

1 Association mapping and haplotype analysis of the pre-
2 harvest sprouting resistance locus *Phs-A1* reveals a causal
3 role of *TaMKK3-A* in global germplasm
4

5

6 Oluwaseyi Shorinola¹, Barbara Balcárková², Jessica Hyles³, Josquin F. G. Tibbits⁴,
7 Matthew J. Hayden⁴, Katarina Holuřova², Miroslav Valárik², Assaf Distelfeld⁵,
8 Atsushi Torada⁶, Jose M. Barrero³ and Cristobal Uauy^{1*}
9

10

11

12 ¹ John Innes Centre, Norwich Research Park, NR4 7UH, UK.

13 ² Institute of Experimental Botany, Centre of the Region Haná for Biotechnological
14 and Agricultural Research, Olomouc, 783 71, Czech Rep.

15 ³ Commonwealth Scientific and Industrial Research Organisation (CSIRO),
16 Canberra, ACT, 2601, Australia.

17 ⁴ Department of Economic Development, Jobs, Transport and Resources, Centre for
18 AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3083, Australia.

19 ⁵ The Institute for Cereal Crop Improvement, Tel Aviv University, Tel Aviv, Israel.

20 ⁶ HOKUREN Agricultural Research Institute, Naganuma, Hokkaido 069-1317,
21 Japan.
22

23

24 OS: oluwaseyi.shorinola@jic.ac.uk

25 BB: balcarkova.bara@seznam.cz

26 JH: jessica.hyles@csiro.au

27 JFGT: josquin.tibbits@ecodev.vic.gov.au

28 MJH: matthew.hayden@ecodev.vic.gov.au

29 KH: holusovak@ueb.cas.cz

30 MV: valarik@ueb.cas.cz

31 AD: adistel@post.tau.ac.il

32 AT: torada-atusi@hokuren.jp

33 JMB: Jose.Barrero@csiro.au

34 CU: cristobal.uauy@jic.ac.uk
35

36

36 * Corresponding author: Cristobal Uauy (cristobal.uauy@jic.ac.uk)

37 Phone: +44-(0)1603-450195
38

39

39 Number of words: 5,493

40 Number of figures: Five main figures and five supplementary figures.
41

41

42 **Abstract**

43 Pre-harvest sprouting (PHS) is an important cause of quality loss in many cereal crops
44 and is particularly prevalent and damaging in wheat. Resistance to PHS is therefore a
45 valuable target trait in many breeding programmes. The *Phs-A1* locus on wheat
46 chromosome arm 4AL has been consistently shown to account for a significant
47 proportion of natural variation to PHS in diverse mapping populations. However the
48 deployment of sprouting resistance is confounded by the fact that different candidate
49 genes, including the tandem duplicated *Plasma Membrane 19 (PM19)* genes and the
50 *mitogen-activated protein kinase kinase 3 (TaMKK3-A)* gene, have been proposed to
51 underlie *Phs-A1*. To further define the *Phs-A1* locus, we constructed a physical map
52 across this interval in hexaploid and tetraploid wheat. We established close proximity
53 of the proposed candidate genes which are located within a 1.2 Mb interval. An
54 association analysis of diverse germplasm used in previous genetic mapping studies
55 suggests that *TaMKK3-A*, and not *PM19*, is the major gene underlying the *Phs-A1*
56 effect in European, North American, Australian and Asian germplasm. We identified
57 the non-dormant *TaMKK3-A* allele at low frequencies within the A-genome diploid
58 progenitor *Triticum urartu* genepool, and show an increase in the allele frequency in
59 modern varieties. In UK varieties, the frequency of the dormant *TaMKK3-A* allele was
60 significantly higher in bread-making quality varieties compared to feed and biscuit-
61 making cultivars. Analysis of exome capture data from 58 diverse hexaploid wheat
62 accessions identified fourteen haplotypes across the extended *Phs-A1* locus and four
63 haplotypes for *TaMKK3-A*. Analysis of these haplotypes in a collection of UK and
64 Australian cultivars revealed distinct major dormant and non-dormant *Phs-A1*
65 haplotypes in each country, which were either rare or absent in the opposing
66 germplasm set. The diagnostic markers and haplotype information reported in the
67 study will help inform the choice of germplasm and breeding strategies for the
68 deployment of *Phs-A1* resistance into breeding germplasm.

69

70 **Keywords:** Dormancy, seed, *PM19*, *TaMKK3-A*, pre-harvest sprouting, *Triticum*
71 *aestivum*, haplotype

72 **Introduction**

73 Pre-harvest sprouting (PHS) refers to the too-early germination of physiologically
74 matured grains while still on the ear, but before harvest. PHS is primarily caused by
75 insufficient levels, or rapid loss, of seed dormancy and is an important cause of quality
76 loss in many cereal crops (Li et al., 2004; Fang and Chu, 2008). This is particularly
77 relevant in wheat due to its detrimental effects on bread-making potential which
78 represents the most common use of wheat grains globally (Simsek et al., 2014). PHS
79 is believed to be a modern phenomenon, as progenitor and wild wheat species
80 generally display high levels of seed dormancy (Gatford et al., 2002; Lan et al., 2005).
81 Selection for reduced seed dormancy during domestication and modern breeding
82 programmes allowed for more uniform seed germination and rapid crop establishment
83 (Nave et al., 2016). However, this also resulted in higher level of susceptibility to PHS
84 in modern wheat varieties (Barrero et al., 2010). In addition to its detrimental effect
85 on quality, PHS also reduces yield and affects seed viability, making resistance to PHS
86 a high priority in many breeding programmes.

87

88 Occurrence of PHS is heavily influenced by the environment. PHS is prevalent in
89 wheat growing regions with high levels of rainfall during the period of grain

90 maturation and after-ripening. Increased ambient temperature during this period can
91 further increase the susceptibility of grains to sprouting (Barnard and Smith, 2009;
92 Mares and Mrva, 2014). This environmental dependency of PHS constitutes a
93 constraint in selecting for PHS resistance in field conditions. In addition, resistance to
94 PHS is highly quantitative and is controlled by numerous quantitative trait loci (QTL)
95 located on all 21 chromosomes of bread wheat (Flintham et al., 2002; Kulwal et al.,
96 2005; Mori et al., 2005; Kottarachchi et al., 2006; Ogbonnaya et al., 2007; Liu et al.,
97 2008; Torada et al., 2008; Xiao-bo et al., 2008; Mohan et al., 2009; Munkvold et al.,
98 2009; Knox et al., 2012; Kulwal et al., 2012; Gao et al., 2013; Lohwasser et al., 2013;
99 Mares and Mrva, 2014; Kumar et al., 2015). This makes resistance to PHS one of the
100 most multi-genic traits in wheat and further highlights the complexity in breeding for
101 this trait.

102

103 Despite the multi-genic control of PHS resistance, a locus on chromosome arm 4AL,
104 designated as *Phs-A1*, has been consistently shown to account for a significant
105 proportion of natural variation to sprouting in diverse mapping populations. The *Phs-*
106 *A1* effect has been identified in at least eleven bi-parental and multi-parent mapping
107 populations derived from diverse germplasm from Australia, UK, Japan, China,
108 Mexico, Canada and Europe (Torada et al., 2005; Ogbonnaya et al., 2007; Chen et al.,
109 2008; Torada et al., 2008; Cabral et al., 2014; Albrecht et al., 2015; Barrero et al.,
110 2015). Physiological evaluation of *Phs-A1* shows that it delays the rate of dormancy
111 loss during seed after-ripening when plants are grown across a wide range of
112 temperatures (13 °C – 22 °C; Shorinola et al., 2016).

113

114 Recently, two independent studies by Barrero et al. (2015) and Torada et al. (2016)
115 identified the tandem duplicated *Plasma Membrane 19 (PM19-A1 and PM19-A2)*
116 genes and a *mitogen-activated protein kinase kinase 3 (TaMKK3-A)* gene,
117 respectively, as candidates for *Phs-A1*. The *PM19* genes were identified through a
118 combined genetic approach using multi-parent mapping populations and
119 transcriptomic analysis of near-isogenic recombinant inbred lines. The *TaMKK3-A*
120 gene was identified through a more traditional positional cloning strategy using bi-
121 parental mapping populations. Each study confirmed the effect of the gene(s) on
122 dormancy through either down-regulation of transcript levels through RNA
123 interference (*PM19*) or transgenic complementation of the susceptible parent with the
124 resistant allele (*TaMKK3-A*).

125

126 It is presently unclear whether the sprouting variation associated with *Phs-A1* across
127 diverse germplasm is due to allelic variation at *PM19* or *TaMKK3-A* alone, or if it's
128 due to a combination of both genes (Torada et al 2016). Fine-mapping studies
129 (Shorinola et al., 2016) defined *Phs-A1* to a genetic interval distal to *PM19* for UK
130 germplasm, consistent with the position of *TaMKK3-A*. However, a comprehensive
131 understanding of *Phs-A1* diversity taking into account both *PM19* and *TaMKK3-A*
132 genes across a wider set of germplasm is lacking.

133

134 In this study, we characterised the *Phs-A1* physical interval in both hexaploid and
135 tetraploid emmer wheat to establish the physical proximity of *PM19* and *TaMKK3-A*.
136 We developed markers for the candidate genes, and showed *TaMKK3-A* alleles to be
137 diagnostic for sprouting resistance in a panel of parental lines from mapping
138 populations in which *Phs-A1* was identified. We used diploid, tetraploid and hexaploid
139 accessions to further trace the origin of the sprouting susceptible *TaMKK3-A* allele

140 and used exome capture data from the wheat HapMap panel (Jordan et al., 2015) to
141 examine the haplotype variation across the *Phs-A1* locus.

142

143 **Materials and Methods**

144

145 **Physical Map Sequence Assembly and Annotation**

146 A fingerprinted Bacterial Artificial Chromosome (BAC) library of flow-sorted 4A
147 chromosome was used for constructing the Chinese Spring *Phs-A1* physical map
148 (accessible at https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_4AL_v2/).
149 Using the high-throughput BAC screening approach described by Cvikova et al.
150 (2015), a sequence database made from three-dimensional pool of BAC clones
151 comprising the Minimum Tilling Path (MTP) was searched for the sequences of
152 *PM19-A1* and *TaMKK3-A*. This identified two positive clones for *PM19-A1*
153 (TaaCsp4AL037H11 and TaaCsp4AL172K12) and three positive clones for
154 *TaMKK3-A* (TaaCsp4AL032F12, TaaCsp4AL012P14 and TaaCsp4AL002F16; Table
155 S1). Using Linear Topology Contig (LTC; Frenkel et al., 2010) BAC clustering
156 information for this library, we identified the BAC clusters (defined as a network of
157 overlapping BACs forming a contiguous sequence) to which these BACs belong. The
158 *PM19-A1*-containing BACs belong to BAC Cluster 16421 which has 20 BACs in its
159 MTP while the *TaMKK3-A*-containing BACs belong to BAC Cluster 285 comprised
160 of four MTP BACs (Table S1).

161

162 DNA of the BACs was extracted using the Qiagen Plasmid Midi Kit (Qiagen, Cat. No.
163 12143). Eleven of the 20 MTP BACs of Cluster 16421 and the four BACs of Cluster
164 285 MTP were sequenced on the Illumina MiSeq with 250 bp paired-end reads. An
165 average of 2,105,488 and 2,752,220 paired-end reads per BAC were produced for
166 Cluster 16421 and 285 BACs, respectively. Illumina reads for each BAC were
167 separately assembled using the CLC Bio genomic software (www.clcbio.com). Before
168 assembly, reads were filtered to remove contaminant sequences by mapping to the
169 BAC vector (pIndigoBAC-5) sequence and the *Escherichia coli* genome. *De novo*
170 assembly of reads after contaminant removal was done with the following assembly
171 parameters: Word size: 64 bp; Bubble size: 250 bp; Mismatch cost: 2; Insertion cost:
172 3; Deletion cost 3; Length fraction: 90%; Similarity fraction: 95%.

173

174 The assembled contigs were repeat-masked by BLASTn analysis against the Triticeae
175 Repeat Element Database (TREP: wheat.pw.usda.gov/ITMI/Repeats; Wicker et al.,
176 2000). Gene annotation was performed using the wheat gene models described by
177 Krasileva et al. (2013) and by BLASTX analysis to NCBI nr
178 (blast.ncbi.nlm.nih.gov/Blast). Gene models were also obtained by *ab-initio* gene
179 prediction with FGENESH (Solovyev et al., 2006). Only FGENESH gene models with
180 protein sequence support from NCBI or *Ensembl* Plant protein databases
181 (plants.ensembl.org) were used. Gene models with greater than 90% protein or
182 nucleotide sequence identity and more than 75% sequence coverage to already
183 annotated genes on NCBI or *Ensembl* databases were considered as high confidence
184 genes. Gene models that did not meet these criteria were considered as low confidence
185 genes, and were not analysed further.

186

187 ***TaMKK3-A* genotyping**

188 A Kompetitive Allelic Specific PCR (KASP; Smith and Maughan, 2015) assay was
189 developed for genotyping the C to A (C>A) causal *TaMKK3-A* mutation reported by

190 Torada et al. (2016). For this, two allele-specific reverse primers (*TaMKK3-A-snp1-*
191 *res*: TTTTGGCTTCGCCCTTAAGG and *TaMKK3-A-snpA1-sus*:
192 TTTTGGCTTCGCCCTTAAGT) each containing the allele-specific SNP at the 3'
193 end, were used in combination with a common A-genome specific forward primer
194 (GCATAGAGATCTAAAGCCAGCA). To distinguish the amplification signal
195 produced from each allele specific primer, FAM and HEX fluorescence dye probes
196 (Ramirez-Gonzalez et al., 2015) were added to the 5' end of *TaMKK3-A-snpA1-res*
197 and *TaMKK3-A-snpA1-sus*, respectively. KASP assays were performed as previously
198 described (Shorinola et al., 2016).

199
200 In addition to the KASP assay, a genome-specific Cleavage Amplified Polymorphism
201 Sequence (CAPS) assay (Konieczny and Ausubel, 1993), designated as *TaMKK3-A-*
202 *caps*, was developed. This CAPS marker is associated with the presence/absence of
203 an *Hpy166II* restriction site which co-localises with the C>A causal polymorphism in
204 the fourth exon of *TaMKK3-A*. Genome-specific primer pairs (Forward:
205 CACCAAAGAATAGAAATGCTCTCT and Reverse:
206 AGGAGTAGTTCTCATTGCGG) were designed to amplify an 887-bp sequence
207 including the fourth exon. PCR was performed with Phusion High Fidelity polymerase
208 (NEB, UK; Cat No: M0530S) in a 50 µL volume containing 20 % buffer, 0.2 mM of
209 dNTP, 5 µM each of *TaMKK3-A-cap* forward and reverse primers, 3 % of DMSO,
210 200 - 400 ng of genomic DNA and 0.5 unit of Phusion polymerase (NEB, UK; Cat
211 No: M0530S). Thermal cycling was done with Eppendorf Mastercycler® Pro Thermal
212 Cyclers with the following programme: initial denaturation at 98 °C for 2 mins; 35
213 cycles of denaturation at 98 °C for 30 s; Annealing at 62 °C for 30 s and extension at
214 72 °C for 60 s; final extension at 72 °C for 10 mins. Following PCR amplification, a
215 25 µL restriction digest reaction containing 21.5 µL of the final PCR reaction, 2.5 µL
216 of CutSmart® Buffer (NEB, UK; Cat No: B7204S) and 10 units of *Hpy166II* was
217 incubated at 37 °C for 1 hr. Digest products were separated on a 1.5 % agarose gel.

218 219 ***PM19-A1* genotyping**

220 Detection of an 18 bp deletion on the promoter region of *PM19-A1* was carried out
221 using primers TaPM19-A1-5F (GAAACAGCTACCGTGTAAGC) and TaPM19-
222 A1-5R (TGGTGAAGTGGAGTGTAGTGG) reported by Barrero et al. (2015). PCR
223 reaction mixture contained template DNA, 2.5 mM MgCl₂, 1.5 mM dNTP, 1.5 µM of
224 each primer, and 1 unit of *Taq* polymerase (NEB). The reaction mixture was made up
225 to a total volume of 10 µl. The PCR conditions were as follows: 3 min at 94°C,
226 followed by 30 cycles of 40 s at 94°C, 40 s at 60°C, and 1 min at 72°C. The last step
227 was incubation for 7 min at 72°C. The PCR products were resolved on a 4% agarose
228 gel and visualized with SYBR green I (Cambrex Bio Science, Rockland, ME.).

229 230 **Germplasm for Association Mapping**

231 We genotyped *PM19* and *TaMKK3-A* across 23 wheat varieties previously reported to
232 segregate for *Phs-A1*, including UK (Alchemy, Robigus, Option, Claire, Boxer,
233 Soleil), Australian (Yitpi, Baxter, Chara, Westonia, Cranbrook, Aus1408, Janz,
234 Cunningham, Halberd), Japanese (Kitamoe, Haruyokoi, OS21-5), Mexican/CIMMYT
235 (W7984, Opata M85), Canadian (Leader), Chinese (SW95-50213) and Swiss varieties
236 (Münstertaler). We also genotyped *TaMKK3-A* in accessions from progenitor species
237 *T. urartu* (41 accessions; A^u genome), *T. turgidum* ssp. *dicoccoides* (151 accessions;
238 AABB genomes), 804 hexaploid accessions from the Watkins landrace collection

239 (Wingen et al., 2014), and 457 modern European bread wheat varieties from the
240 Gediflux collection released between 1945 and 2000 (Reeves et al., 2004).

241

242 **Variant Calling and Haplotype Analysis**

243 We examined the haplotype structure around the *Phs-A1* locus in three different
244 germplasm sets. These included 457 varieties in the Gediflux collection, a panel of
245 200 Australian varieties, and the wheat Haplotype Map (HapMap) panel consisting of
246 62 diverse global accessions (Jordan et al., 2015). For the HapMap panel, we selected
247 polymorphic sites as follows. We extracted SNP information from published variant
248 call files (VCF) produced from whole exome capture (WEC) resequencing dataset of
249 the 62 HapMap lines (accessible at www.wheat-urgi.versailles.inra.fr/Seq-Repository/Variations). For this, the corresponding IWGSC contig information for
250 genes represented in the *Phs-A1* physical map were first obtained and used to filter the
251 HapMap VCF for SNP sites located within these contigs. We kept SNP sites with
252 allele frequencies of >5 % and accessions with >80% homozygous calls across SNPs.
253 Allele information at the selected SNP loci was reconstructed for each line using the
254 reference, alternate and genotype field information obtained from the VCF. Haplotype
255 cluster analysis was done with Network 5.0.0.0 (Fluxus Technology Limited, UK)
256 using the Median Joining Network Algorithm.
257

258

259 **Pedigree Visualisation**

260 Pedigree information was obtained from the Genetic Resources Information System
261 for Wheat and Triticale (GRIS, <http://wheatpedigree.net/>) and the International Crop
262 Information System (ICIS: www.icis.cgiar.org). Pedigree visualisation was performed
263 with Helium (Shaw et al., 2014). The coefficient of parentage (COP) analysis (i.e. the
264 probability that alleles of two individuals are identical by descent) was calculated for
265 all pairwise comparisons of lines within the most prevalent haplotypes (Australian:
266 H1/H2 and H5/H7; UK: H3 and H12). For accuracy, landraces or cultivars with
267 unknown or ambiguous pedigrees were not included in the COP analysis. Diversity
268 within haplotype groups was estimated by the mean calculation of all COPs within
269 each matrix.

270

271 **Results**

272

273 ***TaMKK3-A* and *PM19* are located within a 1.2 Mb physical interval.**

274 We constructed an extended physical map across the *Phs-A1* interval to investigate the
275 physical proximity between the *TaMKK3-A* and *PM19* candidate genes. Using *PM19-
276 A1* and *TaMKK3-A* sequences as queries, we screened *in silico* a BAC library of flow-
277 sorted 4AL chromosome arm of the bread wheat cultivar Chinese Spring (CS). *PM19-
278 A1* and *TaMKK3-A* were found on two independent non-overlapping BAC clone
279 clusters which were anchored on the high resolution radiation hybrid map of
280 chromosome 4A (Balcárková et al., 2016). The MTP of Cluster 16421 (*PM19*) was
281 comprised of eleven BAC clones whereas the MTP of Cluster 285 (*TaMKK3-A*)
282 included four BAC clones (Figure 1A; Table S1).

283

284 Individual BACs were sequenced, assembled, repeat-masked and annotated for coding
285 sequences. Cluster 16421 included nine high-confidence genes in addition to the
286 *PM19-A1* and *PM19-A2* genes. These included *YUCCA3-like*, *Myosin-J Heavy Chain
287 protein*, *Ubiquitin Conjugating Enzyme*, *Amino-Cyclopropane Carboxylate Oxidase 1
288 like (ACC Oxidase-1)*, two *Leucine-Rich Repeat Kinases (LRR kinase 1 and LRR*

289 *kinase 2), Agmatine Coumaroyl Transferase, Malonyl Coenzyme A:anthocyanin 3-O-*
290 *glucoside-6"-O-malonyltransferase* and a gene encoding for a hypothetical protein. In
291 addition to *TaMKK3-A*, Cluster 285 contained four additional genes including *Protein*
292 *Phosphatase1-Like (PPI-Like)*, *Activating Signal Co-integrator 1- Like (ASCI-Like)*,
293 *Ethylene Responsive Factor-1B-Like (ERF-1B-Like)* and a gene fragment showing
294 high sequence similarity to *ERF-1B-Like* and as such designated as *ERF-C*. Together,
295 this highlights the presence of at least 16 protein-coding genes across the *Phs-A1*
296 interval in hexaploid bread wheat (Figure 1).

297

298 We also characterised the *Phs-A1* interval in the recently constructed assembly of a
299 wild emmer wheat, Zavitan (Figure 1B; Hen-Avivi et al., 2016). This allowed
300 comparative analysis of the *Phs-A1* interval in tetraploid and hexaploid wheat species.
301 Fifteen of the 16 genes found in the CSS physical map were located on two Zavitan
302 scaffolds. Nine of these 15 genes were positioned across a 0.93 Mb interval on the
303 Zavitan 4A pseudomolecule. These included 4 genes from BAC Cluster 285 and five
304 genes from BAC Cluster 16421 (Figure 1). The remaining six genes spanned a 0.13
305 Mb interval on an unanchored scaffold. On average, the coding sequence identity
306 between CS and Zavitan was 99.7% across the genes shared by both assemblies. We
307 could not find sequence for *ERF-C* in the Zavitan assembly at similar identity. We
308 annotated two genes encoding for disease resistance protein *RPM1* in the Zavitan
309 sequence corresponding to the gap between the two CS BAC clusters. Combining the
310 CS and Zavitan physical maps, the physical region between *TaMKK3-A* and the *PM19*
311 genes was covered and estimated to be approximately 1.2 Mb (Figure 1).

312

313 ***TaMKK3-A* is most closely associated with *Phs-A1*.**

314 Torada et al. (2016) reported a C>A mutation in position 660 of the *TaMKK3-A* coding
315 sequence (C660A) as being causative of the *Phs-A1* effect. Using alignments of the
316 three wheat genomes, we developed a genome-specific and co-dominant KASP assay
317 for this SNP designated as *TaMKK3-A-snp1*. The *TaMKK3-A-snp1* assay is co-
318 dominant as it distinguished between heterozygotes and homozygotes F₂ progenies in
319 the Alchemy x Robigus population previously reported to segregate for *Phs-A1*
320 (Shorinola et al., 2016; Figure 2A). We also developed a CAPS marker (Konieczny
321 and Ausubel, 1993) for *TaMKK3-A* to enable genotyping of *Phs-A1* using a gel-based
322 assay. This marker, designated as *TaMKK3-A-cap*, amplifies a genome-specific 887
323 bp region and is designed to discriminate for the presence of an *Hpy166II* site
324 (GTNNAC) which is lost by the C660A mutation. Dormant lines with the C allele
325 maintain the *Hpy166II* site which leads to digestion of the 887 bp amplicon into
326 fragments of 605 and 282 bp (Figure 2B). Conversely, non-dormant lines with the A
327 allele lose the *Hpy166II* site and hence remain intact (887 bp) after digestion. As with
328 the KASP assay, the CAPS marker was co-dominant when used to genotype F₂
329 progenies (Figure 2B).

330

331 Using the KASP assay, we genotyped an association panel comprised of the parents
332 of 11 bi-parental mapping populations and a MAGIC population in which *Phs-A1* had
333 previously been reported (Table 1). The *TaMKK3-A-snp1* was polymorphic and
334 diagnostic for *Phs-A1* in all parental lines. Consistent with Torada et al. (2016), non-
335 dormant sprouting-susceptible parents carry the *TaMKK3-A* "A" allele while all the
336 dormant sprouting-resistant parents carry the *TaMKK3-A-snp1* "C" allele (Table 1).
337 We genotyped the same panel for the promoter deletion in *PM19-A1* previously
338 proposed to be causal of PHS susceptibility (Barrero et al., 2015). We found the *PM19-*

339 *AI* deletion to be linked with the non-dormant *TaMKK3-A* A allele in most, but not
340 all, of these populations. The putative linkage was broken in the dormant Kitamoe,
341 OS21-5 and SW95-50213 parents, whose dormancy phenotypes are not consistent
342 with their *PM19-A1* promoter deletion status, but can be explained by their *TaMKK3-*
343 *A* genotype (Table 1). This association was confirmed genetically in the SW95-50213
344 x AUS1408 cross. This population, which did not segregate for the dormancy
345 phenotype in the original work by Mares et al. (2005), is monomorphic for the dormant
346 C allele at *TaMKK3-A*, but segregates for the *PM19-A1* deletion. Similarly, parents of
347 the two populations OS21-5 x Haruyokoi and Kitamoe x Münstertaler segregating for
348 the dormancy phenotype in the work by Torada et al. (2005) are monophormic for the
349 *PM19-A1* deletions, but segregate accordingly for the *TaMKK3-A* causal mutation.
350 These results strongly support *TaMKK3-A* as the most likely causal gene for *Phs-A1*
351 across this highly-informative panel.

352

353 **Origin and distribution of the *TaMKK3-A* alleles in ancestral and modern** 354 **germplasm**

355 To examine the origin, distribution and allele frequencies of the causative *TaMKK3-*
356 *A* C660A SNP, we genotyped a set of 41 *T. urartu* (diploid: AA genome) and 151 *T.*
357 *turgidum* ssp. *dicoccoides* (tetraploid: AABB genome) accessions. These represent the
358 diploid and tetraploid progenitors of the modern bread wheat A genome on which *Phs-*
359 *A1* is located. Torada et al. (2016) previously suggested that the non-dormant A allele
360 was the mutant form since the dormant C SNP was conserved across different species.
361 Across *T. urartu* accessions, the C allele was predominant (39 accessions) while the
362 non-dormant A allele was present in only two accessions (5% allele frequency; Figure
363 2C). Similarly, across *T. dicoccoides* accessions, the dormant C allele frequency was
364 found in 134 accessions while the non-dormant allele was found in 17 accessions (11%
365 allele frequency; Figure 2C). Our results are consistent with Torada et al. (2016) in
366 that the non-dormant A allele is derived from the wild type C allele. In addition, the
367 presence of the A allele across both progenitor species suggests that the mutation
368 predates the hybridization and domestication events that gave rise to modern bread
369 wheat.

370

371 We also genotyped the Watkins Collection representing a set of global bread wheat
372 landraces collected in the 1920 and 1930s (Wingen et al., 2014), as well as the
373 Gediflux collection comprised of modern European bread wheat varieties released
374 between 1945 and 2000 (Reeves et al., 2004). The allele frequency of the non-dormant
375 A allele was 13% in the Watkins landrace collection (Figure 2C; Table S2),
376 comparable to that in *T. dicoccoides* (11%). However, the non-dormant A allele
377 frequency in the Gediflux collection was 48% across 457 varieties (Figure 2C). This
378 represents a marked increase of the non-dormant allele in the more modern European
379 collection when compared to European accessions within the Watkins landraces in
380 which the non-dormant A allele was found at a 15% frequency (Table S2).

381

382 To determine if the *TaMKK3* dormant allele was associated with improved end-use
383 quality, we genotyped 41 UK varieties representing the four UK market classes
384 (Figure 2D, nabim groups 1-4; nabim, 2014). Of the 13 bread-making quality varieties
385 (groups 1 and 2), eleven (85%) had the dormant *TaMKK3* allele. This frequency was
386 significantly higher (Contingency table $\chi^2 = 8.497$; $P < 0.01$) than in the 28 biscuit and
387 animal feed varieties (groups 3 and 4) in which the *TaMKK3* dormant allele was only
388 present in 10 varieties (36%).

389

390 ***Phs-A1* haplotypes in global germplasm**

391 We next examined the allelic diversity across the extended *Phs-A1* interval (including
392 *TaMKK3-A* and *PM19*) with the aim of elucidating the haplotype structure across this
393 region. For this, we used the SNP Haplotype Map (HapMap) dataset obtained from
394 whole exome capture resequencing of 62 diverse germplasm (Jordan et al., 2015).
395 From this SNP dataset, we obtained data for eight of the sixteen genes found in the
396 *Phs-A1* interval (*PPI-like*, *TaMKK3-A*, *ASC1-like*, *ERF-C*, *LRR Kinase 1*, *LRR Kinase*
397 *2*, *PM19-A2* and *PM19-A1*) corresponding to 51 SNPs. To improve the accuracy of
398 the haplotype analysis, we selected accessions with >80% homozygous calls across
399 the selected genes and SNPs with >5% allele frequency. This filtering resulted in 39
400 SNPs across the eight genes in 58 accessions.

401

402 Across the *Phs-A1* interval (*PPI-like* to *PM19-A1*) we identified 14 distinct
403 haplotypes (H1–14; Figure 3A). Haplotypes were comprised of a mix of cultivars,
404 landrace, breeding lines and synthetic population in varying proportion (Figure 3B;
405 Table S3). H1 represented the major haplotype present in 33% of all accessions
406 examined, whereas five haplotypes were relatively infrequent (<5%; H2, H5, H6, H9,
407 H13). Also, we observed haplotype linkage from the *TaMKK3-A* to *LRR Kinase 2* in
408 76% of the accessions, highlighting possible evidence for limited recombination in
409 this 780 kb interval in global germplasm. Similar haplotype linkage was observed at
410 the tandem *PM19* loci in all but one of the 58 lines.

411

412 Five of the selected SNPs were found in *TaMKK3-A* including the causal C660A SNP
413 in the fourth exon and four additional intron SNPs. These five SNPs defined four
414 distinct *TaMKK3-A* haplotypes (Figure 3C, *TaMKK3-A_HapA* - D) in the HapMap
415 collection with only one having the non-dormant A allele (*TaMKK3-A_HapA*). The
416 non-dormant A allele was present in 50% of the HapMap population, consistent with
417 the Gediflux collection (48%).

418

419 **Haplotype structure at the *Phs-A1* interval in UK and Australian germplasm**

420 To characterise a larger set of European (Gediflux) and Australian germplasm, we
421 selected seven informative polymorphisms across seven genes from the HapMap
422 dataset and developed KASP assays for these (Table S4). Using these seven assays,
423 we defined 16 haplotypes in the European Gediflux collection (Table S5). This
424 included eleven haplotypes previously identified in the global HapMap dataset and
425 five haplotypes unique to this European germplasm set, although these were relatively
426 infrequent (Figure 4). The UK subpopulation within the Gediflux collection comprised
427 of 176 varieties and contained 11 of the 15 haplotypes identified. Six haplotypes
428 include the dormant *TaMKK3-A* C allele (63% of UK varieties), with the majority of
429 these varieties sharing haplotype H12 (89 of 110 varieties), consistent with the wider
430 Gediflux population (Figure 4). This suggests one main source of PHS resistance in
431 UK and European germplasm.

432

433 By combining haplotype and pedigree information for these lines we could trace, to a
434 reasonable degree of accuracy, the founder lines for the most common resistant
435 haplotypes in UK germplasm (Figure S1). We identified the origin of the major
436 resistant haplotype in the UK germplasm (H12) as Vilmorin-27, a French winter wheat
437 variety released in the late 1920s (Figure 5, Figure S2). Vilmorin-27 was a direct
438 parent and the donor of haplotype H12 for Cappelle-Desprez, a major founder variety

439 for wheat breeding programmes in Northern France and the UK released in 1948.
440 Haplotype H12 has since remained an important part of UK breeding programmes
441 through varieties such as Rendezvous and Riband (released between 1985-1987).

442

443 Within the 200 Australian varieties we identified twelve haplotypes including ten
444 previously identified HapMap haplotypes, and two Australian-specific haplotypes at
445 low frequency (<1%, Table S6, Figure S3). Eight haplotypes present in 89 varieties
446 (44.5%) have the dormant *TaMKK3-A* C allele while the other four haplotypes present
447 in 111 varieties (55.5%) have the non-dormant *TaMKK3-A* A allele (Figure S4). This
448 represents a near balanced distribution of both alleles in Australian germplasm. In this
449 set, 71% of lines with the dormant *TaMKK3-A* C alleles were traced to Federation (or
450 Purple Straw) ancestry. Across the entire set, the alternative, non-dormant allele was
451 more associated with the presence of cv. Gabo or CIMMYT-derived material in the
452 pedigree. These lines had a more recent average release date of 1976 compared to the
453 lines with the dormant allele (average release date 1941).

454

455 The mean coefficient of parentage (COP) for the Australian and UK Gediflux set of
456 lines was 0.10 and 0.11 respectively (Table 2). Within each germplasm set, the lines
457 with the most prevalent haplotypes had higher COP values, indicating a higher degree
458 of relatedness amongst these lines relative to the entire collection (Table 2).

459

460 Discussion

461

462 *Physical Map*

463 We characterised the *Phs-A1* interval by constructing a 1.5 Mb physical map spanning
464 the *PM19* and *TaMKK3-A* candidate genes (Barrero et al., 2015; Torada et al., 2016)
465 and including 16 protein-coding genes. We observed near perfect sequence and gene
466 content conservation in the interval between hexaploid and tetraploid physical maps.
467 A similar overall collinearity between bread wheat, barley and *Brachypodium* was also
468 observed except for the interval between *ACC Oxidase-1* and *ERF-C* where the gene
469 content in each species diverged (Figure S5). The *PM19* candidates were conserved
470 across these species, whereas *TaMKK3-A* was only present in barley and wheat.

471

472 Sequence information from the BAC-based CS assembly and the whole genome
473 shotgun Zavitan assemblies was used in a complementary manner. Neither assembly
474 was fully contiguous across the *Phs-A1* interval, but the gaps were different in the two
475 assemblies allowing the spanning of the complete interval. This lack of contiguity was
476 also present in the IWGSC CS v0.4 (available at
477 https://urgi.versailles.inra.fr/blast_iwgsc/blast.php), TGAC (Clavijo et al., 2017) and
478 Refeqv1.0 assemblies, where intervals covering *TaMKK3-A* and *ASCI-Like* were
479 unanchored. While the new whole genome assemblies offer major improvements in
480 contiguity, the available BAC physical maps will be of value to assign unanchored
481 scaffolds or solve inconsistencies in regions where contiguity is broken.

482

483 *TaMKK3-A* determines *Phs-A1* effect across diverse germplasm

484 The 1.5 Mb physical interval which defines *Phs-A1* includes the proposed candidates
485 *PM19* and *TaMKK3-A*, as well as other genes with potential roles in
486 dormancy/germination regulation. For example, *ACC Oxidase-1* catalyses the last
487 steps in the biosynthesis of ethylene – a germination promoting hormone (Matilla and
488 Matilla-Vázquez, 2008; Linkies and Leubner-Metzger, 2012; Corbineau et al., 2014).

489 However, using two bi-parental mapping populations we showed linkage of *Phs-A1*
490 to the interval between *PPI-Like – LRR Kinase 2* in UK populations, thereby
491 excluding the *PM19* and *ACC Oxidase-1* loci as candidate genes (Shorinola et al.,
492 2016). This was consistent with Torada et al. (2016) who identified *TaMKK3-A* as the
493 causal gene in their mapping population and work in barley which identified the barley
494 homologue (*MKK3*) as the causal gene for the seed dormancy QTL SD2 (Nakamura
495 et al., 2016).

496

497 In support of this, the causal *TaMKK3-A* C660A SNP is perfectly associated with the
498 phenotypes of 19 diverse parents of eleven mapping population in which *Phs-A1* had
499 previously been identified. This was also the case for the parents of the MAGIC
500 population (Yipti, Chara, Westonia, Baxter) previously used to propose the *PM19* loci
501 as the causal gene (Barrero et al., 2015). Barrero et al (2015) proposed a promoter
502 deletion in *PM19-A1* affecting motifs important for ABA responsiveness as the cause
503 of non-dormancy in sprouting susceptible genotypes. The *PM19-A1* deletion and the
504 non-dormant *TaMKK3-A* A allele are in complete linkage in all the non-dormant
505 parents from the multiple mapping populations. However, the *PM19-A1* promoter
506 deletion did not account for the dormant phenotype of Kitamoe, OS21-5 and SW95-
507 50213 (Table 1). These dormant varieties have the *PM19-A1* promoter deletion
508 associated with low dormancy, but carries the dormant *TaMKK3-A* allele. These
509 natural recombinants suggest that *TaMKK3-A* is the causal *Phs-A1* gene. SW95-50213
510 is a Chinese landrace which is an important source of *Phs-A1*-mediated dormancy in
511 Australian breeding programs. When SW95-50213 was crossed to a line carrying both
512 *TaMKK3-A* and *PM19* dormant alleles (AUS1408), no grain dormancy QTL could be
513 identified (Mares et al., 2005). Despite the segregation of the *PM19-A1* promoter
514 polymorphism in this population, all lines displayed dormant to intermediate
515 dormancy phenotype consistent with the *TaMKK3-A* genotype of their parents. Taken
516 together, this evidence confirms the tight linkage between *TaMKK3-A*, *PM19*, and the
517 *Phs-A1* phenotype, and suggest that *TaMKK3-A*, but not *PM19*, is the causal gene
518 underlying sprouting variation associated with *Phs-A1* in diverse European, North
519 American, Australian and Asian germplasm.

520

521 **Breeding implications**

522 Given the identification of a number of *T. urartu* accessions with the non-dormant A
523 allele, it is likely that the C660A mutation originates from this diploid ancestor and
524 predates the domestication and hybridisation events that gave rise to modern bread
525 wheat. The non-dormant allele frequency was below 15% in accessions and landraces
526 collected previous to 1920, but rose sharply to close to 50% in more modern
527 germplasm. It is tempting to speculate that this could be due to selective pressure by
528 breeders over the past 70 years for the non-dormant A allele in European and
529 Australian environments. This pressure could be driven by selection for genotypes
530 with more rapid and uniform germination that would be associated with the non-
531 dormant allele. However, allele frequencies for both alleles have remained overall
532 balanced given the improved end-use quality associated with the dormant allele. This
533 hypothesis is supported by the fact that 85% of UK bread-making varieties carry the
534 dormant allele, compared to only 35% of feed and biscuit-making varieties.

535

536 To facilitate breeding for PHS resistance, we developed co-dominant KASP and
537 CAPS markers for the causal *TaMKK3-A* mutation, as well as KASP markers for the
538 wider region. We identified fourteen *Phs-A1* haplotypes in a global germplasm panel

539 with four haplotypes for the *TaMKK3* gene itself, of which only one included the
540 C660A non-dormant SNP. Comparison of Australian and UK haplotypes highlighted
541 distinct frequencies in both sets with the most prevalent haplotypes containing the
542 dormant *TaMKK3-A* allele differing in both countries. Haplotype H5/H7 is most
543 frequent in Australian varieties, whereas haplotype H12 dominates in the UK.
544 Interestingly, these haplotypes are either rare (<5% H5/H7 in UK) or absent (H12 not
545 present in Australia) in the other country, suggesting distinct sources of pre-harvest
546 sprouting resistance in Australian and UK breeding programmes. Understanding
547 haplotypes structure across genes of agronomic interest is increasingly possible with
548 the latest advances in wheat genomics (Clavijo et al., 2017; Uauy, 2017). It is also
549 increasingly relevant given potential negative linkage drag associated with major
550 phenology traits (Voss-Fels et al., 2017). The markers and knowledge generated in
551 this study should facilitate the choice of parental genotypes for the deployment of
552 *TaMKK3* in commercial cultivars.

553

554 The earliest line in the Australian set (Golden Drop, released 1840) carries the
555 favourable *TaMKK3-A* 'C' SNP and also the most prevalent haplotype (H5/H7) at
556 this locus. Golden Drop was derived from a Purple Straw/Yandilla cross and its sister
557 line, Federation (released in 1901) become the foundation of many successful
558 Australian cultivars due to earlier maturity and thus ability to avoid drought stress late
559 in the growing season. Not only was Federation wheat better adapted to the Australian
560 climate, it also had improved grain quality for milling, and so become widely adopted
561 by breeders (Eagles et al., 2009).

562

563 The next major introduction of germplasm into Australia occurred in the 1970's, as
564 CIMMYT material was deployed widely by breeders seeking traits affecting height,
565 quality and disease resistance (Brennan and Fox, 1998). Important CIMMYT parents
566 in Australian breeding include Sonora-64, Pitic, Pavon-76, WW15 and WW80.
567 Pedigree analysis suggests that such material could be the source of the most prevalent
568 haplotype in Australia (H1/H2) containing the non-favourable *TaMKK3-A* allele. A
569 high proportion of modern Australian cultivars with the non-dormant haplotype
570 suggests opportunities may exist for the incorporation of favourable alleles at the
571 locus.

572

573 **Future outlook**

574 The dormant *TaMKK3-A* C allele is predominant in all the progenitor and historic
575 germplasm evaluated in this study, suggesting that it represents the ancestral allele as
576 proposed by Torada et al. (2016). The N220K amino acid substitution (C660A
577 mutation) in the kinase domain results in a gain-of-function allele which reduces
578 dormancy in wheat. This is in contrast with barley where the non-dormant *MKK3*
579 allele is ancestral and the N260T substitution in the kinase domain results in a loss-of-
580 function allele leading to increased seed dormancy (Nakamura et al., 2016). This
581 provides an additional example of how for the same biological process, gain-of-
582 function (dominant) mutations have been more readily selected in polyploid wheat
583 compared to recessive variation in diploid barley (Borrill et al., 2015). The fact that
584 the same gene has been selected in both species also suggests that the kinase activity
585 of *TaMKK3-A* can be modulated to fine-tune the level of seed dormancy in temperate
586 cereals. A better understanding of the activity and regulation of *TaMKK3-A* and its
587 homoeologs could allow the identification of mutants (Krasileva et al., 2017) or the

588 creating of gene edited alleles (Zong et al., 2017) with different levels of activity or
589 the design of novel alleles with different degrees of dormancy.

590

591 **Conflict of interest statement**

592 The authors declare no conflict of interest.

593

594 **Author contributions**

595 OS led the genotype and pedigree analysis of the UK varieties, annotation of BAC
596 sequences, developed the KASP and CAPS marker, and analysed the HapMap data;
597 JH performed pedigree analysis of Australian; JFGT and MJH performed genotyping
598 of Australian varieties; MV, BB and KH constructed the 4AL physical map of CS;
599 AD constructed the physical map of tetraploid wheat Zavitan; AT performed
600 genotyping of Japanese varieties; JMB led the work on Australian varieties; OS,
601 CU, JH, JMB contributed to the writing of the manuscript; OS and CU designed the
602 experiments.

603

604 **Funding**

605 This work was supported by the UK Biotechnology and Biological Sciences
606 Research Council (BBSRC), AHDB-HGCA, KWS, Lantmännen, Limagrain and
607 RAGT (BB/I01800X/1, BB/J004588/1, BB/J004596/1, BB/P013511/1,
608 BB/P016855/1). BB and MV were supported by grant LO1204 from the National
609 Program of Sustainability I, by the Czech Science Foundation (14-07164S). OS was
610 supported by the John Innes Foundation.

611

612 **Acknowledgements**

613 We thank Chloe Riviere for help with plant husbandry and Dr Sergio Galvez-Rojas
614 for advice on data visualisation. We also thank Katerina Viduka for the assistance in
615 the genotyping, Dr Ben Trevaskis and Dr Howard Eagles for assistance in the pedigree
616 analysis of the Australian collection, and Dr. Frank Gubler for the helpful discussions
617 and support. We thank the International Wheat Genome Sequencing Consortium
618 (IWGSC) for pre-publication access to the wheat genome RefSeq v0.4 and v1.0.

619 **References**

- 620 Albrecht, T., Oberforster, M., Kempf, H., Ramgraber, L., Schacht, J., Kazman, E., Zechner,
621 E., Neumayer, A., Hartl, L., and Mohler, V. (2015). Genome-wide association
622 mapping of preharvest sprouting resistance in a diversity panel of European winter
623 wheats. *Journal of applied genetics* 56, 277-285.
- 624 Balcárková, B., Frenkel, Z., Škopová, M., Abrouk, M., Kumar, A., Chao, S., Kianian, S.F.,
625 Akhunov, E., Korol, A.B., Doležel, J., and Valárik, M. (2016). A High Resolution
626 Radiation Hybrid Map of Wheat Chromosome 4A. *Frontiers in plant science* 7, 2063.
- 627 Barnard, A., and Smith, M.F. (2009). The effect of rainfall and temperature on the preharvest
628 sprouting tolerance of winter wheat in the dryland production areas of the Free State
629 Province. *Field Crops Research* 112, 158-164.
- 630 Barrero, J.M., Cavanagh, C., Verbyla, K.L., Tibbits, J.F., Verbyla, A.P., Huang, B.E.,
631 Rosewarne, G.M., Stephen, S., Wang, P., Whan, A., Rigault, P., Hayden, M.J., and
632 Gubler, F. (2015). Transcriptomic analysis of wheat near-isogenic lines identifies
633 PM19-A1 and A2 as candidates for a major dormancy QTL. *Genome biology* 16, 93.
- 634 Barrero, J.M., Jacobsen, J., and Gubler, F. (2010). "Seed Dormancy: Approaches for Finding
635 New Genes in Cereals," in *Plant Developmental Biology - Biotechnological*

- 636 *Perspectives: Volume 1*, eds. E.C. Pua & M.R. Davey. (Berlin, Heidelberg: Springer
637 Berlin Heidelberg), 361-381.
- 638 Borrill, P., Adamski, N., and Uauy, C. (2015). Genomics as the key to unlocking the polyploid
639 potential of wheat. *The New phytologist* 208, 1008-1022.
- 640 Brennan, J.P., and Fox, P.N. (1998). Impact of CIMMYT varieties on the genetic diversity of
641 wheat in Australia, 1973-1993. *Australian Journal of Agricultural Research* 49, 175-
642 178.
- 643 Cabral, A.L., Jordan, M.C., McCartney, C.A., You, F.M., Humphreys, D.G., Maclachlan, R.,
644 and Pozniak, C.J. (2014). Identification of candidate genes, regions and markers for
645 pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.). *BMC plant biology*
646 14, 340.
- 647 Chen, C.X., Cai, S.B., and Bai, G.H. (2008). A major QTL controlling seed dormancy and
648 pre-harvest sprouting resistance on chromosome 4A in a Chinese wheat landrace.
649 *Molecular Breeding* 21, 351-358.
- 650 Clavijo, B.J., Venturini, L., Schudoma, C., Accinelli, G.G., Kaithakottil, G., Wright, J.,
651 Borrill, P., Kettleborough, G., Heavens, D., Chapman, H., Lipscombe, J., Barker, T.,
652 Lu, F.H., Mckenzie, N., Raats, D., Ramirez-Gonzalez, R.H., Coince, A., Peel, N.,
653 Percival-Alwyn, L., Duncan, O., Trosch, J., Yu, G., Bolser, D.M., Namaati, G.,
654 Kerhornou, A., Spannagl, M., Gundlach, H., Haberer, G., Davey, R.P., Fosker, C.,
655 Palma, F.D., Phillips, A., Millar, A.H., Kersey, P.J., Uauy, C., Krasileva, K.V.,
656 Swarbreck, D., Bevan, M.W., and Clark, M.D. (2017). An improved assembly and
657 annotation of the allohexaploid wheat genome identifies complete families of
658 agronomic genes and provides genomic evidence for chromosomal translocations.
659 *Genome research* doi:
660 10.1101/gr.217117.116.
- 661 Corbineau, F., Xia, Q., Bailly, C., and El-Maarouf-Bouteau, H. (2014). Ethylene, a key factor
662 in the regulation of seed dormancy. *Frontiers in plant science* 5, 539.
- 663 Cvikova, K., Cattonaro, F., Alaux, M., Stein, N., Mayer, K.F., Dolezel, J., and Bartos, J.
664 (2015). High-throughput physical map anchoring via BAC-pool sequencing. *BMC*
665 *plant biology* 15, 99.
- 666 Eagles, H.A., Cane, K., and Vallance, N. (2009). The flow of alleles of important photoperiod
667 and vernalisation genes through Australian wheat. *Crop and Pasture Science* 60, 646-
668 657.
- 669 Fang, J., and Chu, C. (2008). Abscisic acid and the pre-harvest sprouting in cereals. *Plant*
670 *Signaling & Behavior* 3, 1046-1048.
- 671 Flintham, J. (2000). Different genetic components control coat-imposed and embryo-
672 imposed dormancy in wheat. *Seed Science Research* 10, 43-50.
- 673 Flintham, J., Adlam, R., Bassoi, M., Holdsworth, M., and Gale, M.D. (2002). Mapping genes
674 for resistance to sprouting damage in wheat. *Euphytica* 126, 39-45.
- 675 Frenkel, Z., Paux, E., Mester, D., Feuillet, C., and Korol, A. (2010). LTC: a novel algorithm
676 to improve the efficiency of contig assembly for physical mapping in complex
677 genomes. *BMC Bioinformatics* 11, 584-584.
- 678 Gao, X., Hu, C.H., H.Z., L., Yao, Y.J., Meng, M., Dong, L., Zhao, W.C., Chen, Q.J., and Li,
679 X.Y. (2013). Factors affecting pre-harvest sprouting resistance in wheat (*triticum*
680 *aestivum* l.): A review. *The Journal of Animal and Plant Sciences* 23, 556-565.
- 681 Gatford, K.T., Eastwood, R.F., and Halloran, G.M. (2002). Germination inhibitors in bracts
682 surrounding the grain of *Triticum tauschii*. *Functional Plant Biology* 29, 881-890.
- 683 Hen-Avivi, S., Savin, O., Racovita, R.C., Lee, W.S., Adamski, N.M., Malitsky, S., Almekias-
684 Siegl, E., Levy, M., Vautrin, S., Berges, H., Friedlander, G., Kartvelishvily, E., Ben-
685 Zvi, G., Alkan, N., Uauy, C., Kanyuka, K., Jetter, R., Distelfeld, A., and Aharoni, A.
686 (2016). A Metabolic Gene Cluster in the Wheat W1 and the Barley Cer-cqu Loci
687 Determines beta-Diketone Biosynthesis and Glaucousness. *The Plant cell* 28, 1440-
688 1460.
- 689 Jordan, K.W., Wang, S., Lun, Y., Gardiner, L.J., Maclachlan, R., Hucl, P., Wiebe, K., Wong,
690 D., Forrest, K.L., Sharpe, A.G., Sidebottom, C.H., Hall, N., Toomajian, C., Close, T.,

- 691 Dubcovsky, J., Akhunova, A., Talbert, L., Bansal, U.K., Bariana, H.S., Hayden, M.J.,
692 Pozniak, C., Jeddeloh, J.A., Hall, A., and Akhunov, E. (2015). A haplotype map of
693 allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes.
694 *Genome biology* 16, 48.
- 695 Knox, R.E., Clarke, F.R., Clarke, J.M., Fox, S.L., Depauw, R.M., and Singh, A.K. (2012).
696 Enhancing the identification of genetic loci and transgressive segregants for
697 preharvest sprouting resistance in a durum wheat population. *Euphytica* 186, 193-206.
- 698 Konieczny, A., and Ausubel, F.M. (1993). A procedure for mapping Arabidopsis mutations
699 using co-dominant ecotype-specific PCR-based markers. *The Plant journal : for cell
700 and molecular biology* 4, 403-410.
- 701 Kottarachchi, N.S., Uchino, N., Kato, K., and Miura, H. (2006). Increased grain dormancy
702 in white-grained wheat by introgression of preharvest sprouting tolerance QTLs.
703 *Euphytica* 152, 421-428.
- 704 Krasileva, K., Buffalo, V., Bailey, P., Pearce, S., Ayling, S., Tabbita, F., Soria, M., Wang, S.,
705 Consortium, I., Akhunov, E., Uauy, C., and Dubcovsky, J. (2013). Separating
706 homeologs by phasing in the tetraploid wheat transcriptome. *Genome biology* 14,
707 R66.
- 708 Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L.,
709 Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X., Borrill, P., Fosker, C., Ayling, S.,
710 Phillips, A.L., Uauy, C., and Dubcovsky, J. (2017). Uncovering hidden variation in
711 polyploid wheat. *Proceedings of the National Academy of Sciences* 114, E913-E921.
- 712 Kulwal, P., Ishikawa, G., Benschler, D., Feng, Z., Yu, L.-X., Jadhav, A., Mehetre, S., and
713 Sorrells, M. (2012). Association mapping for pre-harvest sprouting resistance in white
714 winter wheat. *Theoretical and Applied Genetics* 125, 793-805.
- 715 Kulwal, P.L., Kumar, N., Gaur, A., Khurana, P., Khurana, J.P., Tyagi, A.K., Balyan, H.S., and
716 Gupta, P.K. (2005). Mapping of a major QTL for pre-harvest sprouting tolerance on
717 chromosome 3A in bread wheat. *Theoretical and Applied Genetics* 111, 1052-1059.
- 718 Kumar, S., Knox, R., Clarke, F., Pozniak, C., Depauw, R., Cuthbert, R., and Fox, S. (2015).
719 Maximizing the identification of QTL for pre-harvest sprouting resistance using seed
720 dormancy measures in a white-grained hexaploid wheat population. *Euphytica* 205,
721 287-309.
- 722 Lan, X.J., Wei, Y.M., Liu, D.C., Yan, Z.H., and Zheng, Y.L. (2005). Inheritance of seed
723 dormancy in Tibetan semi-wild wheat accession Q1028. *Journal of applied genetics*
724 46, 133-138.
- 725 Li, C., Ni, P., Francki, M., Hunter, A., Zhang, Y., Schibeci, D., Li, H., Tarr, A., Wang, J.,
726 Cakir, M., Yu, J., Bellgard, M., Lance, R., and Appels, R. (2004). Genes controlling
727 seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison.
728 *Functional & integrative genomics* 4, 84-93.
- 729 Linkies, A., and Leubner-Metzger, G. (2012). Beyond gibberellins and abscisic acid: how
730 ethylene and jasmonates control seed germination. *Plant Cell Reports* 31, 253-270.
- 731 Liu, S., Cai, S., Graybosch, R., Chen, C., and Bai, G. (2008). Quantitative trait loci for
732 resistance to pre-harvest sprouting in US hard white winter wheat Rio Blanco.
733 *Theoretical and Applied Genetics* 117, 691-699.
- 734 Lohwasser, U., Rehman Arif, M.A., and Börner, A. (2013). Discovery of loci determining pre-
735 harvest sprouting and dormancy in wheat and barley applying segregation and
736 association mapping. *Biologia Plantarum* 57, 663-674.
- 737 Mares, D., and Mrva, K. (2014). Wheat grain preharvest sprouting and late maturity alpha-
738 amylase. *Planta* 240, 1167-1178.
- 739 Mares, D., Mrva, K., Cheong, J., Williams, K., Watson, B., Storlie, E., Sutherland, M., and
740 Zou, Y. (2005). A QTL located on chromosome 4A associated with dormancy in
741 white- and red-grained wheats of diverse origin. *Theoretical and Applied Genetics*
742 111, 1357-1364.
- 743 Matilla, A.J., and Matilla-Vázquez, M.A. (2008). Involvement of ethylene in seed physiology.
744 *Plant Science* 175, 87-97.

- 745 Mohan, A., Kulwal, P., Singh, R., Kumar, V., Mir, R., Kumar, J., Prasad, M., Balyan, H.S.,
746 and Gupta, P.K. (2009). Genome-wide QTL analysis for pre-harvest sprouting
747 tolerance in bread wheat. *Euphytica* 168, 319-329.
- 748 Mori, M., Uchino, N., Chono, M., Kato, K., and Miura, H. (2005). Mapping QTLs for grain
749 dormancy on wheat chromosome 3A and the group 4 chromosomes, and their
750 combined effect. *Theoretical and Applied Genetics* 110, 1315-1323.
- 751 Munkvold, J., Tanaka, J., Benscher, D., and Sorrells, M. (2009). Mapping quantitative trait
752 loci for preharvest sprouting resistance in white wheat. *Theoretical and Applied
753 Genetics* 119, 1223-1235.
- 754 Nabim (2014). *Wheat Varieties* [Online]. Available: [http://www.nabim.org.uk/wheat/wheat-](http://www.nabim.org.uk/wheat/wheat-varieties)
755 [varieties](http://www.nabim.org.uk/wheat/wheat-varieties) [Accessed 19/04/2017 2017].
- 756 Nakamura, S., Pourkheirandish, M., Morishige, H., Kubo, Y., Nakamura, M., Ichimura, K.,
757 Seo, S., Kanamori, H., Wu, J., Ando, T., Hensel, G., Sameri, M., Stein, N., Sato, K.,
758 Matsumoto, T., Yano, M., and Komatsuda, T. (2016). Mitogen-Activated Protein
759 Kinase Kinase 3 Regulates Seed Dormancy in Barley. *Current biology : CB* 26, 775-
760 781.
- 761 Nave, M., Avni, R., Ben-Zvi, B., Hale, I., and Distelfeld, A. (2016). QTLs for uniform grain
762 dimensions and germination selected during wheat domestication are co-located on
763 chromosome 4B. *Theoretical and Applied Genetics* 129, 1303-1315.
- 764 Ogonnaya, F.C., Imtiaz, M., and Depauw, R.M. (2007). Haplotype diversity of preharvest
765 sprouting QTLs in wheat. *Genome* 50, 107-118.
- 766 Ramirez-Gonzalez, R.H., Segovia, V., Bird, N., Fenwick, P., Holdgate, S., Berry, S., Jack, P.,
767 Caccamo, M., and Uauy, C. (2015). RNA-Seq bulked segregant analysis enables the
768 identification of high-resolution genetic markers for breeding in hexaploid wheat.
769 *Plant Biotechnology Journal* 13, 613-624.
- 770 Reeves, J., Chiapparino, E., Donini, P., Ganai, M., Guiard, J., Hamrit, S., Heckenberger, M.,
771 Huan, X., Van Kaauwen, M., Kochieva, E., Koebner, R., Law, J., Lea, V., Leclerc,
772 V., Van Der Lee, T., Leigh, F., Van Der Linden, G., Malysheva, L., Melchinger, A.,
773 Orford, S., Reif, J., Röder, M., Schulman, A., Vosman, B., Van Der Wiel, C., Wolf,
774 M., and Zhang, D. (Year). "Changes over time in the genetic diversity of four major
775 European crops: a report from the GEDIFLUX Framework 5 Project", in: *Proceedings
776 of the 17th EUCARPIA general congress, Tulln, Austria, 8–11 September 2004*, eds.
777 J. Vollmann, H. Grausgruber & P. Ruckebauer), pp 3–7.
- 778 Shaw, P.D., Graham, M., Kennedy, J., Milne, I., and Marshall, D.F. (2014). Helium:
779 visualization of large scale plant pedigrees. *BMC Bioinformatics* 15, 259.
- 780 Shorinola, O., Bird, N., Simmonds, J., Berry, S., Henriksson, T., Jack, P., Werner, P., Gerjets,
781 T., Scholefield, D., Balcarkova, B., Valarik, M., Holdsworth, M.J., Flintham, J., and
782 Uauy, C. (2016). The wheat Phs-A1 pre-harvest sprouting resistance locus delays the
783 rate of seed dormancy loss and maps 0.3 cM distal to the PM19 genes in UK
784 germplasm. *Journal of experimental botany* 67, 4169-4178.
- 785 Simsek, S., Ohm, J.-B., Lu, H., Rugg, M., Berzonsky, W., Alamri, M.S., and Mergoum, M.
786 (2014). Effect of pre-harvest sprouting on physicochemical properties of starch in
787 wheat. *Foods* 3, 194-207.
- 788 Smith, S.M., and Maughan, P.J. (2015). SNP genotyping using KASPar assays. *Methods in
789 molecular biology (Clifton, N.J.)* 1245, 243-256.
- 790 Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of
791 eukaryotic genes, pseudogenes and promoters. *Genome biology* 7 Suppl 1, S10 11-
792 12.
- 793 Torada, A., Ikeguchi, S., and Koike, M. (2005). Mapping and validation of PCR-based
794 markers associated with a major QTL for seed dormancy in wheat. *Euphytica* 143,
795 251-255.
- 796 Torada, A., Koike, M., Ikeguchi, S., and Tsutsui, I. (2008). Mapping of a major locus
797 controlling seed dormancy using backcrossed progenies in wheat (*Triticum aestivum*
798 L.). *Genome* 51, 426-432.

- 799 Torada, A., Koike, M., Ogawa, T., Takenouchi, Y., Tadamura, K., Wu, J., Matsumoto, T.,
800 Kawaura, K., and Ogihara, Y. (2016). A Causal Gene for Seed Dormancy on Wheat
801 Chromosome 4A Encodes a MAP Kinase Kinase. *Current biology : CB* 26, 782-787.
802 Uauy, C. (2017). Wheat genomics comes of age. *Current Opinion in Plant Biology* 36, 142-
803 148.
804 Voss-Fels, K.P., Qian, L., Parra-Londono, S., Uptmoor, R., Frisch, M., Keeble-Gagnère, G.,
805 Appels, R., and Snowdon, R.J. (2017). Linkage drag constrains the roots of modern
806 wheat. *Plant, Cell & Environment* 40, 717-725.
807 Wicker, T., Matthews, D.E., and Keller, B. (2000). TREP: a database for Triticeae repetitive
808 elements. *Trends in Plant Science* 7, 561-562.
809 Wingen, L.U., Orford, S., Goram, R., Leverington-Waite, M., Bilham, L., Patsiou, T.S.,
810 Ambrose, M., Dicks, J., and Griffiths, S. (2014). Establishing the A. E. Watkins
811 landrace cultivar collection as a resource for systematic gene discovery in bread
812 wheat. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*
813 127, 1831-1842.
814 Xiao-Bo, R., Xiu-Jin, L., Deng-Cai, L., Jia-Li, W., and You-Liang, Z. (2008). Mapping QTLs
815 for pre-harvest sprouting tolerance on chromosome 2D in a synthetic hexaploid
816 wheat×common wheat cross. *Journal of applied genetics* 49, 333-341.
817 Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.-L., Wang, D., and Gao, C.
818 (2017). Precise base editing in rice, wheat and maize with a Cas9- cytidine deaminase
819 fusion. *Nat Biotech* advance online publication.

820

821
822

Tables

Table 1: *TaMKK3-A* and *PM19* alleles in *Phs-A1* association panel

Population	Variety	Origin	Status	<i>TaMKK3-A</i> Allele	<i>PM19-A1</i> promoter InDel	Reference
Alchemy x Robigus	Alchemy	UK	Dormant	C	Insertion	Shorinola et al (2016)
	Robigus	UK	Non-dormant	A	Deletion	
Option x Claire	Option	UK	Dormant	C	Insertion	Shorinola et al (2016)
	Claire	UK	Non-dormant	A	Deletion	
MAGIC Population	Yitpi	AUS	Dormant	C	Insertion	Barrero et al (2015)
	Baxter	AUS	Non-dormant	A	Deletion	
	Chara	AUS	Non-dormant	A	Deletion	
	Westonia	AUS	Non-dormant	A	Deletion	
Opata x W7984	W7984	MEX	Dormant	C	Insertion	Lohwasser et al, (2013)
	Opata	MEX	Non-dormant	A	Deletion	
OS21-5 x Haruyokoi	OS21-5	JPN	Dormant	C	Deletion	Torada et al, (2008)
	Haruyokoi	JPN	Non-dormant	A	Deletion	
Leader x Haruyokoi	Leader	CAN	Dormant	C	Insertion	Torada et al, (2008)
	Haruyokoi	JPN	Non-dormant	A	Deletion	
Kitamoe x Münstertaler	Kitamoe	JPN	Dormant	C	Deletion	Torada et al, (2005)
	Münstertaler	SUI	Non-dormant	A	Deletion	
Cranbrook x Halberd	Halberd	AUS	Dormant	C	Insertion	Mare et al (2005), Zhang et al (2008)
	Cranbrook	AUS	Non-dormant	A	Deletion	
Janz x AUS1408	Aus1408	AUS, SA	Dormant	C	Insertion	Mare et al (2005); Ogbonnaya et al (2007)
	Janz	AUS	Non-dormant	A	Deletion	
SW95-50213 x Cunningham	SW95-50213	CHN	Dormant	C	Deletion	Mares et al (2005)
	Cunningham	AUS	Non-dormant	A	Deletion	
SW95-50213 x AUS1408*	SW95-50213	CHN	Dormant	C	Deletion	Mares et al (2005)
	Aus1408	AUS, SA	Dormant	C	Insertion	
Boxer x Soleil	Soleil	UK	Dormant	C	Insertion	Flintham (2000)
	Boxer	UK	Non-dormant	A	Deletion	

823 **Phs-A1* was not detected in this population as the DH lines were generally dormant. However, limited number of lines showed transgressive
824 segregation relative to the dormant phenotypes of the two parents.

825 **Table 2.** Mean Coefficient of Parentage (COP) within Australian and UK
826 germplasm, and between groups of the most prevalent haplotypes containing
827 dormant (C) and non-dormant (A) SNPs at *TaMKK3-A*.

Germplasm	Haplotype	<i>TaMKK3-A</i> SNP	Mean COP	Comparisons (n)
Australia	All	A/C	0.10	13530
Australia	H5/H7	C	0.17	350
Australia	H1/H2	A	0.15	2700
UK	All	A/C	0.11	1596
UK	H12	C	0.20	496
UK	H3	A	0.21	55

828

829 **Figure Legends:**

830

831 **Figure 1:** Physical map of the *Phs1-A1* interval in bread wheat Chinese Spring (CS)
832 and wild emmer (Zavitan). (A) *Phs1-A1* interval in CS is covered by two non-
833 overlapping BAC clusters: Cluster 285 (4 BACs) and Cluster 16421 (11 BACs). BACs
834 are represented by solid lines while genes found on the BAC are represented by filled
835 ovals. The proposed candidate genes for *Phs-A1* are in red font. (B) Whole genome
836 assembly of Zavitan wild emmer across the *Phs1-A1* interval. Genes present in both
837 assemblies are joined by dotted lines.

838

839 **Figure 2:** Marker development and allele distribution of *MKK3-A* in ancestral and
840 historic germplasm. (A) Genotype plot of varieties and a F₃ population segregating for
841 *Phs-A1* using the *TaMKK3-A* KASP assay. (B) Development of co-dominant CAPS
842 marker based on *Hpy166II* restriction digest of the C660A SNP. Non-dormant
843 varieties Robigus (R) and Clare (C); dormant varieties Alchemy (A) and Option (O).
844 (C) Allele frequency of the causal C660A SNP in *T. urartu* and *T. turgidum* ssp.
845 *dicoccoides* accessions and the Watkins and Gediflux collections. The number of lines
846 genotyped in each germplasm set is in parenthesis. (D) *TaMKK3-A* allele distribution
847 in the four wheat end-use groups (nabim 1-4) in the UK.

848

849 **Figure 3:** *Phs-A1* haplotype analysis. (A) Structure of 14 haplotypes identified in the
850 HapMap population across 39 SNP loci in the *Phs-A1* interval. SNP loci are ordered
851 based on their physical position on the Zavitan assembly. Exons, intron and intergenic
852 regions are represented by filled boxes, solid lines and breaks, respectively. Note that
853 *ERF-C* is not on the Zavitan assembly. (B) Haplotype network of the 14 haplotypes.
854 The size of each circle corresponds to the number of lines in each haplotype. Blue,
855 light blue, amber and red represent cultivars, landraces, breeding and synthetic line in
856 each haplotype, respectively. (C) Geographical distribution of the four *TaMKK3-A*
857 haplotypes. The size of the circle represents the sample size obtained within each
858 country while each section represents the proportion of the country sample size with
859 the specified haplotype.

860

861 **Figure 4:** Relationship between the European, UK, Australian and HapMap *Phs-A1*
862 Haplotypes. The distribution (bar charts) of the HapMap and unique haplotypes found
863 in the Gediflux (European), UK and Australian germplasm using genotype
864 information of seven of the 39 HapMap SNPs within the *Phs-A1* interval.

865

866 **Figure 5:** Pedigree of selected UK and European varieties highlighting the origin of
867 the major resistant haplotype (H12). Each circle represents a variety and colours
868 represent the different haplotypes. Nodes are size based on the number of varieties
869 derived from the node.

870 **Supplemental files:**

871

872 A single excel document includes the following supplemental tables as individual
873 spreadsheets.

874

875 **Table S1:** BAC, Sequence and gene content information for the BAC cluster in *Phs-*
876 *A1* interval.

877 **Table S2:** *TaMKK3-A* Alleles of cultivars in the Watkins collection.

878 **Table S3:** Classification of HapMap lines according to *Phs-A1* haplotypes.

879 **Table S4:** KASP markers designed for haplotype analysis of *Phs-A1* interval.

880 **Table S5:** Haplotype map of the Gediflux Collection

881 **Table S6:** Haplotype map of Australian germplasm

882

883 **Figure S1:** Pedigree of UK and European varieties with their corresponding *Phs-A1*
884 haplotype status. Each circle represents a variety and the colours represent the different
885 haplotypes.

886

887 **Figure S2:** Pedigree of UK and European varieties with their corresponding *TaMKK3-*
888 *A* allele. Each circle represents a variety and the colours represent the different
889 haplotypes.

890

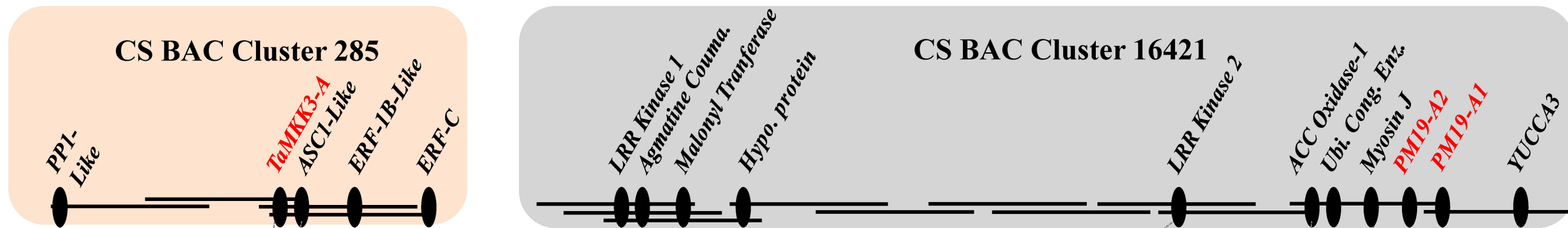
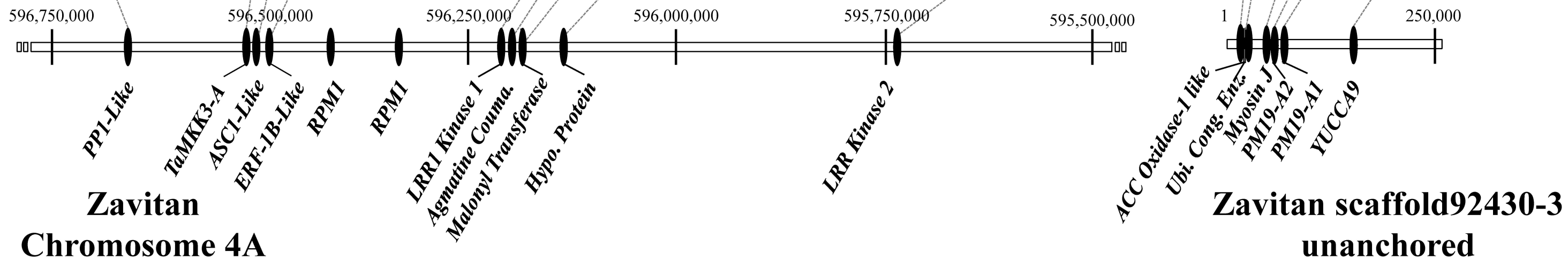
891 **Figure S3:** Pedigree of Australian varieties with their corresponding *Phs-A1* haplotype
892 status. Each circle represents a variety and the colours represent the different
893 haplotypes.

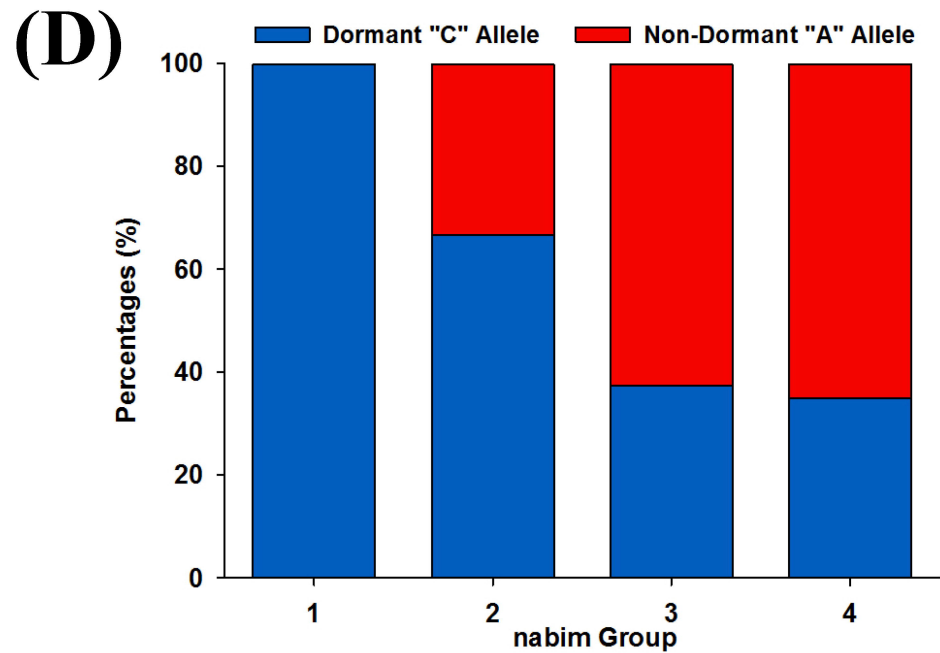
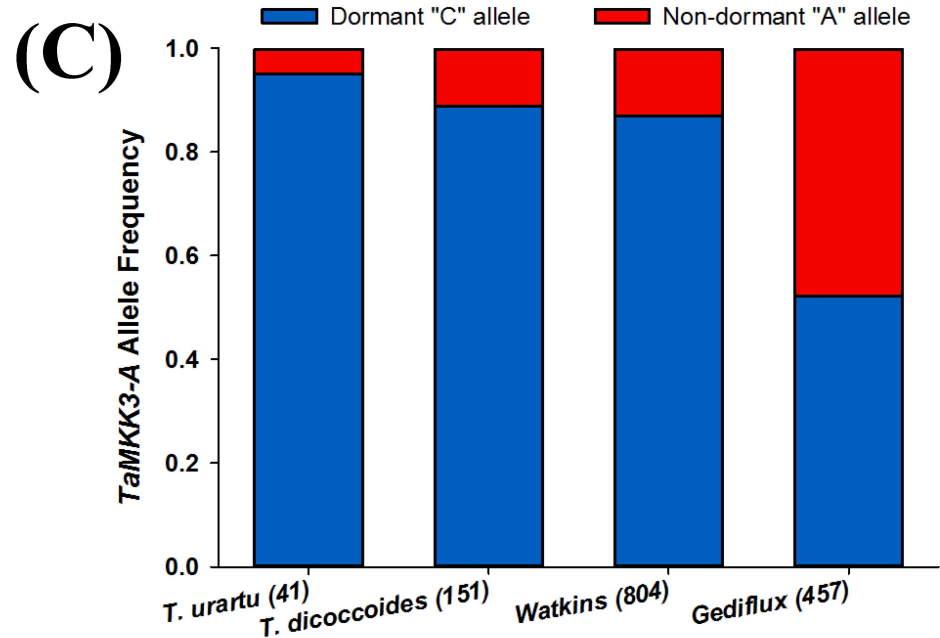
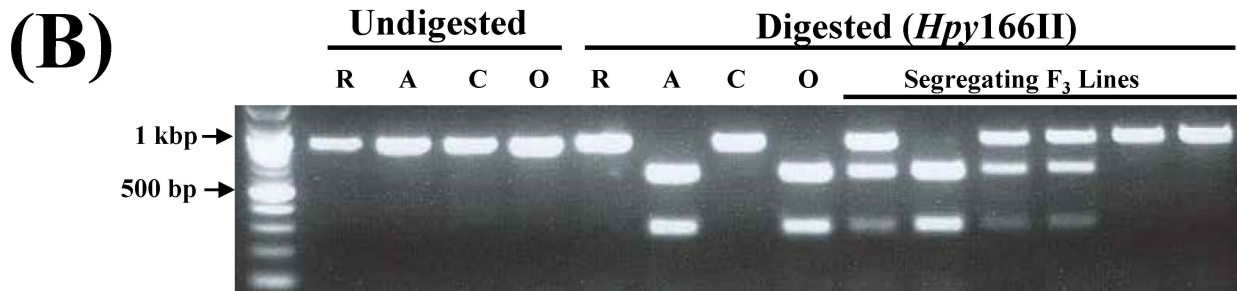
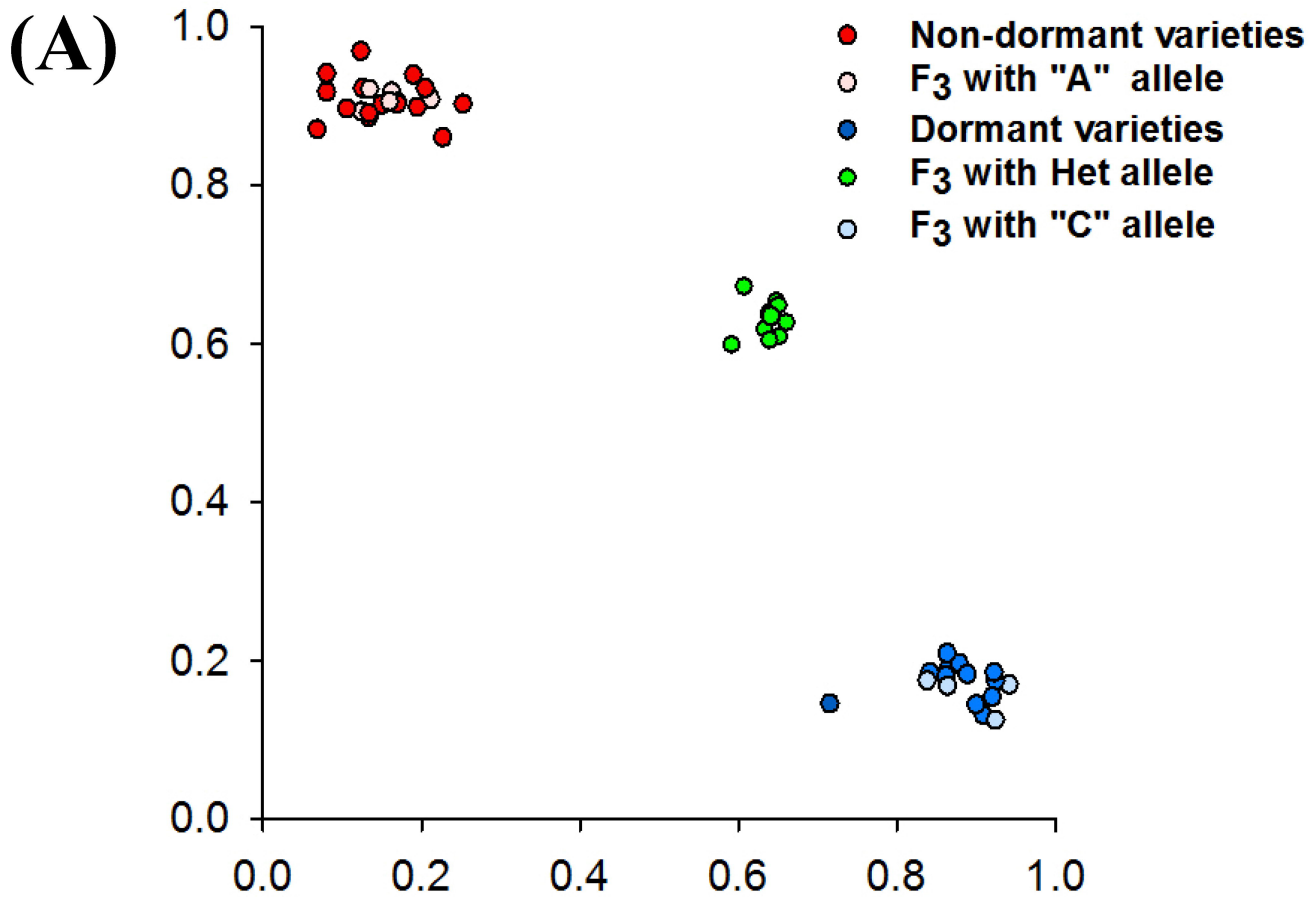
894

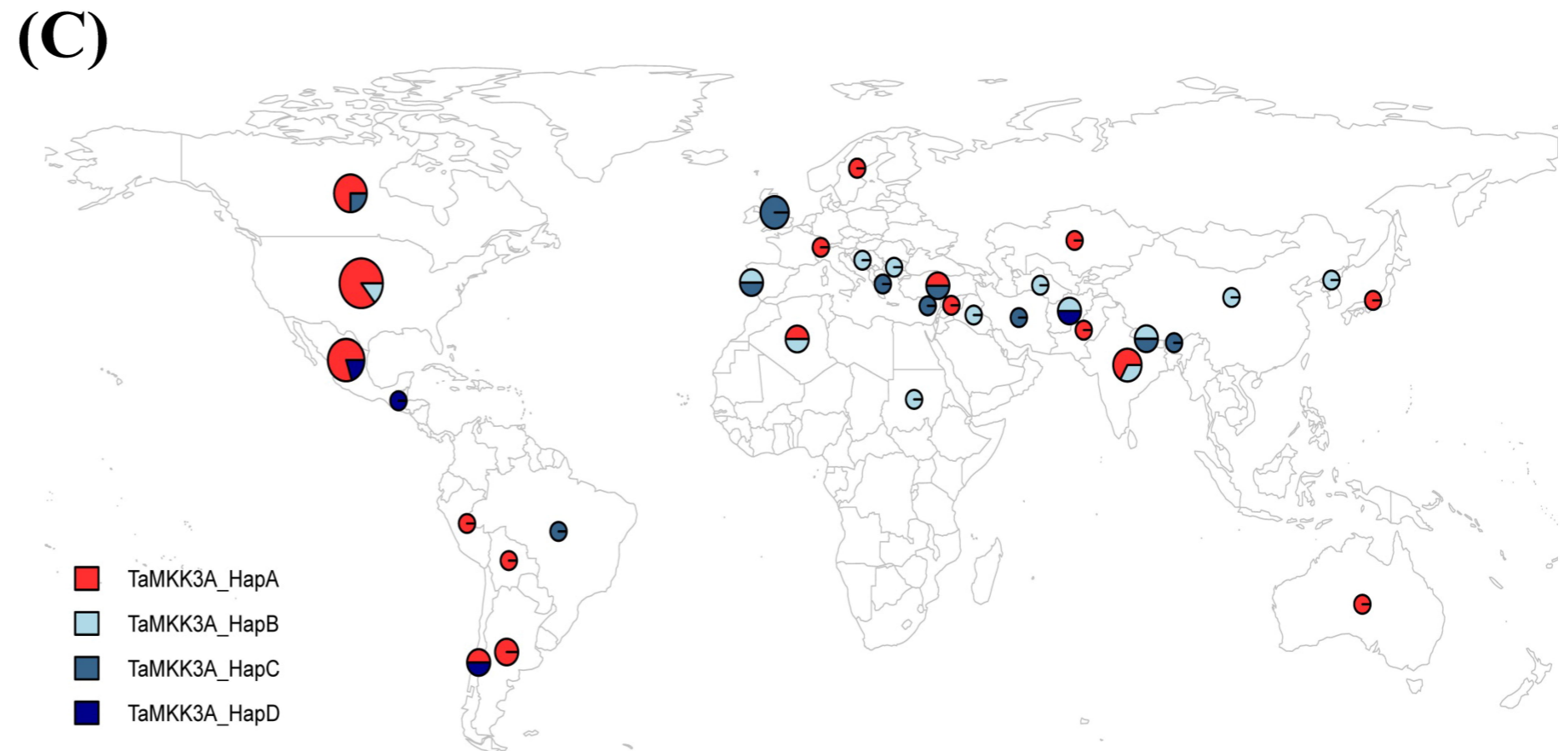
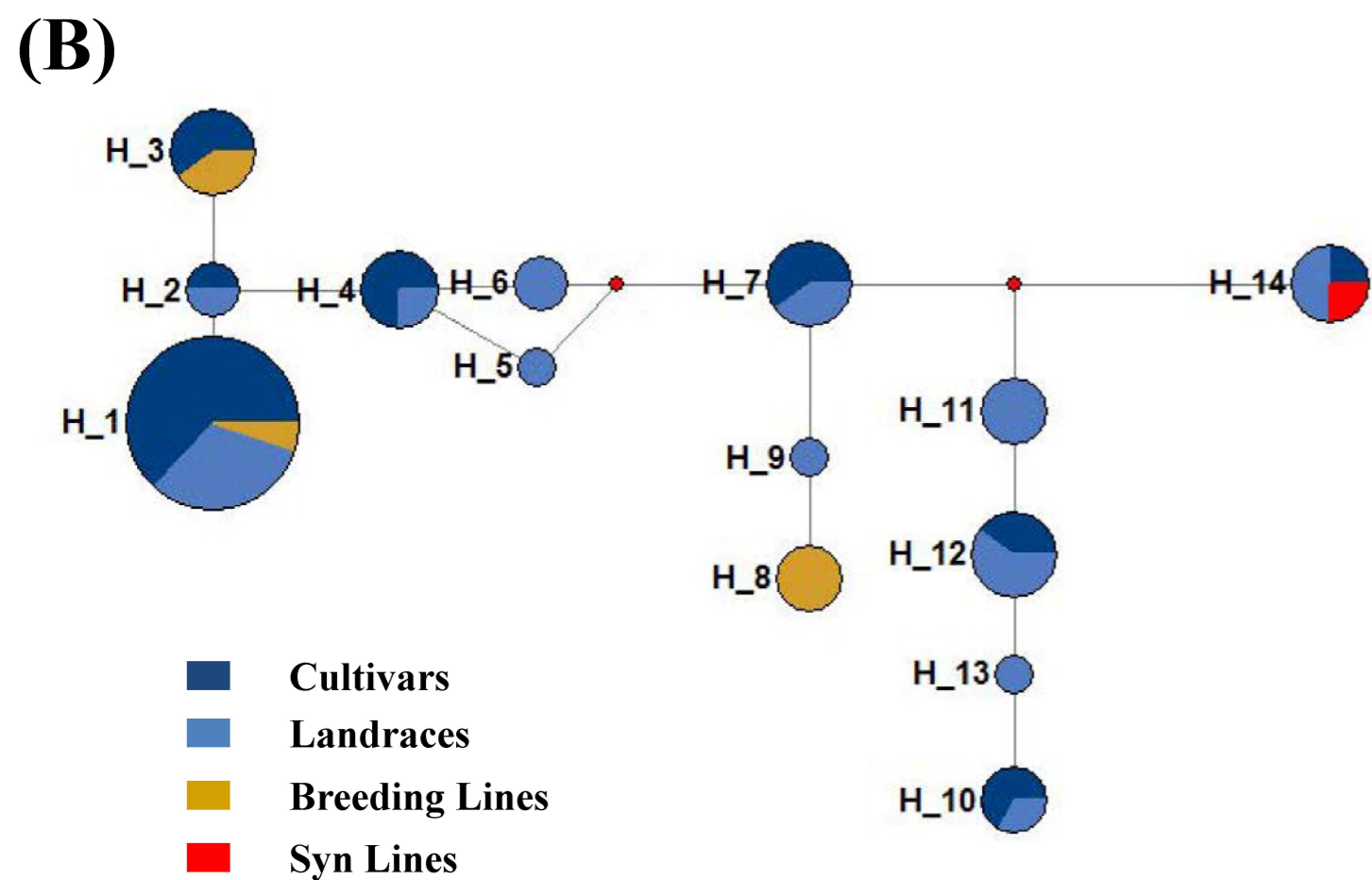
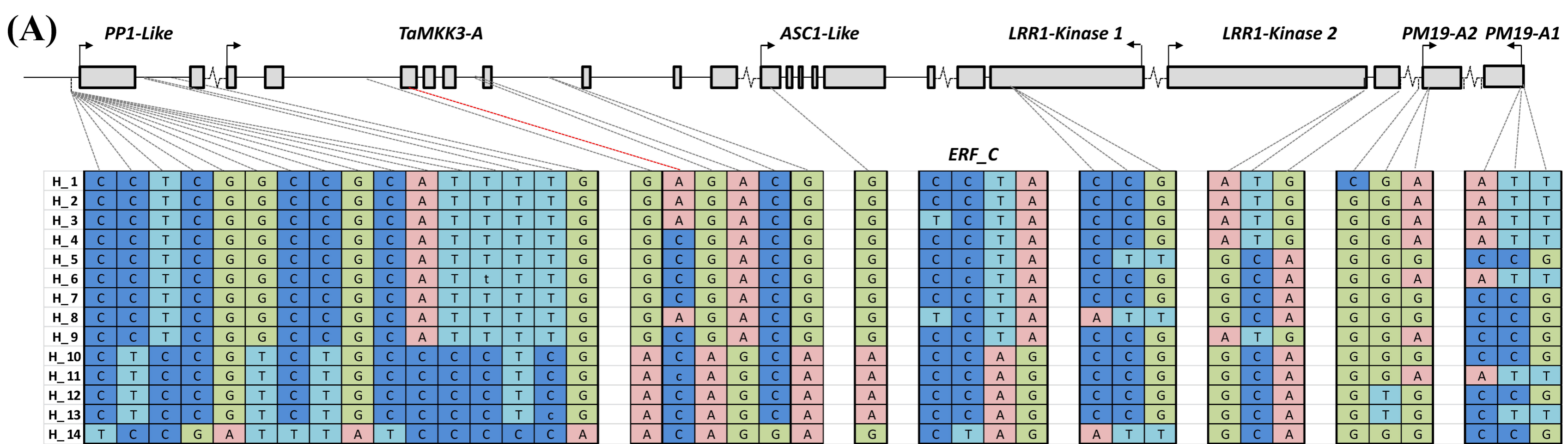
895 **Figure S4:** Pedigree of Australian varieties with their corresponding *TaMKK3-A*
896 allele. Each circle represents a variety and the colours represent the different
897 haplotypes.

898

899 **Figure S5:** Comparison of syntenic *Phs-A1* physical maps and contigs in
900 *Brachypodium*, wheat and barley. Genes in the homologous intervals in *Brachypodium*
901 (amber line), wheat (black lines) and barley (grey line) are compared against each
902 other. Orthologous genes across genomes are joined by lines. The region of high
903 conservation is indicated with grey background while the region of low conservation
904 has plain background.

(A)**(B)**



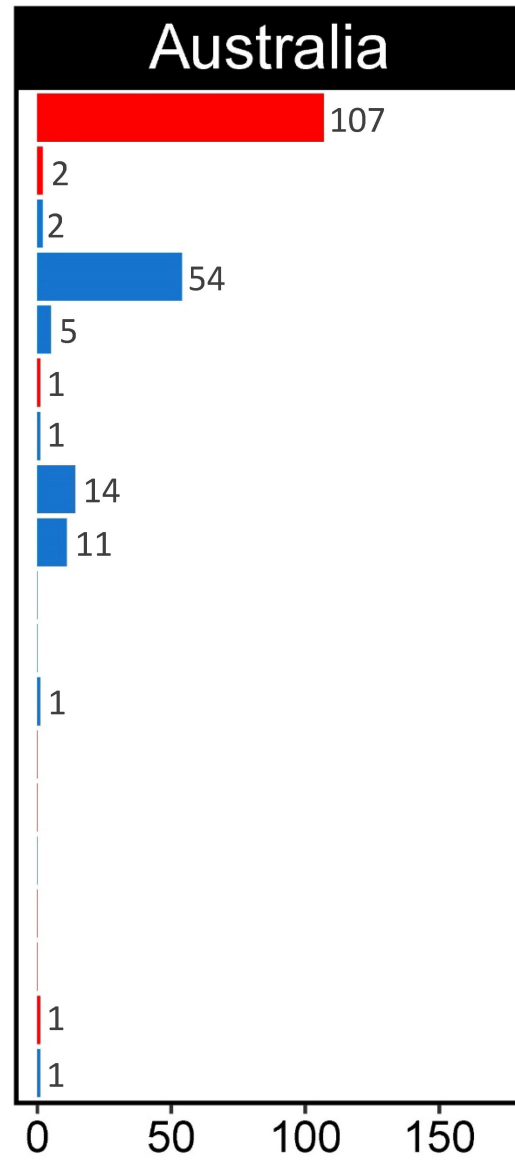
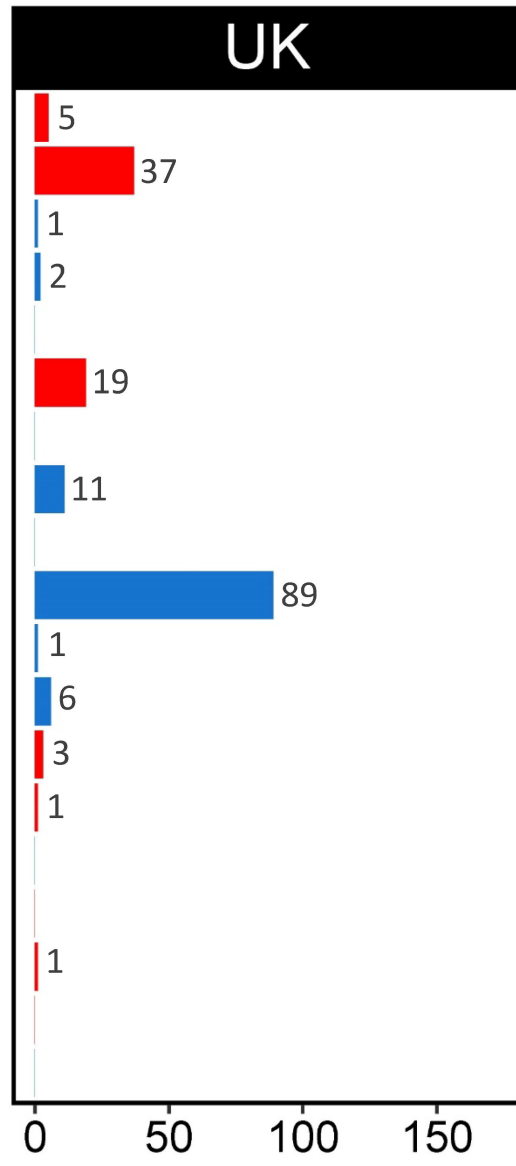
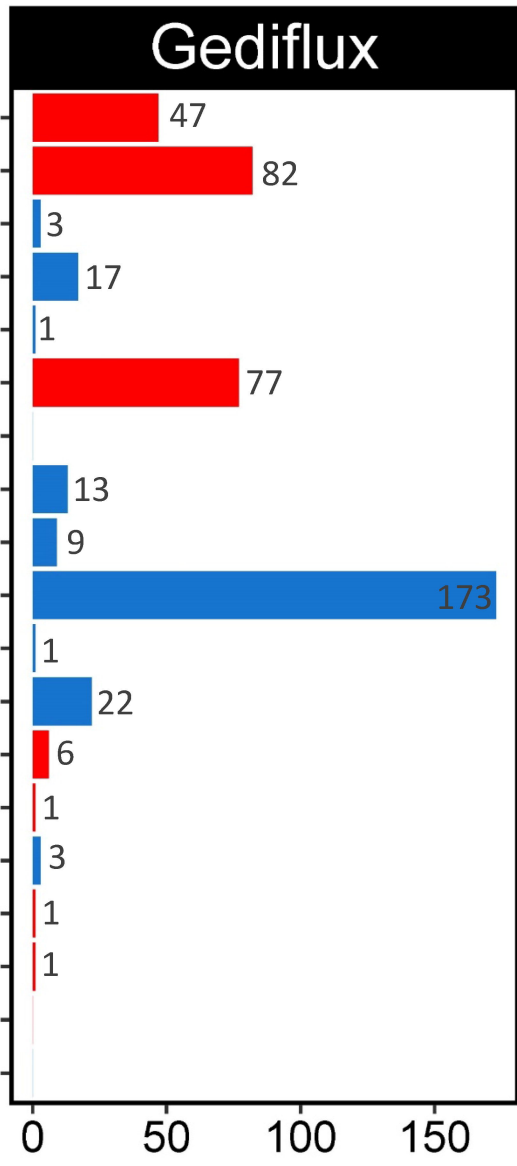


IWGS Contig (SNP position)

4AL_7159962 (713)
4AL_3844919 (137)
 4AL_5621077 (266)
 4AL_7064804 (6018)
 4AL_7158475 (4203)
 4AL_7174272 (20841)
 4AL_7123764 (2216)

■ Dormant ■ Non-Dormant

H1/H2	T	A	G	C	G	G	T
H3	T	A	G	T	G	G	T
H4	T	C	G	C	G	G	T
H5/H7	T	C	G	C	A	G	C
H6	T	C	G	C	A	G	T
H8	T	A	G	T	A	G	C
H9	C	C	G	C	G	G	C
H10	C	C	A	C	A	G	C
H11	C	C	A	C	A	G	T
H12	C	C	G	C	A	G	C
H13	C	C	A	C	A	T	C
H14	C	C	A	C	A	T	T
New Gediflux 1	T	A	G	C	G	T	C
New Gediflux 2	T	A	G	C	A	G	C
New Gediflux 3	T	C	A	C	A	G	C
New Gediflux 4	T	A	G	C	G	G	T
New Gediflux 5	C	A	G	C	G	G	C
New Aus 1	T	A	A	C	G	G	T
New Aus 2	C	C		C	G	G	T



Number of Lines

