Variety Protection: Current Practices and Insights. *Genes*
602 (*Basel*) **12** (2021). [https://doi.org:10.3390/genes12081127](https://doi.org/10.3390/genes12081127)
- 603 42 Engledow, F. Inheritance in barley: I. The lateral florets and the rachilla. *Journal of*
604 *Genetics* **10**, 93-108 (1920).

- 605 43 Cockram, J. *et al.* Genome-wide association mapping to candidate polymorphism
606 resolution in the unsequenced barley genome. *Proceedings of the National Academy*
607 *of Sciences* **107**, 21611-21616 (2010). <https://doi.org/doi:10.1073/pnas.1010179107>
608 44 Beier, S. *et al.* Construction of a map-based reference genome sequence for barley,
609 *Hordeum vulgare* L. *Scientific Data* **4**, 170044 (2017).
610 <https://doi.org:10.1038/sdata.2017.44>
611 45 Kumar, N. *et al.* Functional Conservation in the SIAMESE-RELATED Family of Cyclin-
612 Dependent Kinase Inhibitors in Land Plants. *Plant Cell* **27**, 3065-3080 (2015).
613 <https://doi.org:10.1105/tpc.15.00489>
614 46 Wang, K. *et al.* The CDK Inhibitor SIAMESE Targets Both CDKA;1 and CDKB1
615 Complexes to Establish Endoreplication in Trichomes. *Plant Physiol* **184**, 165-175
616 (2020). <https://doi.org:10.1104/pp.20.00271>
617 47 Knudsen, S. *et al.* FIND-IT: Accelerated trait development for a green evolution.
618 *Science Advances* **8**, eabq2266 (2022). <https://doi.org:doi:10.1126/sciadv.abq2266>
619 48 Nomoto, Y. *et al.* A hierarchical transcriptional network activates specific CDK
620 inhibitors that regulate G2 to control cell size and number in Arabidopsis. *Nature*
621 *Communications* **13**, 1660 (2022). <https://doi.org:10.1038/s41467-022-29316-2>
622 49 Liao, W.-W. *et al.* A draft human pangenome reference. *Nature* **617**, 312-324 (2023).
623 <https://doi.org:10.1038/s41586-023-05896-x>
624 50 Brown, T. A. Is the domestication bottleneck a myth? *Nature Plants* **5**, 337-338
625 (2019). <https://doi.org:10.1038/s41477-019-0404-1>
626
627

628 **Figures**

629



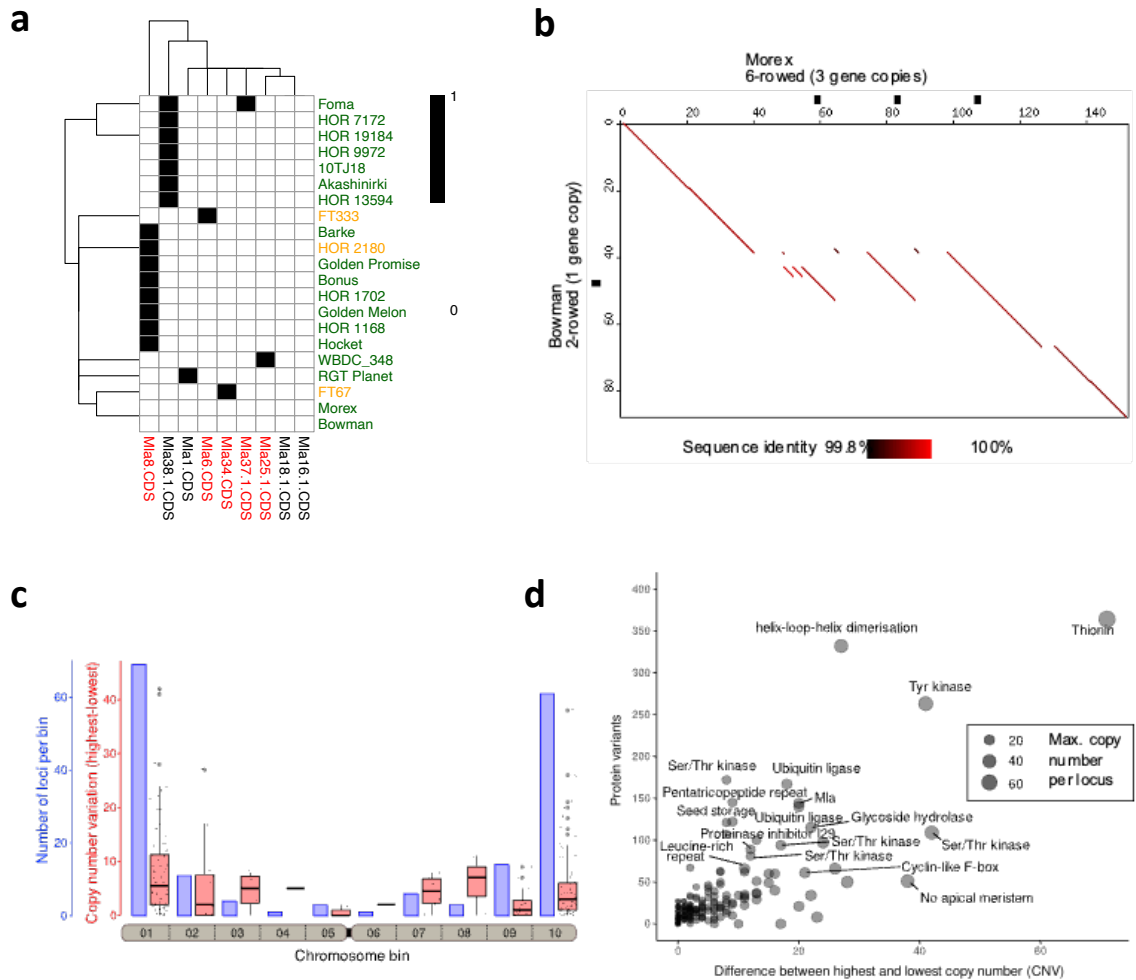
630

631

632 **Figure 1:** A species-wide pangenome of *Hordeum vulgare*. (a) Principal component analysis
633 showing domesticated accessions (n=53) in the pangenome panel in the global diversity
634 space. Regions of origins are color coded. The proportion of variance explained by each PC in
635 panels is given in the axis labels. Other PCs are shown in **Extended Data Fig. 1a**. (b) Example
636 of large structural variants including interchromosomal translocations and inversions
637 between pangenome accessions. (c) Interchromosomal linkage disequilibrium (LD) in
638 segregating offspring derived from a cross between HID055 and Barke. LD is indicated by the
639 intensity of red color. (d) Size of the single-copy pangenome in wild and domesticated
640 barleys as a function of sample size.

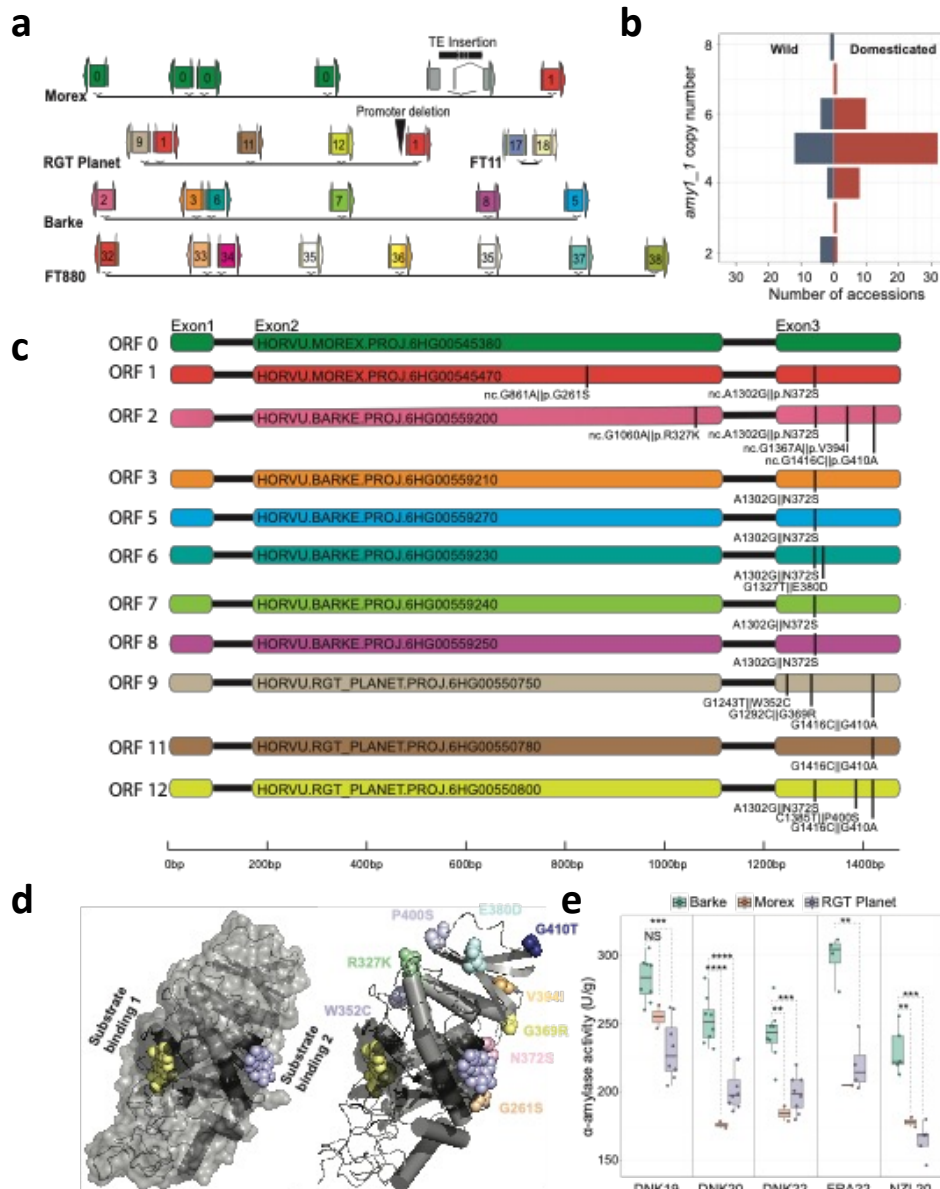
641

642



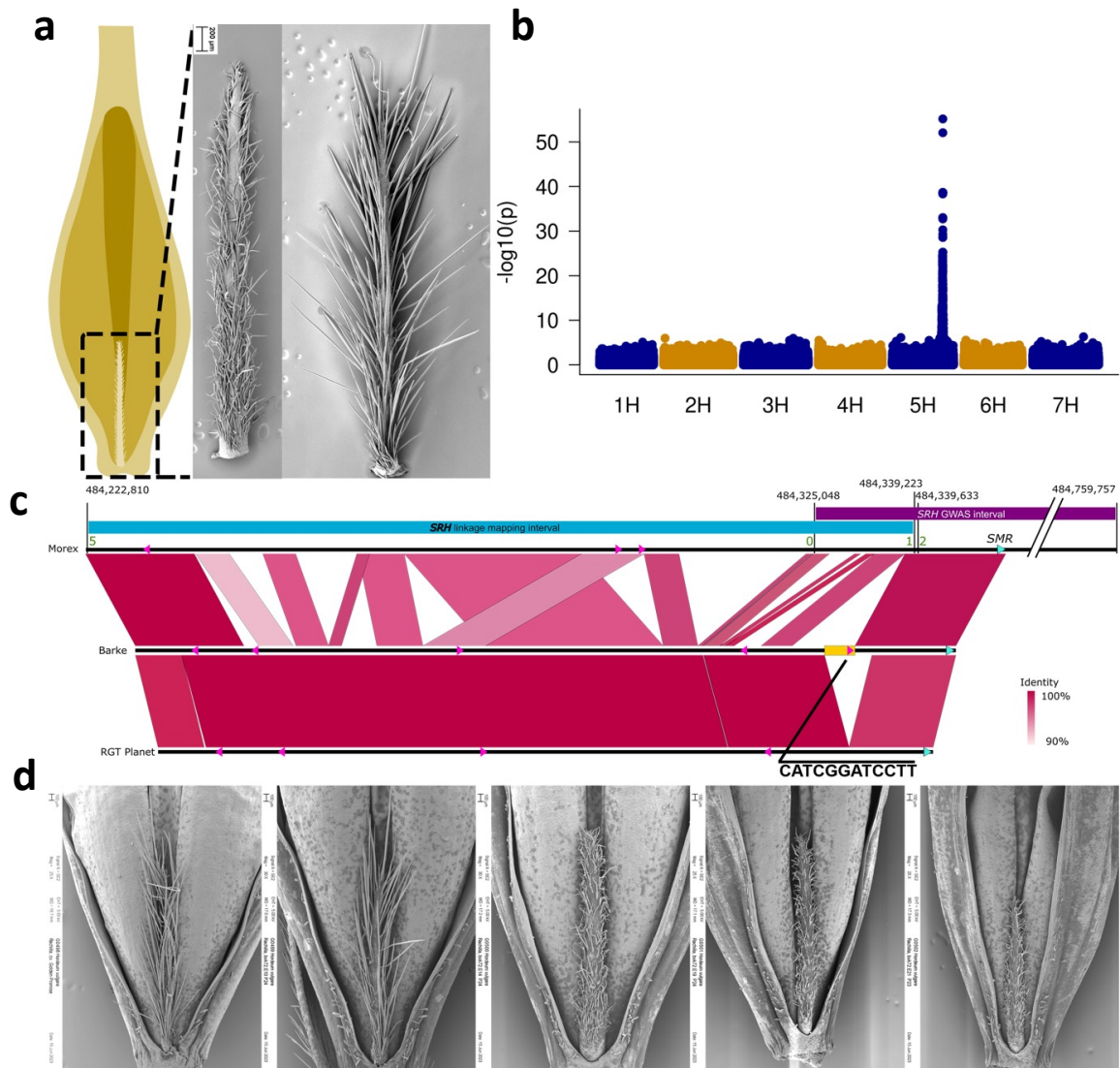
643
644

645 **Figure 2: Structurally complex loci in the barley pangenome.** (a) Presence/absence of
 646 known *Mla* alleles in the barley pangenome. Black and white squares denote presence and
 647 absence, respectively. The names of *Mla* alleles (y-axis) and genotypes (x-axis) are coloured
 648 according to subfamily and domestication statuses, respectively. (green – domesticated;
 649 orange – wild). Only the genomes containing known alleles are displayed. (b) Dot plot
 650 alignment of complex locus Chr04_015772 which contains *Int-c* genes. The plot shows an
 651 alignment of Morex (six-rowed barley) and Bowman (two-rowed barley). In Morex, *Int-c* and
 652 its surrounding sequence is present in three copies. Genes are indicated as black boxes along
 653 the axes of the plot. Individual tandem repeat units are 96-100% identical. (c) Complex loci
 654 are enriched in distal chromosomal regions. The seven barley chromosomes were divided
 655 into ten equally sized bins, and cumulative data for all chromosomes is shown. The bar plot
 656 indicates the number of loci, while the box blot shows the extent of CNV for all loci in the
 657 bin. (d) CNV levels and numbers of encoded protein variants identified in 76 barley
 658 accessions. The x-axis shows the level of CNV (i.e. the difference between the accession with
 659 the fewest copies to that with the most copies for each locus). The y-axis shows the total
 660 number of protein variants identified in all 76 barley accessions. Labels mark genes families
 661 with the highest copy numbers or the highest CNV levels.
 662



663
664

665 **Figure 3. Structural diversity at the *amy1_1* locus and its importance in malting.** (a)
666 Simplified structure of the *amy1_1* locus in selected pangenome assemblies. A detailed
667 depiction of the *amy1_1* locus across all 76 assemblies is shown in **Extended Data Fig. 9a**.
668 Identical ORFs have the same colours in (a) and (c). (b) Distribution of *amy1_1* copy numbers
669 in wild and domesticated accessions of the pangenome. (c) Non-synonymous sequence
670 exchanges in 12 non-redundant *amy1_1* ORFs in the malting barleys Morex, Barke and RGT
671 Planet. The positions of sequence variants and respective amino acid variations are marked
672 by black lines. ORF numbers refer to **Supplementary Table 13**. (d,e) X-ray crystal structure
673 (pdb: 1BG9; ref. ³⁶) of α -amylase bound to acarbose as a substrate analogue (magenta and
674 yellow spheres in panel (d)). In panel (e), *amy1_1* amino acid variants (found in Morex, Barke
675 and RGT Planet, **Supplementary Table 20**) are added as coloured spheres. (f) α -amylase
676 activity of micro-malted near-isogenic lines (NILs) containing *amy1_1*-Morex, Barke and RGT
677 Planet haplotypes.
678



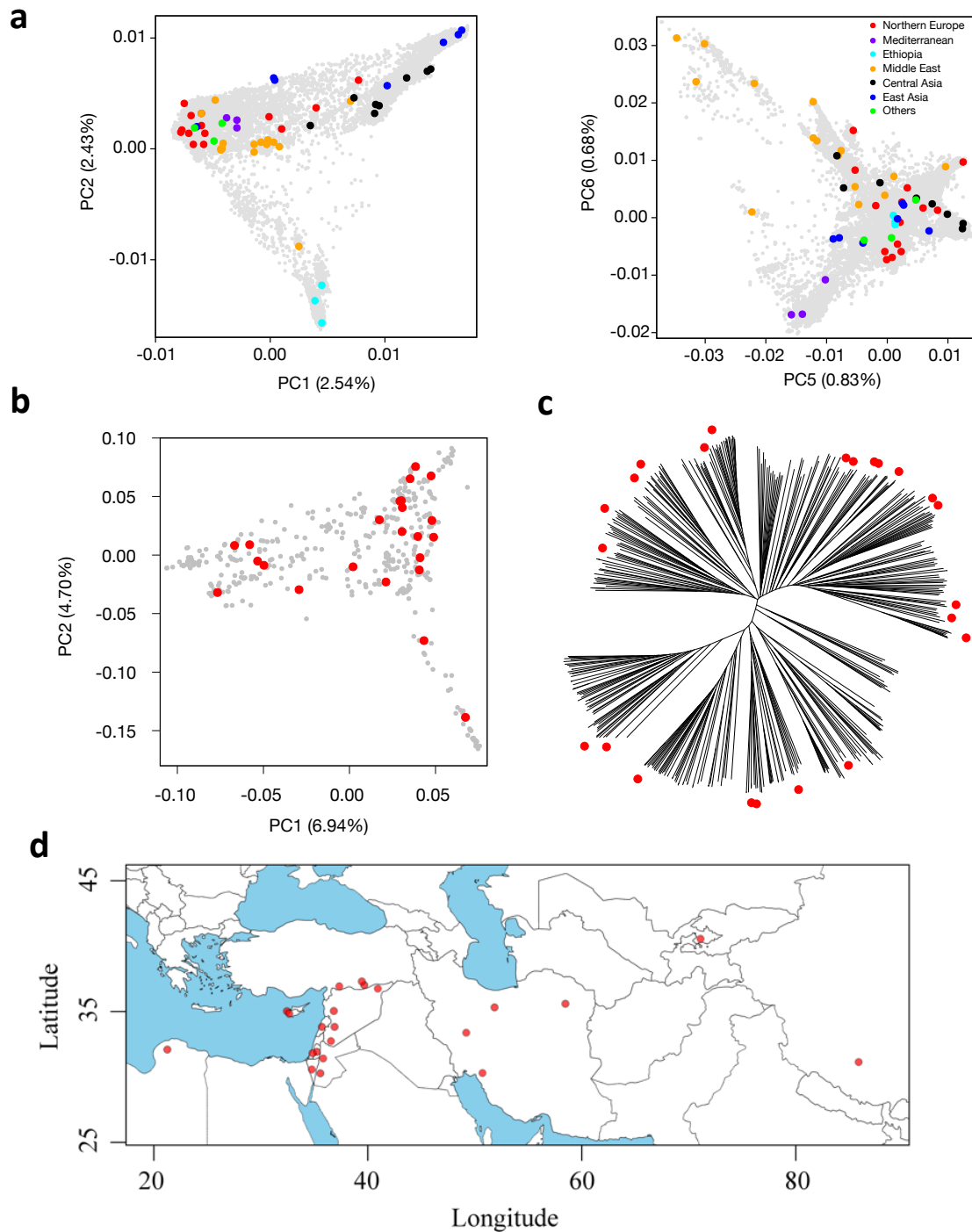
679

680

681 **Figure 4. A deletion in an enhancer motif is associated with trichome branching. (a)**
 682 Schematic drawing of a seed from a hulled and awned barley. The rachilla is a rudimentary
 683 structure attached to the base of the seed, representing reduced lateral branches in the
 684 barley inflorescence. On the right, scanning electron micrographs are shown of a short-
 685 haired and a long-haired rachilla of genotypes Morex and Barke, respectively. **(b)** Genome-
 686 wide association study (GWAS) for rachilla hair phenotype in the core1000. **(c)** Top part:
 687 schematic representation of the high-resolution genetic linkage analysis at the *Srh1* locus.
 688 Blue and purple horizontal bars represent the overlapping biparental and GWAS mapping
 689 intervals in reference to the 160 kb physical interval in the Morex genome (black line below
 690 the colored bars). Note, the SMR-like gene sits outside the high-resolution biparental
 691 mapping interval. Bottom part: connector plot showing orthologous regions in the
 692 genotypes Barke (long hairs) and RGT Planet (short hairs). A region harboring a conserved
 693 enhancer element (yellow rectangle) is present in Barke, but absent in Morex and RGT
 694 Planet. **(d)** Rachilla hair phenotype of the Cas9-induced knock-out mutants of the SMR-like
 695 gene. From left to right: wild-type Golden Promise (GP); wild-type segregant from the
 696 *brhE72P19* family; independent mutant segregants showing the short-hair phenotype.
 697

698 **Extended Data Figures**

699



700

701

702 **Extended Data Figure 1: A globally representative diversity panel of domesticated and wild**

703 **barley. (a)** Higher principal components (PC) of the barley diversity space with pangenome

704 accessions highlighted. **(b)** The first two PCs of the diversity space of 412 wild barley

705 (*Hordeum vulgare* subsp. *spontaneum*) with pangenome accessions highlighted. **(c)**

706 Neighbor-joining phylogenetic tree of those wild barleys. The branch tips corresponding to

707 accessions selected for the pangenome are marked with red circles. The proportion of

708 variance explained by each PC in panels **(a)** and **(b)** is given in the axis labels. **(d)** Map

showing the collection sites of wild accessions (n=23) included in the pangenome panel.

a

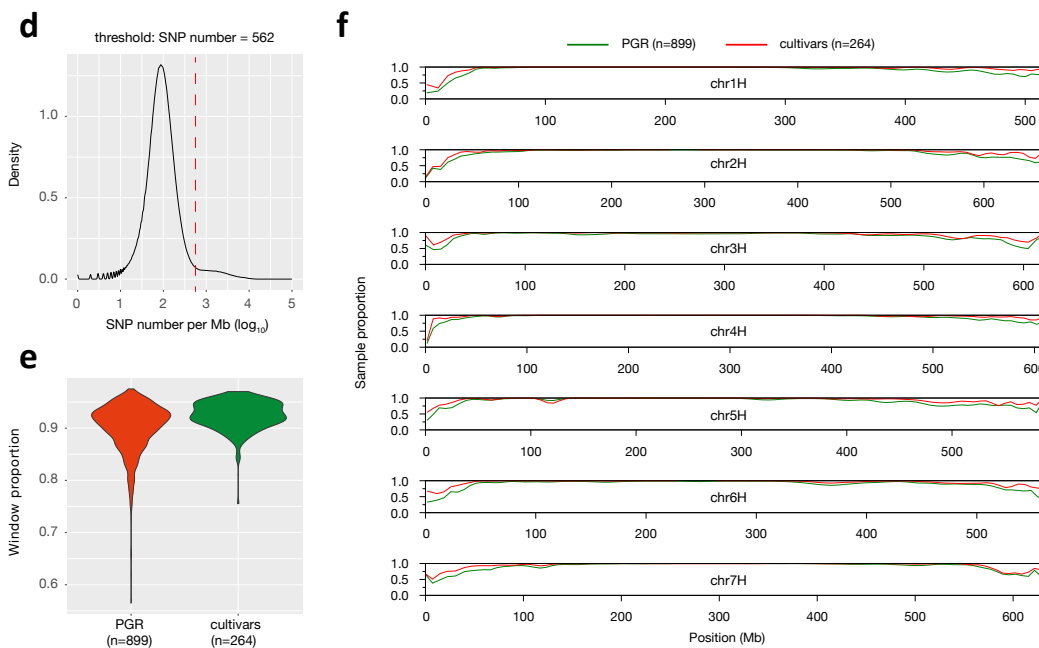
Quality category	Metric	Domesticate (N=53)	Wild (N=23)
Continuity	Avg. contig N50	18	14
	Max. contig N50	37	21
	Min. contig N50	10	8
	Avg. no. of gaps	445	556
Chromosome status	Avg. chromosome anchoring rate (%)	98.0	98.1
	Avg. chromosome anchored size (Gb unanchored size (Mb))	4.19 47	4.21 53
Structural accuracy	False duplications (%)	0.012	0.010
	Curation (Hi-C)	Manual	Manual
Base accuracy	Consensus quality value (QV)	66.0	66.3
	k-mer completeness (%)	97.5	97.6
Functional completeness	BUSCO (%)	96.4	96.5

b

	Summary
No. of PAVs (> 50 nt)	1,703,288
Presence in Morex	787,285
Absence in Morex	916,003
Polymorphic (>2 & < 74)	581,248 (34%)

c

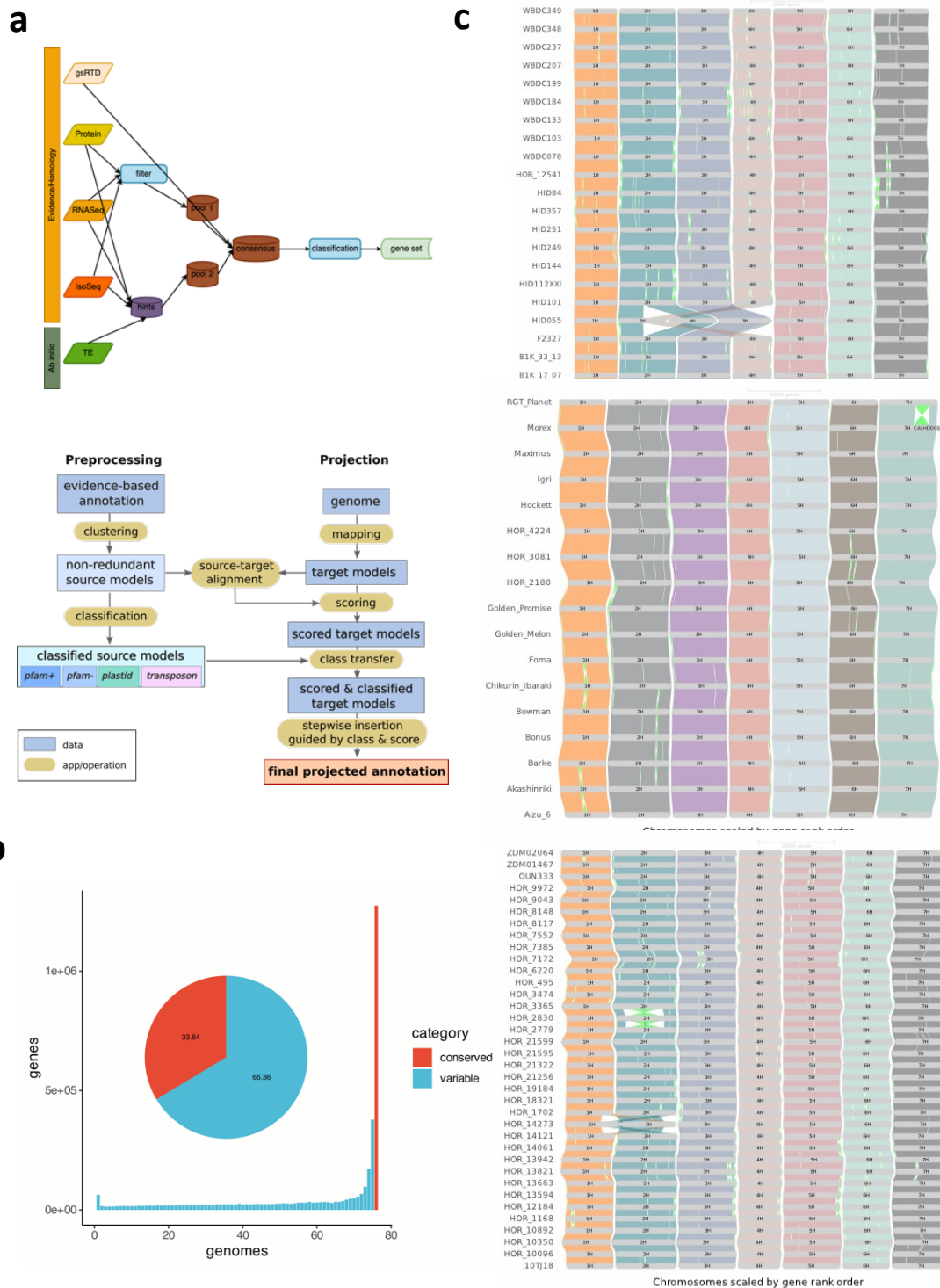
Type	Summary
Inversions (> 2 kb)	3,277
Shared events (>2 & < 74)	548 (17%)
Private to domesticate	197
Private to wild	76



709
710
711
712
713
714
715
716
717

Extended Data Figure 2: A pangenomic diversity map of barley. (a) Assembly statistics of 76 chromosome-scale reference genomes sequences. **(b)** Counts of presence/absence variants. **(c)** Counts of inversion polymorphisms spanning 2 kb or more. **(d)** Selection of threshold based on pairwise differences (number of SNPs per Mb) for the binary classification into similar/dissimilar haplotypes. **(f)** The proportion of samples with a close match to one of the 76 pangenome accessions is shown for plant genetic resources (PGR) and elite cultivars in sliding windows along the genome (size: 1 Mb, shift: 500 kb). **(h)** Distribution of the share of similar windows in individual PGR and cultivar genomes.

718



719
720

721 **Extended Data Figure 3: Gene annotation and orthologous framework.** (a) Workflow for
722 annotating, projecting and clustering gene models. The upper panel describes the workflow
723 for the de-novo gene predictions, the lower panel for the gene projections (b) Histogram
724 showing the number of pagenome genotypes contributing to individual hierarchical
725 orthologous groups (HOGs). The pie chart shows the ratio between conserved and variable
726 genes. (c) GENESPACE alignments of 76 barley genomes, grouped by wild barley, cultivated
727 barley and landraces.

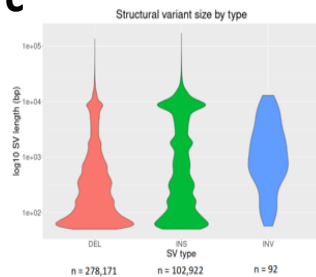
a

chromosome	# nodes	# edges	combined node length (bp)
chr1H	615,439	874,894	1,092,683,108
chr2H	789,027	1,123,085	1,409,260,038
chr3H	778,830	1,106,625	1,332,024,267
chr4H	627,447	890,321	1,094,625,160
chr5H	719,146	1,024,533	1,295,813,787
chr6H	677,475	963,220	1,174,448,066
chr7H	848,394	1,206,148	1,442,759,314
joint graph	5,055,758	7,188,826	8,841,613,740

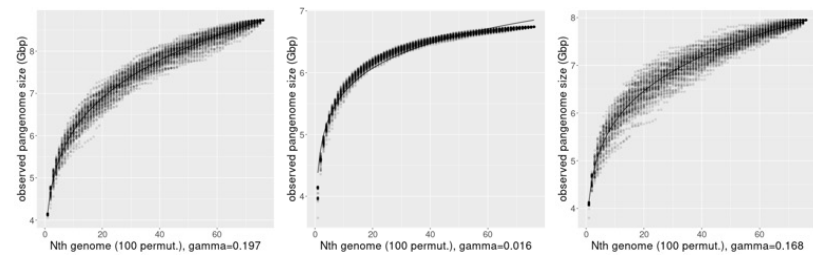
b

accession	mapping to graph	mapping to linear reference	difference (linear - graph)
% reads mapped			
ERR2766176	97.25	98.42	1.17
SRR10200200	98.62	99.81	1.19
SRR5197485	97.19	99.7	2.51
SRR5197496	97.15	99.71	2.56
SRR6281633	93.15	95.66	2.51
% reads properly paired			
ERR2766176	96.54	92.52	-4.02
SRR10200200	96.47	94.03	-2.44
SRR5197485	94.7	93.81	-0.9
SRR5197496	94.46	94.74	0.28
SRR6281633	92.08	90.47	-1.61
% mismatches			
ERR2766176	1.16	1.4	0.24
SRR10200200	1.08	1.28	0.2
SRR5197485	0.83	0.92	0.09
SRR5197496	0.79	0.89	0.1
SRR6281633	1.5	1.73	0.23
# mismatches			
ERR2766176	578,805,187	711,908,069	133,102,882
SRR10200200	1,011,740,607	1,212,059,305	200,318,698
SRR5197485	581,657,638	666,287,736	84,630,098
SRR5197496	569,171,755	656,634,935	87,463,180
SRR6281633	1,496,337,825	1,767,306,648	270,968,823

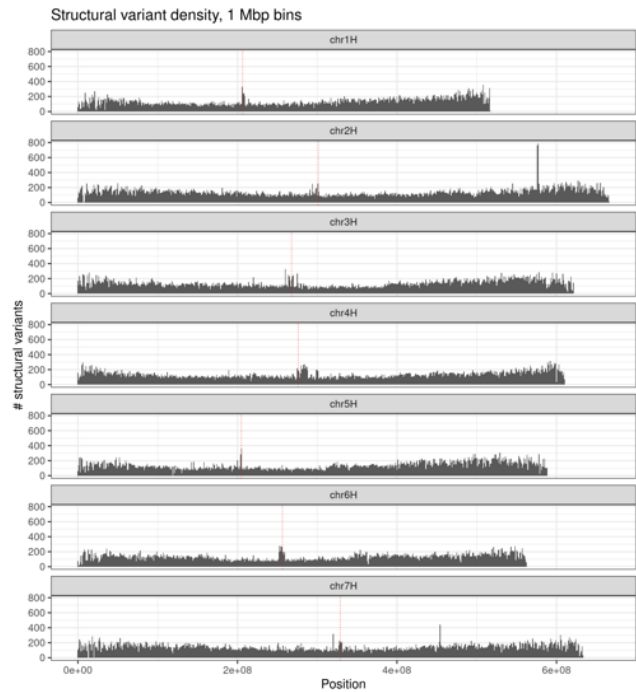
c



e

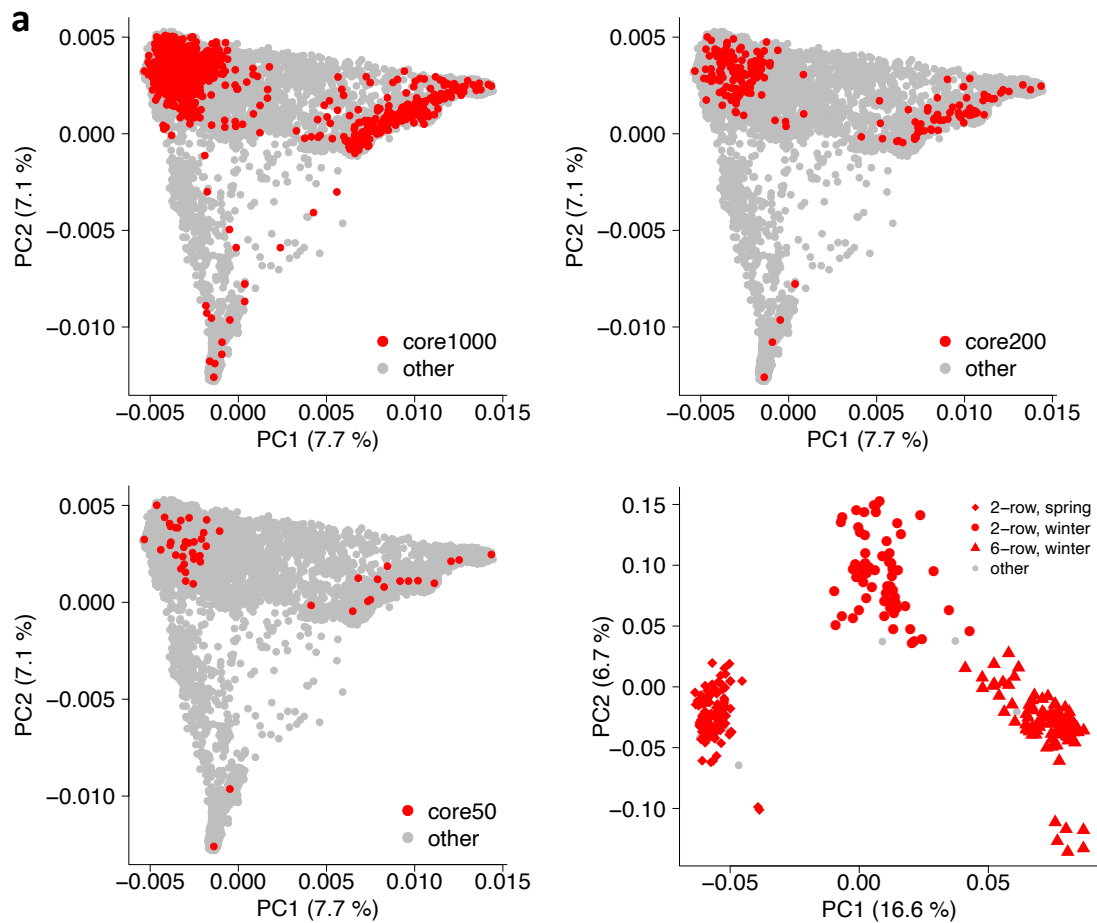


d



728
729

730 **Extended Data Figure 4:** Graph-based pan-genome analysis with Minigraph. **(a)** Descriptive
731 statistics per chromosome and for joint graph. **(b)** Comparative statistics of read mappings
732 from five publicly available Illumina whole genome shotgun sequence read runs against the
733 pan-genome graph and the MorexV3 linear reference sequence. **(c)** Size distribution of
734 structural variants (SVs) in graph. **(d)** Chromosomal distribution of SVs. Centromere positions
735 are indicated by vertical dashed lines in red. **(e)** Pan-genome graph growth curves generated
736 with the odgi heaps tool. One hundred permutations were computed for each number of
737 genomes included. Values of gamma > 0 in Heaps' law indicate an open pan-genome. Plots
738 shown are for all accessions (left, n = 76), domesticated accessions only (cultivars +
739 landraces, centre, n = 53) and *H. spontaneum* accessions (right, n = 23).
740



b

Chromosome	SNPs	indels
chr1H	17,314,288	1,088,174
chr2H	24,995,054	1,569,302
chr3H	24,482,473	1,695,142
chr4H	20,039,734	1,372,443
chr5H	21,590,485	1,411,261
chr6H	22,929,086	1,477,914
chr7H	24,207,458	1,603,509
Total	155,558,578	9,129,571

741

742

743 **Extended Data Figure 5: Short-read data complement the pangenome infrastructure. (a)**

744 Accessions selected for short-read sequencing. Nested coresets of 1000, 200 and 50

745 accessions (core1000, core200, core50) are shown in the global diversity space of barley as

746 represented by a principal component (PCA). The top-right subpanel shows a PCA of 315

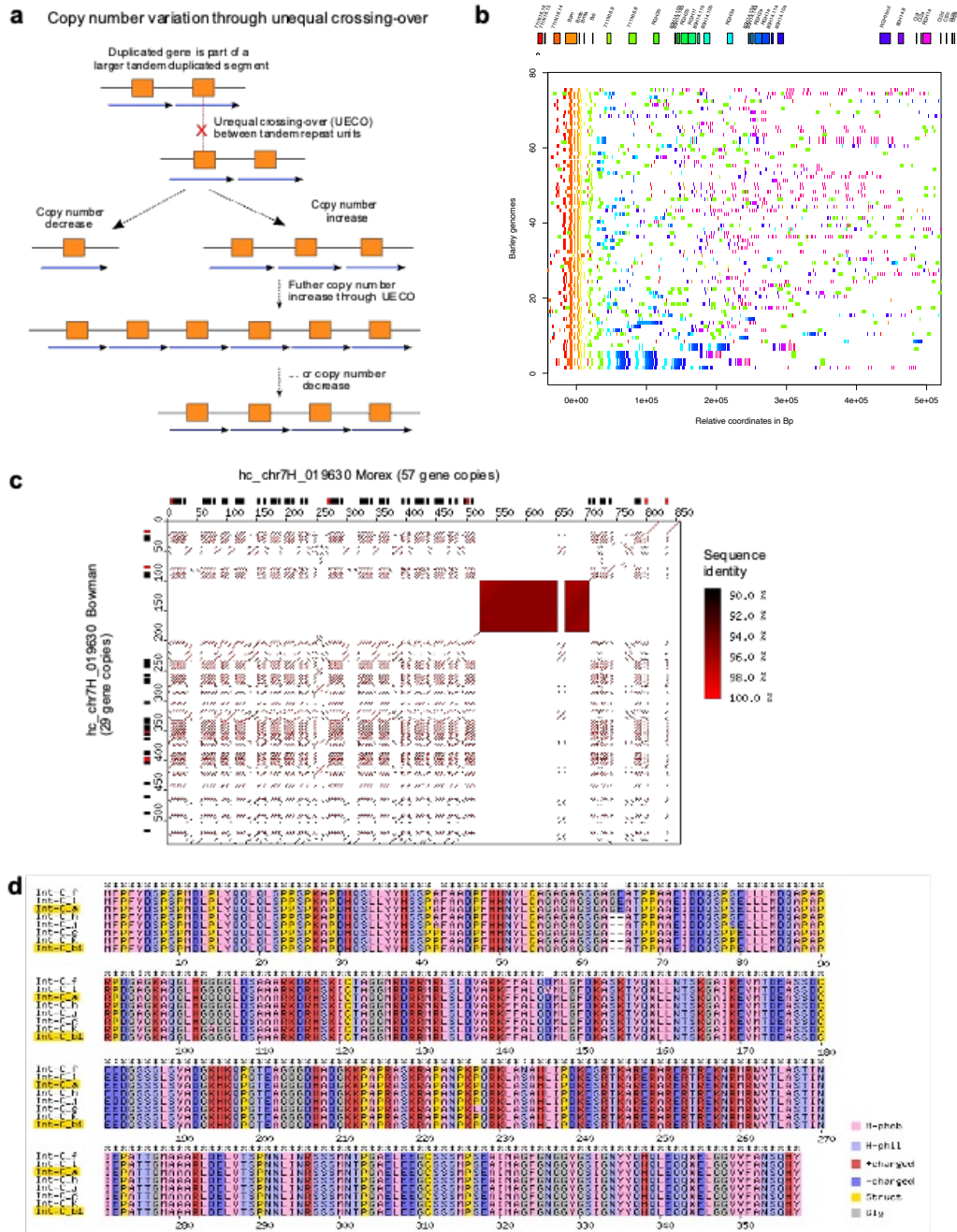
747 elite cultivars. Accessions are according to genepool (2-rowed spring, 2-rowed winter, 6-

748 rowed winter). The proportion of variance explained by the PCA is shown in the axis labels.

749 **(b)** Counts of single-nucleotide polymorphisms (SNPs) and short insertions and deletions

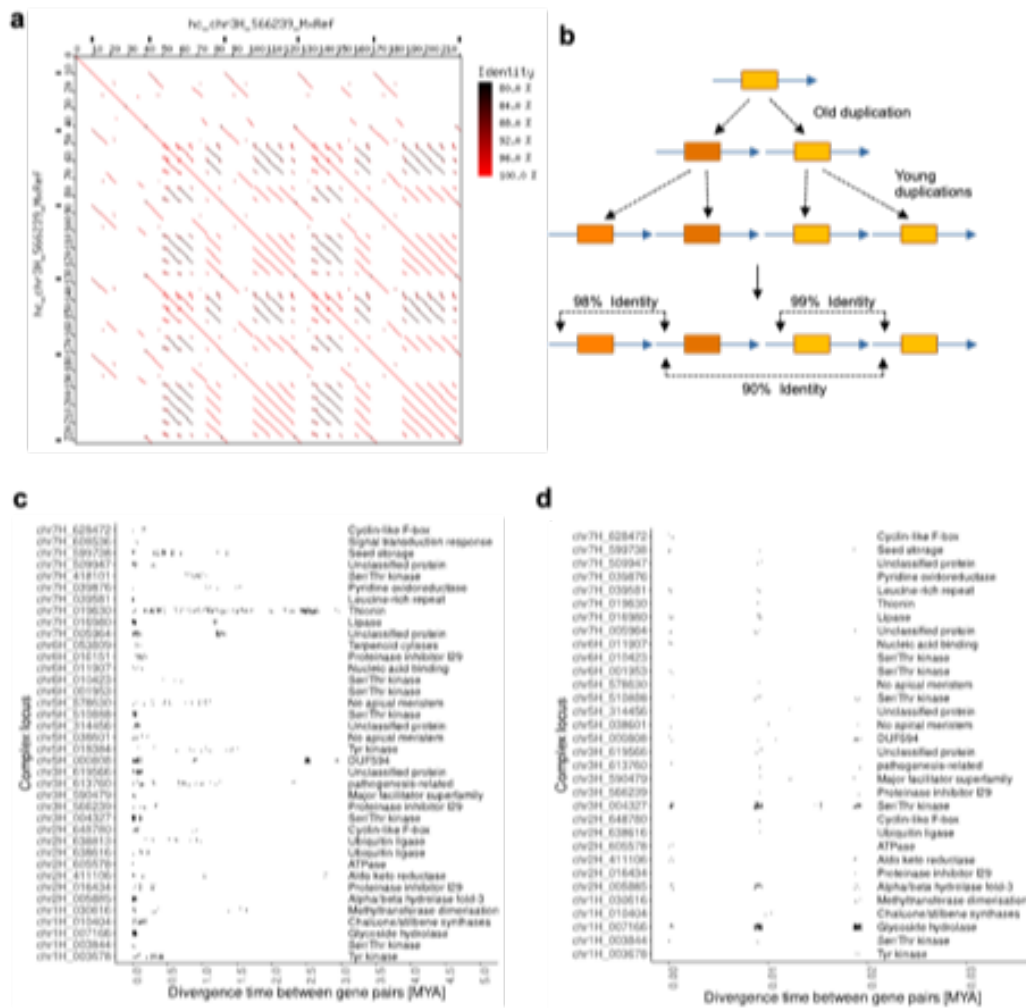
750 (indels) detected in those data.

751

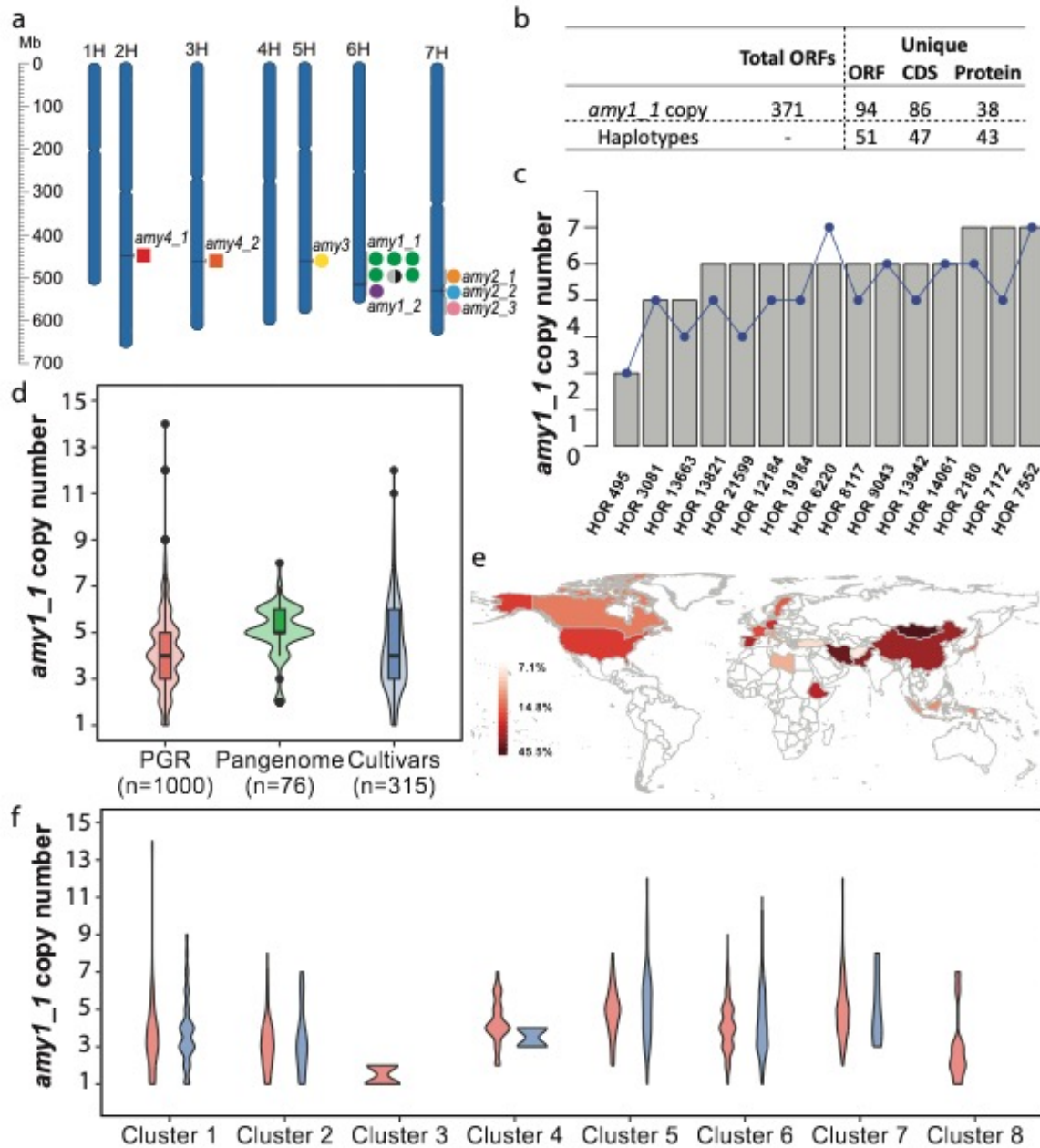


752
753

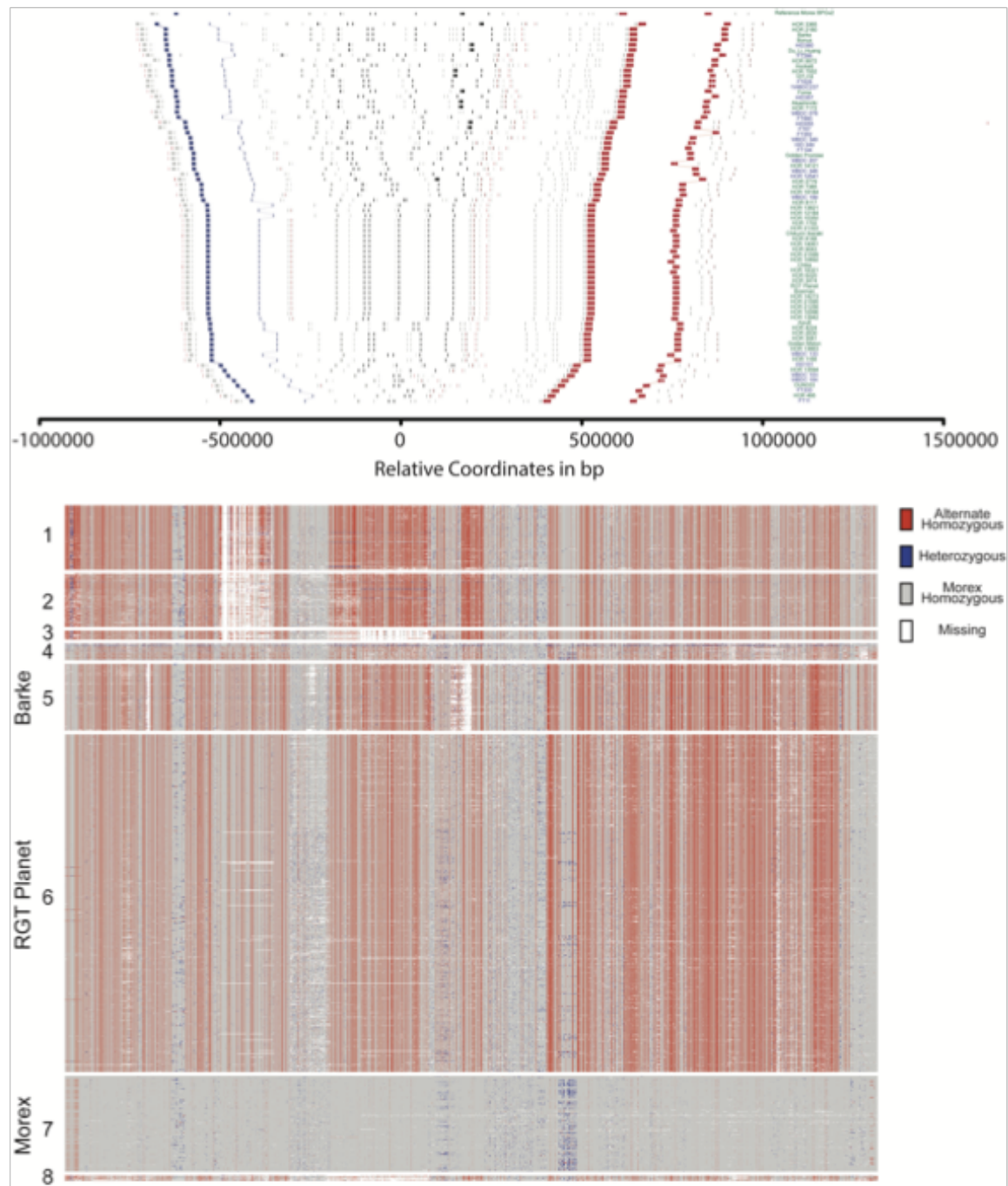
754 **Extended Data Figure 6. Complex loci are hot spots for copy number variation (CNV).** (a) The schematic
755 model shows how, once an initial duplication is established, unequal homologous recombination (unequal
756 crossing-over, UECO) between repeat units can lead to rapid expansion and contraction of the loci, thereby
757 leading to CNV of genes. (b) Structure of the *Mla* region across the 76 pangenome accessions. The gene
758 models present in the Morex genome are shown on top. (c) Dot plot alignment of the example locus
759 chr7H_019630 which contains a cluster of thionin genes. The sequences of cv. Morex (horizontal) and wild
760 barley HID101 (vertical) were aligned. Predicted intact genes are indicated as black boxes along the left and top
761 axes. Predicted pseudogenes are shown in red. The axis scale is kb. The filled rectangle at positions ~520-720
762 kb in Morex represents an array of short tandem repeats which does not contain genes and does not have
763 sequence homology to the gene-containing tandem repeats of the locus. (d) Predicted protein variants of *Int-c*
764 (*HvTB1*) genes. Previously described alleles are highlighted in yellow. Color code: H-phob: Hydrophobic aa, H-
765 phil: Hydrophilic aa, +charged: positively charged aa, - charged: negatively charged aa, Struct: structural aa,
766 Cystein or Prolin, Gly: Glycin.



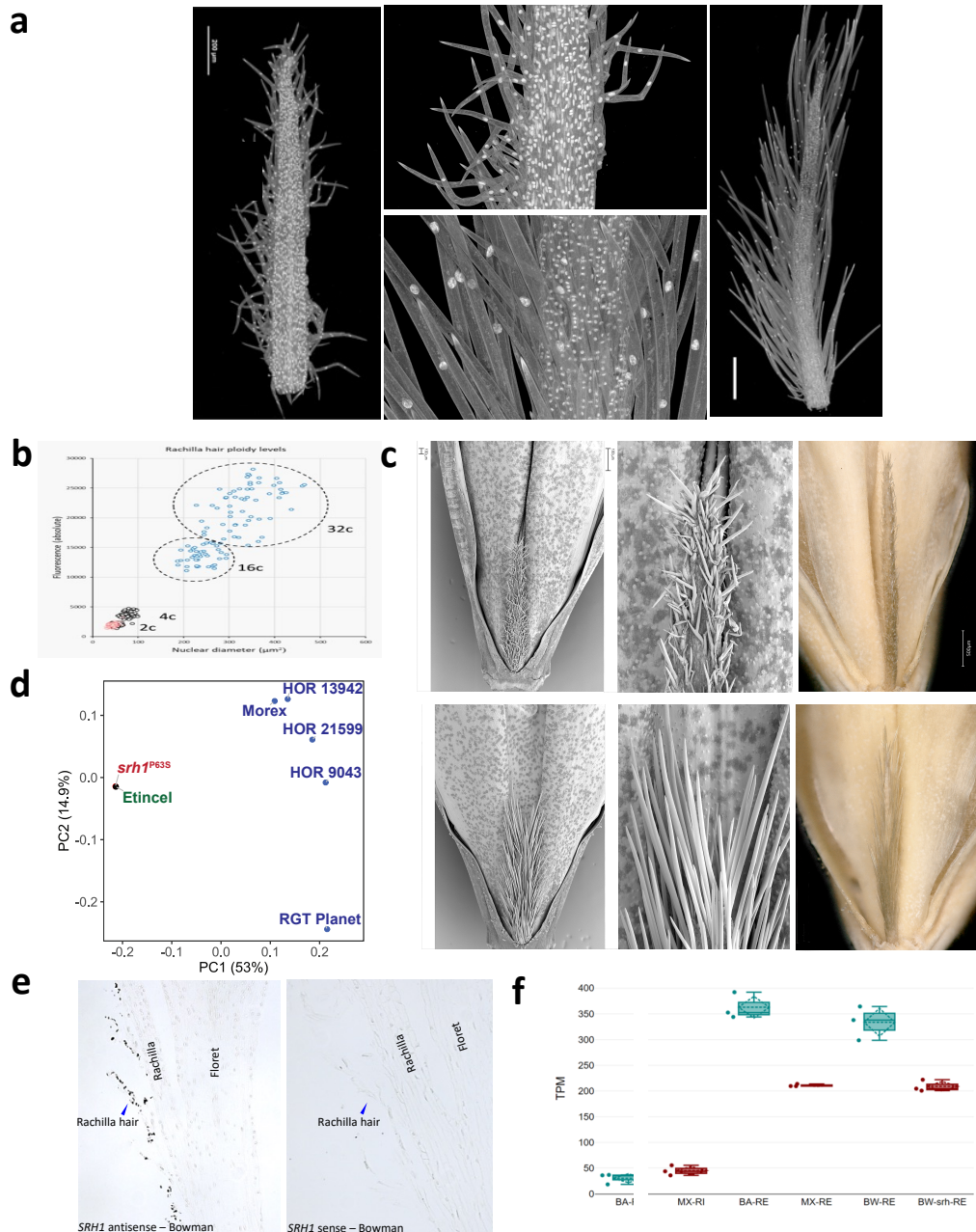
767
 768 **Extended Data Figure 7.** Molecular dating of divergence times between duplicated gene
 769 copies in complex loci. **(a)** Dot plot example of locus `hc_chr3H_566239` which underwent
 770 multiple waves of tandem duplications, which is reflected in varying levels of sequence
 771 identity between tandem repeats (color-coded). **(b)** Schematic mechanism for how different
 772 levels of sequence identity between tandem repeats evolve. In the example, an ancestral
 773 duplication was followed by two independent subsequent duplications, leading to varying
 774 levels of sequence identity between tandem repeat units. Genes are indicated as orange
 775 boxes while blue arrows indicate the tandem repeats they are embedded in. **(c)** Divergence
 776 time estimates between duplicated gene copies in complex loci. Shown are only those
 777 complex loci which have at least six tandem-duplicated genes. Each dot represents one
 778 divergence time estimate for a duplicated gene pair from the respective locus. The x-axis
 779 shows the estimated divergence time in million years. At the right-hand side, classification of
 780 proteins encoded by genes in the locus are shown. Note that several loci had multiple waves
 781 of gene duplications over the past 3 million years. **(d)** Subset of those loci shown in (c) that
 782 had at least one gene duplication within the past 20,000 years. The divergence time
 783 estimates appear in groups, since they represent the presence of 0, 1 and 2 nucleotide
 784 substitutions, respectively, in the approx. 4 kb of aligned sequences that were used for
 785 molecular dating.



786
 787 **Extended Data Figure 8. *amy1_1* locus structure and copy number in 76 assemblies and 1,315**
 788 **whole genome sequenced accessions. (a)** Chromosomal location of 12 α -amylase genes in the
 789 MorexV3 genome assembly. **(b)** Summary of *amy1_1* locus sequence diversity in 76 pangenome
 790 assemblies (**Supplementary Tables 13-16, 18-19**). Total *amy1_1* ORFs in pangenome and unique
 791 copies and haplotypes of ORF, CDS and protein. Haplotype denotes unique combinations of ORF, CDS
 792 and protein in individual accessions. **(c)** Comparison of *amy1_1* copy numbers identified in the
 793 pangenome assemblies versus *k*-mer based estimation from raw reads. Grey bars denote copy
 794 number from pangenome, blue dots denote *k*-mer estimated copy number. **(d)** *amy1_1* copy number
 795 estimation in 76 pangenome assemblies ("Pangenome"), 1,000 whole-genome sequenced plant
 796 genetic resources ("PGR"), and 315 whole-genome sequenced European elite cultivars ("Cultivars")
 797 using *k*-mer based methods. **(e)** Distribution of accessions with *amy1_1* copy numbers >5 per
 798 country (as percentage of total accessions in country for countries with ≥ 10 accessions). **(f)** *amy1_1*
 799 copy number within each haplotype cluster (see **Extended Data Figure 9b**). Red color refers to 1,000
 800 plant genetic resource accessions, green refers to 76 pangenome accessions and blue refers to 315
 801 European elite cultivars. Cluster #5, #6 and #7 contain Barke, RGT Planet and Morex, respectively.



802
803 **Extended Data Figure 9. Haplotype structure of the *amy1_1* locus. (a)** Structural diversity in the vicinity of
804 *amy1_1* in the 76 pangenome assemblies. Each line shows the gene order in the sequence assembly of one
805 genotype. The Morex V3 reference is shown on top. Coloured rectangles stand for gene models extracted from
806 BLAST alignments against the corresponding gene models in MorexV3. Black rectangles represent *amy1_1*
807 homologs and grey rectangles other genes. Blue and red rectangles represent marker genes used to define the
808 synteny, delimit the region, and sort the accessions based on the distance between endpoints. Lines connect
809 genes models between different genomes. Accession names are given on the right axis and are coloured
810 according to type (blue – wild, green – domesticated). In HOR 8148, five copies assigned to 6H are shown. Two
811 copies assigned to an unanchored contig are not shown. **(b)** SNP haplotype clusters at the *amy1_1* locus among
812 1,315 genomes of domesticated and wild barley accessions, including genomes of 315 elite barley cultivars.
813 The 6H:516,385,490-517,116,415 bp in the Morex V3 genome sequence is shown. Haplotype clusters #5, #6
814 and #7 contain the elite malting varieties Barke, RGT Planet and Morex, respectively.



815
 816 **Extended Data Figure 10: Genetic dissection of the *srh1* locus. (a)** Light microscopy of short- and long-haired
 817 rachillae at developmental stage W8.5-9 using DAPI staining to visualize the nuclei. Size differences of nuclei in
 818 epidermal and trichome cells are very obvious. **(b)** Densitometric measurement of DNA content in epidermal and
 819 trichome cells of DAPI stained rachillae of genotypes Morex and Barke, respectively. While trichome cells
 820 in short-haired rachillae undergo only one cycle of endoreduplication, the cells in long haired trichomes show
 821 eight to sixteen-fold higher DNA contents than epidermal cells indicating three to four cycles of
 822 endoreduplication. **(c)** *srh1* mutant discovery. FIND-IT screenings identified a mutant with short-fuzzy hairs
 823 (top) in the background of the long-haired variety Etincel (bottom). The mutants are a P64S non-synonymous
 824 sequence exchange. Scale bar - 1mm. **(d)** Principal coordinate analysis of SNP array genotyping data of different
 825 barley genotypes. Etincel and its mutant cluster together, proving their isogenicity. **(e)** mRNA *in situ*
 826 hybridization of *HvSRH1* in longitudinal spikelet sections of Bowman with anti-sense (left) and sense (right)
 827 probes. The blue arrow indicates the position of a rachilla hair. **(f)** *HvSRH1* transcript abundance in RNA
 828 sequencing data of rachilla tissue in Barke (BA, long-haired), Morex (MX, short-haired), Bowman (BW, long-
 829 haired) and a short-haired near-isogenic line of Bowman (BW-srh). Samples were taken at two development
 830 stages: rachilla hair initiation (RI) and elongation (RE). Abundance was measured as transcripts per million
 831 (TPM).

833 **Plant growth and high molecular weight DNA isolation**

834

835 Twenty-five seeds each from the selected accessions (**Supplementary Tables 1 and 6**) were
836 sown on 16 cm diameter pots with compost soil. Plants were grown under greenhouse
837 conditions with sodium halogen artificial 21°C in the day for 16 hrs and 18°C at night for 8 hrs.
838 Leaves (8 g) were collected from 7-day old seedlings, ground with liquid nitrogen to a fine
839 powder and stored at -80°C.

840 High molecular weight (HMW) DNA was purified from the powder, essentially as described¹.
841 Briefly, nuclei were isolated, digested with proteinase K and lysed with SDS. Here, a standard
842 watercolor brush with synthetic hair (size 8) was used to re-suspend the nuclei for digestion
843 and lysis. HMW DNA was purified using phenol-chloroform extraction and precipitation with
844 ethanol as described¹. Subsequently, the HMW DNA was dissolved in 50 ml TE (pH 8,0) and
845 precipitated by the addition of 5 ml 3 M sodium acetate (pH 5,2) and 100 ml ice-cold ethanol.
846 The suspension was mixed by slow circular movements resulting in the formation of a white
847 precipitate (HMW DNA), which was collected using a wide-bore 5 ml pipette tip and
848 transferred for 30 sec into a tube containing 5 ml 75% ethanol. The washing was repeated
849 twice. The HMW DNA was transferred into a 2 ml tube using a wide-bore tip, collected with a
850 polystyrene spatula, air-dried in a fresh 2 ml tube and dissolved in 500 µl 10 mM Tris-Cl (pH
851 8.0). For quantification the Qubit dsDNA High Sensitivity assay kit (Thermo Fisher Scientific,
852 MA, USA) was used. The DNA size-profile was recorded using the Femto Pulse system and the
853 Genomic DNA 165 kb kit (Agilent Technologies Inc, CA, USA). In typical experiments the peak
854 of the size-profile of the HMW DNA for library preparation was around 165 kb.

855

856 **DNA library preparation and Pacbio HiFi sequencing**

857

858 For fragmentation of the HMW DNA into 20 kb fragments, a Megaruptor 3 device (speed: 30)
859 was used (Diagenode, NJ, USA). A minimum of two HiFi SMRTbell libraries were prepared for
860 each barley genotype following essentially the manufacturer's instructions and the SMRTbell
861 Express Template Prep Kit (Pacific Biosciences, CA, USA). The final HiFi libraries were size
862 selected (narrow size range: 18-21 kb) using the SageELF system with a 0,75% Agarose Gel
863 Cassette (Sage Sciences, MA, USA) according to standard manufacturer's protocols.

864 HiFi CCS reads were generated operating the PacBio Sequel IIe instrument (Pacific Biosciences,
865 CA, USA) following the manufacturer's instructions. Per genotype about four 8 M SMRT cells
866 (average yield: 24 Gb HiFi CCS per 8 M SMART cell) were sequenced to obtain an approximate
867 haploid genome coverage of about 20-fold. In typical experiments the concentration of the
868 HiFi library on plate was 80-95 pM. 30 h movie time, 2 h pre-extension and sequencing
869 chemistry v2.0 were used. The resulting raw data was processed using the CCS4 algorithm
870 (<https://github.com/PacificBiosciences/ccs>).

871

872 **Hi-C library preparation and Illumina sequencing**

873

874 *In situ* Hi-C libraries were prepared from one-week old barley seedlings based on the
875 previously published protocol². Dovetail Omni-C data were generated for Bowman, Aizu6,
876 Golden Melon, 10TJ18 as per manufacturer's instructions
877 (<https://dovetailgenomics.com/products/omni-c-product-page/>). Sequencing and Hi-C raw
878 data processing was performed as described before^{3,4}.

879

880

881 **Genome sequence assembly and validation**

882 PacBio HiFi reads were assembled using hifiasm (v0.11-r302)⁵. Pseudomolecule construction
883 was done with the TRITEX pipeline⁶. Chimeric contigs and orientation errors were identified
884 through manual inspection of Hi-C contact matrices. Genome completeness and consensus
885 accuracy were evaluated using Merqury (v1.3)⁷. Levels of duplication and heterozygosity were
886 assessed with Merqury and FindGSE (v1.94)⁸. BUSCO (Benchmarking Universal Single-Copy
887 Orthologs) (v3.0.2)⁹ with the embryophyta_odb9 data set was run on the final assemblies.

888

889 **Single-copy pangenome construction**

890 The single-copy regions in each chromosome-level assembly were identified by filtering 31-
891 mers occurring more than once in the genomic regions by BBDuk (BBDuk_37.93,
892 <https://jgi.doe.gov/data-and-tools/software-tools/bbtools>). Then, the single-copy regions
893 were obtained in BED format and their sequences were retrieved using BEDTools (v2.29.2)¹⁰.
894 The single-copy sequences were clustered using MMseqs2 (Many-against-Many sequence
895 searching)¹¹ with the parameters “ --cluster-mode ” and setting over 95% sequence identity.
896 A representative from each cluster (the largest in a cluster) was selected to estimate the
897 pangenome size.

898

899 **Illumina resequencing**

900 A total of 1,000 plant genetic resources and 315 elite barley varieties (**Supplementary Table**
901 **5**) were used for whole genome resequencing. Illumina Nextera libraries were prepared and
902 sequenced on an Illumina NovaSeq 6000 at IPK Gatersleben (**Supplementary Table 5**).

903

904 **SNP and SV calling**

905 Reciprocal genome alignment in which each of the pangenome assemblies was aligned to the
906 MorexV3 assembly with the latter acting either as alignment query or reference. Alignment
907 was done with Minimap2 (version 2.20)¹². From the resultant two alignment tables, insertion
908 and deletions were called by Assemblytics (version 1.2.1)¹³ and only deletions were selected
909 in both alignments to convert into presence/absence variants relative to the Morex reference
910 genome. Further, balanced rearrangements (inversions, translocations) were scanned for with
911 SyRI¹⁴. To call SNPs, raw sequencing reads were trimmed using cutadapt (version 3.3)¹⁵ and
912 aligned to the MorexV3 reference genome using Minimap2 (version 2.20)¹². The resulting
913 alignments were sorted with Novosort (V3.09.01) (<http://www.novocraft.com>). BCFtools
914 (version 1.9)¹⁶ was used to call SNPs and short insertions and deletions. GWAS for was done
915 with GEMMA¹⁷.

916

917 **Preparation and Illumina sequencing of narrow-size WGS libraries for core50**

918 10 µg DNA in 130 µl were sheared in tubes (Covaris microTUBE AFA Fiber Pre-Slit Snap Cap)
919 to an average size of approximately 250 bp using a Covaris S220 focused-ultrasonicator (peak
920 incidence power: 175 W, duty factor: 10%; cycles per burst: 200; time: 180 sec) according to
921 standard manufacturer’s protocols (Covaris Ltd., Brighton, UK). The sheared DNA was size
922 selected using a BluePippin device and a 1.5% agarose cassette with internal R2 marker (Sage
923 Sciences, MA, USA). A tight size setting at 260 bp was used for the purification of fragments in
924 the narrow range between 200-300 bp (typical yield: 1-3 µg). The size selected DNA was used
925 for the preparation of PCR-free whole genome shotgun libraries using the Roche KAPA Hyper
926 Prep kit according to the manufacturer’s protocols (Roche Diagnostics Deutschland GmbH,

927 Mannheim, Germany). A total of 10 to 12 libraries were provided with unique barcodes,
928 pooled at equimolar concentrations and quantified by qPCR using the KAPA Library
929 Quantification Kit for Illumina Platforms according to standard protocols (Roche Diagnostics
930 Deutschland GmbH, Mannheim, Germany). The pools were sequenced (2x 151 bp, paired-
931 end) using four S4 XP flowcells and the Illumina NovaSeq 6000 system (Illumina Inc., San
932 Diego, CA, USA) at IPK Gatersleben.

933

934 **Contig assembly of core50 sequencing data**

935 Raw reads were demultiplexed based on index sequences and duplicate reads were removed
936 from the sequencing data using Fastuniq¹⁸. The read1 and read2 sequences were merged
937 based on the overlap using bbmerge.sh from bbmap (v37.28)¹⁹. The merged reads were error
938 corrected using BFC (v181)²⁰. The error corrected merged reads were used as an input for
939 Minia3 (v3.2.0)²¹ to assemble reads into unitigs with the following parameters, -no-bulge-
940 removal -no-tip-removal -no-ec-removal -out-compress 9 -debloom original. The Minia3
941 source was assembled to enable k-mer size up to 512 as described in the Minia3 manual.
942 Iterative Minia3 runs with increasing k-mer sizes (100, 150, 200, 250 and 300) were used for
943 assembly generation as provided in the GATB Minia pipeline ([https://github.com/GATB/gatb-
944 minia-pipeline](https://github.com/GATB/gatb-minia-pipeline)). In the first iteration, k-mer size of 50 was used to assemble input reads into
945 unitigs. In the next runs, the input reads as well as the assembly of the previous iteration were
946 used as input for the minia3 assembler. BUSCO analysis was conducted on the contig
947 assemblies using BUSCO (v3.0.2) with embryophyta_odb9 data set⁹. In addition, high-
948 confidence gene models from Morex V3 reference²² were aligned to the contig assemblies to
949 assess completeness with the parameter of $\geq 90\%$ query coverage and $\geq 97\%$ identity.

950

951 **Pangenome accession in diversity space**

952 Pseudo-FASTQ paired-end reads (10-fold coverage) were generated from the 76 pangenome
953 assemblies with fastq_generator (https://github.com/johanzi/fastq_generator) and aligned to
954 MorexV3 reference genome sequence assembly²² using Minimap2 (version 2.24-r1122, ref.
955 ¹²). SNPs were called together with short-read data (**Supplementary Table 5**) using BCFtools²³
956 version 1.9 with the command “mpileup -q 20 -Q20 --excl-flags 3332”. To plot the diversity
957 space of cultivated barley, the resultant variant matrix was merged with that of 19,778
958 domesticated barleys of Milner et al.²⁴ (genotyping-by-sequencing [GBS] data). SNPs with
959 more than 20 % missing or more than 20 % heterozygous calls were discarded. PCA was done
960 with smartpca²⁵ version 7.2.1. To represent the diversity of wild barleys, we used published
961 GBS and whole-genome sequencing (WGS) data of 412 accessions of that taxon^{26,27}. Variant
962 calling for GBS data was done with BCFtools²³ (version 1.9) using the command “mpileup -q
963 20 -Q20”. The resultant variant matrix was filtered as follows: (1) only bi-allelic SNP sites were
964 kept; (2) homozygous genotype calls were retained if their read depth was ≥ 2 and ≤ 50 and
965 set to missing otherwise; (3) heterozygous genotype calls were retained if the read depth of
966 both alleles was ≥ 2 and set to missing otherwise. SNPs with more than 20 % missing, more
967 than 20 % heterozygous calls or a minor allele frequency below 5 % were discarded. PCA was
968 done with smartpca²⁵ version 7.2.1. A matrix of pairwise genetic distances based on identity-
969 by-state (IBS) was computed with Plink2 (version 2.00a3.3LM, ref.²⁸) and used to construct an
970 NJ tree with Fneighbor (<http://emboss.toulouse.inra.fr/cgi-bin/emboss/fneighbor>) in the
971 EMBOSS package²⁹. The tree was visualized with Interactive Tree Of Life (iTOL)³⁰.

972

973 **Haplotype representation**

974 Pangenome assemblies were mapped to MorexV3 as described above (“Pangenome accession
975 in diversity space”). Read depth was calculated with SAMtools²³ version 1.16.1. Genotype calls
976 were set to missing if they were supported by fewer than two reads. Identity-by-state (IBS)
977 was calculated with Plink2 (version 2.000a3.3LM, ref. ²⁸) in 1 Mb windows (shift: 0.5 Mb) using
978 the using command “--sample-diff counts-only counts-cols=ibs0, ibs1”. Windows which in one
979 of both accessions in the comparison had 2-fold coverage over less than 200 kb were set to
980 missing. The number of differences (d) in a window was calculated as $\text{ibs0} + \text{ibs1} / 2$, where ibs0
981 is the number of homozygous differences and ibs1 that of heterozygous ones. This distance
982 was normalized for coverage by the formula $d / i \times 1 \text{ Mb}$, where i is the size in bp of the region
983 covered in both accessions in the comparison had at least 2-fold coverage. In each window,
984 we determined for each among the PGR and cultivars panel, the closest pangenome accession
985 according to the coverage-normalized IBS distance. Only accessions with fewer than 10 %
986 missing windows due to low coverage were considered, leaving 899 PGRs and 264 cultivars.
987 The distance to the closest pangenome accession was plotted with the R package ggplot2 to
988 determine the threshold for similarity (**Extended Data Fig. 2d**).

989

990 **Transcriptome sequencing for gene annotation**

991 Data for transcript evidence-based genome annotation was provided by the International
992 Barley Pan-Transcriptome Consortium, and a detailed description of sample preparation and
993 sequencing is provided elsewhere. In brief, the 20 genotypes sequenced for the first version
994 of the barley pangenome²⁶ were used for transcriptome sequencing. Five separate tissues
995 were sampled for each genotype. These were: embryo (including mesocotyl and seminal
996 roots), seedling shoot, seedling root, inflorescence and caryopsis. Three biological replicates
997 were sampled from each tissue type, amounting to 330 samples. Four samples failed quality
998 control and were excluded.

999 Preparation of the strand-specific dUTP RNA-Seq libraries, and Illumina paired-end 150 bp
1000 sequencing were carried out by Novogene (UK) Company Limited. In addition, PacBio Iso-Seq
1001 sequencing was carried out using a PacBio Sequel IIe sequencer at IPK Gatersleben. For this,
1002 a single sample per genotype was obtained by pooling equal amounts of RNA from a single
1003 replicate from all five tissues. Each sample was sequenced on an individual 8M SMRT cell.

1004

1005 **De novo gene annotation**

1006 Structural gene annotation was done combining de novo gene calling and homology-based
1007 approaches with RNAseq, IsoSeq, and protein datasets (**Extended Data Fig. 3a**). Using
1008 evidence derived from expression data, RNAseq data were first mapped using STAR³¹ (version
1009 2.7.8a) and subsequently assembled into transcripts by StringTie³² (version 2.1.5, parameters
1010 -m 150-t -f 0.3). Triticeae protein sequences from available public datasets (UniProt³³,
1011 <https://www.uniprot.org>, 05/10/2016) were aligned against the genome sequence using
1012 GenomeThreader³⁴ (version 1.7.1; arguments -startcodon -finalstopcodon -species rice -
1013 gcmincoverage 70 -prseedlength 7 -prhdist 4). Isoseq datasets were aligned to the genome
1014 assembly using GMAP³⁵ (version 2018-07-04). All assembled transcripts from RNAseq, IsoSeq,
1015 and aligned protein sequences were combined using Cuffcompare³⁶ (version 2.2.1) and
1016 subsequently merged with StringTie (version 2.1.5, parameters --merge -m150) into a pool of
1017 candidate transcripts. TransDecoder (version 5.5.0; <http://transdecoder.github.io>) was used
1018 to identify potential open reading frames and to predict protein sequences within the
1019 candidate transcript set.

1020 *Ab initio* annotation was initially done using Augustus³⁷ (version 3.3.3). GeneMark³⁸ (version
1021 4.35) was additionally employed to further improve structural gene annotation. To avoid
1022 potential over-prediction, we generated guiding hints using the above described RNAseq,
1023 protein, and IsoSeq datasets as described before³⁹. A specific Augustus model for barley was
1024 built by generating a set of gene models with full support from RNAseq and IsoSeq. Augustus
1025 was trained and optimized following a published protocol³⁹. All structural gene annotations
1026 were joined using EvidenceModeller⁴⁰ (version 1.1.1), and weights were adjusted according
1027 to the input source: *ab initio* (Augustus: 5, GeneMark: 2), homology-based (10). Additionally,
1028 two rounds of PASA⁴¹ (version 2.4.1) were run to identify untranslated regions and isoforms
1029 using the above described IsoSeq datasets.

1030 We used BLASTP⁴² (ncbi-blast-2.3.0+, parameters -max_target_seqs 1 -evalue 1e-05) to
1031 compare potential protein sequences with a trusted set of reference proteins (Uniprot
1032 Magnoliophyta, reviewed/Swissprot, downloaded on 3 Aug 2016; <https://www.uniprot.org>).
1033 This differentiated candidates into complete and valid genes, non-coding transcripts,
1034 pseudogenes, and transposable elements. In addition, we used PTREP (Release 19;
1035 <http://botserv2.uzh.ch/kelldata/trep-db/index.html>), a database of hypothetical proteins
1036 containing deduced amino acid sequences in which internal frameshifts have been removed
1037 in many cases. This step is particularly useful for the identification of divergent transposable
1038 elements with no significant similarity at the DNA level. Best hits were selected for each
1039 predicted protein from each of the three databases. Only hits with an e-value below 10e-10
1040 were considered. Furthermore, functional annotation of all predicted protein sequences was
1041 done using the AHRD pipeline (<https://github.com/groupschoof/AHRD>).

1042 Proteins were further classified into two confidence classes: high and low. Hits with subject
1043 coverage (for protein references) or query coverage (transposon database) above 80% were
1044 considered significant and protein sequences were classified as high-confidence using the
1045 following criteria: protein sequence was complete and had a subject and query coverage
1046 above the threshold in the UniMag database or no BLAST hit in UniMag but in UniPoa and not
1047 PTREP; a low-confidence protein sequence was incomplete and had a hit in the UniMag or
1048 UniPoa database but not in PTREP. Alternatively, it had no hit in UniMag, UniPoa, or PTREP,
1049 but the protein sequence was complete. In a second refinement step, low-confidence proteins
1050 with an AHRD-score of 3* were promoted to high-confidence.

1051

1052 **Transposon masking for de novo gene detection**

1053 The 20 barley accessions with expression data were softmasked for transposons prior to the
1054 *de novo* gene detection using the REdat_9.7_Triticeae section of the PGSB transposon
1055 library⁴³. Vmatch (<http://www.vmatch.de>) was used as matching tool with the following
1056 parameters: identity>=70%, minimal hit length 75 bp, seedlength 12 bp (vmmatch -d -p -l 75
1057 -identity 70 -seedlength 12 -exdrop 5 -qmaskmatch tolower). The percentage masked was
1058 around 80% and almost identical for all 20 accessions.

1059

1060 **Gene projections**

1061 Gene contents of the remaining 56 barley genotypes were modelled by the projection of high
1062 confidence (HC) genes based on evidence-based gene annotations of the 20 barley genotypes
1063 described above. The approach was similar to and built upon a previously described method²⁶.
1064 To reduce computational load, 760,078 HC-genes of the 20 barley annotations were clustered
1065 by cd-hit⁴⁴ requiring 100% protein sequence similarity and a maximal size difference of four
1066 amino acids. The resulting 223,182 source genes were subsequently used for all downstream

1067 projections as non-redundant transcript set representative for the evidence-based
1068 annotations. For each source, its maximal attainable score was determined by global protein
1069 self-alignment using the Needleman-Wunsch algorithm as implemented in Biopython⁴⁵ v1.8
1070 and the blosum62 substitution matrix⁴⁶ with a gap open and extension penalty of 0.5 and
1071 10.0, respectively.

1072 Next, we surveyed each barley genome sequence using minimap2 (ref. ¹²) with options ‘-ax
1073 splice:hq’ and ‘-uf’ for genomic matches of source transcripts. Each match was scored by its
1074 pairwise protein alignment with the source sequence that triggered the match. Only complete
1075 matches with start and stop codons and a score ≥ 0.9 of the source self-score (see above) were
1076 retained. The source models were classified into four bins by decreasing confidence qualities:
1077 with or without pfam domains, plastid- and transposon-related genes. Projections were
1078 performed stepwise for the four qualities, starting from the highest to the lowest. In each
1079 quality group, matches were then added into the projected annotation if they did not overlap
1080 with any previously inserted model by their coding region. Insertion order progressed from
1081 the top to the lowest scoring match. In addition, we tracked the number of insertions for each
1082 source by its identifier. For the two top quality categories, we performed two rounds of
1083 projections, firstly inserting each source maximally only once followed by rounds allowing one
1084 source inserted multiple times into the projected annotation. To consolidate the 20 evidence-
1085 based, initial annotations for any genes potentially missed, we employed an identical
1086 approach but inserted any non-overlapping matches starting from the prior RNA-seq based
1087 annotation. Phylogenetic Hierarchical Orthogroups (HOGs) based on the primary protein
1088 sequences from 76 annotated barley genotypes were calculated using Orthofinder⁴⁷ version
1089 2.5.5 (standard parameters). Conserved HOGs contain at least one gene model from all 76
1090 barley genotypes. Variable HOGs contain gene models from at least one barley genotypes and
1091 at most 75 barley genotypes. The distribution of all HOG configurations is provided in
1092 **Extended Data Fig. 3b**. GENESPACE⁴⁸ was used to determine syntenic relationships between
1093 the chromosomes of all 76 genotypes.

1094

1095 **Whole-genome pangenome graphs**

1096 Genome graphs were constructed using Minigraph⁴⁹ version 0.20-r559. Other graph
1097 construction tools (PGGB⁵⁰, Minigraph-Cactus⁵¹) turned out to be computationally prohibitive
1098 for a genome of this size and complexity, combined with the large number of accessions used
1099 in this investigation. Minigraph does not support small variants (< 50 bp), thus graph
1100 complexity is lower than with other tools. However, even with Minigraph, graph construction
1101 at the whole genome level was computationally prohibitive and thus graphs had to be
1102 computed separately for each chromosome, precluding detection of interchromosomal
1103 translocations.

1104 Graph construction was initiated using the Morex V3 assembly⁵² as a reference. The remaining
1105 assemblies were added into the graph sequentially, in order of descending dissimilarity to
1106 Morex. Structural variants were called after each iteration using gfatools bubble (v. 0.5-r250-
1107 dirty, <https://github.com/lh3/gfatools>). Following graph construction, the input sequences of
1108 all accessions were mapped back to the graph using Minigraph with the “--call” option
1109 enabled, which generates a path through the graph for each accession. The resulting BED
1110 format files were merged using Minigraph’s mgutils.js utility script to convert them to P lines
1111 and then combined with the primary output of Minigraph in the proprietary RGFA format
1112 (<https://github.com/lh3/gfatools/blob/master/doc/rGFA.md>). Graphs were then converted
1113 from RGFA format to GFA format (<https://github.com/GFA-spec/GFA->

1114 spec/blob/master/GFA1.md) using the “convert” command from the vg toolkit⁵³ version
1115 v1.46.0 "Altamura". This step ensures that graphs are compatible with the wider universe of
1116 graph processing tools, most of which require GFA format as input. Chromosome-level graphs
1117 were then joined into a whole-genome graph using vg combine. The combined graph was
1118 indexed using vg index and vg gbwt, two components of the the vg toolkit⁵³.

1119 General statistics for the whole-genome graph were computed with vg stats. Graph growth
1120 was computed using the heaps command from the ODGI toolkit⁵⁴ version 0.8.2-0-g8715c55,
1121 followed by plotting with its companion script heaps_fit.R. The latter also computes values for
1122 gamma, the slope coefficient of Heap’s law which allows the classification of pangenome
1123 graphs into open or closed pangenomes, i.e. a prediction of whether the addition of further
1124 accessions would increase the size of the pangenome⁵⁵.

1125 Structural variant (SV) statistics were computed based on the final BED file produced after the
1126 addition of the last line to the graph. A custom shell script was used to classify variants
1127 according to the Minigraph custom output format. This allows the extraction of simple, i.e.
1128 non-nested, insertions and deletions (relative to the MorexV3 graph backbone), as well simple
1129 inversions. The remaining SVs fall into the “complex” category where there can be multiple
1130 levels of nesting of different variant types and this precluded further, more fine-grained
1131 classification.

1132 To elucidate the effect of a graph-based reference on short read mapping, we obtained whole
1133 genome shotgun Illumina reads from five barley samples (**Extended Data Fig. 4b**) in the
1134 European Nucleotide Archive (ENA) and mapped these onto the whole genome graph using
1135 vg giraffe⁵⁶. For comparison with the standard approach of mapping reads to a linear single
1136 genome reference, we mapped the same reads to the Morex V3 reference genome sequence
1137 assembly⁵² with bwa mem⁵⁷ version 0.7.17-r1188. Mapping statistics were computed with
1138 vg⁵³ stats and samtools²³ stats (version 1.9), respectively.

1139

1140 **Analysis of the *Mla* locus**

1141

1142 The coordinates and sequences of the 32 genes present at the *Mla* locus were extracted from
1143 the MorexV3 genome sequence assembly⁵². To find the corresponding position and copy
1144 number in each of the 76 genomes, we used BLAST⁴² (-perc_identity: 90, -word_size:11, all
1145 other parameters set as default). The expected BLAST result for a perfectly conserved allele is
1146 a long fragment (exon_1) of 2,015 bp follow by a gap of ~1,000 bp due to the intron and
1147 another fragment (exon_2) of 820 bp. To detect the number of copies, first multiple BLAST
1148 results for a single gene were merged if two different BLAST segments were within 1.1kb. Then
1149 only if the total length of the input was found, this was counted as a copy. To analyse the
1150 structural variation across all 76 accessions, the non-filtered BLAST results were plotted in a
1151 region of -20,000 and +500,000 base pairs around the start of the BPM gene
1152 HORVU.MOREX.r3.1HG0004540 that was used as an anchor (present in all 76 lines,
1153 Supplementary Figure 1. To detection the different *Mla* alleles, three different threshold of -
1154 Perc_identity for the BLAST were used: 100, 99 and 98.

1155

1156 **Scan for structurally complex loci**

1157 We utilised a pipeline developed by Rabanus-Wallace et al. ⁵⁸ that performs sequence-
1158 agnostic identification of long-duplication-prone-regions (L-DPRs) in a reference genome,
1159 followed by identification of gene families with a statistical tendency to occur within L-DPRs.
1160 The pipeline assumes that a candidate L-DPR will contain an elevated concentration of locally

1161 repeated sequences in the kb-scale length range. We first aligned the MorexV3 genome
1162 sequence assembly⁵² against itself using lastz⁵⁹ (v1.04.03; arguments: '--notransition --
1163 step=500 -gapped'). For practicality purposes, this was done in 2 Mb blocks with a 200 kb
1164 overlap, and any overlapping I-DPRs identified in multiple windows were merged. For each
1165 window, we ignored the trivial end-to-end alignment, and of the remaining alignments,
1166 retained only those longer than 5 kb and falling fully within 200 kb of one and another. An
1167 alignment 'density' was calculated over the chromosome by calculating, at 'interrogation
1168 points' spaced equally at 1 kb intervals along the length of the chromosome, an alignment
1169 density score that is simply the sum of all the lengths of any of the filtered alignments
1170 spanning that interrogation point. A Gaussian kernel density (bandwidth 10 kbp) was
1171 calculated over these the interrogation points, weighted by their scores. To allow
1172 comparability between windows, the interrogation point densities were normalised by the
1173 sum of scores in the window. Runs of interrogation points at which the density surpassed a
1174 minimum density threshold were flagged as I-DPRs. A few minor adjustments to these regions
1175 (merging of overlapping regions, and trimming⁵⁹ the end coordinates to ensure the stretches
1176 always begin and end in repeated sequence) yielded the final tabulated list of I-DPR
1177 coordinates (**Supplementary Table 7**). The method was implemented in R making use of the
1178 package data.table. Genes in each I-DPRs were clustered with UCLUST⁶⁰ (v11, default
1179 parameters) using a protein clustering distance cutoff of 0.5 and for each cluster the most
1180 frequent functional description as per the MorexV3 gene annotation⁵² was assigned as the
1181 functional description of the cluster.

1182

1183 **Molecular dating of divergence times of duplicated genes in complex loci**

1184 For molecular dating of gene duplications, we used segments of up to 4 kb, starting 1 kb
1185 upstream of duplicated genes in complex loci. With this, we presumed to only use intergenic
1186 sequences which are free from selection pressure and thus evolve at a neutral rate of 1.3×10^{-8}
1187 substitutions per site per year⁶¹. The upstream sequences of all duplicated genes of
1188 respective complex locus were then aligned pairwise with the program Water from the
1189 EMBOSS package²⁹ (obtained from Ubuntu repositories, ubuntu.com). This was done for all
1190 gene copies of all barley accession for which multiple gene copies were found. Molecular
1191 dating of the pairwise alignments was done as previously described⁶² using the substitution
1192 rate of 1.3×10^{-8} substitutions per site per year⁶¹.

1193

1194 ***Amy1_1* analysis in pangenome assemblies**

1195 The *amy1_1* gene copy HORVU.MOREX.PROJ.6HG00545380 was used was used to BLAST
1196 against all 76 genome assemblies. Full-length sequences with identity over 95% were
1197 extracted and used for further analyses. Unique sequences were identified by clustering at
1198 100% identity using CD-Hit⁴⁴ and were aligned using MAFFT⁶³ v7.490. Sequence variants
1199 among *amy1_1* gene copies at genomic DNA, CDS and respective protein level were collected
1200 and *amy1_1* haplotypes (i.e. the combinations of copies) in each genotype assembly were
1201 summarized using R⁶⁴ v4.2.2. A Barke-specific SNP locus
1202 (GGCGCCAGGCATGATCGGGTGGTGGCCAGCCAAGGCGGTGACCTTCGTGGACAACCACGACACCG
1203 GCTCCACGCAGCACATGTGGCCCTCCCTTCTGACA[A/G]GGTCATGCAGGGATATGCGTACATACTCA
1204 CGCACCCAGGGACGCCATGCATCGTGAGTTCGTGTCGTACCAATACATCACATCTCAATTTCTTTCTTGT
1205 TTCGTTTCATAA) for *amy1_1* haplotype cluster ProtHap3 (**Supplementary Table 20**) was
1206 identified and used for KASP marker development (LGC Biosearch Technologies, Hoddesdon,
1207 United Kingdom).

1208

1209 **Comparative analysis of the *amy1_1* locus structure**

1210 Based on the genome annotation of cv. Morex, 15 gene sequences on either side of *amy1_1*
1211 gene copy HORVU.MOREX.PROJ.6HG00545440 were extracted. The 31 genes were compared
1212 against the 76 genome assemblies using NCBI-BLAST⁴² (BLASTN, word_size of 11 and percent
1213 identity of 90, other parameters as default). Alignment plots were generated from the BLAST
1214 result coordinates by scaling based on the mid-point between
1215 HORVU.MOREX.r3.6HG0617300/HORVU.MOREX.PROJ.6HG00545250 and
1216 HORVU.MOREX.r3.6HG0617710/HORVU.MOREX.PROJ.6HG00545670. All BLAST results in the
1217 region (+/- 1Mb) around this mid-point were plotted using R⁶⁴.

1218

1219 ***Amy1_1* PacBio amplicon sequencing.**

1220 Genomic DNA from one-week old Morex seedling leaves was extracted with DNeasy® Plant
1221 Mini Kit (QIAGEN GmbH, Hilden, Germany). Based on the MorexV3 genome sequence
1222 assembly⁵², *amy1_1* full-length copy-specific primers were designed using Primer3 (ref. ⁶⁵)
1223 (<https://primer3.ut.ee/>): 6F - GTAGCAGTGCAGCGTGAAGTC, 80F -
1224 AGACATCGTTAACCACACATGC, 82F - GTTCTCGTCCCTTGCCTTAA, 82F -
1225 GTTCTCGTCCCTTGCCTTAA, 33R - GATCTGGATCGAAGGAGGGC, 79R -
1226 TCATACATGGGACCAGATCGAG, 80R - ACGTCAAGTTAGTAGGTAGCCC. All forward primers were
1227 tagged with bridge sequence (preceding T to primer name)
1228 [AmC6]gcagtcgaacatgtagctgactcaggtcac, while reverse primers were tagged with
1229 [AmC6]tggatcactgtgcaagcatcacatcgtag to allow annealing to barcoding primers. These bridge
1230 sequence-tagged gene-specific primers were used in pairs with each other targeting 1-2
1231 copies of 3-6 kb *amy1_1* genes, including upstream and downstream 500-1000 bp regions:
1232 T6F + T33R, T6F + T79R, T80F + T80R and T82F + T80R. A two-step PCR protocol was
1233 conducted. The first step PCR reaction was prepared in 25 µl volume using 2 µl DMSO, 0.3 µl
1234 Q5 polymerase (New England Biolab, Massachusetts, United States), 1 µl *amy1_1*-specific
1235 primer pair (10 µM each), 2 µl gDNA, 0.5 µl dNTPs (10mM), 5 µl Q5 buffer and H₂O. The PCR
1236 program was as following: initial denaturation at 98°C/1min followed by 25-28 cycles of
1237 98°C/30 sec, 58°C/30 sec, and 72°C/3 min for extension, with a final extension step of 72°C/2
1238 min. The second PCR step (barcoding PCR) was prepared in the same way using 1 µl of the first
1239 PCR product as DNA template, barcoding primers (Pacific Biosciences of California, Inc.,
1240 California, United States) and the PCR program reduced to 20 cycles. After quality check on
1241 1% agarose gel, all barcoded PCR products were mixed and purified with AMPure® PB (Pacific
1242 Biosciences of California, Inc., California, United States). The SMRT bell library preparation and
1243 sequencing were carried out at BGI Tech Solutions (BGI Tech Solutions Co., Ltd., Hongkong,
1244 China). Sequencing data was analysed using SMRT Link v.10.2. To minimize PCR chimeric noise,
1245 CCS were first constructed for each molecule. Secondly, Long Amplicon Analysis (LAA) was
1246 carried out based on subreads from 50 bp window spanning peak positions of all CCS length.
1247 Final consensus sequences for each *amy1_1* was determined with the aid of size estimation
1248 from agarose gel imaging.

1249

1250 ***Amy1_1* SNP haplotype analysis and k-mer based copy number estimation.**

1251 SNP haplotypes were analyzed in 1,315 plant genetic resources and elite varieties in the
1252 extended *amy1_1*-cluster region (MorexV3 chr6H: 516,385,490 - 517,116,415 bp). SNPs with
1253 >20% missing data among the analyzed lines and minor allele frequency (MAF) < 0.01 were
1254 removed from downstream analyses. The data was converted to 0,1 and 2 format using

1255 VCFtools⁶⁶ and samples were clustered using pheatmap package ([https://cran.r-](https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf)
1256 [project.org/web/packages/pheatmap/pheatmap.pdf](https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf)) from R statistical environment⁵⁸. The
1257 sequential clustering approach was used to achieve the desired separation. At each step, two
1258 extreme clusters were selected and then samples from each cluster were clustered separately.
1259 The process was repeated until the desired separation was achieved based on visual
1260 inspection.

1261 K-mers (k=21) were generated from Morex *amy1_1* gene family member's conserved region
1262 using jellyfish⁶⁷ v2.2.10. After removing k-mers with counts from regions other than *amy1_1*
1263 in the Morex V3 genome assembly, k-mers were counted in the Illumina raw reads
1264 (**Supplementary Table 5**) using Seal (BBtools, [https://jgi.doe.gov/data-and-tools/software-](https://jgi.doe.gov/data-and-tools/software-tools/bbtools/)
1265 [tools/bbtools/](https://jgi.doe.gov/data-and-tools/software-tools/bbtools/)). All k-mer counts were normalized to counts per MorexV3 genome and
1266 *amy1_1* copy number was estimated as the median count of all k-mers from each accession
1267 in R.

1268 Estimation ability was validated by comparing copy number from pangenome assemblies and
1269 short-read sequencing data (**Extended Data Fig. 8c**). For 1,000 plant genetic resources,
1270 countries (≥ 10 accessions) were color shaded based on their proportions of accessions with
1271 *amy1_1* copy number > 5 on a world map using the R package maptools ([https://cran.r-](https://cran.r-project.org/web/packages/maptools/index.html)
1272 [project.org/web/packages/maptools/index.html](https://cran.r-project.org/web/packages/maptools/index.html)).

1273

1274 **AMY1_1 protein structure and protein folding simulation**

1275 The published protein structure of α -amylase AMY1_1 from accession Menuet, in complex
1276 with the pseudo-tetrasaccharide acarbose (PDB:1bg9; ref. ⁶⁸), was used to simulate the
1277 structural context of the amino acid variants identified in barley accessions Morex, Barke and
1278 RGT Planet. The amino acid sequence of the crystalized AMY1_1 protein from Menuet and
1279 the Morex reference copy *amy1_1* HORVU.MOREX.PROJ.6HG00545380 used in this study are
1280 identical. The protein was visualized using PyMol 2.5.5 (Schrödinger Inc. New York, NY, USA).
1281 The Dynamut2 webserver⁶⁹ was used to predict changes in protein stability and dynamics by
1282 introducing amino acid variants identified in the Morex, Barke and RGT Planet genome
1283 assemblies.

1284

1285 **Development of diverse *amy1_1* haplotype barley near-isogenic lines**

1286 Near-isogenic lines (NILs) with different *amy1_1* haplotypes were derived from crosses
1287 between RGT Planet as recipient and Barke or Morex *amy1_1*-cluster donor parents
1288 (ProtHap3, ProtHap4 and ProtHap0, respectively; **Supplementary Table 20**), followed by two
1289 subsequent backcrosses to RGT Planet and one selfing step (BC₂S₁) to retrieve homozygous
1290 plants at the *amy1_1* locus. A total of four *amy1_1*-Barke NILs (ProtHap3) and one *amy1_1*-
1291 Morex NIL (ProtHap0) were developed and tested against RGT Planet (ProtHap4) replicates.
1292 Plants were grown in a greenhouse at 18°C under 16/8-hour light/dark cycles. Foreground and
1293 background molecular markers were used in each generation to assist plant selection.
1294 Respective BC₂S₁ plants were genotyped with the Barley Illumina 15K array (SGS Institut
1295 Fresenius GmbH, TraitGenetics Section, Germany) and grown to maturity. Grains were
1296 harvested and further propagated in field plots in consecutive years in various locations (Nørre
1297 Aaby, Denmark; Lincoln, New Zealand; Maule, France). Grains from field plots were harvested
1298 and threshed using a Wintersteiger Elite plot combiner (Wintersteiger AG, Germany), and
1299 sorted by size (threshold, 2.5 mm) using a Pfeuffer SLN3 sample cleaner (Pfeuffer GmbH,
1300 Germany).

1301

1302 **Micro-malting and α -amylase activity analysis**

1303 Non-dormant barley samples of RGT Planet and respective NILs with different *amy1_1*
1304 haplotypes (50g each, graded >2.5 mm) were micro-malted in perforated stainless-steel
1305 boxes. The barley samples were steeped at 15 °C by submersion of the boxes in water.
1306 Steeping took place for six hours on day one, three hours on day two and one hour on day
1307 three, followed by air rests, to reach 35%, 40% and 45% water content, respectively. The actual
1308 water uptake of individual samples was determined as the weight difference between initial
1309 water content, measured with Foss 1241 NIT instrument (Foss A/S, Hillerød, Denmark), and
1310 the sample weight after surface water removal. During air rest, metal beakers were placed
1311 into a germination box at 15°C. Following the last steep, the barley samples were germinated
1312 for 3 days at 15°C. Finally, barley samples were kiln dried in an MMK Curio kiln (Curio Group
1313 Ltd, Buckingham, England) using a two-step ramping profile. First ramping step started at a set
1314 point of 27°C and a linear ramping at 2°C/h to the breakpoint at 55°C using 100% fresh air.
1315 Second linear ramping was at 4°C/h reaching a maximum at 85°C. This temperature was kept
1316 constant for 90 minutes using 50% air recirculation. The kilned samples were then deculmed
1317 using a manual root removal system (Wissenschaftliche Station für Brauerei, Munich,
1318 Germany). α -amylase activity was measured using the Ceralpha method (Ceralpha Method
1319 MR-CAAR4, Megazyme) modified for Gallery Plus Beermaster (Thermo Fisher Scientific, USA).

1320

1321 **Rachilla hair ploidy measurements**

1322 Ploidy assessment was performed on rachillae harvested from barley spikes at developmental
1323 stage⁷⁰ ~W9.0. Once isolated, rachillae were fixed with 50% Ethanol/10% acetic acid for 16h
1324 after which they were stained with 1 μ M 4',6-Diamidino-2-phenylindol (DAPI) in 50 mM
1325 phosphate buffer (pH 7.2) supplemented with 0.05% Triton X100. Probes were analyzed with
1326 a Zeiss LSM780 confocal laser scanning microscope using a 20x NA 0.8 objective, zoom 4x, and
1327 image size 512 x 512 pixel. DAPI was visualized with a 405 nm laser line in combination with a
1328 405–475 nm bandpass filter. Pinhole was set to ensure the whole nucleus was measured in
1329 one scan. Size and fluorescence intensity of the nuclei were measured with ZEN black (ZEISS)
1330 software. For data normalization small round nuclei of the epidermal proper were used for 2C
1331 calibration.

1332

1333 **Scanning electron microscopy**

1334 Sample preparation and recording by scanning electron microscopy was essentially performed
1335 as described previously⁷¹. In brief, samples were fixed overnight at 4°C in 50 mM phosphate
1336 buffer (pH 7.2) containing 2% v/v glutaraldehyde and 2% v/v formaldehyde. After washing
1337 with distilled water and dehydration in an ascending ethanol series, samples were critical
1338 point-dried in a Bal-Tec critical point dryer (Leica microsystems, [https://www.leica-](https://www.leica-microsystems.com)
1339 [microsystems.com](https://www.leica-microsystems.com)). Dried specimens were attached to carbon-coated aluminium sample
1340 blocks and gold-coated in an Edwards S150B sputter coater (Edwards High Vacuum Inc.,
1341 <http://www.edwardsvacuum.com>). Probes were examined in a Zeiss Gemini30 scanning
1342 electron microscope (Carl Zeiss Microscopy GmbH, <https://www.zeiss.de>) at 5 kV acceleration
1343 voltage. Images were digitally recorded.

1344

1345 **Linkage mapping of *SHORT RACHILLA HAIR 1 (HvSRH1)***

1346 Initial linkage mapping was performed using genotyping-by-sequencing (GBS) data of a large
1347 'Morex' x 'Barke' F₈ RIL population⁷² (ENA project PRJEB14130). The GBS data of 163 RILs,
1348 phenotyped for rachilla hair in the F₁₁ generation, and the two parental genotypes were

1349 extracted from the variant matrix using VCFtools⁶⁶ and filtered as described previously²⁴ for a
1350 minimum depth of sequencing to accept heterozygous and homozygous calls of 4 and 6,
1351 respectively, a minimum mapping quality score of the SNPs of 30, a minimal fraction of
1352 homozygous calls of 30 %, and a maximum fraction of missing data of 25%. The linkage map
1353 was built with the R package ASMap⁷³ using the MSTMap algorithm⁷⁴ and the Kosambi
1354 mapping function, forcing the linkage group to split according to the physical chromosomes.
1355 The linkage mapping was done with R/qtl⁷⁵ using the binary model of the scanone function
1356 with the EM method⁷⁶. The significance threshold was calculated running 1000 permutations
1357 and the interval was determined by a LOD drop of 1. To confirm consistency between the F₈
1358 RIL genotypes and F₁₁ RIL phenotypes, three PCR Allele Competitive Extension (PACE) markers
1359 were designed through 3CR Bioscience (Essex, UK) free assay design service, using
1360 polymorphisms between the genome assemblies of the two parents (**Supplementary Table**
1361 **23**), and PACE genotyping was performed as described earlier⁷⁷. To reduce the *Srh1* interval,
1362 22 recombinant F₈ RILs were sequenced by Illumina whole-genome sequencing (WGS), the
1363 sequencing reads were mapped on MorexV3 reference genome⁵², and the SNP called. The 100
1364 bp region around the flanking SNPs of the *Srh1* interval as well as the sequence of the
1365 candidate gene HORVU.MOREX.r3.5HG0492730 were compared to the pangenome
1366 assemblies using BLASTN⁷⁸ to identify the corresponding coordinates and extract the
1367 respective intervals for comparison. Gene sequences were aligned with Muscle5 (ref. ⁷⁹).
1368 Structural variation between intervals was assessed with LASTZ⁵⁹ version 1.04.03. The motif
1369 search was carried out with the EMBOSS⁸⁰ 6.5.7 tool fuzznuc.
1370

1371 **Cas9-mediated mutagenesis**

1372 Guide RNA (gRNA) target motifs in the 'Golden Promise' *HvSrh1* candidate gene
1373 HORVU.GOLDEN_PROMISE.PROJ.5HG00440000.1 were selected by using the online tool WU-
1374 CRISPR⁸¹ to induce translational frameshift mutations by insertion/deletion of nucleotides
1375 leading to loss-of-function of the gene. One pair of target motifs (gRNA1a:
1376 CCTCGCTGCCCGCCGACGC, gRNA1b: GACAAGACGAAGGCCGCGG) was selected within the
1377 *HvSrh1* candidate gene based on their position within the first half of the coding sequence
1378 and the two-dimensional minimum free energy structures of the cognate single-gRNAs
1379 (NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAGGCUAGUC
1380 CGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU) as modelled by the RNAfold
1381 WebServer⁸² and validated as suggested by Koeppel et al. ⁸³. gRNA-containing transformation
1382 vectors were cloned using the modular CasCADE vector system
1383 (<https://doi.org/10.15488/13200>). gRNA-specific sequences were ordered as DNA
1384 oligonucleotides (**Supplementary Table 24**) with specific overhangs for BsaI-based cloning
1385 into the gRNA-module vectors carrying the gRNA scaffold, driven by the *Triticum aestivum* U6
1386 promoter. Golden Gate assembly of gRNAs and the *cas9* module, driven by the *Zea mays*
1387 *Polyubiquitin 1* (*ZmUbi1*) promoter, were performed according to the CasCADE protocol to
1388 generate the intermediate vector pHP21. To generate the binary vector pHP22, the gRNA and
1389 *cas9* expression units were cloned using SfiI into the generic vector⁸⁴ p6i-2x35S-TE9 that
1390 harbours an *hpt* gene under control of a doubled-enhanced *CaMV35S* promoter in its transfer-
1391 DNA for plant selection. *Agrobacterium*-mediated DNA transfer to immature embryos of the
1392 spring barley Golden Promise was performed as previously described⁸⁵. In brief, immature
1393 embryos were excised from caryopses 12-14 days after pollination and co-cultivated with
1394 *Agrobacterium* strain AGL1 carrying pHP22 for 48 hours. Then, the explants were cultivated
1395 for further callus formation under selective conditions using Timentin and hygromycin, which

1396 was followed by plant regeneration. The presence of T-DNA in regenerated plantlets was
1397 confirmed by *hpt*- and *cas9*-specific PCRs (primer sequences in **Supplementary Table 24**).
1398 Primary mutant plants (M_1 generation) were identified by PCR amplification of the target
1399 region (primer sequences in **Supplementary Table 24**) followed by Sanger sequencing at LGC
1400 Genomics GmbH (Berlin, Germany). Double or multiple peaks in the sequence chromatogram
1401 starting around the Cas9 cleavage site upstream of the target's protospacer-adjacent motif
1402 (PAM) were considered as an indication for chimeric and/or heterozygous mutants. Mutant
1403 plants were grown in a glasshouse until the formation of mature grains. M_2 plants were grown
1404 in a climate chamber under speed breeding conditions (22 h light at 22 °C and 2 h dark at 19
1405 °C, adapted from Watson et al.⁸⁶ and genotyped by Sanger sequencing of PCR amplicons as
1406 given above. M_2 grains were subjected to phenotyping.

1407

1408 **FIND-IT library construction**

1409 We constructed a FIND-IT library in cv. 'Etincel' (6-row winter malting barley; SECOBRA
1410 Recherches) as described in Knudsen et al.⁸⁷. In short, we induced mutations by incubating
1411 2.5 kg of 'Etincel' grain in water overnight at 8°C following an incubation in 0.3 mM NaN₃ at
1412 pH 3.0 for 2 hours at 20°C with continuous application of oxygen. After thoroughly washing
1413 with water, the grains were air-dried in a fume hood for 48 hours. Mutagenized grains were
1414 sown in fields in Nørre Aaby, Denmark, and harvested in bulk using a Wintersteiger Elite plot
1415 combiner (Wintersteiger AG, Germany). In the following generation, 2.5 kg of grain was sown
1416 in fields in Lincoln, New Zealand, and 188 pools of approximately 300 plants each were hand-
1417 harvested and threshed. A representative sample, 25% of each pool, was milled (Retsch
1418 GM200, Haan, Germany), and DNA was extracted from 25 g of the flour by LGC Genomics
1419 GmbH (Berlin, Germany).

1420

1421 **FIND-IT screening**

1422 The FIND-IT 'Etincel' library was screened as described in Knudsen et al.⁸⁷ using a single assay
1423 for the isolation of *srh1*^{P635} variant [ID# CB-FINDit-Hv-014]. Forward primer 5'
1424 AATCCTGCAGTCCTTGG 3', reverse primer 5' GAGGAGAAGAAGGAGCC 3', mutant probe 5'-
1425 FAM/CGTGGACGT/ZEN/CGACG/3'IABkFQ/ Wild type probe
1426 /5'SUN/ACGTGGGCG/ZEN/TCGA/3'IABkFQ/ Integrated DNA Technologies, Inc.

1427

1428 **4K SNP chip genotyping**

1429 Genotyping, including DNA extraction from freeze-dried leaf material, was conducted by
1430 TraitGenetics (SGS - TraitGenetics GmbH, Germany). *srh1*^{P635} mutant, the corresponding wild
1431 type 'Etincel' and *srh1* pangenome accessions Morex, RGT Planet, HOR 13942, HOR 9043 and
1432 HOR 21599 were genotyped for background confirmation. Pairwise genetic distance of
1433 individuals was calculated as the average of their per-locus distances⁸⁸ using R package
1434 stringdist⁸⁹ (v 0.9.8). Principal Coordinate Analysis (PCoA) was done with R⁶⁴ (v 4.0.2) base
1435 function cmdscale based on this genetic distance matrix. The first two PCs were illustrated by
1436 ggplot2 (<https://ggplot2.tidyverse.org>).

1437

1438 **Sanger sequencing**

1439 gDNA of the *srh1*^{P635} variant and 'Etincel' was extracted from one-week old seedling leaves
1440 (DNeasy, Plant Mini Kit, Qiagen, Hilden, Germany). Genomic DNA fragments for sequencing
1441 were amplified by PCR using gene specific primers (forward primer
1442 5'TTGCACGATTCAAATGTGGT 3', reverse primer 5' TCACCGGGATCTCTCTGAAT 3') and Taq DNA

1443 Polymerase (NEB) for 35 cycles (initial denaturation at 94°C/3 min followed by 35 cycles of
1444 94°C/45 s, 55°C/60 s, and 72°C/60 s for extension, with a final extension step of 72°C/10 min).
1445 PCR products were purified using the NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel
1446 GmbH & Co. AG, Düren, Germany) according to the manufacturer's instructions. Sanger
1447 sequencing was done at Eurofins Genomics Germany GmbH using a gene-specific sequencing
1448 primer (5' AGAACGGAGAGGAGAGAAAGAAG 3').
1449

1450 **RNA preparation, sequencing, and data analysis**

1451 Rachilla tissues from two contrast groups Morex (short), Barke (long) and Bowman (long) and
1452 BW-NIL-*srh1* (short) were used for RNA sequencing. The rachilla tissues were collected from
1453 the central spikelets of the respective genotypes at rachilla hair initiation (RI; Waddington 8.0),
1454 and elongation (RE; Waddington 9.5) stages. Total RNA was extracted using TRIzol reagent
1455 (Invitrogen) followed by 2-propanol precipitation. Genomic DNA residues were removed with
1456 DNase I (NEB, M0303L). High-throughput paired-end sequencing was conducted at Novogene
1457 Co., Ltd (Cambridge, UK) with Illumina NovaSeq 6000 PE150 Platform. RNAseq reads were
1458 trimmed for adaptor sequences with Trimmomatic⁹⁰ (version 0.39) and the MorexV3 genome
1459 annotation was used as reference to estimate read abundance with Kallisto⁹¹. The raw read
1460 counts were normalized to Transcripts per kilo base per million (TPM) expression levels.
1461

1462 **mRNA *insitu* hybridization**

1463 *in situ* hybridization was conducted in longitudinal and cross sections derived from whole
1464 spikelet tissues of Bowman and Morex at rachilla hair elongation developmental stage (W9.5)
1465 with HvSRH1 sense & antisense probes (124 bp). The *in situ* hybridization was performed as
1466 described before⁹² with few modifications.
1467

1468 **Code availability**

1469 Scripts for pangenome graphs analyses are available at [https://github.com/mb47/minigraph-](https://github.com/mb47/minigraph-barley)
1470 [barley](https://github.com/mb47/minigraph-barley). The scripts for calculation of core/shell and cloud genes are deposited to the
1471 repository <https://github.com/PGSB-HMGU/>. The pipeline for identifying structurally
1472 complex loci is available at <https://github.com/mtrw/DGS>.
1473

1474 **Data availability**

1475 All the sequence data collected in this study have been deposited at the European Nucleotide
1476 Archive (ENA) under BioProjects PRJEB40587, PRJEB57567 and PRJEB58554 (raw data for
1477 pangenome assemblies), PRJEB64639 (pan-transcriptome Illumina data), PRJEB64637
1478 (transcriptome Isoseq data), PRJEB53924 (Illumina resequencing data), PRJEB45466-511 (raw
1479 data for gene space assemblies), PRJEB65284 (*srh1* transcriptome data). Accession codes for
1480 individual genotypes are listed in supplementary tables: **Supplementary Table 1** (pangenome
1481 assemblies and associated raw data), **Supplementary Table 2** (transcriptome data),
1482 **Supplementary Table 5** (Illumina resequencing), **Supplementary Table 6** (gene space
1483 assemblies).
1484

1485 **References for Methods**

1486
1487 1 Dvorak, J., McGuire, P. E. & Cassidy, B. Apparent sources of the A genomes of wheats
1488 inferred from polymorphism in abundance and restriction fragment length of

- 1489 repeated nucleotide sequences. *Genome* **30**, 680-689 (1988).
1490 <https://doi.org/10.1139/g88-115>
- 1491 2 Padmarasu, S., Himmelbach, A., Mascher, M. & Stein, N. In Situ Hi-C for Plants: An
1492 Improved Method to Detect Long-Range Chromatin Interactions. *Methods Mol Biol*
1493 **1933**, 441-472 (2019). https://doi.org/10.1007/978-1-4939-9045-0_28
- 1494 3 Himmelbach, A., Walde, I., Mascher, M. & Stein, N. Tethered Chromosome
1495 Conformation Capture Sequencing in Triticeae: A Valuable Tool for Genome
1496 Assembly. *Bio-protocol* **8**, e2955 (2018). <https://doi.org/10.21769/BioProtoc.2955>
- 1497 4 Himmelbach, A. *et al.* Discovery of multi-megabase polymorphic inversions by
1498 chromosome conformation capture sequencing in large-genome plant species. *The*
1499 *Plant Journal* (2018).
- 1500 5 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo
1501 assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170-175
1502 (2021). <https://doi.org/10.1038/s41592-020-01056-5>
- 1503 6 Marone, M. P., Singh, H. C., Pozniak, C. J. & Mascher, M. A technical guide to TRITEX,
1504 a computational pipeline for chromosome-scale sequence assembly of plant
1505 genomes. *Plant Methods* **18**, 128 (2022). [https://doi.org/10.1186/s13007-022-00964-](https://doi.org/10.1186/s13007-022-00964-1)
1506 [1](https://doi.org/10.1186/s13007-022-00964-1)
- 1507 7 Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality,
1508 completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**,
1509 245 (2020). <https://doi.org/10.1186/s13059-020-02134-9>
- 1510 8 Sun, H., Ding, J., Piednoël, M. & Schneeberger, K. findGSE: estimating genome size
1511 variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics* **34**,
1512 550-557 (2017). <https://doi.org/10.1093/bioinformatics/btx637>
- 1513 9 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M.
1514 BUSCO: assessing genome assembly and annotation completeness with single-copy
1515 orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
1516 <https://doi.org/10.1093/bioinformatics/btv351>
- 1517 10 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
1518 genomic features. *Bioinformatics* **26**, 841-842 (2010).
1519 <https://doi.org/10.1093/bioinformatics/btq033>
- 1520 11 Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching
1521 for the analysis of massive data sets. *Nature Biotechnology* **35**, 1026-1028 (2017).
1522 <https://doi.org/10.1038/nbt.3988>
- 1523 12 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **1**, 7
1524 (2018).
- 1525 13 Nattestad, M. & Schatz, M. C. Assemblytics: a web analytics tool for the detection of
1526 variants from an assembly. *Bioinformatics* **32**, 3021-3023 (2016).
1527 <https://doi.org/10.1093/bioinformatics/btw369>
- 1528 14 Goel, M., Sun, H., Jiao, W.-B. & Schneeberger, K. SyRI: finding genomic
1529 rearrangements and local sequence differences from whole-genome assemblies.
1530 *Genome Biology* **20**, 277 (2019). <https://doi.org/10.1186/s13059-019-1911-0>
- 1531 15 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
1532 reads. *2011* **17**, 3 (2011). <https://doi.org/10.14806/ej.17.1.200>
- 1533 16 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping
1534 and population genetical parameter estimation from sequencing data. *Bioinformatics*
1535 **27**, 2987-2993 (2011). <https://doi.org/10.1093/bioinformatics/btr509>

- 1536 17 Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association
1537 studies. *Nature Genetics* **44**, 821-824 (2012). [https://doi.org:10.1038/ng.2310](https://doi.org/10.1038/ng.2310)
- 1538 18 Xu, H. *et al.* FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads.
1539 *PLOS ONE* **7**, e52249 (2012). [https://doi.org:10.1371/journal.pone.0052249](https://doi.org/10.1371/journal.pone.0052249)
- 1540 19 Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging
1541 via overlap. *PLOS ONE* **12**, e0185056 (2017).
1542 [https://doi.org:10.1371/journal.pone.0185056](https://doi.org/10.1371/journal.pone.0185056)
- 1543 20 Li, H. BFC: correcting Illumina sequencing errors. *Bioinformatics* **31**, 2885-2887
1544 (2015). [https://doi.org:10.1093/bioinformatics/btv290](https://doi.org/10.1093/bioinformatics/btv290)
- 1545 21 Chikhi, R. & Rizk, G. Space-efficient and exact de Bruijn graph representation based
1546 on a Bloom filter. *Algorithms for Molecular Biology* **8**, 22 (2013).
1547 [https://doi.org:10.1186/1748-7188-8-22](https://doi.org/10.1186/1748-7188-8-22)
- 1548 22 Mascher, M. *et al.* Long-read sequence assembly: a technical evaluation in barley.
1549 *Plant Cell* **33**, 1888-1906 (2021). [https://doi.org:10.1093/plcell/koab077](https://doi.org/10.1093/plcell/koab077)
- 1550 23 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10** (2021).
1551 [https://doi.org:10.1093/gigascience/giab008](https://doi.org/10.1093/gigascience/giab008)
- 1552 24 Milner, S. G. *et al.* Genebank genomics highlights the diversity of a global barley
1553 collection. *Nat Genet* **51**, 319-326 (2019). [https://doi.org:10.1038/s41588-018-0266-](https://doi.org/10.1038/s41588-018-0266-x)
1554 [x](https://doi.org/10.1038/s41588-018-0266-x)
- 1555 25 Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLOS*
1556 *Genetics* **2**, e190 (2006). [https://doi.org:10.1371/journal.pgen.0020190](https://doi.org/10.1371/journal.pgen.0020190)
- 1557 26 Jayakodi, M. *et al.* The barley pan-genome reveals the hidden legacy of mutation
1558 breeding. *Nature* **588**, 284-289 (2020). [https://doi.org:10.1038/s41586-020-2947-8](https://doi.org/10.1038/s41586-020-2947-8)
- 1559 27 Sallam, A. H. *et al.* Genome-Wide Association Mapping of Stem Rust Resistance in
1560 *Hordeum vulgare* subsp. *spontaneum*. *G3 (Bethesda, Md.)* **7**, 3491-3507 (2017).
1561 [https://doi.org:10.1534/g3.117.300222](https://doi.org/10.1534/g3.117.300222)
- 1562 28 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
1563 richer datasets. *GigaScience* **4** (2015). [https://doi.org:10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8)
- 1564 29 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open
1565 software suite. *Trends in genetics* **16**, 276-277 (2000).
- 1566 30 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
1567 tree display and annotation. *Nucleic Acids Research* **49**, W293-W296 (2021).
1568 [https://doi.org:10.1093/nar/gkab301](https://doi.org/10.1093/nar/gkab301)
- 1569 31 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1570 (2013). [https://doi.org:10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635)
- 1571 32 Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with
1572 StringTie2. *Genome Biology* **20**, 278 (2019). [https://doi.org:10.1186/s13059-019-](https://doi.org/10.1186/s13059-019-1910-1)
1573 [1910-1](https://doi.org/10.1186/s13059-019-1910-1)
- 1574 33 Consortium, T. U. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic*
1575 *Acids Research* **51**, D523-D531 (2022). [https://doi.org:10.1093/nar/gkac1052](https://doi.org/10.1093/nar/gkac1052)
- 1576 34 Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for
1577 gene structure prediction in higher organisms. *Information and Software Technology*
1578 **47**, 965-978 (2005). [https://doi.org:https://doi.org/10.1016/j.infsof.2005.09.005](https://doi.org/10.1016/j.infsof.2005.09.005)
- 1579 35 Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for
1580 mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).
1581 [https://doi.org:10.1093/bioinformatics/bti310](https://doi.org/10.1093/bioinformatics/bti310)

- 1582 36 Ghosh, S. & Chan, C. K. Analysis of RNA-Seq Data Using TopHat and Cufflinks.
1583 *Methods Mol Biol* **1374**, 339-361 (2016). [https://doi.org/10.1007/978-1-4939-3167-](https://doi.org/10.1007/978-1-4939-3167-5_18)
1584 [5_18](https://doi.org/10.1007/978-1-4939-3167-5_18)
- 1585 37 Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic*
1586 *Acids Research* **34**, W435-W439 (2006). <https://doi.org/10.1093/nar/gkl200>
- 1587 38 Ter-Hovhannisyanyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction
1588 in novel fungal genomes using an ab initio algorithm with unsupervised training.
1589 *Genome Res* **18**, 1979-1990 (2008). <https://doi.org/10.1101/gr.081612.108>
- 1590 39 Hoff, K. J. & Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Current*
1591 *Protocols in Bioinformatics* **65**, e57 (2019). <https://doi.org/10.1002/cpbi.57>
- 1592 40 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using
1593 EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**,
1594 R7 (2008). <https://doi.org/10.1186/gb-2008-9-1-r7>
- 1595 41 Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal
1596 transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654-5666 (2003).
1597 <https://doi.org/10.1093/nar/gkg770>
- 1598 42 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
1599 search tool. *J Mol Biol* **215**, 403-410 (1990). [https://doi.org/10.1016/S0022-](https://doi.org/10.1016/S0022-2836(05)80360-2)
1600 [2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 1601 43 Spannagl, M. *et al.* PGSB PlantsDB: updates to the database framework for
1602 comparative plant genome research. *Nucleic Acids Res* **44**, D1141-1147 (2016).
1603 <https://doi.org/10.1093/nar/gkv1130>
- 1604 44 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
1605 generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
1606 <https://doi.org/10.1093/bioinformatics/bts565>
- 1607 45 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular
1608 biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).
1609 <https://doi.org/10.1093/bioinformatics/btp163>
- 1610 46 Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks.
1611 *Proceedings of the National Academy of Sciences* **89**, 10915-10919 (1992).
1612 <https://doi.org/10.1073/pnas.89.22.10915>
- 1613 47 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for
1614 comparative genomics. *Genome Biology* **20**, 238 (2019).
1615 <https://doi.org/10.1186/s13059-019-1832-y>
- 1616 48 Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number
1617 variation across multiple genomes. *eLife* **11**, e78526 (2022).
1618 <https://doi.org/10.7554/eLife.78526>
- 1619 49 Li, H., Feng, X. & Chu, C. The design and construction of reference pangenome graphs
1620 with minigraph. *Genome Biology* **21**, 265 (2020). [https://doi.org/10.1186/s13059-](https://doi.org/10.1186/s13059-020-02168-z)
1621 [020-02168-z](https://doi.org/10.1186/s13059-020-02168-z)
- 1622 50 Garrison, E. *et al.* Building pangenome graphs. *bioRxiv*, 2023.2004.2005.535718
1623 (2023). <https://doi.org/10.1101/2023.04.05.535718>
- 1624 51 Hickey, G. *et al.* Pangenome graph construction from genome alignments with
1625 Minigraph-Cactus. *Nature Biotechnology* (2023). [https://doi.org/10.1038/s41587-](https://doi.org/10.1038/s41587-023-01793-w)
1626 [023-01793-w](https://doi.org/10.1038/s41587-023-01793-w)
- 1627 52 Mascher, M. *et al.* Long-read sequence assembly: a technical evaluation in barley.
1628 *Plant Cell* (2021). <https://doi.org/10.1093/plcell/koab077>

- 1629 53 Garrison, E. *et al.* Variation graph toolkit improves read mapping by representing
1630 genetic variation in the reference. *Nature Biotechnology* **36**, 875-879 (2018).
1631 <https://doi.org/10.1038/nbt.4227>
- 1632 54 Guarracino, A., Heumos, S., Nahnsen, S., Prins, P. & Garrison, E. ODGI: understanding
1633 pangenome graphs. *Bioinformatics* **38**, 3319-3326 (2022).
1634 <https://doi.org/10.1093/bioinformatics/btac308>
- 1635 55 Park, S.-C., Lee, K., Kim, Y. O., Won, S. & Chun, J. Large-Scale Genomics Reveals the
1636 Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for
1637 Estimating Pan-Genome Size. *Frontiers in Microbiology* **10** (2019).
1638 <https://doi.org/10.3389/fmicb.2019.00834>
- 1639 56 Sirén, J. *et al.* Pangenomics enables genotyping of known structural variants in 5202
1640 diverse genomes. *Science* **374**, abg8871 (2021).
1641 <https://doi.org/doi:10.1126/science.abg8871>
- 1642 57 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
1643 MEM. *arXiv preprint arXiv:1303.3997* (2013).
- 1644 58 Rabanus-Wallace, M. T., Wicker, T. & Stein, N. Replicators, genes, and the C-value
1645 enigma: High-quality genome assembly of barley provides direct evidence that self-
1646 replicating DNA forms 'cooperative' associations with genes in arms races. *bioRxiv*,
1647 2023.2010.2001.560391 (2023). <https://doi.org/10.1101/2023.10.01.560391>
- 1648 59 Harris, R. S. *Improved pairwise alignment of genomic DNA*. (The Pennsylvania State
1649 University, 2007).
- 1650 60 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
1651 *Bioinformatics* **26**, 2460-2461 (2010). <https://doi.org/10.1093/bioinformatics/btq461>
- 1652 61 Ma, J. & Bennetzen, J. L. Rapid recent growth and divergence of rice nuclear
1653 genomes. *Proceedings of the National Academy of Sciences* **101**, 12404-12410
1654 (2004). <https://doi.org/doi:10.1073/pnas.0403715101>
- 1655 62 Buchmann, J. P., Matsumoto, T., Stein, N., Keller, B. & Wicker, T. Inter-species
1656 sequence comparison of Brachypodium reveals how transposon activity corrodes
1657 genome colinearity. *The Plant Journal* **71**, 550-563 (2012).
1658 [https://doi.org:https://doi.org/10.1111/j.1365-313X.2012.05007.x](https://doi.org/https://doi.org/10.1111/j.1365-313X.2012.05007.x)
- 1659 63 Kuraku, S., Zmasek, C. M., Nishimura, O. & Katoh, K. aLeaves facilitates on-demand
1660 exploration of metazoan gene family trees on MAFFT sequence alignment server
1661 with enhanced interactivity. *Nucleic Acids Research* **41**, W22-W28 (2013).
1662 <https://doi.org/10.1093/nar/gkt389>
- 1663 64 R Core Team. R: A language and environment for statistical computing. R Foundation
1664 for Statistical Computing, Vienna, Austria. 2022. (2022).
- 1665 65 Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids*
1666 *Research* **40**, e115-e115 (2012). <https://doi.org/10.1093/nar/gks596>
- 1667 66 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158
1668 (2011). <https://doi.org/10.1093/bioinformatics/btr330>
- 1669 67 Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of
1670 occurrences of k-mers. *Bioinformatics* **27**, 764-770 (2011).
1671 <https://doi.org/10.1093/bioinformatics/btr011>
- 1672 68 Kadziola, A., Sjøgaard, M., Svensson, B. & Haser, R. Molecular structure of a barley
1673 alpha-amylase-inhibitor complex: implications for starch binding and catalysis. *J Mol*
1674 *Biol* **278**, 205-217 (1998). <https://doi.org/10.1006/jmbi.1998.1683>

- 1675 69 Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut2: Assessing changes in
1676 stability and flexibility upon single and multiple point missense mutations. *Protein*
1677 *Science* **30**, 60-69 (2021). [https://doi.org:https://doi.org/10.1002/pro.3942](https://doi.org/10.1002/pro.3942)
- 1678 70 WADDINGTON, S. R., CARTWRIGHT, P. M. & WALL, P. C. A Quantitative Scale of Spike
1679 Initial and Pistil Development in Barley and Wheat. *Annals of Botany* **51**, 119-130
1680 (1983). [https://doi.org:10.1093/oxfordjournals.aob.a086434](https://doi.org/10.1093/oxfordjournals.aob.a086434)
- 1681 71 Poursarebani, N. *et al.* The Genetic Basis of Composite Spike Form in Barley and
1682 'Miracle-Wheat'. *Genetics* **201**, 155-165 (2015).
1683 [https://doi.org:10.1534/genetics.115.176628](https://doi.org/10.1534/genetics.115.176628)
- 1684 72 Beier, S. *et al.* Construction of a map-based reference genome sequence for barley,
1685 *Hordeum vulgare* L. *Scientific Data* **4**, 170044 (2017).
1686 [https://doi.org:10.1038/sdata.2017.44](https://doi.org/10.1038/sdata.2017.44)
- 1687 73 Taylor, J. & Butler, D. R Package ASMap: Efficient Genetic Linkage Map Construction
1688 and Diagnosis. *Journal of Statistical Software* **79**, 1 - 29 (2017).
1689 [https://doi.org:10.18637/jss.v079.i06](https://doi.org/10.18637/jss.v079.i06)
- 1690 74 Wu, Y., Bhat, P. R., Close, T. J. & Lonardi, S. Efficient and Accurate Construction of
1691 Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. *PLOS Genetics* **4**,
1692 e1000212 (2008). [https://doi.org:10.1371/journal.pgen.1000212](https://doi.org/10.1371/journal.pgen.1000212)
- 1693 75 Broman, K. W., Wu, H., Sen, Ś. & Churchill, G. A. R/qtl: QTL mapping in experimental
1694 crosses. *Bioinformatics* **19**, 889-890 (2003).
1695 [https://doi.org:10.1093/bioinformatics/btg112](https://doi.org/10.1093/bioinformatics/btg112)
- 1696 76 Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data
1697 via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*
1698 **39**, 1-22 (1977).
- 1699 77 Pidon, H. *et al.* High-resolution mapping of Rym14Hb, a wild relative resistance gene
1700 to barley yellow mosaic disease. *Theoretical and Applied Genetics* **134**, 823-833
1701 (2021). [https://doi.org:10.1007/s00122-020-03733-7](https://doi.org/10.1007/s00122-020-03733-7)
- 1702 78 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
1703 (2009). [https://doi.org:10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421)
- 1704 79 Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased
1705 assessments of sequence homology and phylogeny. *Nature Communications* **13**,
1706 6968 (2022). [https://doi.org:10.1038/s41467-022-34630-w](https://doi.org/10.1038/s41467-022-34630-w)
- 1707 80 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open
1708 Software Suite. *Trends Genet* **16**, 276-277 (2000). [https://doi.org:10.1016/s0168-](https://doi.org/10.1016/s0168-9525(00)02024-2)
1709 [9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2)
- 1710 81 Chen, Y. & Wang, X. Evaluation of efficiency prediction algorithms and development
1711 of ensemble model for CRISPR/Cas9 gRNA selection. *Bioinformatics* **38**, 5175-5181
1712 (2022). [https://doi.org:10.1093/bioinformatics/btac681](https://doi.org/10.1093/bioinformatics/btac681)
- 1713 82 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna
1714 RNA websuite. *Nucleic Acids Res* **36**, W70-74 (2008).
1715 [https://doi.org:10.1093/nar/gkn188](https://doi.org/10.1093/nar/gkn188)
- 1716 83 Koepfel, I., Hertig, C., Hoffie, R. & Kumlehn, J. Cas Endonuclease Technology—A
1717 Quantum Leap in the Advancement of Barley and Wheat Genetic Engineering.
1718 *International Journal of Molecular Sciences* **20**, 2647 (2019).
- 1719 84 Gerasimova, S. V. *et al.* Conversion of hulled into naked barley by Cas endonuclease-
1720 mediated knockout of the NUD gene. *BMC Plant Biology* **20**, 255 (2020).
1721 [https://doi.org:10.1186/s12870-020-02454-9](https://doi.org/10.1186/s12870-020-02454-9)

- 1722 85 Hensel, G., Kastner, C., Oleszczuk, S., Riechen, J. & Kumlehn, J. Agrobacterium-
1723 mediated gene transfer to cereal crop plants: current protocols for barley, wheat,
1724 triticale, and maize. *Int J Plant Genomics* **2009**, 835608 (2009).
1725 <https://doi.org/10.1155/2009/835608>
- 1726 86 Watson, A. *et al.* Speed breeding is a powerful tool to accelerate crop research and
1727 breeding. *Nature Plants* **4**, 23-29 (2018). <https://doi.org/10.1038/s41477-017-0083-8>
- 1728 87 Knudsen, S. *et al.* FIND-IT: Accelerated trait development for a green evolution.
1729 *Science Advances* **8**, eabq2266 (2022). <https://doi.org/doi:10.1126/sciadv.abq2266>
- 1730 88 Witherspoon, D. J. *et al.* Genetic similarities within and between human populations.
1731 *Genetics* **176**, 351-359 (2007). <https://doi.org/10.1534/genetics.106.067355>
- 1732 89 Van der Loo, M. P. The stringdist package for approximate string matching. *R J.* **6**, 111
1733 (2014).
- 1734 90 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
1735 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
1736 <https://doi.org/10.1093/bioinformatics/btu170>
- 1737 91 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
1738 quantification. *Nature biotechnology* **34**, 525-527 (2016).
- 1739 92 Poursarebani, N. *et al.* COMPOSITUM 1 contributes to the architectural simplification
1740 of barley inflorescence via meristem identity signals. *Nature Communications* **11**,
1741 5138 (2020). <https://doi.org/10.1038/s41467-020-18890-y>
- 1742 93 Arend, D. *et al.* PGP repository: a plant phenomics and genomics data publication
1743 infrastructure. *Database* **2016** (2016).
1744
1745
1746