bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome-wide screen of genomic imprinting in endosperm and population-level analysis reveal allelic variation for imprinting in flax

Haixia Jiang[1, #], Dongliang Guo[1, #], Jiali Ye[2], Yanfang Gao[1], Huiqing Liu[1], Yue Wang[1], Min Xue[1], Qingcheng Yan[1], Jiaxun Chen[1], Lepeng Duan[1], Gongzhe Li[1], Xiao Li[1] and Liqiong Xie[1, *]

[1]National Center of Melon Engineering and Technology, Molecular Breeding Laboratory, College of Life Science and Technology, Xinjiang University, Urumqi, Xinjiang, China.

[2]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, China.

[#]These authors have contributed equally to this work.

[*]**Correspondence:** Liqiong Xie, picea@sina.com.

Abstract

Genomic imprinting is an epigenetic phenomenon caused by the biased expression of maternally and paternally inherited alleles. In flowering plants, genomic imprinting predominantly occurs in triploid endosperm and plays a vital role in seed development. In this study, we identified 241 candidate imprinted genes including 143 maternally expressed imprinted genes (MEGs) and 98 paternally expressed imprinted genes (PEGs) in flax (*Linum usitatissimum* L.) endosperm using deep RNA sequencing. The conservation of imprinting in plants is very limited and imprinting clustering is not a general feature. MEGs tends to be endosperm expression specific, while PEGs are non-tissue specific. Imprinted SNPs differentiated 200 flax cultivars into oil flax, oil-fiber dual purpose flax (OF) and fiber flax subgroups, suggesting that genomic imprinting contributes to intraspecific variation in flax. The nucleotide diversity (π) of imprinted genes in oil flax subgroup is significantly higher than that in fiber flax subgroup, indicating that some imprinted genes undergo positive selection during flax domestication from oil flax to fiber flax. Imprinted genes undergo positive selection is related to the functions. Eleven imprinted genes related to seed size and weight were identified using the candidate gene-based association study. Our study provides information for further exploring the function and genomic variation of imprinted genes in flax population.

Keywords: genomic imprinting, flax endosperm, intraspecific variation, seed size, positive selection

Introduction

Genomic imprinting is an epigenetic phenomenon occurring in both mammals and flowering plants (Hutter *et al.*, 2010; Waters *et al.*, 2013). Imprinting occurs in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38 placenta and embryo as well as in adult tissues in mammals (Renfree *et al.*, 2012),
39 while in flowering plants, imprinting predominantly occurs in endosperm, and rarely
40 in embryo and seed coat (Yuan *et al.*, 2017; Meng *et al.*, 2018). The diploid embryo
41 transmits genetic information to the next generation, while the triploid endosperm,
42 with 2:1 ratio of maternal to paternal genomes (2m:1p), provides nutrition and signals
43 for embryo development (Sabelli and Larkins, 2009; Mei *et al.*, 2015). The deviation
44 of the 2m:1p ratio in endosperm has adverse effect on seed development (Wang *et al.*,
45 2018; Scott *et al.*, 1998; Lu *et al.*, 2012; Sekine *et al.*, 2013), implying the important
46 role of imprinting in endosperm development. Imprinted genes were studied only in a
47 limited plant species, including monocots of rice (*Oryza sativa*) (Yuan *et al.*, 2017;
48 Luo *et al.*, 2011), maize (*Zea mays*) (Waters *et al.*, 2013; Meng *et al.*, 2018; Zhang *et*
49 *al.*, 2011; Waters *et al.*, 2011; Dong *et al.*, 2017), sorghum (*Sorghum bicolor* L.
50 Moench) (Zhang *et al.*, 2016a) and dicots of *Arabidopsis* (Gehring *et al.*, 2011; Wolff
51 *et al.*, 2011; Hsieh *et al.*, 2011), castor bean (*Ricinus communis*) (Xu *et al.*, 2014). In
52 most dicots, the endosperm is transient and consumed by the embryo at the later stage
53 of seed development (Sreenivasulu and Wobus, 2013), while the endosperm of most
54 monocots remains persistently and serves as a source of nutrition for seed germination
55 (Luo *et al.*, 2011). There is no extensive conservation of imprinted genes across
56 species (Waters *et al.*, 2011; Zhang *et al.*, 2016a; Dong, 2017), suggesting different
57 species may require a unique set of imprinted genes for seed development.

58 In plants, imprinted genes were detected to be involved in the regulation of seed
59 development, seed dormancy and postzygotic reproductive isolation (Sun *et al.*, 2017;
60 Piskurewicz *et al.*, 2016; Kradolfer *et al.*, 2013; Wolff *et al.*, 2015). The loss of
61 function of some imprinted genes leads to seed abortion (Joanis and Lloyd, 2002;
62 Chaudhury *et al.*, 1997; Berger *et al.*, 2006). In *Arabidopsis*, the maternally expressed
63 imprinted gene *MEA* encodes a Polycomb Repressive Complex 2 (PRC2) subunit in
64 Polycomb group (PcG) complex. The seeds with maternal *MEA* allele developed
65 normally, while those carrying maternal *mea* allele were aborted, regardless of the
66 genotypes of paternal allele (Joanis and Lloyd, 2002). Imprinted genes influence seed
67 size by regulating the development of embryo and endosperm (Yuan *et al.*, 2017;
68 Scott *et al.*, 1998; Köhler *et al.*, 2005; Chen *et al.*, 2016). In rice, the loss-of-function
69 mutants of *MEG2* and *MEG3* (MEGs) showed a significant reduction in seed size and
70 weight, and the loss-of-function of *PEG1*, *PEG2*, *PEG3* (PEG) decreased starch
71 content, seed size and yield (Yuan *et al.*, 2017). Regarding to another imprinted gene
72 *OsFIE1* in rice, RNAi lines and homozygous T-DNA insertion mutant *osfie1* lines all
73 showed delayed embryo development and reduction of seeds fertility, grain size, grain
74 weight and aleurone layer cells (Huang *et al.*, 2016). Even a few imprinted genes
75 were shown to be important in seed development, there are many more, of which the
76 function is yet to be determined.

77 Some imprinted genes were shown to be under positive selection and intraspecific
78 variation features (Hutter *et al.*, 2010; Berger *et al.*, 2012; Pignatta *et al.*, 2014). *MEA*

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

79 originated a block duplication 35 to 85 million years ago owing to a whole-genome
80 duplication within the Brassicaceae lineage (Spillane *et al.*, 2007). After duplication,
81 *MEA* underwent positive selection consistent with neo-functionalization and the
82 parental conflict theory (Spillane *et al.*, 2007; Miyake *et al.*, 2009). In maize,
83 conservative imprinting genes increased the substitution rate of nonsynonymous to
84 synonymous (dN/dS) compared with non-conservative imprinting genes and more
85 likely to undergo positive selection (Waters *et al.*, 2013). PEGs were more likely to be
86 under positive selection and rapidly evolve than MEGs in *Arabidopsis thaliana*
87 (Tuteja *et al.*, 2019). Imprinting showed evidence of intraspecific variation in
88 *Arabidopsis* and maize (Waters *et al.*, 2013; Pignatta *et al.*, 2014). The existence of
89 intraspecific variation of imprinting was associated with epigenetic variation (Pignatta
90 *et al.*, 2014).

91 Flax (*Linum usitatissimum* L.) is an important economic crop due to its stem fiber
92 and seed oil (Cloutier *et al.*, 2012; Guo *et al.*, 2020). Flax is a strict annual self
93 pollination crop with a smaller genome size (～373 Mb) (Wang *et al.*, 2012), which
94 are good for biological research. Cultivated flax were domesticated from a pale flax
95 (*Linum bienne*) for oil usage at 10,000 years ago in the Near East and differentiated
96 into fiber, OF and oil subgroups during the domestication of flax (Guo *et al.*, 2020;
97 Allaby *et al.*, 2005; Fu and Allaby, 2010). Some genes underlying important traits
98 including plant architecture, flowering, dehiscence, oil production and yield
99 underwent strong artificial selection in the domestication process of flax (Guo *et al.*,
100 2020; Zhang *et al.*, 2020). Whether imprinted genes are also undergone artificial
101 selection during flax domestication has not been reported.

102 In this study, we have performed RNA-seq analysis of flax endosperm isolated
103 from the reciprocal crosses between CIli2719 and Z11637. Based on the
104 parent-of-origin biased expression of the parental alleles in the endosperm from both
105 crosses, we identified 241 moderately imprinted genes including 143 MEGs and 98
106 PEGs, 67 strongly imprinted genes including 63 MEGs and 4 PEGs and 19
107 completely MEGs. The analysis of imprinted genes at population level demonstrated
108 that imprinted genes divided the 200 flax germplasms into oil flax, OF and fiber flax
109 subgroups, which suggested that imprinting promoted intraspecific variation of flax.
110 The nucleotide diversity analysis showed that some imprinted genes were shown to be
111 under positive selection consistent with function. Furthermore, we identified 11
112 imprinted genes associated to seed size and weight. Our results will provide a
113 theoretical basis for further study of gene imprinting and provide some insights for
114 understanding the diversity of imprinted genes.

115 **Results**

116 **Identification of imprinted genes in flax endosperm**

117 To understand the parental origin of gene expression in flax, we performed RNA-seq
118 analysis in endosperm isolated from the F_1 generation of CIli2719×Z11637 (CZ) and

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

119 Z11637×CIli2719 (ZC) at 7 days after pollination (DAP). Average 7.21 Gb of 125bp
120 paired-end clean reads were obtained from each biological replicate with the Illumina
121 novaseq6000 platform. The clean reads from CZ and ZC were aligned to parental
122 genomes CIli2719 and Z11637 (Guo *et al.*, 2020) to identify the reads specifically
123 originated from one of parents at each SNP site for allelic expression (Figure S1). In
124 total, 52,794 SNPs in both reciprocal crosses with at least ten reads could be assigned
125 to a parental allele and were used for allele-specific expression analysis in hybrid
126 endosperm. About 37,000 SNP loci showed statistically significant deviation ($p<0.05$,
127 χ^2 test) from the expected 2m:1p ratio in both CZ and ZC endosperm.

128 Three different thresholds were used to identify genes showing parent-of-origin
129 biased expression at three different levels (see "Materials and methods"). Among
130 these SNP loci, 498 loci were considered as moderately imprinted loci, including 319
131 maternally expressed imprinted SNP loci (ME-SNPs) corresponding to 143 MEGs
132 and 179 paternally expressed imprinted SNP loci (PE-SNPs) corresponding to 98
133 PEGs. And 141 loci were identified as strongly imprinted loci, including 135
134 ME-SNPs and 6 PE-SNPs which correspond to 63 MEGs and 4 PEGs, respectively.
135 In addition, 36 loci were identified as completely imprinted loci and all of them were
136 ME-SNPs corresponding to 19 MEGs (Figure 1, Table S1). Among 241 imprinted
137 genes, 229 genes were protein-coding genes (Table S2).

138 **Validation of imprinted genes**

139 For experimental confirmation, thirteen genes including five MEGs, five PEGs, and
140 three non-imprinted genes, which represented the whole transcripts by RNA
141 sequencing in this study, were randomly selected to validate the gene expression level
142 of the high throughput sequencing with qRT-PCR analysis. The analysis showed that
143 the gene expression level of the selected genes by qRT-PCR analysis was consistent
144 with the RNA-seq data (Figure S2).

145 For further verification of the imprinting status, nine MEGs and three PEGs were
146 used to perform RT-PCR on the hybrid endosperm and parents endosperm followed
147 by Sanger sequencing (Figure S3). A 400-800bp fragment of each gene with at least
148 one imprinted SNP site was selected for PCR amplification. The results showed that
149 nine MEGs were predominantly expressed from the maternal alleles and three PEGs
150 were preferentially expressed from the paternal alleles in reciprocal crosses, which
151 were consistent with the RNA-seq data.

152 **Characterization analysis of imprinted genes identified in flax**

153 We carried out gene ontology (GO) analysis for the 229 imprinted protein-coding
154 genes of flax, including 135 MEGs and 94 PEGs (Table S2). The background genes
155 for GO analysis were 12,395 endosperm-expressed genes with at least ten reads could
156 be assigned to a specific allele in both CZ and ZC. Categories with a significant level
157 ($P<0.05$) were defined as enriched. Compared with the whole transcripts in

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

158 endosperm, imprinted genes were significantly enriched in catalytic activity or
159 transferase activity according to their molecular function and metabolic process
160 according to their biological processes (Figure S4A, Table S3).

161 To evaluate the interspecific conservation of imprinted genes, we compared
162 imprinted genes identified in this study with the imprinted genes reported in
163 *Arabidopsis* (Gehring *et al.*, 2011), rice (Yuan *et al.*, 2017; Luo *et al.*, 2011), maize
164 (Waters *et al.*, 2013; Dong, 2017), sorghum (Zhang *et al.*, 2016a), and castor bean
165 (Xu *et al.*, 2014). The analysis showed that there were 115, 89 and 59 genes found to
166 be homologous in at least one of the five species at different confidence levels
167 (<1E-10, <1E-20 and <1E-50, respectively) (Figure 2, Table S4-S5). At
168 E-value<1E-10, there were 32, 26, 43, 52, 39 flax imprinted genes conserved in
169 *Arabidopsis*, castor bean, rice, sorghum and maize, respectively (Figure 2A, Table
170 S4-S5). And 19, 21, 35, 46, 27 imprinted genes conserved with these plants at
171 E-value<1E-20, respectively (Figure 2B, Table S4-S5). In addition, less imprinted
172 genes 13, 11, 21, 22 and 15 conserved with the five species at E-value<1E-50,
173 respectively (Figure 2C, Table S4-S5). Some imprinted genes had imprinted
174 homologs in up to four species, while no imprinted gene in flax was conserved in all
175 species (Figure 2, Table S4-S5). These results suggested that the conservation of
176 imprinting in plants was quite limited.

177 Intriguingly, the expression of some conserved flax imprinted genes showed
178 different parental origin in other species. For example, the flax gene *Lus10022747*
179 encoding a serine/threonine-protein kinase WNK5 and its homologues in *Arabidopsis*,
180 castor bean and rice showed maternally preferential expression, while its maize
181 homolog displayed preferentially paternal expression (Table S4). A PAS domain
182 tyrosine kinase family protein-coding gene *Lus10040540* and its homolog of maize
183 were PEGs, but its homologues in castor bean, rice and sorghum were MEGs (Table
184 S4). Among 115 conserved flax imprinted genes, only 59% (58 MEGs and 10 PEGs)
185 remain the same preference of parental expression with other species (Table S4).

186 GO enrichment analysis of 115 conserved imprinted genes in flax displayed that
187 MEGs significantly enriched in catalytic activity and metabolic process according to
188 their molecular function and biological processes, respectively. PEGs enriched in
189 compound binding according to their molecular function (Figure S4B, Table S6).

190 **Clustering of the flax imprinted genes**

191 To study the genomic distribution of flax imprinted genes, 229 imprinted genes were
192 mapped to fifteen chromosomes for cluster analysis. The 229 imprinted genes were
193 scattered distribution across fifteen chromosomes. By analyzing the genomic distance
194 between the imprinted genes, we found that most of them were not co-localized in a
195 cluster, and only 24 were fall into 12 clusters, where two imprinted genes of each
196 cluster were within 10 kb (Figure 3, Table S7). The finding was similar to the results
197 of *Arabidopsis* (Gehring *et al.*, 2011; Wolff *et al.*, 2011), maize (Waters *et al.*, 2011),

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

198 rice (Luo *et al.*, 2011), castor bean (Xu *et al.*, 2014) and sorghum (Zhang *et al.*,
199 2016a), showing that clustering of imprinted gene is not a common phenomenon in
200 plants.

201 **Endosperm-specific expression of the flax imprinted genes**

202 We analyzed the expression specificity of the imprinted genes in various tissues. The
203 majority of MEGs (60%) preferentially expressed in endosperm, and only a minority
204 (25.5%) of PEGs were preferentially expressing in endosperm (Figure 4A-B). The
205 expression level of endosperm-preferred MEGs (endo-MEGs) and PEGs (endo-PEGs)
206 were significantly (P<0.05) higher than that of all genes, whereas there was no
207 evidence that endo-MEGs or endo-PEGs exhibited unusually high or low expression
208 levels than other MEGs or PEGs which also expressed in other tissues (Figure 4C-D).
209 We also analyzed the tissue specificity of 115 (75 MEGs and 40 PEGs) conserved
210 imprinted genes and 38 (36 MEGs and 2 PEGs) conserved strong imprinted genes.
211 Among 115 conserved imprinted genes, 49 (65.3%) MEGs and 12 (30%) PEGs
212 showed endosperm-preferred expression (Figure S5A-B), and among the 38
213 conserved strong imprinted genes, 29 MEGs (80.6%) and all PEGs were
214 preferentially expressed in endosperm (Figure S5C), respectively. These results
215 suggested that MEGs and the conserved imprinted genes are more likely to be
216 preferentially expressed in endosperm.

217 **Flax imprinted genes can differentiate flax subgroups**

218 To investigate whether the variation in imprinted genes reflects genetic diversity
219 among 200 natural flax varieties, we detected individual kinship of imprinted SNPs
220 (498) or genome-wide SNPs (674,074) (Guo *et al.*, 2020). The kinships between
221 imprinted SNPs or genome-wide SNPs were significantly correlated (R^2=0.8457)
222 (Figure 5A). Phylogenetic tree constructed based on all imprinted SNPs, ME-SNPs or
223 PE-SNPs separated 200 accessions into three different subgroups which correspond to
224 oil flax, OF and fiber flax subgroups (Figure 5B, Figure S6). The flax population
225 could also be separated into three subgroups by principle component analysis (PCA)
226 (Figure 5C). These results indicated that the allele frequency of imprinted SNPs was
227 significantly different among subgroups, suggesting that imprinted genes may be
228 selected differently in subgroups and contribute to domestication.

229 **Selective sweep signals in imprinted genes**

230 To test the hypothesis that the diversity of imprinted genes are different in flax
231 population, we collated all SNPs (3,191 SNPs) of 241 imprinted genes and compared
232 the nucleotide diversity between oil flax (78 germplasms) and fiber flax subgroups
233 (51 germplasms) which represented two primary morphotypes of cultivated flax
234 (Zhang *et al.*, 2020). The π values of all imprinted genes, MEGs or PEGs decreased
235 significantly in fiber flax compared with oil flax (P < 0.0001, t test) (Figure 6A-C).
236 We focused on two imprinted genes, *Lus10010350* (PEG) and *Lus10024230* (MEG),

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

237 which contained more SNPs (31 SNPs and 43 SNPs, respectively) in genomic
238 sequences. The allele frequency distribution of imprinted SNPs in *Lus10010350* was
239 significantly different in the two subgroups, and the alternate allele 'G' (position
240 333406) was primarily found in the oil subgroup and rarely in fiber subgroup (Figure
241 6D). Significant reduction of π was also observed at the *Lus10010350* locus in fiber
242 subgroup compared with that of oil subgroup and 90.32% of the SNPs were identified
243 with a signature of purifying selection (Figure 6E-F). Similarly, the allele frequency
244 distribution of *Lus10024230* was obvious different between the two subgroups and
245 the alternate allele 'A' (position 492363) accounted for 61.67% in oil subgroup but
246 only 6.98% in fiber subgroup (Figure 6G). Compared with oil subgroup, the π value
247 of the SNPs in *Lus10024230* was dramatically decreased in fiber subgroup, and
248 95.35% of the SNPs were identified with a signature of purifying selection (Figure
249 6H-I). Taken together, these findings suggested that some imprinted genes may have
250 been subjected to artificial selection during flax domestication.

251 **Candidate gene-based association study for seed size using flax imprinted genes**

252 Imprinted genes played an important role in the regulation of endosperm development
253 and seed size (Yuan *et al.*, 2017; Luo *et al.*, 2000; Guitton *et al.*, 2004; Huang *et al.*,
254 2017; Kinoshita *et al.*, 1999; Kiyosue *et al.*, 1999). To investigate whether imprinting
255 is associated with seed size in flax, imprinted genes were used to perform candidate
256 gene-based association study of seed size-related traits including seed length (SL),
257 seed width (SW) and 1,000-seed weight (1000-SW). Using the general linear model
258 (GLM) and mixed linear model (MLM) in TASSEL 5.0 (Bradbury *et al.*, 2007), 33
259 imprinted genes containing 63 associated loci (SNPs) were detected to be associated
260 with seed size and weight. Among them, 11 imprinted genes were repeatedly detected
261 at least two environments or traits (Figure 7A-B, Table S8).

262 One of the significant signal peaks on chromosome 15 contained 9 repetitive SNPs
263 (Figure 7C, Table S8), which located in the *Lus1001035*0 (PEG), encoding a
264 bifunctional arginine demethylase and lysine hydroxylase jmjd6 protein. This gene
265 contained 31 SNPs, of which 12 induced nonsynonymous mutations and formed 11
266 haplotypes (Figure 7D-E). We classified 200 accessions into two groups including
267 haplotype A (reference alleles) and haplotypes B–K (alternate alleles) based on gene
268 structural variation (Figure 7F). We found that flax accessions in haplotypes B–K had
269 significantly longer seed length and width, and larger 1,000-seed weight than those in
270 haplotype A (Figure 7G-J, Figure S7). These results suggested that the PEG
271 *Lus10010350* may be involved in the seed size regulation in flax.

272 **Discussion**

273 **Characterization of imprinted genes in flax endosperm**

274 GO analysis for the 229 imprinted protein-coding genes of flax revealed that a
275 majority of imprinted genes were significantly enriched in catalytic activity and

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

276 metabolic process (Figure S4A, Table S3), similar to the results in castor bean (Xu *et*
277 *al.*, 2014) and sorghum (Zhang *et al.*, 2016a), suggesting imprinted genes affected
278 various aspects of endosperm development. However, imprinted genes have limited
279 conservation across plant species, in contrast to those in mammals (Waters *et al.*,
280 2011; Zhang *et al.*, 2016a; Xu *et al.*, 2014; Dong, 2017; Zhang *et al.*, 2003). In maize
281 and sorghum endosperm, only about 10% and 33% imprinted genes were conserved
282 with other species, respectively (Waters *et al.*, 2011; Zhang *et al.*, 2016a). Among 165
283 imprinted genes in rice, only 33% of them were conserved in maize (Dong, 2017).
284 Twenty-five (12%) imprinted genes identified in castor bean were conserved in
285 *Arabidopsis*, rice or maize (Xu *et al.*, 2014). In this study, the conservation of flax
286 imprinted genes was evaluated with those in other species (*Arabidopsis*, castor bean,
287 rice, sorghum and maize) at three levels of stringency. There were 115 (50.2%), 89
288 (38.9%), 59 (25.8%) imprinted genes in flax which were homologous to those in at
289 least one of the five species at E-value<1E-10, 1E-20 and 1E-50, respectively (Figure
290 2, Table S4-S5) while none imprinted genes of flax were found having imprinted
291 homologs with all five species (Figure 2, Table S4-S5). Those results suggested that
292 some common pathways in different flowering plants may need to be regulated by
293 imprinting to modulate endosperm development but different genes in the pathway
294 are selected to be imprinted in different species. This explains why the conservation
295 of imprinting in plants is quite limited.

296 An interesting observation was that among 115 conserved imprinted genes, only 68
297 genes (58 MEGs and 10 PEGs) showed the same origin of parental expression as in
298 other species (Table S4). The remaining 47 genes (17 MEGs and 30 PEGs) had
299 opposite origin of parental expression. For instance, the gene *Lus10041031* is a
300 complete MEG identified in our study, but its maize homolog is a PEG (Waters *et al.*,
301 2013). *Lus10021926* is a strong PEG in flax, while in *Arabidopsis* and castor bean its
302 homologies are MEGs (Gehring *et al.*, 2011; Xu *et al.*, 2014). The homologous genes
303 of the strong PEG *Lus10016563* are PEG in maize and MEGs in rice and sorghum
304 (Waters *et al.*, 2013; Yuan *et al.*, 2017; Zhang *et al.*, 2016a). This suggested that those
305 genes have opposite mode of parental expression in different species may be subject
306 to gene dosage, a mechanism thought to be important for endosperm development
307 (Wang *et al.*, 2018; Scott *et al.*, 1998; Lu *et al.*, 2012; Sekine *et al.*, 2013). It was
308 noteworthy to note that the gene *Lus10041386* encoding a histone-lysine
309 N-methyltransferase Enhancer of Zeste homolog 2 (EZH2) is homologous to the
310 FERTILIZATION-INDEPENDENT SEED Polycomb Repressive Complex 2
311 (FIS-PRC2) class gene *MEA* in *Arabidopsis* and the imprinted gene
312 *GRMZM2G157820* (*EZH2*) in maize (Table S4**)**. In rice and other species, other
313 members of PRC2 genes are also imprinted, suggesting that imprinting of PRC2
314 genes is a conserved mechanism in flowering plants. Several lines of evidence
315 suggested that PRC2 repressed the replication of central-cell nuclear before
316 fertilization likely by the maternally expressed alleles and regulated endosperm
317 proliferation, suggesting a vital role in seed development (Chaudhury *et al.*, 1997;

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

318 Kiyosue *et al.*, 1999; Zhang *et al.*, 2005; Ohad *et al.*, 1996; Ohad *et al.*, 1999; Luo *et*
319 *al.*, 1999; Moreno-Romero *et al.*, 2019).

320 **Imprinted genes are not extensively clustered**

321 Physical clustering of imprinted genes is a conserved feature in mammals (Gehring *et*
322 *al.*, 2011; Gregg *et al.*, 2010), while there is little evidence of clustering in plant
323 species. Imprinted genes identified in maize (Waters *et al.*, 2011), *Arabidopsis*
324 (Gehring *et al.*, 2011; Wolff *et al.*, 2011), rice (Luo *et al.*, 2011), castor bean (Xu *et*
325 *al.*, 2014) and sorghum (Zhang *et al.*, 2016a) were not shown to be extensive
326 clustered. Using a clustering criterion consistent with that in *Arabidopsis* (～125 Mb)
327 and castor bean (～350 Mb) which have comparable genome size to flax (～373 Mb)
328 (Gehring *et al.*, 2011; Xu *et al.*, 2014), we found that 24 of 229 flax imprinted genes
329 were fall into 12 clusters (Figure 3, Table S7), similar to the proportion in *Arabidopsis*
330 (Gehring *et al.*, 2011; Wolff *et al.*, 2011), rice (Luo *et al.*, 2011), maize (Waters *et al.*,
331 2011), castor bean (Xu *et al.*, 2014) and sorghum (Zhang *et al.*, 2016a), suggesting
332 that imprinting clustering may be not a general feature in plants. Whether the
333 clustered imprinted genes are coordinately regulated as those genes in animal clusters
334 remains to be investigated.

335 **Endosperm-specific expression of flax imprinted genes**

336 According to previous reports, imprinted genes in plants were mainly restricted to
337 express in endosperm (Berger *et al.*, 2012). But more and more studies had shown
338 that only some imprinted genes are preferentially expressed in endosperm, while
339 others are also expressed in other tissues (Waters *et al.*, 2013; Waters *et al.*, 2011;
340 Dong, 2017). The proportions of endo-MEGs and endo-PEGs were dramatically
341 different (68% MEGs versus 26% PEGs, 51% MEGs versus 24% PEGs in maize;
342 50% MEGs versus 16% PEGs in rice; 50% MEGs versus 20% PEGs in sorghum)
343 (Waters *et al.*, 2013; Dong, 2017). In our study, we found 81 endo-MEGs (60%) and
344 only 24 endo-PEGs (25.5%) in flax (Figure 4A-B), similar to the proportion in rice,
345 sorghum and maize (Waters *et al.*, 2013; Zhang *et al.*, 2016a; Dong, 2017). The
346 expression level of endo-MEGs and endo-PEGs was significantly higher than that of
347 all genes (Figure 4C-D). Compared with all imprinted genes (60% endo-MEGs versus
348 25.5% endo-PEGs), the proportion of endo-MEGs (65%) and endo-PEGs (30%) of
349 conserved imprinted genes increased (Figure S5A-B). In the 38 (36 MEGs and 2
350 PEGs) conserved strong imprinted genes, 29 MEGs (80.6%) and all PEGs were
351 endosperm-preferred expression (Figure S5C). These findings suggested that MEGs
352 tends to be endosperm preferentially expressed, while PEGs are inclined to non-tissue
353 specific expression. It also implied that the conserved imprinted genes are more likely
354 to be preferentially expressed in endosperm and play an important role in seed
355 development.

356 **Candidate gene-based association study reveals that some imprinted genes are**
357 **involved in flax seed size regulation**

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

358 Previous studies have shown that imprinted genes play an important role in seed
359 development by regulating the development of endosperm (Yuan *et al.*, 2017; Köhler
360 *et al.*, 2005; Chen *et al.*, 2016; Luo *et al.*, 2000; Guitton *et al.*, 2004; Huang *et al.*,
361 2017; Zhang *et al.*, 2016b). Eleven imprinted genes related to seed size and
362 1,000-seed weight were obtained based on candidate gene-based association study.
363 Among 11 imprinted genes, the gene *Lus10036044* (MEG) encoding a plant AT-rich
364 sequence- and zinc-binding (PLATZ) transcription factor had significantly associated
365 signal peaks on chromosome 1 (Table S8). PLATZ transcription factor is a novel class
366 of plant-based zinc ion and DNA binding proteins, reported to regulate the seed size
367 and weight (Azim *et al.*, 2020). *ZmPLATZ12* (*Fl3*) is a maternally expressed
368 imprinted gene specifically expressing in the starchy cells of endosperm in maize. The
369 semi dominant negative *fl3* mutant resulted in severe defects of endosperm and
370 dramatically reduced the weight of seeds (Li *et al.*, 2017). In rice, the PLATZ
371 transcription factor *GL6* positively controlled grain length through promoting cell
372 proliferation in grains. The null *gl6* mutant led to short grains, whereas
373 overexpression the *GL6* produced large grains (Wang *et al.*, 2019). Another PLATZ
374 gene *SHORT GRAIN6* (*SG6*) determined grain size by regulating the cell division of
375 spikelet hull. The grain size and weight was significantly enlarged in the *SG6*
376 overexpression lines and reduced in *sg6* mutant lines in rice (Zhou and Xue, 2020).

377 Another candidate gene *Lus10037040* (MEG) located on chromosome 1, which
378 belongs to the MADS-box genes (Table S8). MADS-box genes had important
379 functions in the development of seed by epigenetic mechanism including DNA
380 methylation and histone modifications (Zhang *et al.*, 2016b). In rice, *OsMADS87*
381 (MEG) affected seed size by regulating endosperm cellularization during syncytial
382 stage. Over expression the *OsMADS87* led to larger seeds, and *OsMADS87*-RNAi
383 resulted in smaller seeds (Chen *et al.*, 2016). The MADS-box gene *PHE1* (PEG)
384 regulated seed size in *Arabidopsis thaliana* via influencing the expression of *AGL62*
385 which might affect the endosperm cellularization (Sun *et al.*, 2017). *OsMADS29*
386 regulated seed development though regulating cell degeneration of maternal tissues.
387 *OsMADS29*-RNAi resulted in aborted and/or shriveled seeds with deficient starch
388 accumulation in endosperm (Yang *et al.*, 2012). Heterologous expression the
389 *CnMADS* gene significantly increased the seed size of *Arabidopsis* (Sun, 2018).

390 *Lus10024230* annotated as flavonol synthase (FLS) is also potentially involved in
391 seed size control (Table S8). In the lines of *FLS*-RNAi of tobacco, the pods and seed
392 development was arrested and the height, pods size, pods weight, seeds number were
393 significantly reduced (Mahajan *et al.*, 2011). Furthermore, the alternative alleles at
394 *Lus10010350* (haplotypes B–K) had significantly longer seed length and width, and
395 larger 1,000-seed weight than those in reference allele (haplotype A) in 200 flax
396 accessions (Figure 7G-J, Figure S7). Together, our study identified a few candidate
397 imprinted gene which are potentially involved in seed development and modulate the

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

398 seed size. The genetic variation of these genes between flax lines may be harnessed as
399 breeding tool for enhance seed yield.

400 **Intraspecific variation of flax imprinted genes**

401 DNA methylation, histone modification and non-coding small RNAs caused
402 genomic imprinting (Zhang *et al.*, 2016a; Sha, 2008; Hanna and Kelsey, 2017).
403 Epigenetic modification often varied across different individuals of the same species
404 (Pignatta *et al.*, 2014; Xu *et al.*, 2019). In maize, differentially methylated regions
405 (DMRs) were changed in different subgroups and genotypes (Xu *et al.*, 2019; Li *et al.*,
406 2015). In *Arabidopsis*, DNA methylation and small RNAs differentiated in natural
407 populations and contributed to phenotypic diversity (Pignatta *et al.*, 2014; Schmitz *et*
408 *al.*, 2011; Becker *et al.*, 2011; Graaf *et al.*, 2015; Schmitz *et al.*, 2013). Genomic
409 imprinting, as the functional product of epigenetic modification, varied within a same
410 species and the intraspecific variation of imprinted genes was associated with
411 epigenetic variation (Waters *et al.*, 2013; Pignatta *et al.*, 2014). In this research, the
412 analysis of phylogenetic tree and PCA for imprinted SNPs showed that imprinted
413 SNPs effectively divided the 200 flax germplasms into oil, OF and fiber flax
414 subgroups (Figure 5, Figure S6), suggesting that genomic imprinting changed in
415 different subgroups and contributed to phenotypic diversity in flax.

416 **Some imprinted genes show evidence of positive selection**

417 According to previous reports, some imprinted genes showed positive selection
418 features (Hutter *et al.*, 2010; Berger *et al.*, 2012). *MEA* as a component of FIS-PRC2
419 was a very important conserved imprinted gene in seed development underwent
420 positive selection in the out-crossing lineages but not in the self-fertilizing species of
421 *Arabidopsis* (Spillane *et al.*, 2007; Miyake *et al.*, 2009). Conserved imprinted genes
422 displayed higher dN/dS rates than non-conservative imprinted genes between maize,
423 rice and sorghum, suggesting conserved imprinted genes showing greater evidence of
424 positive selection (Waters *et al.*, 2013). Compared with MEGs, PEGs exhibited
425 elevated dN/dS values and more likely to under positive darwinian selection in
426 *Arabidopsis thaliana* (Tuteja *et al.*, 2019). Our data showed that the nucleotide
427 diversity of imprinted genes in oil flax subgroup was significantly higher than that in
428 fiber flax subgroup (Figure 6A-C). The π values of some imprinted genes, such as
429 *Lus10010350* (PEG, Figure 6D-F), *Lus10024230* (MEG, Figure 6G-I) and
430 *Lus10041386* (MEG, Figure S8) were also significant difference between oil and fiber
431 flax subgroup. Our results revealed that imprinted genes have been undergone
432 artificial selection in the process of flax domestication from oil flax to fiber flax (Guo
433 *et al.*, 2020).

434 By analyzing the nucleotide diversity of imprinted genes in different flax subgroups,
435 we found that the π values of imprinted genes in oil flax subgroup were significantly
436 higher than those in fiber flax subgroup no matter what parental origin they were
437 (Figure 6A-C). Meanwhile, we also discovered that the imprinted genes related to

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

438 seed size and weight contained MEGs and PEGs (Table S8). It seemed that MEGs
439 and PEGs were same shaped by selective force in flax population differentiation
440 although the number of MEGs was larger than that of PEGs. So, we expected that
441 imprinted genes undergo positive selection is related to the functions, but not to the
442 parental origin which was different from the previous report (Tuteja *et al.*, 2019).
443 Compared with parental conflict theory, the imprinting under relaxed selection theory
444 that genomic imprinting evolves consistent with neo-functionalization (Rodrigues and
445 Zilberman, 2015) can better explain the intraspecific imprinting variation in flax
446 subgroups.

447 **Materials and methods**

448 **Plant Material and Tissue collection**

449 The two parental lines of flax (*Linum usitatissimum* L.) for reciprocal crosses,
450 CIli2719 (C) and Z11637 (Z), were grown at the Miquan Experiment filed in Urumqi,
451 Xinjiang. The large seed line CIli2719 which 1000-seed weight was about 10.5g
452 originated from France and the small seed line Z11637 that 1000-seed weight was
453 about 3.7g originated from the United States. The seeds of CIli2719×Z11637 (CZ)
454 and Z11637×CIli2719 (ZC) were collected at 7 DAP (day after pollination).
455 Endosperm tissues were collected from at least 50 seeds by manual dissection in each
456 replicate and were immediately frozen in liquid nitrogen. Three biological repeats
457 were set up for each line. For phenotyping, the 200 accessions were planted in four
458 environment comprising Dali in Yunnan Province in 2016 (2016DL), Urumqi in
459 Xinjiang autonomous region in 2017 and 2019 (2017UR, 2019UR), and YiLi in
460 Xinjiang autonomous region in 2019 (2019YL). Planting and phenotyping of the 200
461 accessions were performed using a same strategy as described in our previous study
462 (Guo *et al.*, 2020).

463 **Library construction for RNA-Seq**

464 Total RNA was extracted using a RNAprep Pure Plant Kit (Tiangen Biotechnology of
465 Beijing, http://www.tiangen.com/). The quantification and qualification of RNA was
466 checked by 1% agarose gels, NanoPhotometer® spectrophotometer (IMPLEN, CA,
467 USA), Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA,
468 USA) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent
469 Technologies, CA, USA). The RNA-seq libraries were generated using NEBNext®
470 Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) according to the
471 manufacturer's instructions and the high-throughput sequencing was performed with
472 the Illumina NovaSeq6000 platform. Then, the quality and quantity of these libraries
473 were assessed by using the Agilent Bioanalyzer 2100 system and Q-PCR. A data size
474 of 301.98 million 125bp paired-end raw reads was obtained from CZ and ZC.

475 **Read mapping and gene expression analysis**

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

476 After removing the reads containing adapter, reads containing ploy-N (> 10%) and
477 low quality reads ($Q_{phred} \leq 20$) from raw data, a total of 288.37 million clean reads
478 (43.26 Gb) were obtained for the following analysis. The clean reads were aligned to
479 flax reference genome
480 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Lusitatissimum)
481 (Wang *et al.*, 2012) using Hisat2 v2.0.4. HTSeq v0.9.1 to count the reads numbers
482 mapped to each gene. And then the expected number of Fragments Per Kilo base of
483 transcript sequence per Millions base pairs sequenced (FPKM) of each gene was
484 calculated based on the length of the gene and reads count mapped to this gene
485 (Trapnell *et al.*, 2010). In the three biological replicates, the gene with an average
486 expression level of FPKM > 1 was identified as "expressed" (Meng *et al.*, 2018).

487 **Identification of imprinted genes**

488 The clean reads from CZ and ZC were aligned to parental genomes CIli2719 and
489 Z11637 from our previous research (Guo *et al.*, 2020) to obtain the reads of C and Z
490 alleles at each SNP site for parental allelic expression analysis. Theoretically, the
491 allelic ratio of the maternal to paternal is 2 to 1 in hybrid endosperm. Based on the
492 2m:1p ratio, SNP loci with more than 10 alleles reads in reciprocal crosses were used
493 to perform a two-tailed chi square (χ^2) test. Moderately imprinted SNP loci had
494 significant allelic bias ($\chi^2<0.05$) and >80% of the transcripts from the maternal allele
495 for maternally expressed imprinted SNP loci or >60% of the transcripts coming from
496 the paternal allele for paternally expressed imprinted SNP loci in both reciprocal
497 hybrids. Strong maternally or paternally expressed imprinted SNP loci were defined
498 as having significant allelic bias ($\chi^2<0.01$) and >90% of transcripts derived from the
499 maternal allele or paternal allele, respectively. Complete maternally/paternally
500 expressed imprinted SNP loci had >99% of the transcripts from the maternal/paternal
501 allele (Waters *et al.*, 2013; Meng *et al.*, 2018). And the genes containing at least one
502 imprinted SNP loci were identified as imprinted genes.

503 **Validation of imprinted gene and expression analysis**

504 Thirteen genes were used to perform Quantitative RT-PCR (qRT-PCR) analysis
505 (Table S9) and twelve imprinted genes were detected using a PCR-sequencing
506 method (Table S10) (Meng *et al.*, 2018; Xu *et al.*, 2014). The endosperm cDNA
507 samples at 7-DAP were collected for RNA isolation with three biological repeats for
508 each sample.

509 The extraction, quantification and identification of total RNA were the same as that
510 of library construction for RNA-Seq. First-strand cDNA synthesis was performed
511 using 5×All-In-One RT MasterMix (with AccuRT Genomic DNA Removal Kit)
512 according to the manufacturer recommended protocol for qRT-PCR and RT-PCR
513 (abm, Cat. No.G492, http://www.abmGood.com/). Each qRT-PCR reaction of CZ and
514 ZC was performed by the manufacturer's instructions of EvaGreen Express 2×qPCR
515 MasterMix (abm, Cat. No.MasterMix-ES, http://www.abmGood.com/) and

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

516 BioRad®CFX96 Real-Time PCR system (Bio-Rad). Relative expression was
517 quantified with the geometric mean of internal reference genes *ETIFI* (eukaryotic
518 translation initiationfactor 1), *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase),
519 and *ETIF5A* (eukaryotic translation initiationfactor 5A) (Hobson and Deyholos, 2013;
520 Huis *et al.*, 2010). For RT-PCR, a 400-800bp amplification fragment of each gene
521 was amplified by different primers with four endosperm cDNA samples: CC (inbred
522 lines of CIli2719), ZZ (inbred lines of Z11637), CZ and ZC of 7-DAP endosperm.
523 The RT-PCR amplified products contained at least one imprinted SNP sites were
524 analyzed on agarose gels and then sequenced.

525 **Functional characterization of imprinted features**

526 Gene annotation of imprinted genes in flax endosperm was downloaded from the
527 reference genome (https://phytozome.jgi.doe.gov/pz/portal.html), and GO enrichment
528 analysis was carried out with WEGO (http://wego.genomics.org.cn/) (Xu *et al.*, 2014;
529 Ye *et al.*, 2006).

530 The imprinted genes of flax were investigated for sequence homology in
531 *Arabidopsis* (Gehring *et al.*, 2011), rice (Yuan *et al.*, 2017; Luo *et al.*, 2011), maize
532 (Waters *et al.*, 2013; Dong, 2017), sorghum (Zhang *et al.*, 2016a),and castor bean (Xu
533 *et al.*, 2014) using blast. The peptide sequences of flax imprinted genes were obtained
534 from the flax database in Phytozome v12.1
535 (https://phytozome.jgi.doe.gov/pz/portal.html). Then, the peptide sequences were
536 aligned to the *Arabidopsis* genome (*Arabidopsis thaliana* TAIR10) and high-scoring
537 (E-value<1E-10, 1E-20 and 1E-50) blasts hits were ordered by increasing E-value. If
538 the *Arabidopsis* imprinted genes were identified amongst the blast hits, and the gene
539 with the smallest E-value was recorded. Similarly, candidate genes from this study in
540 flax were aligned to the rice (*Oryza sativa* v7_JGI), maize (*Zea mays* Ensembl-18),
541 sorghum (*Sorghum bicolor* v3.1.1) and castor bean genome (*Ricinus communis* v0.1).
542 The Venn diagrams were drawn by the draw venn diagram online software
543 (http://bioinformatics.psb.ugent.be/webtools/Venn/).

544 For clustering analysis of imprinted genes, a standard was applied that imprinted
545 genes within 10 kb of one another in the flax genome was a candidate cluster
546 (Gehring *et al.*, 2011; Xu *et al.*, 2014). Positions of imprinted genes on chromosome
547 were mapped using the MapChart software (Voorrips, 2002).

548 **The tissue-specific expression analysis of imprinted genes in endosperm**

549 The gene-expression patterns for MEGs and PEGs in various flax tissues in reciprocal
550 hybrids were identified based on RNA-seq analysis. The endosperm and embryo
551 tissues were harvested at 7 DAP and the leaf tissues were collected at 2 weeks after
552 planting. For each sample, three biological replicates were used. The FPKM
553 expression values of all genes and imprinted genes in CZ and ZC were
554 log-transformed. All genes with FPKM>1 in endosperm were used in this study. The
555 heat map and hierarchical clustering of normalized expression levels (FPKM) were

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

556 performed with the MeV4.9.0 software (Multi Experiment Viewer,
557 https://sourceforge.net/projects/mevtm4/files/mev-tm4/) (Guo *et al.*, 2020).

558 **Phylogenetic tree and population structure analysis using imprinted SNPs**

559 To test the relationship between SNP variation at population level and population
560 structure, an individual-based neighbor-joining tree was generated based on the all
561 imprinted SNPs (498 SNPs), ME-SNPs (319 ME-SNPs) or PE-SNPs (179 PE-SNPs)
562 by TASSEL's Cladogram function (Bradbury *et al.*, 2007). We compared the kinship
563 of 200 accessions calculated by imprinted SNPs (498) and genome-wide SNPs
564 (674,074) obtained from our previous research (Guo *et al.*, 2020). Principal
565 component analysis (PCA) was conducted based on the imprinted SNPs by using the
566 software TASSEL 5.0 (Bradbury *et al.*, 2007).

567 **Identifying selection signatures of imprinted genes**

568 A set of 78 oil flax and 51 fiber flax accessions which represent two primary
569 morphotypes of cultivated flax were used for selective sweeps analysis. To test the
570 genetic diversity of imprinted genes in different subgroups, 3,191 SNPs were obtained
571 by mapping the imprinted genes to our previously constructed variation map (Guo *et*
572 *al.*, 2020). The π values of all imprinted genes (241), MEGs (143) and PEGs (98)
573 between oil and fiber subgroups were calculated at the gene level using all SNPs
574 within each imprinted gene by the software DnaSP 5.1 (Librado and Rozas, 2009).
575 Furthermore, the π values were also calculated at the SNP level in oil subgroup and
576 fiber subgroup for detecting the selection signatures in a single imprinted gene using
577 DnaSP 5.1 (Librado and Rozas, 2009).

578 **Candidate gene-based association study for seed size-related traits using flax**
579 **imprinted genes**

580 To analyze the association between imprinted genes and seed size-related traits,
581 imprinted genes were used to perform candidate gene-based association study by the
582 general linear model (GLM) and mixed linear model (MLM) in TASSEL 5.0
583 (Bradbury *et al.*, 2007). The 3,191 SNPs in the imprinted genes and seed size-related
584 traits including seed length (SL), seed width (SW) and 1,000-seed weight (1,000-SW)
585 of 200 flax accessions were obtained from our previous research (Guo *et al.*, 2020).
586 For GLM analysis, the top two principal components (PC) were used to generate the
587 population structure matrix and the threshold was set as 0.1/total SNPs ($\log_{10}(P)$ =
588 -4.50). For MLM analysis, P matrix and Kinship (K) matrix need to be considered,
589 and the suggestive threshold was set as 1/total SNPs ($\log_{10}(P)$ = -3.50). The imprinted
590 genes repeatedly detected for at least two environments or traits were considered to be
591 associated with seed size and weight.

592 **Acknowledgements**

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

593 We thank Chinese Crop Germplasm Resources Information System (CGRIS), Plant
594 Gene Resource Centre in Canada (PGRC) and United States National Plant
595 Germplasm System (U.S.NPGS) for providing us with germplasm resources. This
596 study was supported by the National Natural Science Foundation of China
597 (32060426), Resource Platform Project of Xinjiang Uygur Autonomous Region of
598 China (PT1808) and Science and Technology Innovation Project for Doctoral
599 Students of Xinjiang University of China (XJUBSCX-2017017). We are grateful to
600 professors Ming Luo and Donghe Xu for constructive comments to improve the
601 manuscript. The authors declare no competing interests of this work.

602 **Conflicts of interest statement**

603 The authors declare no competing financial interests.

604 **Availability of data and materials**

605 Raw resequencing sequence data of the two parental lines CIli2719 and Z11637 of
606 flax for reciprocal crosses are available at NCBI under accession PRJNA590636
607 (ncbi.nlm.nih.gov/bioproject/PRJNA590636) (Guo *et al.*, 2020).

608 **References**

609 Allaby RG, Peterson GW, Merriwether DA, Fu Y-B (2005). Evidence of the
610 domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the
611 *sad2* locus. *Theoretical and Applied Genetics* **112**, 58-65.
612 https://doi.org/10.1007/s00122-005-0103-3
613 Azim JB, Khan MFH, Hassan L, Robin AHK (2020). Genome-wide characterization
614 and expression profiling of plant-specific PLATZ transcription factor family genes
615 in *Brassica rapa* L. *Plant Breeding and Biotechnology* **8**, 28-45.
616 Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011).
617 Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*
618 **480**, 245–249. https://doi.org/10.1038/nature10555
619 Berger F, Grini PE, Schnittger A (2006). Endosperm: an integrator of seed growth and
620 development. *Current Opinion in Plant Biology* **9**, 664–670.
621 https://doi.org/10.1016/j.pbi.2006.09.015
622 Berger F, Vu TM, Li J, Chen B (2012). Hypothesis: selection of imprinted genes is
623 driven by silencing deleterious gene activity in somatic tissues. *Cold Spring Harbor*
624 *Symposia on Quantitative Biology* **77**, 23-29.
625 https://doi.org/10.1101/sqb.2012.77.014514
626 Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007).
627 TASSEL: software for association mapping of complex traits in diverse samples.
628 *Bioinformatics* **23**, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308
629 Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997).
630 Fertilization-independent seed development in *Arabidopsis thaliana*. *PNAS* **94**,
631 4223–4228. https://doi.org/10.1073/pnas.94.8.4223
632 Chen C, Begcy K, Liu K, Folsom JJ, Wang Z, Zhang C, Walia H (2016). Heat stress

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

633 yields a unique MADS box transcription factor in determining seed size and
634 thermal sensitivity. *Plant Physiology* **171**, 606-622.
635 https://doi.org/10.1104/pp.15.01992
636 Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E, Walichnowski A,
637 Ward K, Rowland G, Duguid S, Banik M (2012). Integrated consensus genetic and
638 physical maps of flax (*Linum usitatissimum* L.). *Theoretical and Applied Genetics*
639 **125**, 1783-1795. https://doi.org/10.1007/s00122-012-1953-0
640 Dong X 2017. The regulation of genetic imprinting in maize endosperm and its
641 stability and conservation. China Agricultural University.
642 Dong X, Zhang M, Chen J, Peng L, Zhang N, Wang X, Lai J (2017). Dynamic and
643 antagonistic allele-specific epigenetic modifications controlling the expression of
644 imprinted genes in maize endosperm. *Molecular Plant* **10**, 442–455.
645 https://doi.org/10.1016/j.molp.2016.10.007
646 Fu Y-B, Allaby RG (2010). Phylogenetic network of *Linum* species as revealed by
647 non-coding chloroplast DNA sequences. *Genetic Resources and Crop Evolution* **57**,
648 667-677. https://doi.org/10.1007/s10722-009-9502-7
649 Gehring M, Missirian V, Henikoff S (2011). Genomic analysis of parent-of-origin
650 allelic expression in *Arabidopsis thaliana* seeds. *Plos One* **6**, e23687.
651 https://doi.org/10.1371/journal.pone.0023687
652 Graaf Avd, Wardenaar R, Neumann DA, Taudt A, Shaw RG, Jansen RC, Schmitz RJ,
653 Colomé-Tatché M, Johannes F (2015). Rate, spectrum, and evolutionary dynamics
654 of spontaneous epimutations. *PNAS* **112**, 6676-6681.
655 https://doi.org/10.1073/pnas.1424254112
656 Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C (2010).
657 High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Mouse
658 Brain. *Science* **329**, 643-648. https://doi.org/10.1126/science.1190830
659 Guitton A-E, Page DR, Chambrier P, Lionnet C, Faure J-E, Grossniklaus U, Berger F
660 (2004). Identification of new members of Fertilisation Independent Seed Polycomb
661 Group pathway involved in the control of seed development in *Arabidopsis*
662 *thaliana*. *Development* **131**, 2971-2981. https://doi.org/10.1242/dev.01168
663 Guo D, Jiang H, Yan W, Yang L, Ye J, Wang Y, Yan Q, Chen J, Gao Y, Duan L, Liu H,
664 Xie L (2020). Resequencing 200 Flax Cultivated Accessions Identifies Candidate
665 Genes Related to Seed Size and Weight and Reveals Signatures of Artificial
666 Selection. *Frontiers in Plant Science* **10**, 1682.
667 https://doi.org/10.3389/fpls.2019.01682
668 Hanna CW, Kelsey G (2017). Genomic imprinting beyond DNA methylation: a role
669 for maternal histones. *Genome Biology* **18**, 177.
670 https://doi.org/10.1186/s13059-017-1317-9
671 Hobson N, Deyholos MK (2013). Genomic and expression analysis of the flax (*Linum*
672 *usitatissimum*) family of glycosyl hydrolase 35 genes. *BMC Genomics* **14**, 344.
673 https://doi.org/10.1186/1471-2164-14-344
674 Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

675 RC, Harada JJ, Zilberman D, Fischer RL (2011). Regulation of imprinted gene
676 expression in *Arabidopsis* endosperm. *PNAS* **108**, 1755-1762.
677 https://doi.org/10.1073/pnas.1019273108
678 Huang F, Zhu Qh, Zhu A, Wu X, Xie L, Wu X, Helliwell C, Chaudhury A, Finnegan
679 EJ, Luo M (2017). Mutants in the imprinted *PICKLE RELATED 2* gene suppress
680 seed abortion of fertilization independent seed class mutants and paternal excess
681 interploidy crosses in *Arabidopsis*. *The Plant Journal* **90**, 383–395
682 Huang X, Lu Z, Wang X, Ouyang Y, Chen W, Xie K, Wang D, Luo M, Luo J, Yao J
683 (2016). Imprinted gene *OsFIE1* modulates rice seed development by influencing
684 nutrient metabolism and modifying genome H3K27me3. *The Plant Journal* **87**,
685 305–317.
686 Huis R, Hawkins S, Neutelings G (2010). Selection of reference genes for quantitative
687 gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant*
688 *Biology* **10**, 71. https://doi.org/10.1186/1471-2229-10-71
689 Hutter B, Bieg M, Helms V, Paulsen M (2010). Divergence of imprinted genes during
690 mammalian evolution. *BMC Evolutionary Biology* **10**, 116.
691 https://doi.org/10.1186/1471-2148-10-116
692 Joanis V, Lloyd VK (2002). Genomic imprinting in *Drosophila* is maintained by the
693 products of *Suppressor of variegation* and *trithorax* group, but not *Polycomb* group,
694 genes. *Molecular Genetics and Genomics* **268**, 103-112.
695 https://doi.org/10.1007/s00438-002-0731-0
696 Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999). Imprinting of
697 the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. *The Plant Cell* **11**,
698 1945-1952. https://doi.org/10.2307/3871089
699 Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny JR, Wells D, Katz A,
700 Margossian L, Harada JJ, Goldberg RB, Fischer RL (1999). Control of
701 fertilization-independent endosperm development by the *MEDEA* polycomb gene
702 in *Arabidopsis*. *PNAS* **96**, 4186-4191. https://doi.org/10.1073/pnas.96.7.4186
703 Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005). The *Arabidopsis thaliana*
704 MEDEA Polycomb group protein controls expression of PHERES1 by parental
705 imprinting. *Nature Genetics* **37**, 28-30. https://doi.org/10.1038/ng1495
706 Kradolfer D, Wolff P, Jiang H, Siretskiy A, Köhler C (2013). An imprinted gene
707 underlies postzygotic reproductive isolation in *Arabidopsis thaliana*.
708 *Developmental Cell* **26**, 525–535. https://doi.org/10.1016/j.devcel.2013.08.006
709 Li Q, Song J, West PT, Zynda G, Eichten SR, Vaughn MW, Springer NM (2015).
710 Examining the causes and consequences of context-specific differential DNA
711 methylation in maize. *Plant Physiology* **168**, 1262-1274.
712 https://doi.org/10.1104/pp.15.00052
713 Li Q, Wang J, Ye J, Zheng X, Xiang X, Li C, Fu M, Wang Q, Zhang Z, Wu Y (2017).
714 The maize imprinted gene *Floury3* encodes a PLATZ protein required for tRNA
715 and 5S rRNA transcription through interaction with RNA polymerase III. *The Plant*
716 *Cell* **29**, 2661-2675. https://doi.org/10.1105/tpc.17.00576

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

717 Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA
718 polymorphism data. *Bioinformatics* **25**, 1451-1452.
719 https://doi.org/10.1093/bioinformatics/btp187
720 Lu J, Zhang C, Baulcombe DC, Chen ZJ (2012). Maternal siRNAs as regulators of
721 parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds.
722 *PNAS* **109**, 5529-5534. https://doi.org/10.1073/pnas.1203094109
723 Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000). Expression and
724 parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of
725 developing *Arabidopsis* seeds. *PNAS* **97**, 10637–10642.
726 https://doi.org/10.1073/pnas.170292997
727 Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM (1999).
728 Genes controlling fertilization-independent seed development in *Arabidopsis*
729 *thaliana*. *PNAS* **96**, 296-301. https://doi.org/10.1073/pnas.96.1.296
730 Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A
731 (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting
732 primarily occurs in the endosperm. *PLoS Genetics* **7**, e1002125.
733 https://doi.org/10.1371/journal.pgen.1002125
734 Mahajan M, Ahuja PS, Yadav SK (2011). Post-transcriptional silencing of flavonol
735 synthase mRNA in tobacco leads to fruits with arrested seed set. *Plos One* **6**,
736 e28315. https://doi.org/10.1371/journal.pone.0028315
737 Mei X, Liu C, Yu T, Liu X, Xu D, Wang J, Wang G, Cai Y (2015). Identification and
738 characterization of paternal-preferentially expressed gene *NF-YC8* in maize
739 endosperm. *Molecular Genetics and Genomics* **290**, 1819–1831.
740 https://doi.org/10.1007/s00438-015-1043-5
741 Meng D, Zhao J, Zhao C, Luo H, Jin W (2018). Sequential gene activation and gene
742 imprinting during early embryo development in maize. *The Plant Journal* **93**,
743 445-459.
744 Miyake T, Takebayashi N, Wolf DE (2009). Possible diversifying selection in the
745 imprinted gene, *MEDEA*, in *Arabidopsis*. *Molecular Biology and Evolution* **26**,
746 843-857. https://doi.org/10.1093/molbev/msp001
747 Moreno-Romero J, León GDT-D, Yadav VK, Santos-González J, Köhler C (2019).
748 Epigenetic signatures associated with imprinted paternally expressed genes in the
749 *Arabidopsis* endosperm. *Genome Biology* **20**, 41.
750 https://doi.org/10.1186/s13059-019-1652-0
751 Ohad N, Margossian L, Hsu Y-C, Williams C, Repetti P, Fischer RL (1996). A
752 mutation that allows endosperm development without fertilization. *PNAS* **93**,
753 5319-5324. https://doi.org/10.1073/pnas.93.11.5319
754 Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB,
755 Fischer RL (1999). Mutations in *FIE*, a WD polycomb group gene, allow
756 endosperm development without fertilization. *The Plant Cell* **11**, 407-415.
757 https://doi.org/10.1105/tpc.11.3.407
758 Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M (2014). Natural

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

759 epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene
760 imprinting. *eLife* **3**, e03198. https://doi.org/10.7554/eLife.03198
761 Piskurewicz U, Iwasaki M, Susaki D, Megies C, Kinoshita T, Lopez-Molina L (2016).
762 Dormancy-specific imprinting underlies maternal inheritance of seed dormancy in
763 *Arabidopsis thaliana*. *eLife* **5**, e19573. https://doi.org/10.7554/elife.19573
764 Renfree MB, Suzuki S, Kaneko-Ishino T (2012). The origin and evolution of genomic
765 imprinting and viviparity in mammals. *Philosophical Transactions of the Royal*
766 *Society B Biological Sciences* **368**, 20120151.
767 https://doi.org/10.1098/rstb.2012.0151
768 Rodrigues JA, Zilberman D (2015). Evolution and function of genomic imprinting in
769 plants. *Genes and Development* **29**, 2517–2531.
770 https://doi.org/10.1101/gad.269902.115
771 Sabelli PA, Larkins BA (2009). The development of endosperm in grasses. *Plant*
772 *Physiology* **149**, 14–26. https://doi.org/10.1104/pp.108.129437
773 Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork
774 NJ, Ecker JR (2011). Transgenerational epigenetic instability is a source of novel
775 methylation variants. *Science* **334**, 369–373.
776 https://doi.org/10.1126/science.1212959
777 Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A,
778 McCosh RB, Chen H, Schork NJ, Ecker JR (2013). Patterns of population
779 epigenomic diversity. *Nature* **495**, 193–198. https://doi.org/10.1038/nature11968
780 Scott RJ, Spielman M, Bailey J, Dickinson HG (1998). Parent-of-origin effects on
781 seed development in *Arabidopsis thaliana*. *Development* **125**, 3329-3341.
782 Sekine D, Ohnishi T, Furuumi H, Ono A, Yamada T, Kurata N, Kinoshita T (2013).
783 Dissection of two major components of the post-zygotic hybridization barrier in
784 rice endosperm. *The Plant Journal* **76**, 792-799. https://doi.org/10.1111/tpj.12333
785 Sha K (2008). A mechanistic view of genomic imprinting. *Annu Rev Genomics Hum*
786 *Genet* **9**, 197-216. https://doi.org/10.1146/annurev.genom.122007.110031
787 Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo J-M, Baroux C,
788 Gagliardini V, Page DR, Wolfe KH, Grossniklaus U (2007). Positive darwinian
789 selection at the imprinted *MEDEA* locus in plants. *Nature* **448**, 349-352.
790 https://doi.org/10.1038/nature05984
791 Sreenivasulu N, Wobus U (2013). Seed-development programs: A systems
792 biology–based comparison between dicots and monocots. *Annual Review of Plant*
793 *Biology* **64**, 189-217. https://doi.org/10.1146/annurev-arplant-050312-120215
794 Sun C, Tang T, Tang W, Sui L, Zhang H, Xia Z, Zhao H, Han Y, Lin L (2017). Plant
795 genomic imprinting and its effect on seed development. *Agricultural Science and*
796 *Technology* **18**, 984-987. https://doi.org/10.16175/j.cnki.1009-4229.2017.06.006
797 Sun R-h 2018. Characterization and Functional Analvsis of MADS-box from coconut
798 (*Cocos nucifera* L .) endosperm. [Hainan University.
799 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJv, Salzberg SL,
800 Wold BJ, Pachter L (2010). Transcript assembly and quantification by RNA-Seq

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

801 reveals unannotated transcripts and isoform switching during cell differentiation.
802 *Nature Biotechnology* **28**, 511-515. https://doi.org/10.1038/nbt.1621
803 Tuteja R, McKeown PC, Ryan P, Morgan CC, Donoghue MTA, Downing T,
804 O'Connell MJ, Spillane C (2019). Paternally expressed imprinted genes under
805 positive Darwinian selection in *Arabidopsis thaliana*. *Molecular Biology and*
806 *Evolution* **36**, 1239–1253. https://doi.org/10.1093/molbev/msz063
807 Voorrips RE (2002). MapChart: software for the Graphical Presentation of Linkage
808 Maps and QTLs. *Journal of Heredity* **93**, 77-78.
809 https://doi.org/10.1093/jhered/93.1.77
810 Wang A, Hou Q, Si L, Huang X, Luo J, Lu D, Zhu J, Shangguan Y, Miao J, Xie Y,
811 Wang Y, Zhao Q, Feng Q, Zhou C, Li Y, Fan D, Lu Y, Tian Q, Wang Z, Han B
812 (2019). The PLATZ transcription factor GL6 affects grain length and number in
813 rice. *Plant Physiology* **180**, 2077-2090. https://doi.org/10.1104/pp.18.01574
814 Wang L, Yuan J, Ma Y, Jiao W, Ye W, Yang D, Yi C, Chen ZJ (2018). Rice interploidy
815 crosses disrupt epigenetic regulation, gene expression, and seed development.
816 *Molecular Plant* **11**, 300-314. https://doi.org/10.1016/j.molp.2017.12.006
817 Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S,
818 Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk
819 QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK-S, Wang J,
820 Deyholos MK (2012). The genome of flax (*Linum usitatissimum*) assembled *de*
821 *novo* from short shotgun sequence reads. *The Plant Journal* **72**, 461–473.
822 https://doi.org/10.1111/j.1365-313X.2012.05093.x
823 Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer
824 NM (2013). Comprehensive analysis of imprinted genes in maize reveals allelic
825 variation for imprinting and limited conservation with other species. *PNAS* **110**,
826 19639–19644. https://doi.org/10.1073/pnas.1309182110
827 Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C-T, Xu W,
828 Schnable PS, Vaughn MW, Gehring M, Springer NM (2011). Parent-of-origin
829 effects on gene expression and DNA methylation in the maize endosperm. *The*
830 *Plant Cell* **23**, 4221-4233. https://doi.org/10.1105/tpc.111.092668
831 Wolff P, Jiang H, Wang G, Santos-González J, Köhler C (2015). Paternally expressed
832 imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*.
833 *eLife* **4**, e10074. https://doi.org/10.7554/eLife.10074
834 Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C,
835 Nordborg M, Rehmsmeier M, Köhler C (2011). High-resolution analysis of
836 parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genetics* **7**,
837 e1002126. https://doi.org/10.1371/journal.pgen.1002126
838 Xu J, Chen G, Hermanson PJ, Xu Q, Sun C, Chen W, Kan Q, Li M, Crisp PA, Yan J,
839 Li L, Springer NM, Li Q (2019). Population-level analysis reveals the widespread
840 occurrence and phenotypic consequence of DNA methylation variation not tagged
841 by genetic variation in maize. *Genome Biology* **20**, 243.
842 https://doi.org/10.1186/s13059-019-1859-0

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

843 Xu W, Dai M, Li F, Liu A (2014). Genomic imprinting, methylation and
844 parent-of-origin effects in reciprocal hybrid endosperm of castor bean. *Nucleic*
845 *Acids Research* **42**, 6987-6998. https://doi.org/10.1093/nar/gku375
846 Yang X, Wu F, Lin X, Du X, Chong K, Gramzow L, Schilling S, Becker A, Theißen G,
847 Meng Z (2012). Live and let die - the B_{sister} MADS-box gene *OsMADS29* controls
848 the degeneration of cells in maternal tissues during seed development of rice
849 (*Oryza sativa*). *Plos One* **7**, e51435. https://doi.org/10.1371/journal.pone.0051435
850 Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L,
851 Wang J (2006). WEGO : a web tool for plotting GO annotations. *Nucleic Acids*
852 *Research* **34**, W293-W297. https://doi.org/10.1093/nar/gkl031
853 Yuan J, Chen S, Jiao W, Wang L, Wang L, Ye W, Lu J, Hong D, You S, Cheng Z, Yang
854 D, Chen ZJ (2017). Both maternally and paternally imprinted genes regulate seed
855 development in rice. *New Phytologist* **216**, 373–387.
856 Zhang J, Qi Y, Wang L, Wang L, Yan X, Dang Z, Li W, Zhao W, Pei X, Li X, Liu M,
857 Tan M, Wang L, Long Y, Wang J, Zhang X, Dang Z, Zheng H, Liu T (2020).
858 Genomic comparison and population diversity analysis provide insights into the
859 domestication and improvement of flax. *iScience* **23**, 100967.
860 https://doi.org/10.1016/j.isci.2020.100967
861 Zhang M, Li N, He W, Zhang H, Yang W, Liu B (2016a). Genome-wide screen of
862 genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine
863 methylation. *The Plant Journal* **85**, 424-436.
864 Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y,
865 Song W, Lai J (2011). Extensive, clustered parental imprinting of protein-coding
866 and noncoding RNAs in developing maize endosperm. *PNAS* **108**, 20042–20047.
867 https://doi.org/10.1073/pnas.1112186108
868 Zhang S, Feng D, Tian X, Yang X (2003). Progress on imprinted genes in mammals.
869 *Progress in Biotechnology* **23**, 48-54.
870 https://doi.org/10.3969/j.issn.1671-8135.2003.12.011
871 Zhang W, Cao S, Jiang L, Zhu S, Wan J (2005). Genomic imprinting and seed
872 development. *Hereditas* **27**, 665-670.
873 https://doi.org/10.3321/j.issn:0253-9772.2005.04.029
874 Zhang X-J, Jiang W-B, Pang Y-Z (2016b). Advances in the regulation mechanism of
875 plant seed size. *Plant Physiology Journal* **52**, 998-1010.
876 https://doi.org/10.13592/j.cnki.ppj.2016.0159
877 Zhou S-R, Xue H-W (2020). The rice PLATZ protein SHORT GRAIN6 determines
878 grain size by regulating spikelet hull cell division. *Journal of Integrative Plant*
879 *Biology* **62**, 847-864. https://doi.org/10.1111/jipb.12851

880 **Figure legends**

881 **Figure 1. Identification of imprinted genes in flax endosperm**. (A) The proportion
882 of parental transcripts in both CZ and ZC was plotted for 52,794 SNPs with at least
883 ten reads could be assigned to a specific allele. The shaded areas indicated moderate

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

884 (pink), strong (blue), or complete (arrows) maternally expressed imprinted SNP loci
885 (upper right) or paternally expressed imprinted SNP loci (lower left). (B) The number
886 of moderate (pink), strong (blue), and complete (gray) imprinted SNP loci and
887 imprinted genes in endosperm. ME-SNPs, maternally expressed imprinted SNP loci;
888 PE-SNPs, paternally expressed imprinted SNP loci; MEG, maternally expressed
889 imprinted genes; PEG, paternally expressed imprinted gene.

890 **Figure 2. The conservation of flax imprinted genes between other species.** (A)
891 Venn diagram showing overlaps of imprinted genes at E-value<1E-10 between flax
892 and maize, rice, *Arabidopsis,* castor bean, sorghum, respectively. (B) Venn diagram
893 showing overlaps of imprinted genes at E-value<1E-20. (C) Venn diagram showing
894 overlaps of imprinted genes at E-value<1E-50.

895 **Figure 3. Some flax imprinted genes are located in mini-clusters.** The MEGs (red)
896 and PEGs (blue) were mapped to the 15 flax chromosomes. Genes clustered within 10
897 kb are boxed in green.

898 **Figure 4. Expression of flax imprinted genes over various flax tissues in both**
899 **reciprocal hybrids based on RNA-seq analysis.** (A, B) The gene-expression
900 patterns for MEGs (A) and PEGs (B). endo-MEGs, MEGs that expressed
901 preferentially in endosperm; con-MEGs, MEGs that also expressed in other tissues;
902 endo-PEGs, PEGs that expressed preferentially in endosperm; con-PEGs, PEGs that
903 also expressed in other tissues. The normalized values were used for hierarchical
904 clustering and the heat map indicated relative levels of expression. The endosperm
905 and embryo tissues were harvested at 7 DAP and the leaf tissues were collected at 2
906 weeks after planting. For each sample, three biological replicates were used. (C, D)
907 The Log_{10} (FPKM) values of CZ (C) and ZC (D). The Log_{10} (FPKM) values of CZ (C)
908 and ZC (D). All genes with FPKM>1 in endosperm were used in this study. The
909 values listed above box plots were the number of genes in each group. Asterisks
910 indicate the significance level (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

911 **Figure 5. Imprinted genes differentiate flax subgroups.** (A) The correlation of
912 kinships between imprinted SNPs or all SNPs. (B) Phylogenetic tree of 200 flax
913 accessions inferred from imprinted SNPs. (C) PCA plots of all imprinted SNPs,
914 ME-SNPs and PE-SNPs. Fiber flax, oil-fiber dual purpose flax (OF), and Oil flax
915 were represented in red, blue and green colors, respectively.

916 **Figure 6. Distribution of nucleotide diversity (π) within imprinted genes and**
917 **allele frequency differences of two genes across oil and fiber subgroups.** (A-C)
918 Boxplots for nucleotide diversity of all imprinted genes (A), MEGs (B) and PEGs (C)
919 across Oil (green) and Fiber (red) groups. (D, G) The distribution of allele frequency
920 of SNPs located in *Lus10010350* (D, PEG) and *Lus10024230* (G, MEG) in Oil and
921 Fiber subgroups. The alternate alleles and reference alleles were shown in purple and
922 orange, respectively. (E, H) The nucleotide diversity distribution of *Lus10010350* on
923 chromosome 15 (E) and *Lus10024230* on chromosome 12 (H) among Oil and Fiber

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

924 subgroups. (F, I) Boxplots for nucleotide diversity of *Lus10010350* (F) and
925 *Lus10024230* (I) among Oil and Fiber subgroups. The difference was analyzed by
926 two-tailed *t* tests.

927 **Figure 7. Imprinted gene-based association study for seed size and weight, and**
928 **identification of a causal gene for the peak on chromosome 15.** (A-B) The
929 overlapping Manhattan plots for seed length, seed width and 1,000-seed weight in
930 four environments (including 2016DL, 2017UR, 2019UR, 2019YL) using GLM (A)
931 and MLM (B) models. Imprinted genes repeatedly detected at least two environments
932 or traits related to seed size and weight were marked by red arrows. (C) Local
933 Manhattan plots for seed width in 2016DL of imprinted gene-based association
934 analysis surrounding the peak on chromosome 15. The position of *Lus10010350* was
935 highlighted by a shaded pink column. (D) Gene structure of *Lus10010350.* (E) DNA
936 polymorphism in *Lus10010350.* (F) Schematic representation of the structural
937 variation in *Lus10010350.* (G-J) Boxplots for 1,000-seed weight based on the
938 haplotypes (Hap.) for *Lus10010350* in 2016DL (G), 2017UR (H), 2019UR (I) and
939 2019YL (J). In the box plots, the center line represented the median, box limits
940 indicated the upper and lower quartiles, whiskers marked the range of the data and
941 points showed outliers. *n* indicated the number of accessions with the same genotype.
942 The difference between haplotypes was analyzed by two-tailed *t* tests.

943 **Supplementary information**

944 **Figure S1.** Flow chart for identification of imprinted genes in flax endosperm.

945 **Figure S2.** Verification of thirteen genes in flax endosperm based on qRT-PCR
946 analysis. Thirteen genes were chosen for the qRT-PCR analyses. Among these genes,
947 five were MEGs, five were PEGs, and others were not imprinted genes. So, the gene
948 expression level between qRT-PCR and RNA sequencing of these thirteen genes
949 represented the whole types of genes in this study.

950 **Figure S3.** Validation of the imprinted genes in flax endosperm by PCR sequencing.
951 Twelve imprinted genes including nine MEGs and three PEGs were selected for
952 validation. Each gene was designed by a pair of primers with a 400-800bp
953 amplification fragment which was a part of the corresponding CDS sequence of CC
954 (endosperm of CIli2719 self-cross), ZZ (endosperm of Z11637 self-cross), CZ
955 (endosperm of CIli2719×Z11637), and ZC (endosperm of Z11637×CIli2719) and the
956 amplification fragment contained at least one imprinted SNP site.

957 **Figure S4.** Gene ontology analysis of 229 identified imprinted genes and 115
958 conserved imprinted genes. (A) Gene ontology analysis of identified imprinted genes.
959 MEGs represented 135 maternally expressed genes (red), PEGs represented 94
960 paternally expressed genes (purple), All imprinted genes represented 229 moderate
961 imprinted genes (blue), All genes represented all endosperm-expressed genes with at
962 least ten reads could be assigned to a specific allele in both CZ and ZC (yellow). (B)

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

963 Gene ontology analysis of conserved imprinted genes. Conserved MEGs represented
964 75 conserved maternally expressed genes (red), Conserved PEGs represented 40
965 conserved paternally expressed genes (purple), Conserved imprinted genes
966 represented 115 conserved imprinted genes (blue), All genes were presented as A.

967 **Figure S5.** Expression of conserved imprinted genes in different tissues of flax in
968 both reciprocal hybrids CZ and ZC based on RNA-seq analysis. (A, B) The
969 gene-expression patterns for MEGs (A) and PEGs (B) of 115 conserved imprinted
970 genes. (C) The gene-expression patterns for MEGs and PEGs of 38 conserved strong
971 imprinted genes. endo-MEGs, MEGs that expressed preferentially in endosperm;
972 con-MEGs, MEGs that expressed in many tissues; endo-PEGs, PEGs that expressed
973 preferentially in endosperm; con-PEGs, PEGs that expressed in many tissues. The
974 normalized values were used for hierarchical clustering and the heat map indicates
975 relative levels of expression. The endosperm and embryo tissues were harvested at 7
976 DAP and the leaf tissues were collected at 2 weeks after planting. For each sample,
977 three biological replicates were used.

978 **Figure S6.** MEGs and PEGs can differentiate flax subgroups. (A) Phylogenetic tree of
979 200 flax accessions inferred from ME-SNPs. (B) Phylogenetic tree of 200 flax
980 accessions inferred from PE-SNPs. ME-SNPs, maternally expressed imprinted SNP
981 loci; PE-SNPs, paternally expressed imprinted SNP loci. Fiber flax, oil-fiber dual
982 purpose flax (OF), and Oil flax were represented in red, blue and green colors,
983 respectively.

984 **Figure S7.** Boxplots for seed length and seed width based on the haplotypes (Hap.)
985 for *Lus10010350.* (A-B) The seed length (A) and seed width (B) in 2016DL. (C-D)
986 The seed length (C) and seed width (D) in 2017UR. (E-F) The seed length (E) and
987 seed width (F) in 2019UR. (G-H) The seed length (G) and seed width (H) in 2019YL.
988 In the box plots, the center line represented the median, box limits indicated the upper
989 and lower quartiles, whiskers marked the range of the data and points showed outliers.
990 *n* indicates the number of accessions with the same genotype. The difference between
991 haplotypes was analyzed by two-tailed *t* tests.

992 **Figure S8.** The nucleotide diversity distribution of *Lus10041386*. (A) The nucleotide
993 diversity distribution of *Lus10041386* on chromosome 15 among Oil and Fiber
994 subgroups. (B) Boxplots for nucleotide diversity of *Lus10041386* among Oil and
995 Fiber subgroups. The difference was analyzed by two-tailed *t* tests.

996 **Table S1.** Imprinted genes in both hybrid endosperms and associated SNPs.

997 **Table S2.** Functional annotations of 229 imprinted protein-coding genes.

998 **Table S3.** Gene ontology enrichment analysis of 229 flax imprinted genes.

999 **Table S4.** The conservation of imprinted genes detected in our endosperm samples.

1000 **Table S5.** The number of conserved flax imprinted genes with other species.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1001 **Table S6.** Gene ontology enrichment analysis of 115 conserved imprinted genes in
1002 flax.

1003 **Table S7.** Clusters of imprinted genes.

1004 **Table S8.** Imprinted genes related to seed size and 1000-seed weight based on
1005 candidate gene-based association study.

1006 **Table S9.** Primers for qRT-PCR.

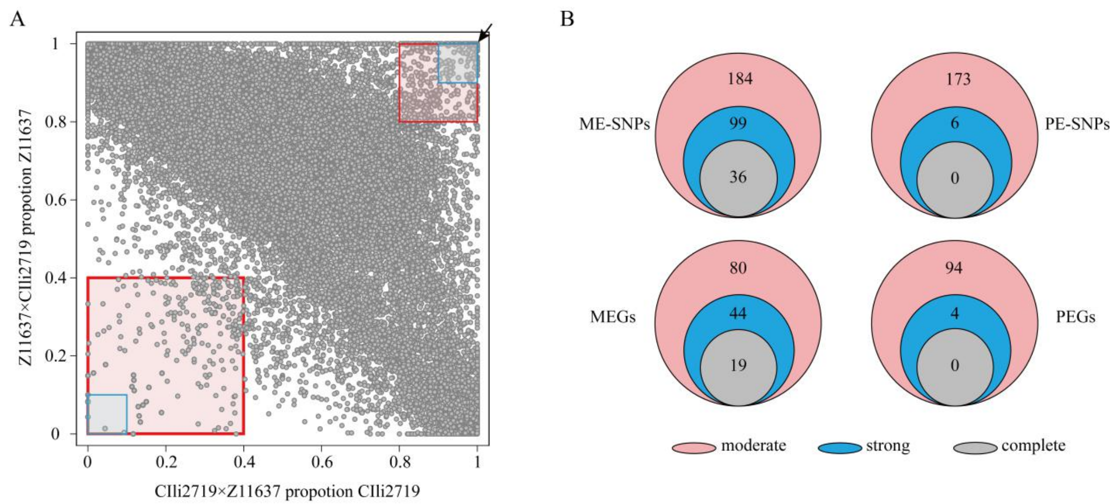
1007 **Table S10.** Primers for imprinting validation by PCR-sequencing.

1008

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1009

1010



1011

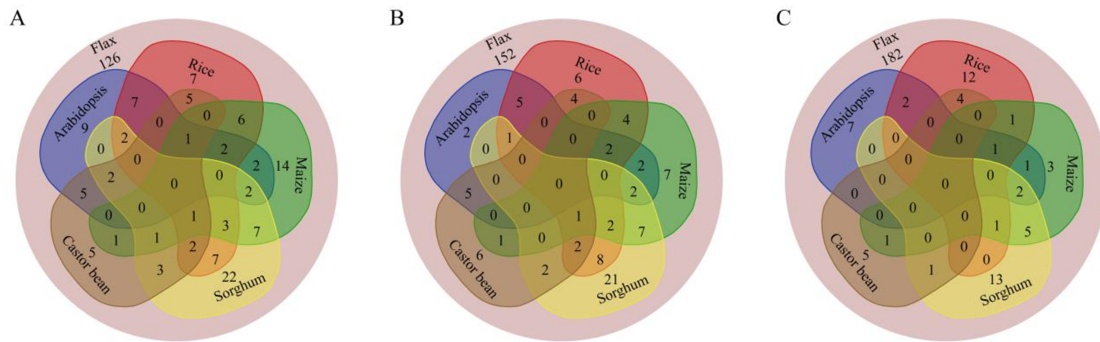
1012 **Figure 1.** Identification of imprinted genes in flax endosperm. (A) The proportion of parental
1013 transcripts in both CZ and ZC was plotted for 52,794 SNPs with at least ten reads could be assigned to
1014 a specific allele. The shaded areas indicated moderate (pink), strong (blue), or complete (arrows)
1015 maternally expressed imprinted SNP loci (upper right) or paternally expressed imprinted SNP loci
1016 (lower left). (B) The number of moderate (pink), strong (blue), and complete (gray) imprinted SNP loci
1017 and imprinted genes in endosperm. ME-SNPs, maternally expressed imprinted SNP loci; PE-SNPs,
1018 paternally expressed imprinted SNP loci; MEG, maternally expressed imprinted genes; PEG, paternally
1019 expressed imprinted gene.

1020

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1021

1022



1023

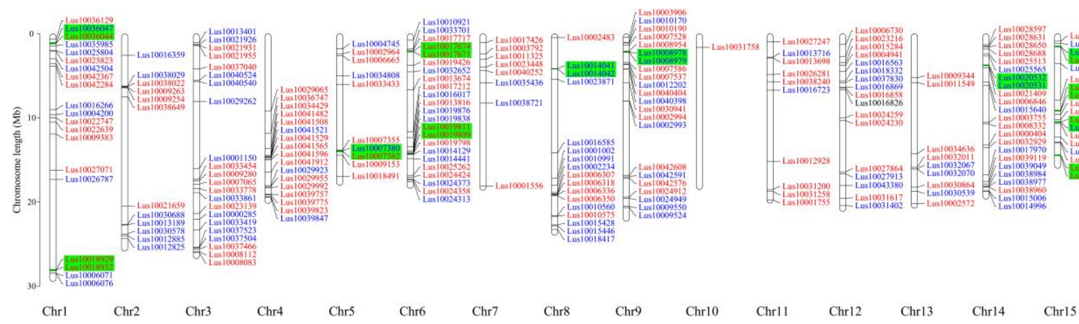
1024 **Figure 2.** The conservation of flax imprinted genes between other species. (A) Venn diagram showing
1025 overlaps of imprinted genes at E-value<1E-10 between flax and maize, rice, *Arabidopsis,* castor bean,
1026 sorghum, respectively. (B) Venn diagram showing overlaps of imprinted genes at E-value<1E-20. (C)
1027 Venn diagram showing overlaps of imprinted genes at E-value<1E-50.

1028

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1029

1030

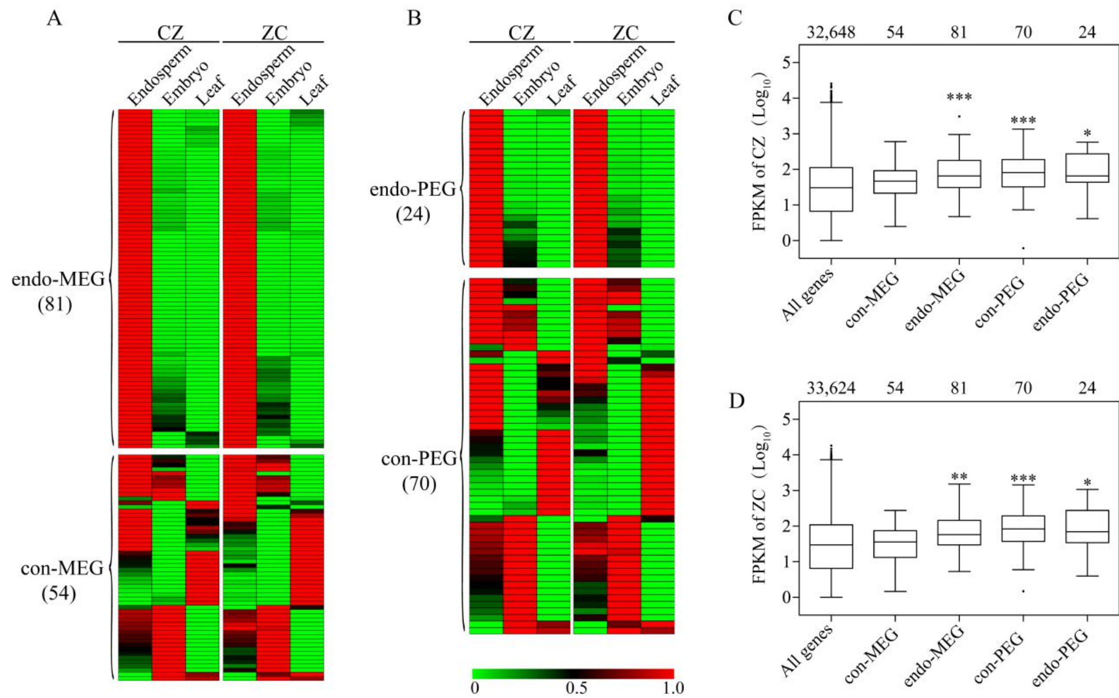


1031

1032 **Figure 3.** Some flax imprinted genes are located in mini-clusters. The MEGs (red) and PEGs (blue)
1033 were mapped to the 15 flax chromosomes. Genes clustered within 10 kb are boxed in green.

1034

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

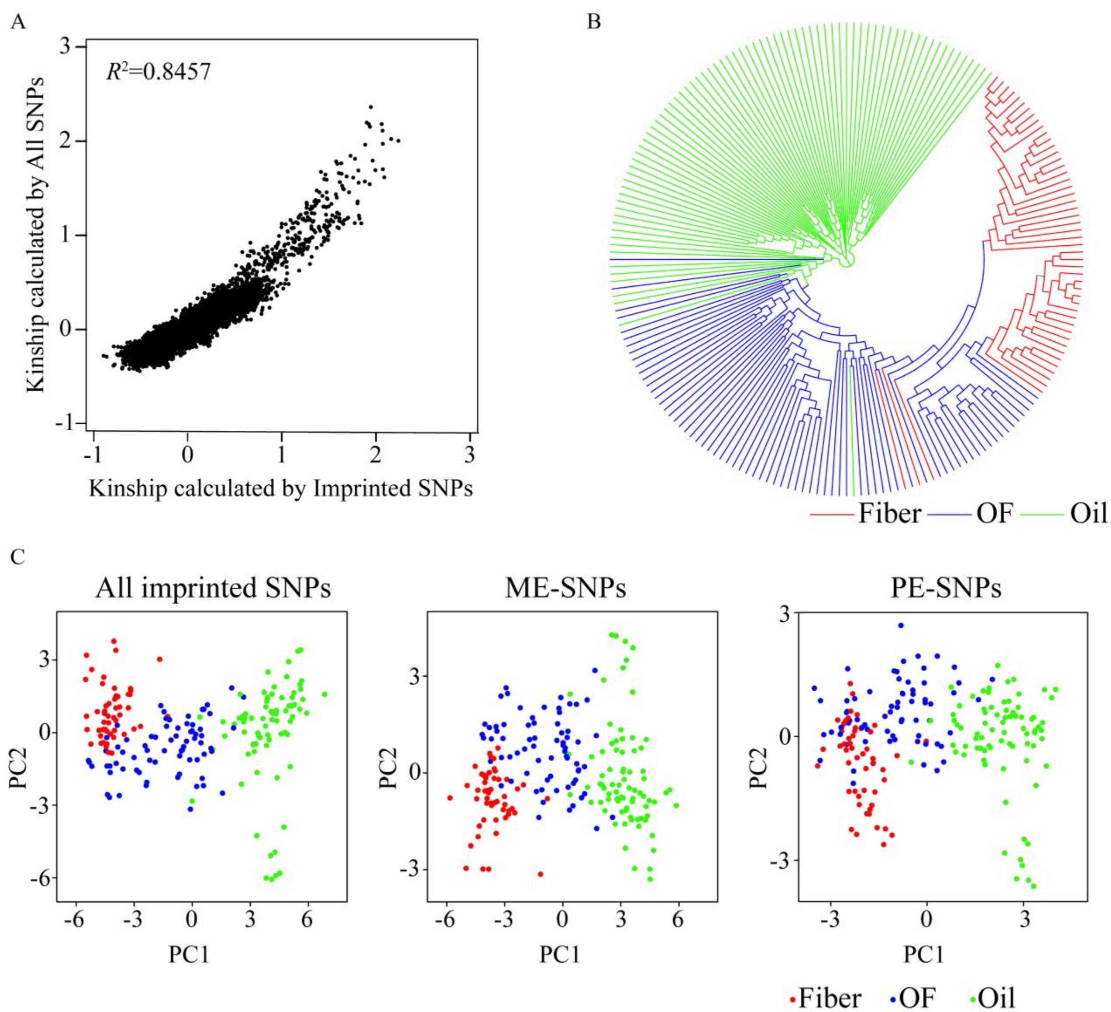


1035

1036 **Figure 4.** Expression of flax imprinted genes over various flax tissues in both reciprocal hybrids based
1037 on RNA-seq analysis. (A, B) The gene-expression patterns for MEGs (A) and PEGs (B). endo-MEGs,
1038 MEGs that expressed preferentially in endosperm; con-MEGs, MEGs that also expressed in other
1039 tissues; endo-PEGs, PEGs that expressed preferentially in endosperm; con-PEGs, PEGs that also
1040 expressed in other tissues. The normalized values were used for hierarchical clustering and the heat
1041 map indicated relative levels of expression. The endosperm and embryo tissues were harvested at 7
1042 DAP and the leaf tissues were collected at 2 weeks after planting. For each sample, three biological
1043 replicates were used. (C, D) The Log_{10} (FPKM) values of CZ (C) and ZC (D). The Log_{10} (FPKM)
1044 values of CZ (C) and ZC (D). All genes with FPKM>1 in endosperm were used in this study. The
1045 values listed above box plots were the number of genes in each group. Asterisks indicate the
1046 significance level (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

1047

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

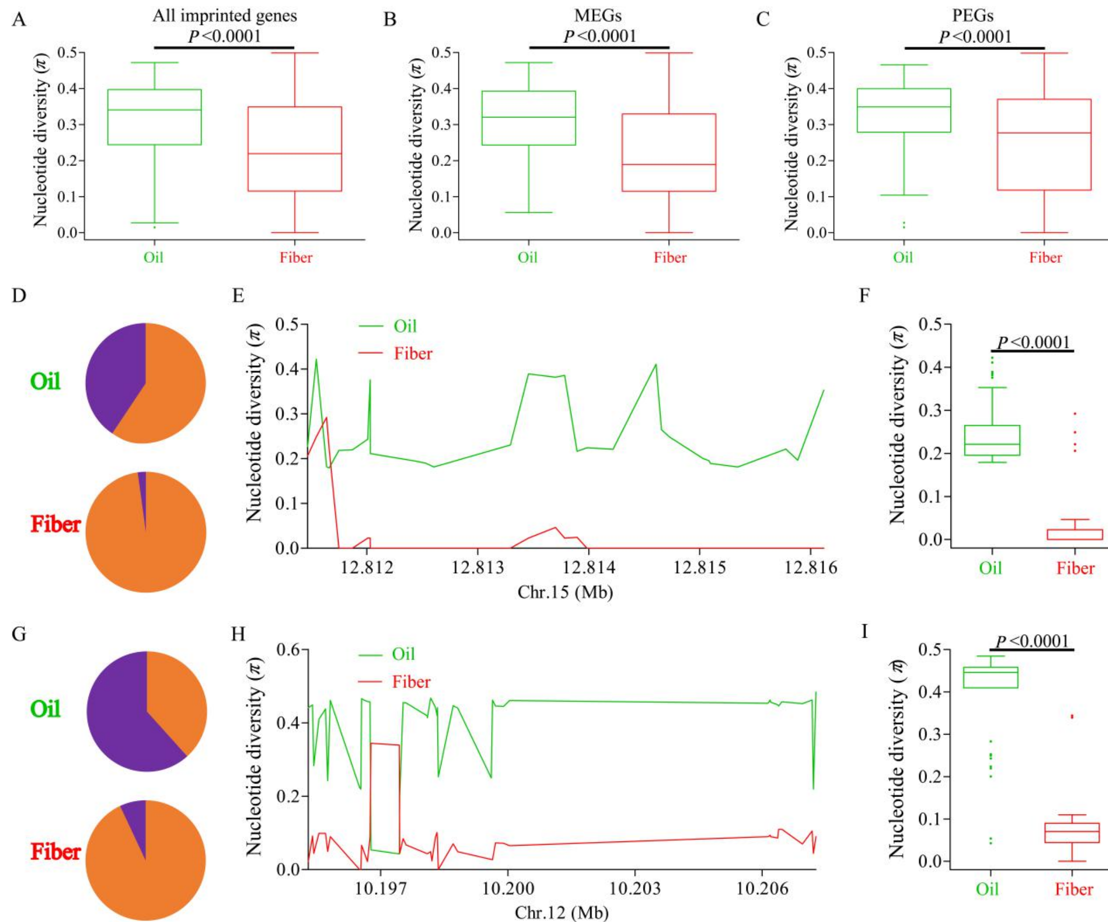


1048

1049 **Figure 5.** Imprinted genes differentiate flax subgroups. (A) The correlation of kinships between
1050 imprinted SNPs or all SNPs. (B) Phylogenetic tree of 200 flax accessions inferred from imprinted
1051 SNPs. (C) PCA plots of all imprinted SNPs, ME-SNPs and PE-SNPs. Fiber flax, oil-fiber dual purpose
1052 flax (OF), and Oil flax were represented in red, blue and green colors, respectively.

1053

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

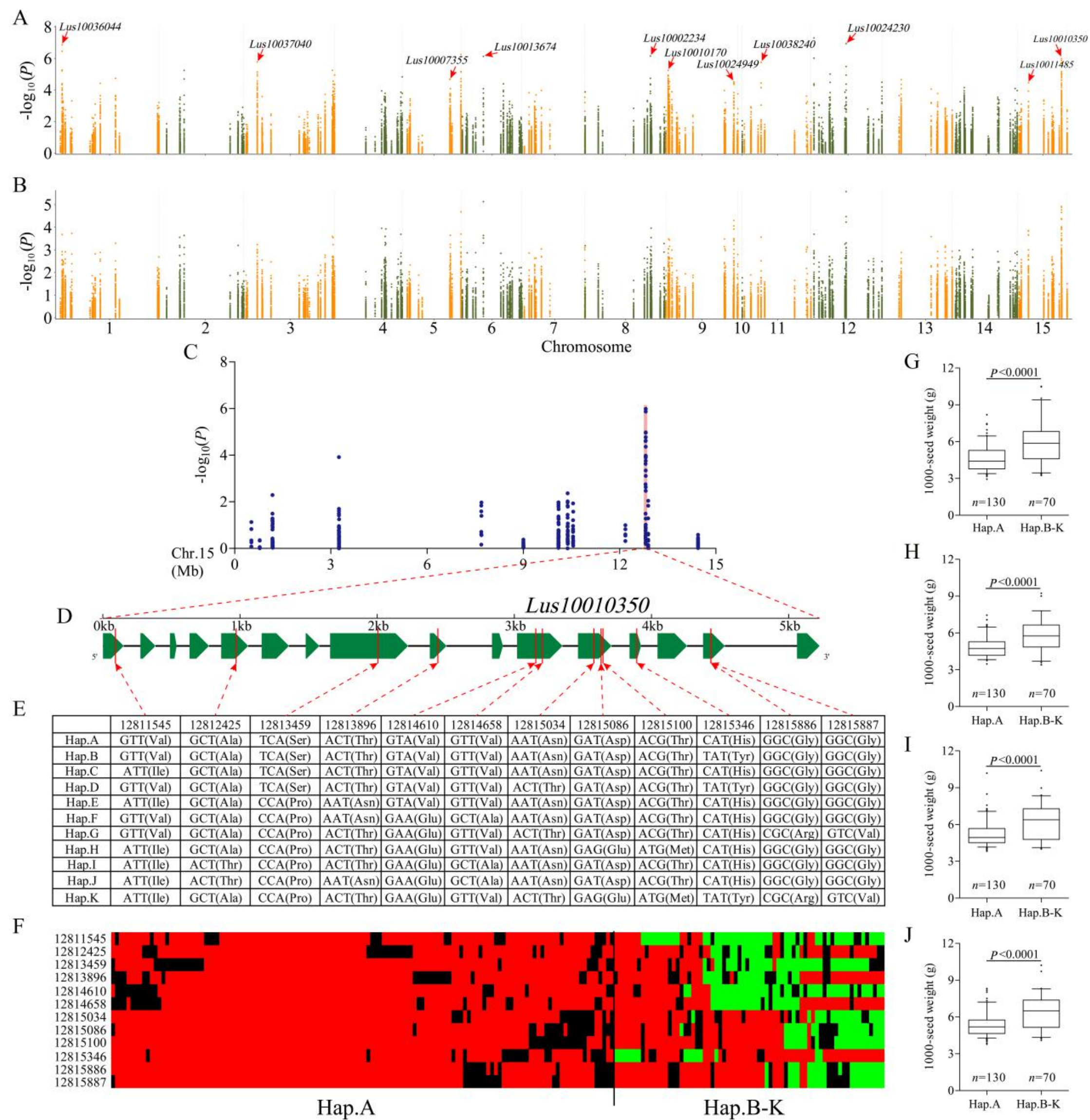


1054

1055 **Figure 6.** Distribution of nucleotide diversity (*π*) within imprinted genes and allele frequency
1056 differences of two genes across oil and fiber subgroups. (A-C) Boxplots for nucleotide diversity of all
1057 imprinted genes (A), MEGs (B) and PEGs (C) across Oil (green) and Fiber (red) groups. (D, G) The
1058 distribution of allele frequency of SNPs located in *Lus10010350* (D, PEG) and *Lus10024230* (G, MEG)
1059 in Oil and Fiber subgroups. The alternate alleles and reference alleles were shown in purple and orange,
1060 respectively. (E, H) The nucleotide diversity distribution of *Lus10010350* on chromosome 15 (E) and
1061 *Lus10024230* on chromosome 12 (H) among Oil and Fiber subgroups. (F, I) Boxplots for nucleotide
1062 diversity of *Lus10010350* (F) and *Lus10024230* (I) among Oil and Fiber subgroups. The difference was
1063 analyzed by two-tailed *t* tests.

1064

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.20.390799; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



	12811545	12812425	12813459	12813896	12814610	12814658	12815034	12815086	12815100	12815346	12815886	12815887
Hap.A	GTT(Val)	GCT(Ala)	TCA(Ser)	ACT(Thr)	GTA(Val)	GTT(Val)	AAT(Asn)	GAT(Asp)	ACG(Thr)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.B	GTT(Val)	GCT(Ala)	TCA(Ser)	ACT(Thr)	GTA(Val)	GTT(Val)	AAT(Asn)	GAT(Asp)	ACG(Thr)	TAT(Tyr)	GGC(Gly)	GGC(Gly)
Hap.C	ATT(Ile)	GCT(Ala)	TCA(Ser)	ACT(Thr)	GTA(Val)	GTT(Val)	AAT(Asn)	GAT(Asp)	ACG(Thr)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.D	GTT(Val)	GCT(Ala)	TCA(Ser)	ACT(Thr)	GTA(Val)	GTT(Val)	ACT(Thr)	GAT(Asp)	ACG(Thr)	TAT(Tyr)	GGC(Gly)	GGC(Gly)
Hap.E	ATT(Ile)	GCT(Ala)	CCA(Pro)	AAT(Asn)	GTA(Val)	GTT(Val)	AAT(Asn)	GAT(Asp)	ACG(Thr)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.F	GTT(Val)	GCT(Ala)	CCA(Pro)	AAT(Asn)	GAA(Glu)	GCT(Ala)	AAT(Asn)	GAT(Asp)	ACG(Thr)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.G	GTT(Val)	GCT(Ala)	CCA(Pro)	ACT(Thr)	GAA(Glu)	GTT(Val)	ACT(Thr)	GAT(Asp)	ACG(Thr)	CAT(His)	CGC(Arg)	GTC(Val)
Hap.H	ATT(Ile)	GCT(Ala)	CCA(Pro)	ACT(Thr)	GAA(Glu)	GTT(Val)	AAT(Asn)	GAG(Glu)	ATG(Met)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.I	ATT(Ile)	ACT(Thr)	CCA(Pro)	ACT(Thr)	GAA(Glu)	GCT(Ala)	AAT(Asn)	GAT(Asp)	ACG(Thr)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.J	ATT(Ile)	ACT(Thr)	CCA(Pro)	AAT(Asn)	GAA(Glu)	GCT(Ala)	AAT(Asn)	GAT(Asp)	ACG(Thr)	CAT(His)	GGC(Gly)	GGC(Gly)
Hap.K	ATT(Ile)	GCT(Ala)	CCA(Pro)	ACT(Thr)	GAA(Glu)	GTT(Val)	ACT(Thr)	GAG(Glu)	ATG(Met)	TAT(Tyr)	CGC(Arg)	GTC(Val)

1065

1066 **Figure 7.** Imprinted gene-based association study for seed size and weight, and identification of a
1067 causal gene for the peak on chromosome 15. (A-B) The overlapping Manhattan plots for seed length,
1068 seed width and 1,000-seed weight in four environments (including 2016DL, 2017UR, 2019UR,
1069 2019YL) using GLM (A) and MLM (B) models. Imprinted genes repeatedly detected at least two
1070 environments or traits related to seed size and weight were marked by red arrows. (C) Local Manhattan
1071 plots for seed width in 2016DL of imprinted gene-based association analysis surrounding the peak on
1072 chromosome 15. The position of *Lus10010350* was highlighted by a shaded pink column. (D) Gene
1073 structure of *Lus10010350.* (E) DNA polymorphism in *Lus10010350.* (F) Schematic representation of
1074 the structural variation in *Lus10010350.* (G-J) Boxplots for 1,000-seed weight based on the haplotypes
1075 (Hap.) for *Lus10010350* in 2016DL (G), 2017UR (H), 2019UR (I) and 2019YL (J). DL, Dali; UR,
1076 Urumqi; YL, YiLi. In the box plots, the center line represented the median, box limits indicated the
1077 upper and lower quartiles, whiskers marked the range of the data and points showed outliers. *n*
1078 indicated the number of accessions with the same genotype. The difference between haplotypes was
1079 analyzed by two-tailed *t* tests.