LYS3 encodes a prolamin-box-binding transcription factor that controls embryo growth in barley and wheat

Beata Orman-Ligeza[a], Philippa Borrill[b], Tansy Chia[a], Marcella Chirico[a], Jaroslav Doležel[c], Sinead Drea[d], Miroslava Karafiátová[c], Nicole Schatlowski[a], Charles U. Solomon[d,f], Burkhard Steuernagel[e], Brande B. H. Wulff[e], Cristobal Uauy[e], Kay Trafford[a]

[a] NIAB, Genetics and Breeding, Huntington Road, Cambridge, CB3 0LE, UK.

[b] School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

[c] Institute of Experimental Botany, Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic.

[d] Department of Genetics, University of Leicester, Adrian Building, University Road, Leicester, LE1 7RH, UK.

[e] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK.

[f] Department of Plant Science and Biotechnology, Abia State University, PMB 2000, Uturu, Nigeria.

Present addresses:

Nicole Schatlowski: Wellcome Genome Campus, Hinxton, CB10 1SA, UK

*****Corresponding author. Kay Trafford, The National Institute of Agricultural Botany, Huntingdon Road, Cambridge, Cambridgeshire CB3 0LE, UK. Email address: kay.trafford@niab.com. Tel +44 (0)1223 34249.

Keywords: high lysine, PBF, large embryo, shrunken endosperm.

Abbreviations: Days after flowering (DAF), DNA binding with one zinc finger (DOF), prolamin-box binding factor (PBF), targeting induced local lesions in genomes (TILLING), sorting intolerant from tolerant (SIFT).

Declarations of interest: none

Number of words: 5428

Number of Tables: 1

Number of Figures: 5

Number of supplementary Tables: 4

Number of supplementary Figures: 2

HIGHLIGHTS

- *LYS3* encodes a transcription factor called Prolamin Binding Factor (PBF) that is expressed in grains only.
- Wheat and barley *LYS3*/*PBF* mutants have enlarged embryos suggesting that this gene suppresses embryo growth.
- The down-stream target genes of *PBF* in wheat are predicted to be involved in a wide range of biological processes including organ development and starch metabolism.

ABSTRACT

Mutations at the *LYS3* locus in barley have multiple effects on grain development, including an increase in embryo size and a decrease in endosperm starch content. The gene underlying *LYS3* was identified by genetic mapping and mutations in this gene were identified in all four barley *lys3* alleles. *LYS3* encodes a transcription factor called Prolamin Binding Factor (PBF). Its role in controlling embryo size was confirmed using wheat TILLING mutants. To understand how *PBF* controls embryo development, we studied its spatial and temporal patterns of expression in developing grains. The *PBF* gene is expressed in both the endosperm and the embryos, but the timing of expression in these organs differs. *PBF* expression in wild-type embryos precedes the onset of embryo enlargement in *lys3* mutants, suggesting that *PBF* suppresses embryo growth. We predicted the down-stream target genes of *PBF* in wheat and found them to be involved in a wide range of biological processes, including organ development and starch metabolism. Our work suggests that *PBF* may influence embryo size and endosperm starch synthesis via separate gene control networks.

A. Introduction

In the 1960-70s, in an attempt to improve the lysine content of barley (*Hordeum vulgare* L.) for animal feed, mutagenized barley germplasm at the Risø National Laboratory, Denmark, was screened for lysine content and a number of high-lysine (*lys*) mutants were identified (Doll *et al.*, 1974; Doll, 1976). The *lys* mutant with the highest lysine content (44 percent higher than wild type) was Risø1508 (*lys3a*) (Ingversen *et al.*, 1973; Mossberg, 1969). In addition to *lys3a*, there are three other barley lines with mutations at the same locus: Risø18 (*lys3b*), Risø19 (*lys3c*) and M1460 (*lys3d*) (Aastrup, 1983; Munck, 1992) (Supplementary Table S1). All four *lys3* mutations are recessive.

Studies have shown that the endosperm of *lys3* mutants, like that of most other high-lysine barley mutants, contains less starch and lysine-poor protein (hordein), but has more lysine-rich protein and free lysine (Brandt, 1976; Ingversen *et al.*, 1973). However, uniquely amongst high-lysine barley mutants, the embryos of all four *lys3* mutants are larger than normal (Tallberg, 1977, Deggerdal *et al.*, 1986, Cook *et al.*, 2018). Most of the barley *lys* mutants, including those known to have lesions in genes encoding components of the starch biosynthesis pathway (Trafford and Fincher, 2014), have reduced embryo weight (as well as shrivelled-endosperm and reduced starch content; Cook *et al.*, 2018). The enlargement of the embryos in *lys3* mutants is therefore a specific response to mutation at the *LYS3* locus and not due to the diversion of resources from the shrivelled endosperm to the embryo. As well as increased size, the embryos of *lys3* mutants have higher-than-normal starch content (Deggerdal *et al.*, 1986; Olsen *et al.*, 1984) and larger-than-normal cells in the scutellum (Olsen *et al.*, 1984; Deggerdal *et al.*, 1986). The mutant embryos also show reduced dormancy (Cook *et al.*, 2018).

Risø1508 (*lys3a*) has been the subject of a breeding programme to improve the nutritional quality of barley, particularly for pig feed (Munck and Jespersen, 2009). Animal feeding

trials, using the Risø1508-derived lines Piggy and Lysimax, showed that *lys3a* mutant lines are a more effective source of protein for maximum growth rate in pigs than the wild-type barley varieties from which they are derived (Munck, 1972; Mortensen *et al.*, 1988; Gabert *et al.*, 1995; Gabert *et al.*, 1996). However, the success of Risø1508-derived barley lines as animal feed is hampered by their reduced starch content which leads to low grain weight and consequently to low yield.

To separate the favourable (nutritional enhancements) and unfavourable (yield depression) traits associated with *lys3* mutations, it is necessary to identify the gene responsible and then to understand how it functions in developing embryos and endosperm. We report here the identification of the *LYS3* gene by genetic mapping using embryo size as the selection phenotype. Whilst our work on this was in progress, the gene underlying the *LYS3* locus was independently identified by Moehs *et al.* (2019) using variation in hordein content as the selection phenotype. Both studies agree that *LYS3* encodes a previously-identified and well-studied transcription factor in barley: prolamin-box binding factor (PBF). To understand PBF function, we studied its gene expression patterns in barley and found that contrary to previous reports (Mena *et al.*, 1998; Mena *et al.*, 2002), *PBF* is expressed in developing embryos as well as in the endosperm. The emphasis of our current work is to identify the down-stream target genes of PBF and here we present our initial studies of its predicted targets in wheat.

2. **Materials and methods**

2.1. Barley germplasm

Grains of Bomi, Morex and Risø1508 were obtained from the Germplasm Resources Unit, John Innes Centre, Norwich, UK and Risø18, Risø19, M1460 and Minerva were kindly supplied by Birthe Møller Jespersen, University of Copenhagen, Denmark.

2.2. Plant growth

For mapping experiments, individual grains were germinated in Petri dishes on moist filter paper. After over-night incubation at 4 °C, plates were transferred to room temperature. When roots and shoots were established, each seedling was transplanted into a 1 L pot containing Levington M2 compost (Scotts Professional, Ipswich, UK) and grown in a glasshouse. In winter, additional lighting was provided by sodium lamps for 16 h per day and temperatures were maintained between 15 °C (night) and 20 °C (day). In summer, plants were grown in a glasshouse under ambient conditions.

Wheat TILLING mutants were sown directly into M2 compost, incubated at 4 °C for 3 days and then transferred to a glasshouse with a 22-hour photoperiod and temperatures of 21 °C (night) and 18 °C (day). Supplementary lighting was provided by a mixture of high-pressure sodium lamps and both far red and white LED lights (Conviron, Winnipeg, US).

2.3. Analysis of grain and embryo development

Anthesis occurred whilst the ear was enveloped in the flag leaf so the exact day of anthesis was difficult to determine without damaging the developing spike. Accordingly, flowering time was defined as the day on which the awns of the developing ear protruded more than 1 cm above the leaf sheath and grain/embryo age was measured in days after flowering (DAF).

127 *2.4. DNA extraction*

128 For barley genotyping, DNA was extracted as follows: leaf material was harvested at seedling
129 stage to 1.5-ml tubes each containing a 5-mm diameter steel ball and frozen at -80 °C. Frozen
130 leaf material was homogenised using a Geno/Grinder (SPEX SamplePrep LLC) and then
131 600 µl of extraction buffer (200 mM Tris pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% [w/v]
132 SDS) was added. The homogenized leaf material was incubated at 65 °C for one hour. Leaf
133 debris was pelleted by centrifugation and the supernatant was transferred to a fresh 1.5-ml
134 tube, mixed with an equal volume of isopropanol and centrifuged. The pellet was washed
135 with 500 ml 70% [v/v] ethanol and resuspended in 200 µl H_2O. DNA was quantified using a
136 spectrometer (NanoDrop 1000, Thermo Scientific) and the concentration adjusted to
137 10 ng/µl.

138 For wheat genotyping, DNA was extracted from seedling leaf material using the method
139 of Fulton *et al.* (1995).

140

141 *2.5. Chromosome sequencing*

142 For each barley line, suspensions of intact mitotic metaphase chromosomes were prepared
143 from synchronized root tip cells of barley seedlings as described by Lysák *et al.* (1999).
144 Chromosomes in suspension were stained with 2 µg.ml^{-1} DAPI (4', 6-diamidino-2-
145 phenylindole) and chromosome 5H was sorted using a FACSAria II SORP flow cytometer
146 and sorter (Becton Dickinson Immunocytometry Systems, San José, USA). Purity in the
147 sorted 5H fractions was determined microscopically after FISH with a probe for GAA
148 microsatellite (Kubaláková *et al.*, 2003). DNA of the sorted chromosomes was purified and
149 amplified by multiple displacement amplification according to Šimková *et al.* (2008). Three
150 independent amplification products were combined in each cultivar to reduce amplification
151 bias.

The samples of pooled amplified chromosomal DNA were subjected to Illumina HiSeq2500 sequencing (The Genome Analysis Centre, Norwich, UK; now the Earlham Institute). Paired-end read size was 250 bp for the wild-type samples and 125 bp for the mutant samples. The total number of paired-reads obtained was 172,974,201 (Bomi), 252,692,421 (Risø1508) and 226,936,729 (Risø19). Data were submitted to the European Nucleotide Archive (www.ebi.ac.uk) with accession number PRJEB33709.

2.6. Genotyping the mapping population

The Bomi chromosome sequencing data was assembled using the published Morex genome sequence (as described below for MutChromSeq). Genes flanking the *LYS3* locus with SNPs between Bomi and Morex were selected for KASP marker design. For primer sequences see Supplementary Table S2A. Genotyping was performed on a Quant Studio 7 (Applied Biosystems) using KASP technology (LGC) following the manufacturers' instructions.

2.7. Determination of embryo size

For fine mapping in barley, the embryo phenotype of recombinant lines was visually determined and categorized as normal or large embryo. For determination of relative embryo dry weight, all barley or wheat plants for each experiment were grown together in one batch and grains were harvested at maturity. To extract the embryos, grains were soaked overnight in sterile water at 4 °C in the dark and then dissected into embryo and non-embryo portions. Both portions were dried to constant weight by heating in an oven at 65 °C for 3 days. Relative embryo weight (%) = (embryo dry weight / (embryo + non-embryo dry weight)) x 100.

177 *2.8. Identification of a candidate gene using MutChromSeq*

178 The chromosome sequencing data was analysed using the MutChromSeq pipeline according
179 to Steuernagel *et al.* (2017). The raw data were trimmed using sickle
180 (https://github.com/najoshi/sickle) with default parameters. Trimmed data from wild types
181 Bomi and Minerva were assembled using CLC Assembly Cell
182 (https://www.qiagenbioinformatics.com/products/clc-assembly-cell/), version 5.1 with default
183 parameters. Trimmed data from mutants and wild types were mapped to wild-type assemblies
184 using bwa sampe (Li and Durbin, 2009), version 0.7.12 with default parameters. Further
185 processing of mappings was done using samtools (Li *et al.*, 2009) version 0.1.19. For running
186 MutChromSeq, we used release 2 (https://github.com/steuernb/MutChromSeq/releases/tag/2),
187 which introduces the filtering of contigs using a mapping interval, as described in Dracatos *et*
188 *al.* (2019). These sequences and those of the wild-type controls have been submitted to
189 GenBank (accession numbers: MN715383, MN715384, MN715385, MN715386,
190 MN715387).
191
192 2.9. *Analysis of barley PBF/LYS3 genes*
193 The SNP mutations in the *PBF/LYS3* genes that were identified by chromosome sequencing
194 were confirmed by PCR amplification and Sanger sequencing. The primer sequences and
195 PCR conditions are given in Supplementary Table S2B. Mutations in the mapping lines were
196 confirmed using KASP assays (Supplementary Table S2C).
197
198 *2.10. Large-embryo TILLING mutants of wheat*
199 Lines of the hexaploid wheat cultivar Cadenza were selected from the *in silico* wheat
200 TILLING resource (www.wheat-tilling.com; Krasileva et al., 2017) (Supplementary Table
201 S2D). The supplied seed was sown, pairs of homozygous mutant lines were selected and

cross-pollinated, and the resulting F_1 plants were confirmed to be heterozygous. A homozygous triple mutant was constructed by crossing Cadenza0903 (*TaPBF-B1*) to Cadenza0904 (*TaPBF-D1*) and homozygous F_2 double mutants were then crossed to Cadenza1807 or Cadenza1553 (*TaPBF-A1* mutants).

2.11. Analysis of expression by Reverse Transcription PCR

Tissue was harvested, weighed, frozen in liquid nitrogen and stored at -20 ºC. Total RNA was extracted using Tri Reagent (Sigma Aldrich, UK) according to the manufacturer's instructions, its concentration was measured with a spectrometer (NanoDrop 1000, Thermo Scientific) and it was stored at -80 °C. An aliquot containing 4 µg RNA was treated for 45 min with 2 µl of DNase RQ1 (1 µg/µl) at 37 °C (Promega, UK), and purified using an RNAeasy spin column (Qiagen, UK). cDNA was prepared from 0.5 µg RNA using a SuperScript III reverse transcriptase kit and oligo(dT)18 primers (Thermo Fisher Scientific, UK) according to the manufacturer's instructions and stored at -20 °C prior to PCR. The primer sequences and PCR conditions are given in Supplementary Table S2E.

2.12. Analysis of expression by in situ hybridization

Developing barley grains were harvested and processed for mRNA *in situ* hybridization as described in Drea *et al*., (2005) and Opanowicz *et al.*, (2010). The probe template consisted of a *PBF* cDNA fragment amplified with gene-specific primers from a barley grain cDNA library (primer sequences are given in Supplementary Table S2F). For amplification of the antisense probe, a T7 polymerase promoter sequence was attached to the 3' end of the reverse primer. For amplification of the sense probe, a T7 promoter sequence was attached to the 5' end of the forward primer. The amplified products were transcribed *in vitro* with T7 RNA polymerase (Bioline, UK).

227

228 *2.13. Downstream targets of wheat PBF*

229 The promotor regions (-2 kb) of 29 wheat starch biosynthesis genes were searched for
230 sequences conforming to the consensus prolamin and pyrimidine binding sequences
231 (TGTAAAG and CCTTTT, respectively). The predicted TaPBF downstream target genes
232 were investigated using the RefSeqv1.0 wheat genome assembly and associated wheat gene
233 networks. One network (referred to as the GENIE3 network) had been constructed to predict
234 transcription factor targets using 850 RNA-seq samples (Ramírez-González *et al.* (2018) and
235 four separate co-expression (Weight Gene Correlation Network Analysis, WGCNA)
236 networks for grain, leaf, root and spike tissues were used to identify co-expressed genes
237 (Ramírez-González et al., 2018). Gene ontology term enrichment analysis was performed as
238 described in Ramírez-González *et al.* (2018) using GOseq (Young et al., 2010).

3. Results

3.1. Fine mapping LYS3 identified a 64-gene region on chromosome arm 5HL controlling embryo size

Our previous work suggested that the *LYS3* gene in barley controls embryo size. To test this and to identify the *LYS3* gene, we developed a mapping population by crossing each of the three Risø *lys3* large-embryo mutants to Morex, a cultivar with normal embryo size and with available genome sequence data. Following self-pollination, the F_2 grains were genotyped with polymorphic markers in genes close to the *LYS3* locus on chromosome arm 5HL (Franckowiak, 1997). Lines with chromosomal recombination points close to *LYS3* were selected and allowed to self-pollinate. No lines with recombination events close to *lys3* were found for progeny from the Risø18 x Morex crosses and so further work involved progeny from crosses between Risø19 and Risø1508 only. Homozygous recombinant lines (and controls) were selected from the progeny for further genotypic and phenotypic analysis.

To design new markers for genotyping within the region of interest, we purified by flow cytometric sorting the 5H chromosomes from both *lys3* mutants and the parental control cultivar, Bomi, and we sequenced DNA amplified from them using Illumina HiSeq2500. The sequencing data was assembled using the MutChromSeq pipeline according to Steuernagel *et al.* (2017), and single nucleotide polymorphisms (SNPs) in the region of interest were identified. These SNPs were used to design KASP primers for further genotypic analysis (Supplementary Table S2).

To phenotype the homozygous recombinant lines, we first assessed embryo size visually (Table 1; Fig. 1A). The phenotype of selected critical lines was also assessed by measurement of relative embryo weight and β-amylase activity (Fig. 1B). In our previous work, we showed that all four *lys3* mutants had increased absolute and relative embryo weights (Cook *et al.*, 2018). When tested again in different growth conditions (Fig. 2A), although the relative

264 embryo size was consistently increased for all *lys3* mutants, the increase in absolute embryo
265 weight was significant for Risø1508 and M1460 only. For this reason, we used the relative
266 embryo weight to phenotype the critical mapping lines. We also used β-amylase activity to
267 phenotype these critical lines because it is known to be strongly affected by *lys3* mutations
268 (Allison, 1978).
269 The visual phenotyping together with the genotyping data suggested that the gene that is
270 responsible for the large-embryo phenotype lies between HORVU5Hr1G048110 and
271 HORVU5Hr1G048950 (Table 1, Fig. 1A). Quantitative measurement of relative embryo size
272 in critical lines confirmed the visual phenotyping results (Fig. 1B). The embryo weights for
273 lines visually assessed as 'mutant' were significantly greater than the weights of those
274 assessed as 'wild-type' ($p < 0.05$, Student's *t*-test). The phenotyping results for β-amylase
275 activity were less clear due to a large amount of variation between the values within each
276 group. Although collectively, the large-embryo lines had lower β-amylase activity than the
277 wild-type lines, not all of the comparative differences between individual mutant and wild
278 type lines were statistically significant (Fig. 1B).
279
280 *3.2. LYS3 encodes the transcription factor, PBF*
281 To identify the gene responsible for controlling embryo size within the region of interest
282 identified by mapping, we isolated and sequenced the 5H chromosome of the two other *lys3*
283 mutants, Risø18 and M1460 and from an additional wild-type line, Minerva which is the
284 parent of M1460. This sequencing information was combined with that of the three lines
285 sequenced previously and bioinformatically assessed using a method based on MutChromSeq
286 (a method that was used previously e.g. for the identification of disease-resistance genes in
287 mutant barleys; Sanchez-Martin *et al.*, 2016). Only one gene in the region of interest (Fig.

1C) had mutations in all four mutant lines relative to the wild-type controls (Fig. 1D) and that was HORVU5Hr1G048700 which encodes the PBF transcription factor.

The *PBF* genes from all four *lys3* mutants were examined and all contained defects likely to be deleterious (Supplementary Table S1 and Fig. 1D). The sequences of three of the four *lys3 PBF* alleles are disrupted by SNPs. Risø1508 and M1460 both have the same SNP (A173T) which causes an amino acid substitution (Q58L) in the DOF (DNA binding with one zinc finger) DNA binding domain, consistent with Moehs *et al.* (2019). The glutamine residue, Q58 is highly conserved amongst orthologous proteins (Ensembl plants), with 27 out of 28 sequences examined having glutamine in this position. Risø19 has a single nucleotide polymorphism (G109T) which results in a nonsense mutation (E to STOP), which would prevent production of a full-length protein. In Risø18, various lines of evidence suggest that the entire *PBF* gene together with several additional genes on either side are deleted. Firstly, analysis of the chromosome 5H sequence data for Risø18 showed that sequence coverage around *PBF* was lacking (data not shown). Secondly, *PBF* could not be amplified by PCR with Risø18 DNA as template (Supplementary Figure S1). The presence of a large deletion around *LYS3* may also explain why we were unable to find any recombination close to this locus in the mapping lines derived from Risø18 (see 3.1).

3.3 Wheat PBF mutant grains have large embryos

To confirm that *PBF* is the gene responsible for controlling embryo size, we selected wheat cv. Cadenza TILLING lines (Krasileva et al., 2017) each affected in one of the wheat *PBF* homoeologs (TraesCS5A02G155900, TraesCS5B02G154100 and TraesCS5D02G61000) (Supplementary Table S2D). For *PBF-A*, no nonsense (premature termination codon) mutations were available and so we selected two lines (Cadenza1533 and Cadenza1807) with missense mutations (F48Y and T46I, respectively) in the DOF domain that are likely to be

313 deleterious (SIFT score = 0.00). For the B- and D-genome *PBF* genes, we selected lines with
314 nonsense mutations (Cadenza0903, Cadenza0904, respectively). Two triple mutant lines,
315 Line 1 and Line 2, were created that contained the same B- and D-genome nonsense
316 mutations but different A-genome missense mutations (Fig. 3A). Lines 2A and 2B contain
317 the same three *PBF* mutations but were derived from different F_1 grains.

318 Grains and embryos from the triple mutant plants, wild-type sibling plants and
319 Cadenza were compared (Fig. 2B and Fig. 3). This showed that the grains of all TILLING
320 lines (both wild-type and mutant) were more shrivelled than those of the non-mutant parent,
321 Cadenza grown at the same time (Fig. 2B). We assume that the shrivelled appearance of the
322 wild-type TILLING lines is due to mutations other than those in *PBF*.

323 The Line 1 triple mutant grains and embryos were indistinguishable from wild-type
324 sibling controls in appearance and weight (Fig. 2B and Fig. 3). We assume that the A-genome
325 missense mutant (F48Y) used to create Line 1 (Cadenza1533) had no deleterious impact on
326 PBF functionality. In contrast, both Lines 2A and 2B (using Cadenza1807 as the A genome
327 missense mutant, T46I) had increased absolute and relative embryo weights compared to
328 their wild type sibling controls and the grains were more shrivelled than their wild-type
329 sibling controls (Figs. 2B and Fig. 3). Thus, in wheat as in barley, PBF suppresses embryo
330 growth.

331

332 *3.4. Patterns of expression of PBF in barley*

333 The Barley eFB Browser (www.bar.utoronto.ca/efp_barley; Druka *et al*., 2006; Winter *et al*.,
334 2007) showed that *PBF* (HORVU5Hr1G048700) in cv. Morex is expressed in the developing
335 grains (caryopsis) (Fig. 4A). Data for the developing embryo are available for one time point
336 only: 22 DAF. *PBF* is expressed in the embryo at this time point, but at a low level compared
337 with its expression in the rest of the caryopsis. *PBF* expression in other tissue types is very

low or absent. A similar pattern of expression was seen for wheat cv. Azhurnaya. The wheat eFB Browser (www.bar.utoronto.ca/efp_wheat; Ramírez-González *et al.*, 2018; Winter *et al.*, 2007) showed that expression for all three homoeologs was strong in the endosperm and that, for the one stage of development for which there is data, there was no expression in the embryo.

To compare with the data available for developing Morex embryos, and to expand upon it, the pattern of expression of *PBF* in developing embryos of the *LYS3* wild type, Bomi was determined using RT-PCR (Fig. 4B). Expression of *PBF* was detected in embryos at 23 and 26 DAF but the expression at 17 and 20 DAF was low and comparable to that seen in some of the control reactions suggesting that it is expressed at higher levels after 20 DAF. In comparison, *PBF* expression was detected in the endosperm at all stages of development tested. To investigate the temporal and tissue specific patterns of expression of *PBF* in young developing barley grains of cv Bomi further, we used *in situ* hybridization (Fig. 4C). This showed that at 8 DAF, *PBF* is expressed intensely in starchy endosperm cells and sparsely in aleurone cells. In the embryo at 12 DAF, *PBF* is expressed in the scutellum, the coleoptile and the tip of the coleorhiza.

To determine when during grain development the suppression of embryo growth occurs, we compared the fresh weights of developing Bomi embryos with those of the *lys3* mutant Risø19 (Fig. 4D). This showed that *lys3* mutant embryos were significantly larger than wild type embryos only after 23 DAF (Student's *t*-test, p= 0.01). Thus, in wild-type embryos, the onset of *PBF* expression precedes the suppression of embryo growth.

3.5. In silico prediction of the downstream targets of wheat PBF

To understand how mutations in *PBF* effect both endosperm and embryo development and how these might be manipulated individually, we investigated the targets of this transcription

363 factor. PBF is known to activate or repress the transcription of a number of genes by binding
364 to specific sequences in their promotor regions called prolamin (TGTAAAG or CTTTACT)
365 or pyrimidine (CCTTTT or AAAAGG) boxes (Mena *et al.*, 1998; Mena *et al.*, 2002). In
366 developing barley endosperm, PBF target genes include B-hordein *(Hor2)* and trypsin-
367 inhibitor BTI-CMe *(Itr1)* and in germinating barley grains, cathepsin B-type protease *(Al21)*
368 and α-Amylase *(Amy2/32b)* are PBF targets (Mena *et al.*, 1998; Diaz *et al.*, 2002; Mena *et al.*,
369 2002). Nothing is yet known about PBF targets in developing barley embryos.

370 At present, the data and tools available for target gene analysis are more advanced for
371 wheat than for barley. Wheat transcription-factor targets can be predicted using a GENIE3
372 network, which was created using 850 diverse RNAseq samples from wheat (Ramírez-
373 González et al, 2018). Using the GENIE3 network to predict the downstream targets of the
374 wheat *PBF* homoeologs, we identified >450 gene targets. We found that these target genes
375 were enriched for a wide range of gene ontology (GO) annotations (Supplementary Fig. S2
376 and Fig. 5) but the two categories that were the most significantly enriched were organ
377 development and starch metabolism.

378 To investigate whether PBF directly regulates genes known to be involved in
379 controlling embryo size, we looked at the downstream targets of PBF for homologues of
380 three genes known to give rise to large embryo/small endosperm mutant phenotypes in rice
381 (*Oryza sativa* L.). The genes examined were *GIANT EMBRYO* (*GE*), encoding a cytochrome
382 P450 (CYP78A13) (Satoh and Omura, 1981; Nagasawa *et al.*, 2013); *BIGE1*, encoding a
383 MATE (Multidrug-And-Toxic-compound-Extrusion) type transporter protein (Suzuki *et al.*,
384 2015); and *LARGE EMBRYO*, encoding a C3HC4-type RING (Really Interesting New Gene)
385 finger protein of unknown function (Lee *et al.*, 2019). None of these three genes was a
386 predicted target of the wheat *PBF* ortholog.

To determine whether *PBF* participates in the same regulatory networks (or modules) as any of the other large-embryo genes (*GE, BIGE1* and *LARGE EMBRYO*), we used co-expression networks built using WGCNA (Ramírez-González *et al.*, 2018) (Supplementary Table S3). For the grain-specific network, the wheat homoeologs of *PBF* are in module 13, *BIGE1* in module 2 and *GE* in module 2 (B genome) or 9 (A and D genome). The *LARGE EMBRYO* homoeologs are in module 0, suggesting that the expression of these genes is invariant across all grain samples. Thus, for wheat it is likely that *PBF* regulates embryo size independently of the other three genes whilst *BIGE1* and the B genome of *GE* both operate within the same regulatory network.

PBF was so named because it was found to recognise a conserved cis-element, the prolamin box, in the promotor regions of prolamin seed storage proteins (Wu and Messing, 2012). In barley, PBF was also shown to bind to the pyrimidine box of genes expressed in aleurone cells during seed germination (Mena et al., 2002). To investigate whether PBF regulates genes involved in starch metabolism by binding to prolamin or pyrimidine boxes in their promotor regions, as is the case for other genes regulated by PBF, we analysed the promotor regions of 27 genes involved in starch synthesis in the endosperm (Supplementary Table S4). This showed that ten of the starch genes have a prolamin box within the 2-kbp upstream regulatory (promotor) region of at least one homoeologue, all 27 genes have a pyrimidine box and 11 genes were predicted by GENIE3 to be downstream targets of PBF. Ten starch genes have both prolamin and pyrimidine boxes of which only four were predicted targets. Seven additional genes were predicted targets of PBF and these lacked a prolamin box. Thus, for this limited data set, there is little or no correlation between the presence of the two known PBF regulatory sequences in starch gene promotors and their interactions with PBF as predicted from gene expression studies. Despite its name, PBF, like other transcription factors, probably interacts with a range of DNA regulatory sequences.

3. Discussion

We showed by genetic mapping that the *LYS3* gene encodes a transcription factor called Prolamin Binding Factor (PBF), and that one of the many effects of mutations in this gene is an increase in embryo size. All four *lys3* alleles have deleterious mutations in the *PBF* gene (Supplementary Table S1) and wheat TILLING mutants with mutations in *PBF* also have enlarged embryos (Fig. 3), confirming a role for *PBF* in determining organ size in Triticeae grains. The *LYS3* gene was also identified as *PBF* independently by Moehs *et al.* (2019). However, they studied the effect of *LYS3* on hordein content in one mutant, *lys3a* and they did not measure embryo size.

The severity of the effects of *lys3* alleles on embryo size varies, with Risø1508 and M1460 having larger embryos than either Risø18 or Risø19 (Fig. 1A). The reasons for this are currently unknown but the two least-severe mutants, Risø18 or Risø19 both lack the PBF protein (due to nonsense mutations), whilst the two mutants with the largest embryos, Risø1508 and M1460 have the same missense mutation (A173T) affecting their DOF domains. Proteins encoded by genes with missense mutations can be partially defective, rather than entirely absent or entirely defective. It is therefore possible that the missense mutation in Risø1508 and M1460 weakens or eliminates the ability of PBF to bind to DNA, but does not affect its interactions with other regulatory proteins. This hypothesis requires investigation.

Previous studies have shown that *PBF* is expressed in starchy endosperm and aleurone cells during barley grain development (Mena *et al*., 1998; Mena *et al*., 2002) and in the aleurone during germination (Mena *et al*., 2002) but no expression in developing embryos was detected using Northern blots (Mena *et al*., 2002). In this study, we confirmed the expression of *PBF* in starchy endosperm and aleurone cells in developing barley grains. We also showed using RT-PCR and *in situ* hybridization that *PBF* is expressed in developing

437 embryos (Fig. 4B, C). *PBF* expression in the embryo increased during grain development but
438 expression was detected prior to the time point when embryo growth in wild-type and mutant
439 grains diverged. These data suggest that control of embryo growth could be mediated by *PBF*
440 expression in the embryo and not indirectly via its expression in the endosperm. Thus, gene
441 regulation by PBF in embryos may involve a regulatory pathway that is independent from the
442 PBF pathway operating in the endosperm.

443 To understand the regulatory pathways downstream of PBF in developing Triticeae
444 grains, we used *in silico* prediction of target genes based on wheat RNAseq data (Ramírez-
445 González et al 2018; Harrington et al 2019). This indicated that PBF in developing grains is
446 involved in the regulation of a wide range of processes but the two most significant
447 categories were organ development and starch metabolism (mirroring its principal roles in
448 embryo growth and endosperm starch synthesis, respectively). Other genes known to cause
449 large embryos when mutated in rice were considered as potential downstream targets of PBF
450 but no such interactions were predicted from analysis of published RNAseq data. We also
451 looked for evidence of co-expression of *PBF* with these other genes in wheat grains. No such
452 evidence was found. Thus, the PBF transcriptional network that controls embryo growth
453 remains unknown. However, it must be noted that the expression data used for this *in silico*
454 analysis included three embryo samples only (Ramírez-González et al 2018). More RNAseq
455 data will need to be gathered for developing embryos to test these ideas further.

456 In maize, as in wheat and barley, there is a *PBF* gene that regulates starch metabolism
457 suggesting that the regulation of endosperm starch synthesis by *PBF*s may be a common
458 feature amongst grasses. Maize PBF controls the expression of four starch biosynthetic genes,
459 *Sh2* and *Bt2* that encode, respectively, the cytosolic large and small subunits of ADP glucose
460 pyrophosphorylase (AGPase); *SBEI* encoding starch branching enzyme I and *Su1* that
461 encodes isoamylase 1. The expression of both AGPase subunit genes was down-regulated in

462 maize *PBF* RNAi lines whilst the other two genes were up-regulated (Zhang *et al.*, 2016).
463 Orthologs of all four genes in wheat are also predicted target genes of wheat *PBF*
464 (Supplementary Table S4) suggesting that the *PBF* transcription factors in maize and the
465 Triticeae species interact with similar sets of downstream genes. However, there is no
466 information to date to suggest that *PBF* controls embryo size in maize.

467 Large-embryo size in cereals is associated with increased nutritional value because
468 embryos are rich in protein, vitamins, oil and non-starch carbohydrates. There is also
469 evidence of human health benefits from eating large-embryo rice grain products (Zhang *et*
470 *al.*, 2005; Lee *et al.*, 2016; Jung *et al.*, 2017). As wholegrain products are becoming
471 increasingly popular for human consumption, cereal grains with large embryos offer new
472 opportunities for cereal grain improvement. We have shown that barley and wheat with
473 mutations in *PBF* have enlarged embryos but unfortunately, they also have reduced grain size
474 due to endosperm defects. However, the possibility that the regulatory networks controlled by
475 PBF in the endosperm and embryo are independent provides an opportunity to explore their
476 independent manipulation. Further work is in progress to pursue this idea.
477

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC), UK [grant numbers BB/L023156/1 and BB/P016855/1] and an Anniversary Future Leader Fellowship BB/M014045/1 to PB. CUS thanks TETFUND Nigeria for a PhD training grant. JD and MK were supported by ERDF project "Plants as a tool for sustainable global development" (No. CZ.02.1.01/0.0/0.0/16_019/0000827). We thank Zdeňka Dubská, Romana Šperková, Jan Vrána and Jitka Weiserová for sample preparation, chromosome sorting and DNA amplification.

AUTHOR CONTRIBUTIONS

KT designed the project, with advice from CU, PB and BBHW. NS generated and analysed the mapping populations. MK managed chromosome sorting and performed purity checks under the supervision of JD. Bioinformatic analysis of the sequenced chromosome was done by BS with help and advice from CU and BBHW. BOL carried out the bulk of the other experimental work on barley and BOL and PB carried out the wheat *PBF* target gene predictions. TC and MC were responsible for the production and analysis of the wheat TILLING mutants. CUS performed the in situ analysis, and SD supervised this work. BOL and KT wrote the paper together and all authors proof-read the paper and provided feedback. Authors, apart from the first and last, are listed alphabetically.

497 **References**

498 Aastrup, S., 1983. Selection and characterization of low beta-glucan mutants from barley.
499 Carlsberg Research Communications 48, 307-316.

500 Alabdullah, A.K., Borrill, P., Martin, A.C., Ramírez-González, R.H., Hassani-Pak, K., Uauy,
501 C., Shaw, P., Moore, G. 2019. A co-expression network in hexaploid wheat reveals
502 mostly balanced expression and lack of significant gene loss of homeologous meiotic
503 genes upon polyploidization. Frontiers in Plant Science 10, 1325.

504 Allison, M.J. 1978. Amylase activity and endosperm hardness of high lysine barleys. Journal
505 of the Institute of Brewing, 84, 231-232.

506 Brandt, A. 1976. Endosperm protein formation during kernel development of wild type and a
507 high-lysine barley mutant. Cereal Chem. 53, 890.

508 Cook, F., Hughes, N., Nibau, C., Orman-Ligeza, B., Schatlowski, N., Uauy C., Trafford, K.
509 2018. Barley *lys3* mutants are unique amongst shrunken-endosperm mutants in having
510 abnormally large embryos. Journal of Cereal Science 82, 16–24.

511 Deggerdal, A., Klemsdal, S.S., Olsen, O.A., 1986. The effect of the high-lysine genes of the
512 barley mutants Risø1508 and 527 on embryo development. Physiologia Plantarum 68,
513 410-418.

514 Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martínez, M., Isabel-La Moneda, I., Carbonero,
515 P. 2002. The GAMYB protein from barley interacts with the DOF transcription factor
516 BPBF and activates endosperm-specific genes during seed development. Plant Journal
517 29, 453-64.

518 Doll, H. 1976. Genetic studies of high lysine barley mutants. In: Gaul, H. (Ed.) Barley
519 Genetics III. Proceedings of the Third International Barley Genetics Symposium.
520 Verlag Karl Thiemig, Munich, pp. 542-546.

521 Doll, H., Koie, B., Eggum, B.O., 1974. Induced high lysine mutants in barley. Radiation
522 Botany 14, 73-80.
523 Dracatos, P.M., Bartoš, J., Elmansour, H., Singh, D., Karafiátová, M., Zhang, P., Steuernagel,
524 B., Svačina, R., Cobbin, J.C.A., Clark, B., Hoxha, S., Khatkar, M.S., Doležel, J.,
525 Wulff, B.B., Park, R.F., 2019. The coiled-coil NLR *Rph1*, confers leaf rust resistance
526 in barley cultivar Sudan. Plant Physiology 179, 1362-1372.
527 Drea S, Corsar J, Crawford B, Shaw P, Dolan L, and Doonan JH. 2005. A streamlined
528 method for systematic, high-resolution *in situ* analysis of mRNA distribution in
529 plants. Plant Methods 1, 8.
530 Druka, A. Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., Schreiber, A.,
531 Wise, R., Close, T., Kleinhofs, A., Graner, A., Schulman, A., Langridge, P., Sato, K.,
532 Hayes, P., McNicol, J., Marshall, D., Waugh, R. 2006. An atlas of gene expression
533 from seed to seed through barley development. Functional and Integrative Genomics
534 6, 202-211.
535 Fulton, T.M, Chunwongse, J., Tanksley, S.D. 1995. Microprep protocol for extraction of
536 DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter 13,
537 207–209.
538 Frankowiak, J. 1997. Revised linkage maps for morphological markers in barley, *Hordeum*
539 *vulgare*. Barley Genetics Newletter 26: 9-21.
540 Gabert, V.M., Brunsgaard, G., Eggum, B.O., Jensen, J. 1995. Protein quality and digestibility
541 of new high-lysine barley varieties in growing rats. Plant Foods for Human Nutrition
542 48, 169-179.
543 Gabert, V.M., Jorgensen, H., Brunsgaard, G., Eggum, B.O., Jensen, J. 1996. The nutritional
544 value of new high-lysine barley varieties determined with rats and young pigs.
545 Canadian Journal of Animal Science 76, 443-450.

546 Harrington, S.A., Backhaus, A.E., Singh, A., Hassani-Pak, K., Uauy, C. 2019. Validation and
547 characterisation of a wheat GENIE3 network using an independent RNA-Seq dataset.
548 doi: http://dx.doi.org/10.1101/684183.
549 Ingversen, J., Koie, B., Doll, H., 1973. Induced seed protein mutant of barley. Experientia 29,
550 1151-1152.
551 Jung, W.Y., Kim, S.G., Lee, J.S., Kim, H.K., Son, B.G., Kim, J.W., Suh, J.W. 2017. Effect of
552 feeding high gamma-aminobutyric acid-containing giant embryo black sticky rice
553 (*Oryza sativa* L.) on anxiety-related behavior of C57BL/6 Mice. J Med Food 20, 777-
554 781.
555 Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L.,
556 Simmonds, J., Ramírez-González, R.H., Wang, X., Borrill, P., Fosker, C., Ayling, S.,
557 Phillips, A.L. Uauy, C., Dubcovsky, J. 2017. Uncovering hidden variation in
558 polyploid wheat. PNAS 114 (6) E913-E921, doi.org/10.1073/pnas.1619268114
559 Kubaláková, M., Valárik, M., Bartoš, J., Vrána, J., Číhalíková, J., Molnár-Láng, M.,
560 Doležel, J. 2003. Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow
561 cytometry. Genome 46: 893-905.
562 Lee, G., Piao, R., Lee, Y. *et al.* 2019. Identification and characterization of LARGE
563 EMBRYO, a new gene controlling embryo size in rice (*Oryza sativa* L.). Rice 12, 22.
564 Li, H., Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
565 transform. Bioinformatics 25, 1754-1760.
566 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,
567 G., Durbin, R. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence
568 Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

569 Lysák, M.A., Číhalíková, J., Kubaláková, M., Šimková, H., Künzel, G., Doležel, J. 1999.
570 Flow karyotyping and sorting of mitotic chromosomes of barley (*Hordeum vulgare*
571 L.). Chrom. Res. 7: 431-444.
572 Mena, M., Vicente-Carbajosa, J., Schmidt, R.J., Carbonero, P. 1998. An endosperm-specific
573 DOF protein from barley, highly conserved in wheat, binds to and activates
574 transcription from the prolamin-box of a native B-hordein promoter in barley
575 endosperm. Plant Journal 16, 53–62.
576 Mena, M., Cejudo, F.J., Isabel-Lamoneda, I., Carbonero, P., 2002. A role for the DOF
577 transcription factor BPBF in the regulation of gibberellin-responsive genes in barley
578 aleurone. Plant Physiology 130, 111–119.
579 Moehs, C.P., Austill, W.J., Holm, A., Large, T.A.G., Loeffler, D., Mullenberg, J., Schnable,
580 P.S., Skinner, W., Boxtel, J., Wu, L., McGuire, C. 2019. Development of decreased-
581 gluten wheat enabled by determination of the genetic basis of *lys3a* barley. Plant
582 Physiology 179, 1692–1703.
583 Mossberg, R., 1969. Evaluation of protein quality and quantity by dye-binding capacity: a
584 tool in plant breeding. In: New approaches to breeding for improved plant protein.
585 International Atomic Energy Agency, Vienna, pp. 151-161.
586 Munck, L., 1992. The case of high-lysine barley breeding. In: Shewry, P.R. (Ed.), Barley:
587 Genetics, Biochemistry, Molecular Biology and Biotechnology. CAB International,
588 Wallingford Oxon, pp. 573–601.
589 Munck, L., Jespersen, B.M. 2009. The multiple uses of barley endosperm mutants in plant
590 breeding for quality and for revealing functionality in nutrition and food technology.
591 In: Shu, Q.Y. (Ed), Induced plant mutations in the genomics era. Proceedings of an
592 International Joint Food and Agriculture Organisation / International Atomic Energy

593 Agency Symposium, 2008. Food and Agriculture Organisation of the United Nations,
594 Rome, pp. 182-186.
595 Munck, L. 1972. High lysine barley - a summary of the present research development in
596 Sweden. Barley Genetics Newsletter 2, 54-60.
597 Mortensen, H.P., Madsen, A., Hall, D.D., Munck, L., Bang-Olsen, K., Stilling, B. 1988.
598 Protein and amino acids for growing pigs. High-lysine barley (in Danish). Beretning
599 fra Statens Husdyrbrugsforsøg: 29 pp.
600 Nagasawa N, Hibara K, Heppard EP, Vander Velden KA, Luck S, Beatty M, Nagato Y, Sakai
601 H. 2013. GIANT EMBRYO encodes CYP78A13, required for proper size balance
602 between embryo and endosperm in rice. Plant J. 75, 592-605.
603 Olsen, O.A., Krekling, T., Nissen, O., Ayaz, M., 1984. An ultrastructural approach to the
604 problem of barley high lysine gene action. In: Cereal grain protein improvement.
605 International Atomic Energy Agency, Vienna, pp. 241-257.
606 Opanowicz M, Hands P, Betts D, Parker ML, Toole GA, Mills EN, Doonan JH, Drea S.
607 2011. Endosperm development in *Brachypodium distachyon*. Journal of Experimental
608 Botany 62, 735-748.
609 Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L.,
610 Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S.J., Cory, A.T.,
611 Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout,
612 C.J., Chalhoub, B., Mayer, K.F.X., Benhamed, M., Latrasse, D., Bendahmane, A.
613 International Wheat Genome Sequencing Consortium, Wulff, B.B.H., Appels, R.,
614 Tiwari, V., Datla, R., Choulet, F., Pozniak, C.J., Provart, N.J., Sharpe, A.G., Paux, E.,
615 Spannagl, M., Bräutigam, A., Uauy, C. 2018. The transcriptional landscape of
616 polyploid wheat. Science 17, 361.

617 Satoh, H. and Omura, T. 1981. New endosperm mutations induced by chemical mutagens in
618 rice. Japanese Journal of Breeding 131, 316-326.
619 Sanchez-Martin, J., Steuernagel, B., Ghosh, S., Herren, G., Hurni, S., Adamski, N., Vrána, J.,
620 Kubaláková, M., Krattinger, S.G., Wicker, T., Doležel, J., Keller, B., Wulff, B.B.H.
621 2016. Rapid gene isolation in barley and wheat by mutant chromosome sequencing.
622 Genome Biol 17, 221.
623 Šimková, H., Svensson, J.T., Condamine, P., Hřibová, E., Suchánková, P., Bhat, P.R., Bartoš,
624 J., Šafář, J., Close, T.J., Doležel, J. 2008. Coupling amplified DNA from flow-sorted
625 chromosomes to high-density SNP mapping in barley. BMC Genomics 9, 294.
626 Steuernagel, B., Vrána, J., Karafiátová, M., Wulff, B.B.H., Doležel, J. 2017. Rapid gene
627 isolation using MutChromSeq. Methods in Molecular Biology 1659, 231-243.
628 Suzuki, M., Sato, Y., Wu, S., Kang, B-H., McCarty, D.R. 2015. Conserved functions of the
629 MATE transporter BIG EMBRYO1 in regulation of lateral organ size and initiation
630 rate. Plant Cell 27, 2288-2300.
631 Tallberg, A., 1977. Amino acid composition in endosperm and embryo of a barley variety
632 and its high lysine mutant. Hereditas 87, 43-46.
633 Trafford, K., and Fincher, G.B. 2014. Barley grain carbohydrates: starch and cell walls. In: P.
634 R. Shewry and S. E Ullrich (Eds.) Barley: chemistry and technology. Minnesota, MN:
635 AACC International press, pp. 71-95.
636 Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., Provart, N.J. 2007. An
637 “electronic fluorescent pictograph” browser for exploring and analyzing large-scale
638 biological data sets. PLoS ONE 2(8): e718.
639 https://doi.org/10.1371/journal.pone.0000718
640 Wu, Y., Messing, J. 2016. Rapid divergence of prolamin gene promotors of maize after gene
641 amplification and dispersal. Genetics 192, 507-519.

642 Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A. 2010. Gene ontology analysis for
643 RNA-seq: accounting for selection bias. Genome Biology 11, R14.
644 Zhang, Z., Zheng, X., Yang, J., Messing, J., Wu, Y. 2016. Maize endosperm-specific
645 transcription factors *O2* and *PBF* network the regulation of protein and starch
646 synthesis. PNAS 113, 10842-10847.
647 Zhang, L.L., Hu, P.S., Tang, S.Q., Zhao, H.J., Wu, D.X. 2005. Comparative studies on major
648 nutritional components of rice with a giant embryo and a normal embryo. Journal of
649 Food Biochemistry 29, 653-661.

Table 1. Fine-mapping *LYS3* in barley.

The genotypes of genes close to *LYS3* are shown: A = wild type (Morex), B = Bomi and *lys3* mutant. The embryo-size phenotype was assessed visually: WT = wild type (normal-sized embryos), MUT = mutant (large embryos). The two genes defining the *LYS3* region are shown in bold. The *PBF* gene, HORVU5Hr1G048700 is underlined.

Line	*lys3* parent	Visual embryo phenotype	Number of plants analysed	Genotype (Gene ID = HORVU5Hr1G04xxxx)								
				7540	7890	**8110**	8380	8630	8700	**8950**	9180	9330
Morex		WT (control)		A	A	A	A	A	A	A	A	A
Line 92	Risø19	WT	2	B	B	A	A	A	A	A	A	A
Line 93	Risø1508	WT	3	B	B	B	A	A	A	A	A	A
Line 123	Risø1508	WT	2	B	B	B	A	A	A	A	A	A
Line 94	Risø19	MUT	1	B	B	B	B	B	B	A	A	A
Line 125	Risø1508	MUT	1	B	B	B	B	B	B	B	B	A
Line 115	Risø1508	WT	1	A	A	A	A	A	A	A	B	B
Line 121	Risø1508	WT	8	A	A	A	A	A	A	B	B	B
Line 122	Risø19	WT	6	A	A	A	A	A	A	B	B	B
Line 124	Risø1508	WT	5	A	A	A	A	A	A	B	B	B
Line 116	Risø19	MUT	10	A	A	B	B	B	B	B	B	B
Line 117	Risø19	MUT	3	A	A	B	B	B	B	B	B	B
Risø19		MUT (control)		B	B	B	B	B	B	B	B	B
Risø1508		MUT (control)		B	B	B	B	B	B	B	B	B
Bomi		WT		B	B	B	B	B	B	B	B	B

Figure 1. Map-based cloning of the gene controlling embryo size.

A. Representative images of grains and embryos from the lines described in Table 1 are shown. Top row = dry, mature grains. Middle row = excised embryos, abaxial surface uppermost. Lower row = excised embryos, adaxial surface uppermost. The scale bars are 5 mm for grains and 2 mm for embryos.

B. Relative embryo weight and β-amylase activity were determined for selected lines shown in Table 1. The values for relative embryo weight are the means ± SE for 5 plants and 10 grains were measured per plant. The values for β-amylase activity are the means ± SE for three plants. One grain from the middle of the ear was assayed per plant. All assays were triplicated.

C. The gene controlling embryo size maps to a region on chromosome arm 5HL. The two markers/genes flanking this region are shown in bold. The *LYS3*/*PBF* gene controlling embryo size was identified as HORVU5Hr1G048700. Gene names are abbreviated: xxxx = HORVU5Hr1G04xxxx.

D. An alignment of PBF amino acid sequences showing the positions of the DOF domain and the SNP mutations. (Risø18 lacks *PBF* and is therefore not shown).

Figure 2. *PBF* affects embryo weight in barley and wheat.

The absolute and relative dry weights of the embryo and non-embryo grain parts of wild-types (black bars) and mutants (white bars) are shown. Mean values ±SE are for 11 to 13 (barley) or 25 (wheat) replicate biological samples. Each sample contained 10 grains. Values for mutants that are significantly different from the values for their relevant wild-type controls (Students t-test, $p<0.05$) are indicated by an asterisk.

A. Barley mutants. Risø1508, Risø18 and Risø19 and M1460 are *lys3* mutants. Bomi is the wild-type parent of the three Risø mutants.

B. Wheat TILLING mutants. Lines are as described in Figure 3.

Figure 3. Wheat *PBF TILLING* mutants.

Two wheat *PBF* triple mutant (aabbdd) lines were created by combining single A, B and D-homoeolog TILLING mutations in a single plant. Line 1 = Cadenza1533 x Cadenza0903 x Cadenza0904 (A missense x B nonsense x D nonsense). Line 2 = Cadenza1807 x Cadenza0903 x Cadenza0904 (A missense x B nonsense x D nonsense). Line 2A and 2B were derived from different F_1 grains.

A. A diagram of the *PBF* A, B and D gene structures. The exons are shown as bars and intron as lines. The coding regions are black and the DNA-binding DOF domain is red. The positions of the mutations are indicated by arrows. The two missense mutations are separated by 5 bp. C = Cadenza, NS = nonsense, MS = missense.

B. Representative wild type (AABBDD) and mutant (aabbdd) F_3 grains, and grains of the non-mutant, parental cultivar Cadenza are shown. All five grains of each type are from a single plant. All F_2 plants were grown as a single batch.

C. Representative wild type (AABBDD) and mutant (aabbdd) embryos are shown. Embryos are from F_3 grains, as shown in Figure 3B.

Figure 4. Patterns of expression of *PBF/LYS3* in barley.

A. Tissue-specific expression according to the Barley eFP Browser 2.0 (www.bar.utoronto). Samples are from barley cv. Morex. Caryopsis (no Emb) means caryopsis without embryo.

B. Temporal and tissue-specific patterns of *LYS3*/*PBF* expression in developing grains of wild type (Bomi) assessed by RT–PCR. Each cDNA sample was from a pool of 10-75 tissue samples each from an individual grain. Samples were collected from at least

five spikes, each from a different plant. Amplicons were visualized on 1% agarose gels stained with SYBR Safe (Invitrogen, UK). Numbers above panels are DAF. *PBF* = *Prolamin Binding Factor* (HORVU5Hr1G048700). *ACT* = *ACTIN*.(HORVU1Hr1G047440), a constitutively expressed gene. Control reactions varied from the test reactions as follows. C1: contained cDNA from the barley *PBF* deletion mutant Risø18 at 26 DAF. No product was observed showing that the primers used were specific for *PBF*. C2: no DNase, no reverse transcriptase (RT). C3: no DNase. C4: no RT. C5: contained water used for PCR instead of cDNA.

C. *In situ* localization of *PBF* in developing wild type (Bomi) grains at 8 and 12 DAF. (**i**) Longitudinal section of 8 DAF grain stained with antisense probe. *PBF* is expressed mostly in the starchy endosperm (En) cells. (**ii**) Longitudinal section of 12 DAF grain (including the embryo) stained with an antisense probe. *PBF* is expressed in the scutellum (Sc), coleoptile (Co) and coleorhiza tip (Cr). *PBF* is not expressed in the radicle (Ra). (**iii**) Longitudinal section of 12 DAF as in (ii) but stained with a sense probe (negative control). Scale bars are 0.5 mm (i) and 20 mm (ii and iii).

D. Embryo fresh weight in developing grains of wild type (Bomi) and *lys3* mutant Risø19. Embryos were extracted from developing grains and immediately weighed. Values are means ± SE ($n > 4$) for 10 embryos extracted from the middle of at least three spikes. Values are significantly different ($p<0.05$, Tukey's HSD after one-way Anova, mutant compared to wild type) for 26, 29 and 32 DAF (as indicated by asterisk). Error bars are included for all data and if not visible are smaller than the marker.

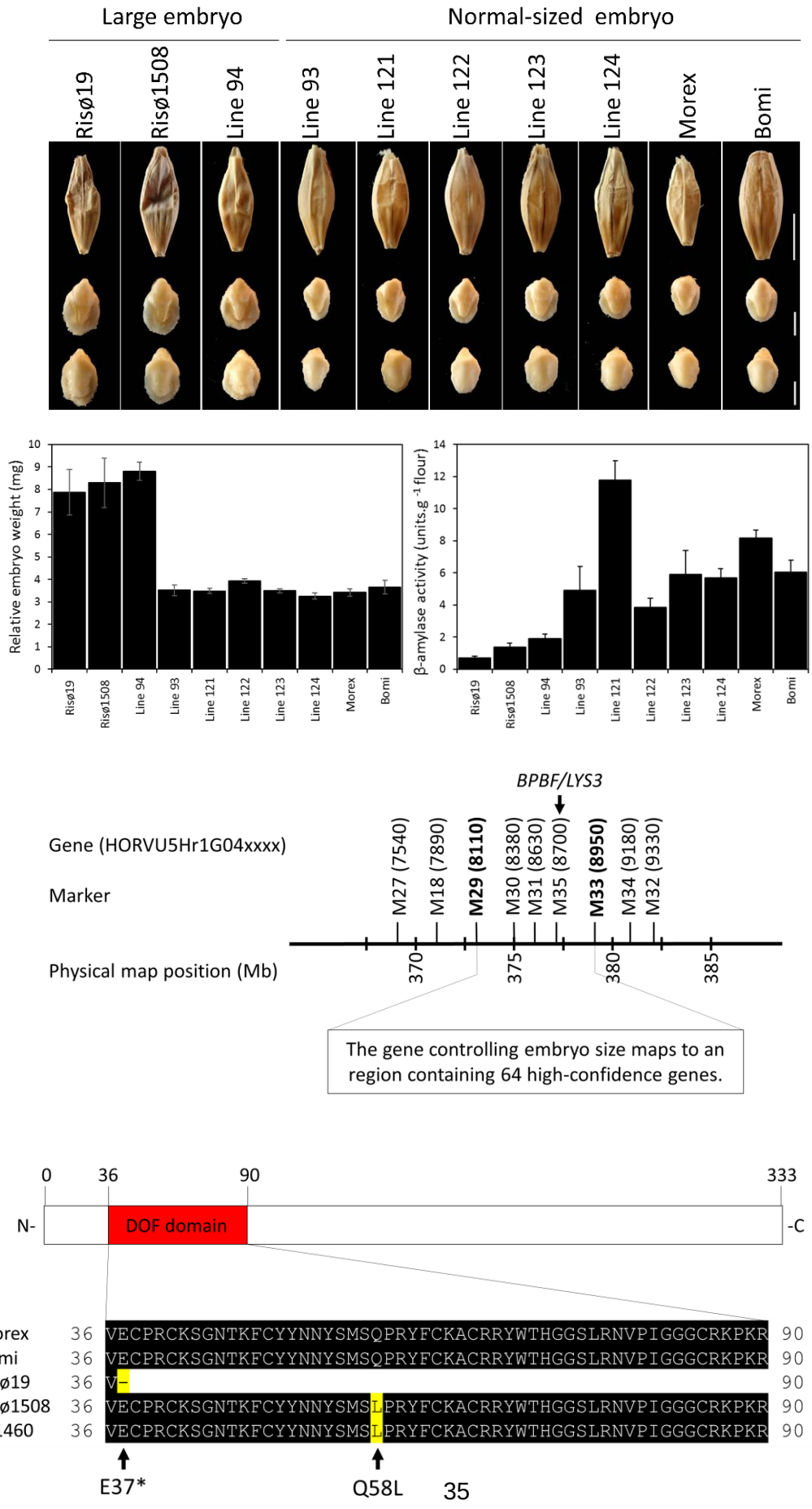
Figure 5. **Downstream targets of wheat *PBF*.**

The downstream target genes of *PBF* were identified using a GENIE3 network constructed using 850 RNA-seq samples (Ramírez-González et al., 2018).

A. A comparison of the target genes of each of the wheat *PBF* homoeologs. The number of predicted target genes common to all three homoeologs is 226.

B. The 226 common target genes were analysed for evidence of enrichment with respect to molecular function. The top-ranking functional categories (adjusted p value <0.05) are presented with the significance of the enrichment [−log(adjP-value)]. In brackets are the numbers of genes detected in each category with respect to all GO terms annotated so far for wheat RefSeqv1 gene models (Ramírez-González *et al.*, 2018). GO terms related to starch metabolism are in orange, organ development in yellow, and others in grey.

Figure 1



763 **Figure 2.**

764

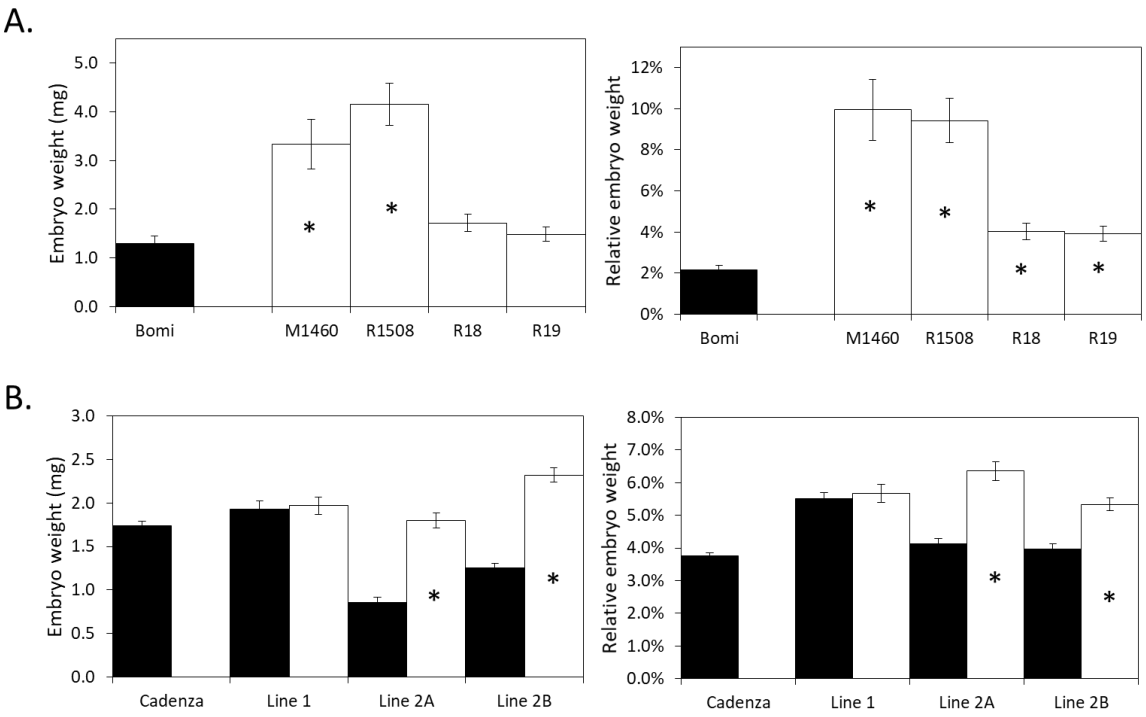
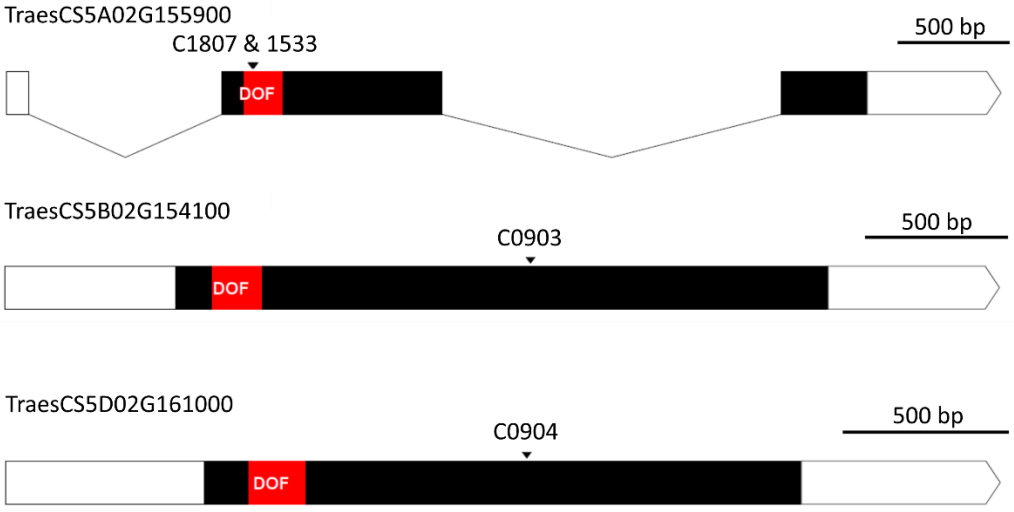
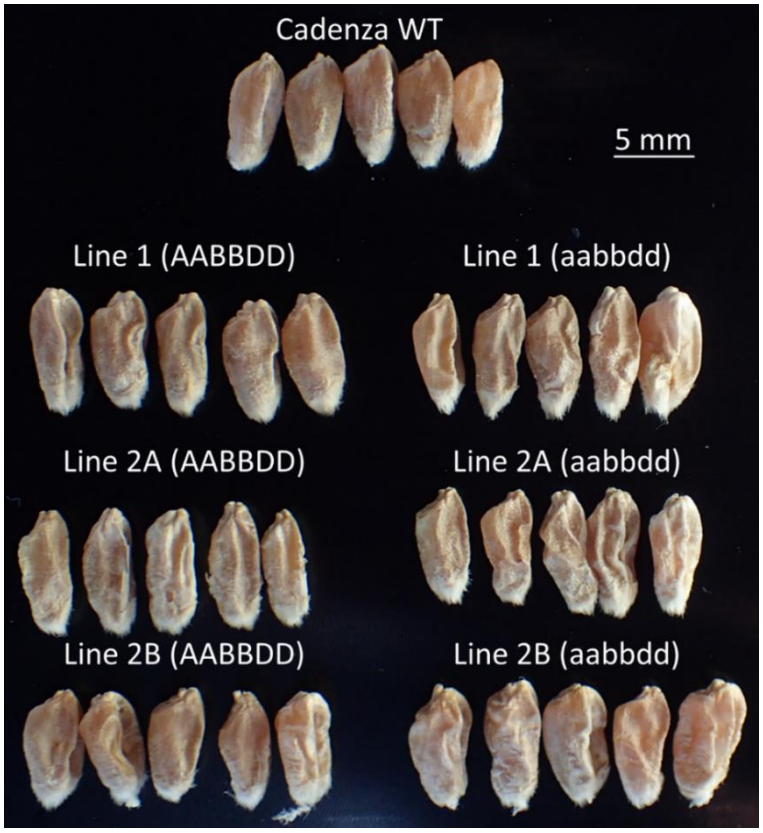


Figure 3.

A.



B.



C.

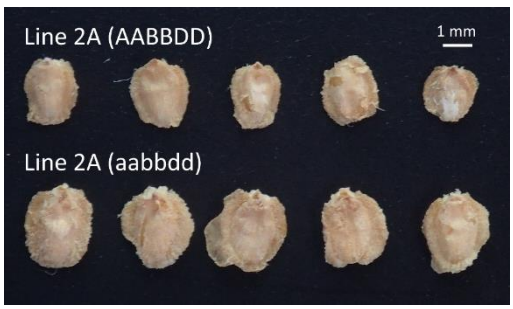
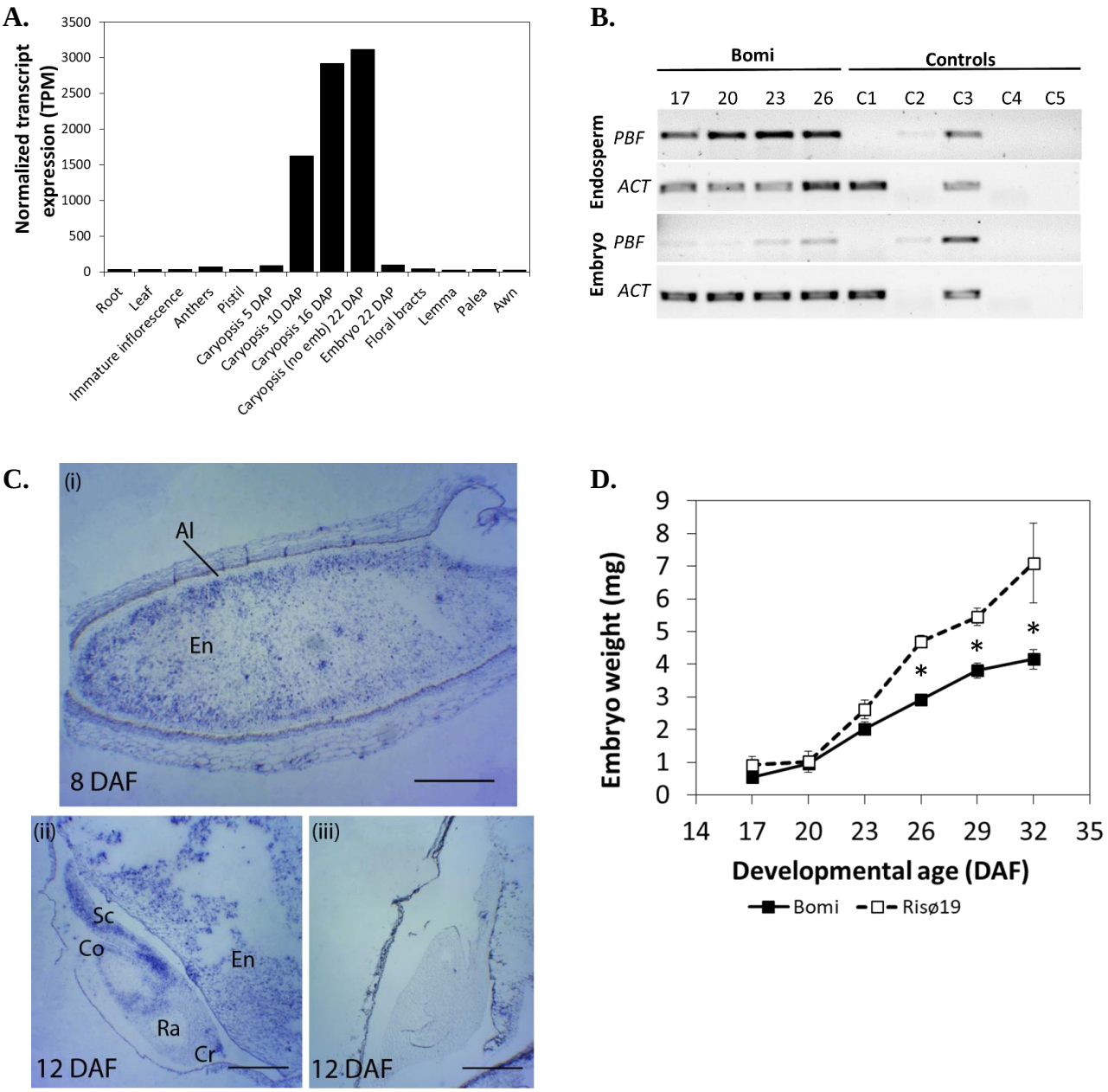
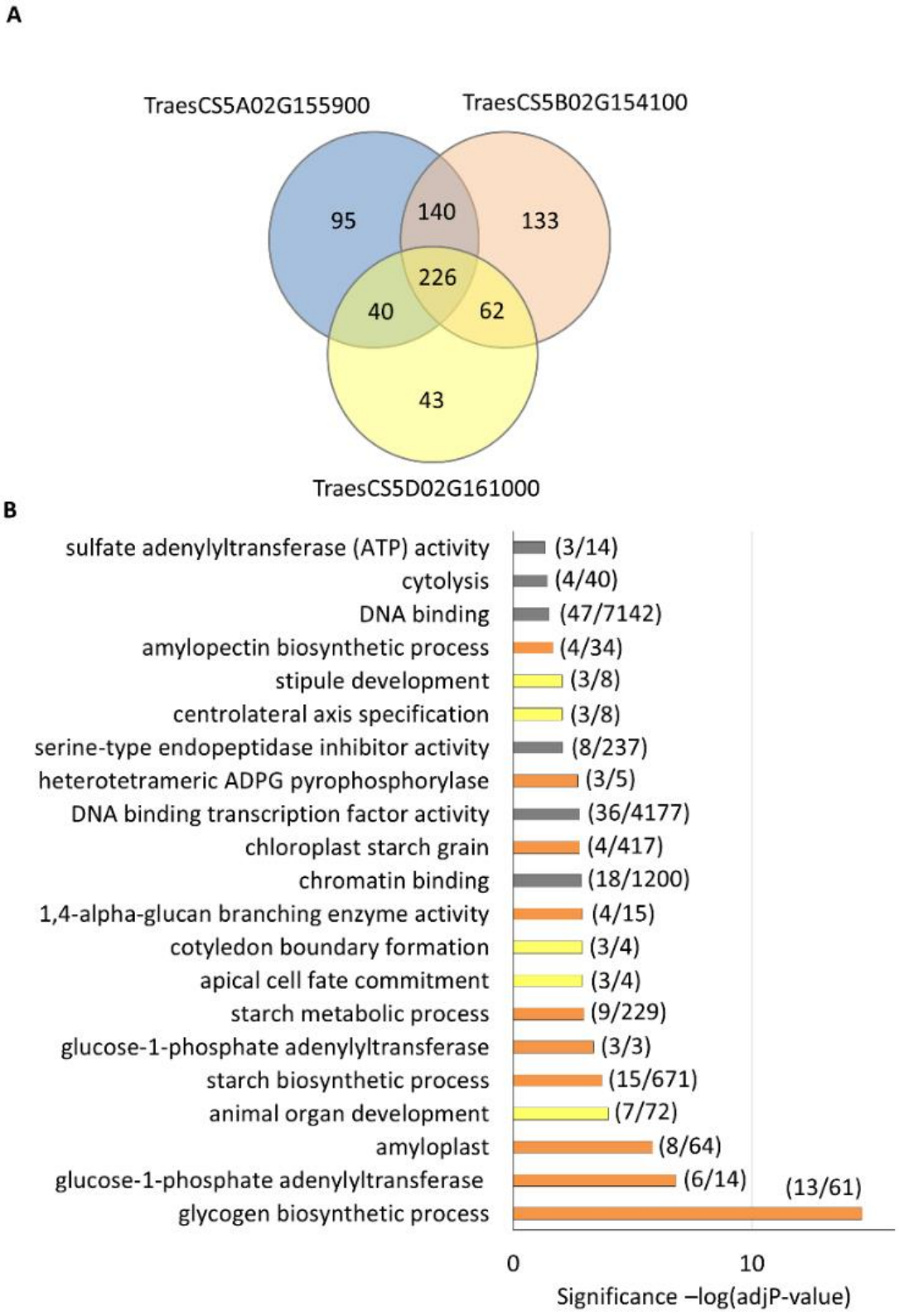


Figure 4.



796 **Figure 5. Downstream targets of *PBF* in wheat**
797



Supplementary Table S1. The *lys3* mutants of barley.

Comparison of the four known *lys3* mutant lines: their origins and the effects of the mutations on *PBF* gene and protein. See also Fig. 1D.

Mutant	**Wild type**	**Mutagen**	**Original selection criteria**	**Reference**	Change to cDNA	position	Effect	Change to protein
Risø18	Bomi	Na azide	Unknown	Munck *et al.*, 1992	-	-	Deletion	-
Risø19	Bomi	Na azide	Unknown	Munck *et al.*, 1993	G/T	109	Nonsense/Stop	E/*
Risø1508	Bomi	Ethylenimine	High lysine	Tallberg *et al.*, 1973	A/T	173	Missense	Q/L
M1460	Minerva	Na azide	Low β-glucan	Aastrup *et al.*, 1983	A/T	173	Missense	Q/L

Supplementary Table S2. Primers and assay conditions.

A. KASP primers for barley genotyping.

Primers were preceded by the VIC sequence (GAAGGTCGGAGTCAACGGATT) (Morex and *LYS3*) or by FAM sequence (GAAGGTGACCAAGTTCATGCT) (Bomi and *lys3*). The discriminatory bases are small case. The ratio of FAM/VIC/Common primers used was 12/12/30, except for the genotyping primers for HORVU5Hr1G048380, which were in the ratio 12/6/30. The KASP reaction (in a final volume of 5 µl) contained: 2.5 µl 2xKASP reagent, 0.07 µl of primer mix and 2.5 µl of gDNA template (1 ng/µl). The PCR conditions were: 94 °C for 15 min, 10 cycles of (94 °C for 20 s, 65 °C for 1 min and -0.8 °C/cycle to 57 °C), 10 cycles of (94 °C for 20 s, 57 °C for 1 min) and then reprise 5 to 10 cycles of (94 °C for 20 s, 57 °C for 1 min). Abbreviated gene names (xxxx) are as in Table 1 (i.e. HORVU5Hr1G04xxxx).

Gene	Morex (5'-3')	Bomi (5'-3')	Common (5'-3')
7540	TGCCAGTGAGTTTGTCCAGt	TGCCAGTGAGTTTGTCCAGg	GACATCCGAGCTCTTTGGAC
7890	TCAGACTGAGCGGCCTTGa	TCAGACTGAGCGGCCTTGc	TCATCGTGTCGGTGTCAAAT
8110	GTGCTACTGCCTTGGCg	GTGCTACTGCCTTGGCa	ATGGACGACACCTGGACAC
8380	ATGAGCATTCATGGTGCAg	ATGAGCATTCATGGTGCAc	AGCTCCTTCCCACATAGCTC
8630	ACAGCATCAACACGGCt	ACAGCATCAACACGGCc	TCAGCTTCGATCAAGAGTTTG
8700	GGATCTACAAGAGTGAAGATCGc	GGATCTACAAGAGTGAAGATCGg	GGTGAGCTTCTTCTTCCTTTGA
8950	ATCATTGACGCATCGCTAa	ATCATTGACGCATCGCTAg	TGGACATGAGTTGCTGGACT
9180	TTAATGTCTCATCAAGGCAATGa	TTAATGTCTCATCAAGGCAATGc	GGCACGCCGATAAAAAGAT
9330	AGCAGCGCAAGAATAATTTc	AGCAGCGCAAGAATAATTTt	GGCGTTGTACTGCAACAAGT

B. **Primers used to amplify barley *PBF* genomic sequence.**

The PCR reaction (in a final volume of 20 µl) contained: 1 µl of gDNA template (2 ng), 0.5 µl of 10 mM forward primer, 0.5 µl of 10 mM reverse primer, 0.6 µl of 10 mM dNTPs, 2 µl

of 10 x PCR buffer, 0.2 µl of Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, UK) and water. The PCR conditions were: 98 °C for 30 s, 35 cycles of (98 °C for 10 s, 58 °C for 60 s and 72 °C for 20 s) and then 72 °C for 5 min.

Primer name	Sequence (5'-3')	Gene
BPBF_F1	ATGGAGGAAGTGTTTTCGTC	HORVU5Hr1G048700
BPBF_R1	TTACATCAGGGAGGTGCTG	HORVU5Hr1G048700

C. **Barley *PBF* (HORVU5Hr1G048700) allele-specific KASP primers.**

Primers were preceded by the VIC sequence (GAAGGTCGGAGTCAACGGATT) (*lys3*) or by FAM sequence (GAAGGTGACCAAGTTCATGCT) (Bomi). The discriminatory bases are small case. The ratio of FAM/VIC/Common primers used was 12/12/30. The PCR reactions and conditions were as for (A). Note that no primers are given for Risø18 because the genomic region containing this allele is deleted.

Allele	Wild type (5'-3')	Mutant (5'-3')	Common (5'-3')
Risø1508	GCAGAAGTAGCGCGGCt	GCAGAAGTAGCGCGGCa	CGTCCAACTCCAAGAGCAAG
M1460	GCAGAAGTAGCGCGGCt	GCAGAAGTAGCGCGGCa	CGTCCAACTCCAAGAGCAAG
Risø19	TGCACCGAGGGCACTc	TGCACCGAGGGCACTa	CGTCCAACTCCAAGAGCAAG
Risø18	N/A	N/A	N/A

D. Wheat TILLING mutants and KASP primers.

Primers were designed for KASP assays and were preceded by the VIC sequence (GAAGGTCGGAGTCAACGGATT) (wild-type) or by FAM sequence (GAAGGTGACCAAGTTCATGCT) (mutant). The discriminatory bases are small case. The ratio of FAM/VIC/Common primers used was 12:12:15 (Cadenza1807), 6:12:30 (Cadenza1533) or 12:12:30 (Cadenza0903 and Cadenza0904). The KASP reaction was carried out as described by the manufacturer (https://www.lgcgroup.com).

Chr 5	Mutant ID	Change to gDNA	position	Effect	Change to protein	position	Wild type primer	Mutant primer	Common primer
A	Cadenza1807	aCc/aTc	137	Missense	T/I	46	CGGTGCAAGTCTGGCAACAc	CGGTGCAAGTCTGGCAACAt	GGCGTCGGAGGTCCCCGG
A	Cadenza1533	tTc/tAc	143	Missense	F/Y	48	AGTCTGGCAACACCAAGTt	AGTCTGGCAACACCAAGTa	GCAGGCCTTGCAGAAGTAG
B	Cadenza0903	Cag/Tag	1844	Stop	Q/*	182	CACCATAGTAGTAGCCATTGTTCTg	CACCATAGTAGTAGCCATTGTTCTa	GAGGGCTTCTTGATGGCAA
D	Cadenza0904	tgG/tgA	1574	Stop	W/*	255	GTGATGGGGCTCCAATGg	GTGATGGGGCTCCAATGa	GCCGCTGTCGTTATTGTTG

836

837 **E. Primers for analysis of gene expression by RT-PCR.**

838 The RT-PCR reaction (in a final volume of 20 µl) contained: 1 µl of cDNA template, 0.5 µl

839 of 10 mM forward primer, 0.5 µl of 10 mM reverse primer, 0.6 µl of 10 mM dNTPs, 2 µl of

840 10 x PCR buffer, 0.2 µl of FastStart *Taq* DNA polymerase (Roche, 04659163103) and water.

841 The PCR conditions were: 95 °C for 3 min, 35 cycles of (95 °C for 30 s, 58 °C for 30 s and

842 72 °C for 30 s or 1 min) and then 72 °C for 5 min. Reactions were performed in triplicate and

843 *ACTIN*, a constitutively-expressed gene, was used as a control.

Primer name	**Sequence (5'-3')**	**Gene**
BPBF_F1	GTCTCGTGGTTGGTGGAAAT	HORVU5Hr1G048700
BPBF_R2	TTACATCAGGGAGGTGCTG	HORVU5Hr1G048700
ACTIN_F	TTCAGTGGTCCCTGCCAATGTA	HORVU1Hr1G047440
ACTIN_R	CCTACCGGCAAGATCAAGAC	HORVU1Hr1G047440

844

845 **F. Primers for *in situ* expression analysis.**

846 *In situ* hybridization was as described previously (Drea, *et al*., 2005; Opanowicz *et al.*, 2010).

PBF_insitu_F	GGGACTGGGAACAATGTGAC
PBF_insitu_R	CCGAGCGAAATACTTCAAGC
PBF_insitu_RT7 (antisense)	GAATTGTAATACGACTCACTATAGGGCCGAGCGAAATACTTCAAGC
PBF_insitu_FT7 (sense)	GGGACTGGGAACAATGTGACGAATTGTAATACGACTCACTATAGGG

847

Supplementary Table S3. WGCNA co-expression analysis.

Analysis of co-expression of *PBF* and other large-embryo genes (*GE*, *BIGE1* and *LARGE EMBRYO*) (Ramírez-González *et al.*, 2018). The co-expression Module (network) associated with each gene for four different tissues is indicated. NA = not analysed (because the gene is not expressed in this tissue and so was not included in the network).

Gene type	Gene ID	Module			
		Grain	Leaf	Root	Spike
GE	TraesCS2A01G175700	9	0	11	1
	TraesCS2B01G201900	2	1	18	1
	TraesCS2D01G183000	9	1	18	0
Lys3/PBF	TraesCS5A01G155900	13	NA	NA	NA
	TraesCS5B01G154100	13	NA	NA	NA
	TraesCS5D01G161000	13	NA	NA	NA
LE	TraesCS4A01G271200	0	0	2	0
	TraesCS4B01G042900	0	0	2	17
	TraesCS4D01G040200	0	0	2	0
BIGE1	TraesCS4A01G350200	2	0	7	12
	TraesCS5B01G522900	2	0	1	12
	TraesCS5D01G521600	2	0	0	0

Supplementary Table S4. Analysis of the promotor regions of wheat starch-related genes.

A set of wheat genes involved in starch synthesis was selected and their promotor regions (2 kb upstream sequences) were identified in the IWGSC RefSeqv1 wheat cv Chinese Spring (CS) genome sequence as part of the WGIN Wheat Promotome Capture project (http://www.wgin.org.uk/). The presence of *PBF* regulatory motifs (prolamin box, TGTAAAG or CTTTACT; pyrimidine box, CCTTTT or AAAAGG) in the promotor regions is indicated. Genes predicted by GENIE3 analysis to be down-stream targets of *lys3 PBF* are also indicated (see Fig. 5).

Gene function or product	Gene designation	IWGSC gene ID	Homoeolog present in CS	Prolamin box present	Pyrimidine box present	Predicted target genes of *lys3 PBF* (GENIE3).
ADPG pyrophosphorylase large subunit	*APL1/AGPL3*	TraesCS5A02G472000	Yes	No	Yes	No
		TraesCS5B02G484700	Yes	No	Yes	No
		TraesCS5D02G484500	Yes	No	Yes	No
ADPG pyrophosphorylase large subunit	*APL3/AGPL1*	TraesCS1A02G419600	Yes	Yes	Yes	Yes
		TraesCS1B02G449700	Yes	Yes	Yes	Yes
		TraesCS1D02G427400	Yes	Yes	Yes	Yes
ADPG pyrophosphorylase small subunit	*APS2/AGPS2*	TraesCS5A02G292900LC	Yes	No	Yes	No
		TraesCS5B02G175800	Yes	No	Yes	No
		TraesCS5D02G182600	Yes	No	No	No
ADPG pyrophosphorylase small subunit	*APS1/AGPS1/Bt2*	TraesCS7A02G287400	Yes	Yes	Yes	Yes
		TraesCS7B02G183300	Yes	No	Yes	Yes
		TraesCS7D02G284900	Yes	No	Yes	Yes
Disproportionating enzyme I	*DPE1*	TraesCS2A02G159300	Yes	No	Yes	No
		TraesCS2B02G184900	Yes	No	Yes	No
		TraesCS2D02G166600	Yes	No	Yes	No
Disproportionating enzyme II	*DPE2*	TraesCS2A02G123800	Yes	No	Yes	No
		TraesCS2B02G145700	Yes	No	Yes	No
		TraesCS2D02G126600	Yes	No	Yes	No

Floury endosperm6	*FLO6*	TraesCS4A02G284000	Yes	No	No	No
		TraesCS4B02G029700	Yes	No	Yes	No
		TraesCS4D02G026700	Yes	No	Yes	No
Granule-bound starch synthase I	*GBSSI/waxy*	TraesCS7A02G070100	Yes	No	Yes	Yes
		TraesCS7D02G064300	No	No	No	Yes
			Yes	No	Yes	No
Granule-bound starch synthase II	*GBSSII*	TraesCS2A02G373600	Yes	No	Yes	No
		TraesCS2B02G390700	Yes	Yes	Yes	No
		TraesCS2D02G369800	Yes	No	Yes	No
Isoamylase I	*ISA1*	TraesCS7A02G251400	Yes	No	Yes	Yes
		TraesCS7B02G139700	Yes	No	Yes	Yes
		TraesCS7D02G249500	Yes	No	Yes	Yes
Isoamylase II	*ISA2*	TraesCS1A02G247100	Yes	No	Yes	No
		TraesCS1B02G257700	Yes	Yes	Yes	No
		TraesCS1D02G246300	Yes	No	Yes	No
Isoamylase III	*ISA3*	TraesCS5A02G248700	Yes	No	Yes	No
		TraesCS5B02G246400	Yes	No	Yes	No
		TraesCS5D02G255800	Yes	No	Yes	No
Limit dextrinase	*PUL*	TraesCS7A02G133500	Yes	Yes	Yes	Yes
		TraesCS7B02G034600	Yes	No	Yes	Yes
		TraesCS7D02G133100	Yes	No	Yes	Yes
Maltose transporter	*MT*	TraesCS2A02G436000	Yes	No	Yes	No
		TraesCS2B02G457100	Yes	No	Yes	No
		TraesCS2D02G433900	Yes	No	Yes	No
Plastidial ADPG transporter	*OsBT1-1*	TraesCS6A02G175100	Yes	No	Yes	Yes
		TraesCS6B02G210000	Yes	No	Yes	Yes
		TraesCS6D02G168200	Yes	No	Yes	Yes
Starch branching enzyme I	*BEI*	TraesCS7A02G549300	Yes	No	Yes	Yes
		TraesCS7B02G472300	Yes	No	Yes	Yes
		TraesCS7D02G535600	Yes	No	Yes	Yes
Starch branching enzyme IIa	*BEIIa*	TraesCS2A02G293400	Yes	No	Yes	No
		TraesCS2B02G309500	Yes	No	Yes	No
		TraesCS2D02G290800	Yes	Yes	Yes	No
Starch branching enzyme IIb	*BEIIb/Ae*	TraesCS2A02G310300	Yes	No	Yes	Yes
		TraesCS2B02G327300	Yes	No	Yes	Yes
		TraesCS2D02G308600	Yes	No	Yes	Yes

Starch phosphorylase (cytosol)	*PHOH*	TraesCS3A02G366300	Yes	No	Yes	No
		TraesCS3B02G397900	Yes	No	Yes	No
		TraesCS3D02G359200	Yes	No	Yes	No
Starch phosphorylase (plastidial)	*PHOL*	TraesCS5A02G395200	Yes	Yes	Yes	Yes
		TraesCS5B02G400000	Yes	No	Yes	Yes
		TraesCS5D02G404500	Yes	Yes	Yes	No
Starch synthase I	*SSI*[c]	TraesCS7A02G120300	Yes	No	No	No
		TraesCS7B02G018600	Yes	No	No	No
		TraesCS7D02G117800	Yes	No	Yes	Yes
Starch synthase IIa	*SSIIa/SSII-3*	TraesCS7A02G189000	Yes	No	Yes	No
		TraesCS7B02G093800	Yes	Yes	Yes	No
		TraesCS7D02G190100	Yes	Yes	No	No
Starch synthase IIb	*SSIIb/SSII-2*	TraesCS6A02G307800	Yes	No	Yes	No
		TraesCS6B02G336400	Yes	No	Yes	No
		TraesCS6D02G287000	Yes	No	Yes	No
Starch synthase IIc	*SSIIc/SSII-1*	TraesCS1A02G137200	Yes	No	Yes	No
		TraesCS1B02G155700	Yes	No	Yes	No
		TraesCS1D02G138100	Yes	No	Yes	No
Starch synthase IIIa	*SSIIIa/SSIII-2*	TraesCS1A02G091500	Yes	No	Yes	Yes
		TraesCS1B02G119300	Yes	No	Yes	Yes
			Yes	No	No	No
Starch synthase IIIb	*SSIIIb/SSIII-1*	TraesCS2A02G468800	Yes	Yes	Yes	No
		TraesCS2B02G491700	Yes	No	Yes	No
		TraesCS2D02G468900	Yes	Yes	Yes	No
Starch synthase IV	*SSIVa/SSIV-1*	TraesCS1A02G353300	Yes	No	Yes	No
		TraesCS1B02G368500	Yes	No	Yes	No
		TraesCS1D02G356900	Yes	Yes	Yes	No

861

862 **Supplementary Figure S1. Comparison of the *PBF* alleles in all four *lys3* mutant lines.**

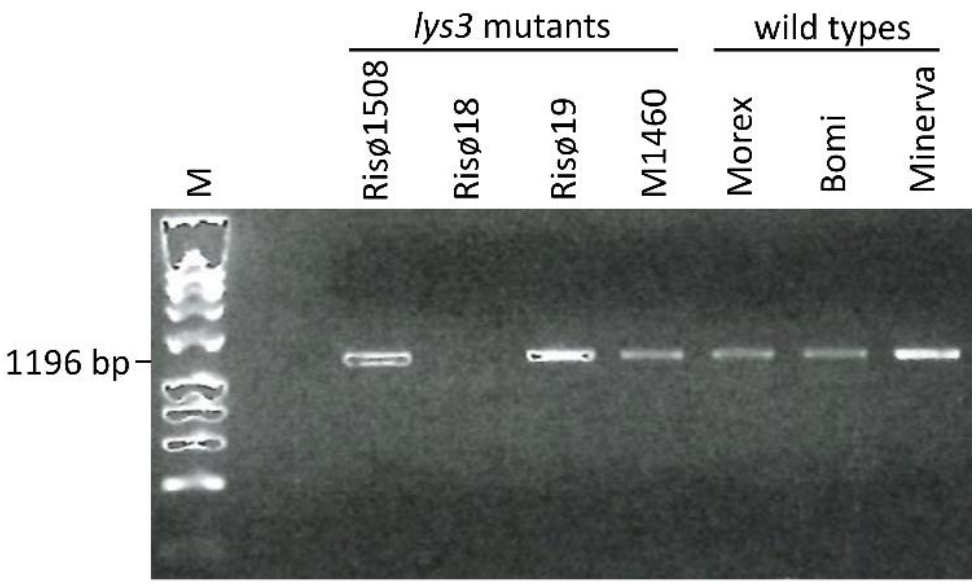
863 Barley *PBF* genomic sequences were amplified by PCR and amplicons were separated by

864 electrophoresis. M = molecular weight ladder (Bioline Ladder 1 kbp). The size of the wild

865 type *PBF* amplicon is indicated.

866

867



868 **Supplementary Figure S2. Predicted downstream targets of *PBF* in wheat.**

869 GENIE3 analysis of 850 wheat RNASeq samples revealed putative downstream targets of

870 *PBF* (Fig. S2). The gene ontologies (GOs) were reduced and visualized using enrichment

871 analysis (http://revigo.irb.hr/revigo.jsp; Ramírez-González *et al.*, 2018). Bubble colour

872 indicates p-value and bubble size the frequency of the GO term in the GO database. Highly

873 similar GO terms are linked by edges in the graph.