Supplemental Information

Genomic Comparison and Population Diversity Analysis Provide Insights into the Domestication and Improvement of Flax

Jianping Zhang, Yanni Qi, Limin Wang, Lili Wang, Xingchu Yan, Zhao Dang, Wenjuan Li, Wei Zhao, Xinwu Pei, Xuming Li, Min Liu, Meilian Tan, Lei Wang, Yan Long, Jing Wang, Xuewen Zhang, Zhanhai Dang, Hongkun Zheng, and Touming Liu
Figure S1. The morphology of Longya-10, Heiya-14, and pale flax. (a) Whole plant morphology of Longya-10, Heiya-14, and pale flax. (b) Seeds of Longya-10, Heiya-14, and pale flax. Related to Figure 2.
Figure S2. DNA interactions in 15 flax chromosomes. Each heat map shows a normalized contact matrix, with strong contacts in red and weak contacts in yellow. Related to Figure 1.
Figure S3. Congruence analysis of Longya-10 Hi-C assembly with a published genetic linkage map. Related to Figure 1.
Figure S4. Phylogenetic tree of the eleven species and their divergence time. Related to Figures 1 and 2.
Figure S5. Distribution of $K_s$ between the collinear genes. Related to Figure 1.
Figure S6. Phylogeny of fourteen *Linum* species. *L. bienne* and *L. usitatissimum* are represented by pale flax and Longya-10, respectively. The *L. grandiflorum* information was obtained from our resequencing data. Related to Figure 1.
Figure S7. Verifications of InDels identified between two cultivars and pale flax using Sanger sequencing. Red indicates InDels in Longya-10 and Heiya-14 compared to pale flax. Related to Figure 2.
Figure S8. Expression analysis of the candidate genes including InDels and flax homolog of MYB46/83 genes between Longya-10, Heiya-14, and pale flax by qRT-PCR. Data are represented as mean ± SEM. Related to Figure 2.
Figure S9. Geographical distributions of the 83 re-sequenced flax accessions. Related to Figure 3.
**Figure S10. Population structure of flax accessions.** The 83 flax accessions were divided into three groups: landrace group, oil flax group, and fiber flax group. Related to Figure 3.
Figure S11. Demographic History inferred with smc++. Estimates of effective population size over time are shown for landrace, oil and fiber population. Synonymous mutation rate per base per year of 1.5 x 10-8 and generation time of 1 year are assumed. Related to Figure 3.
Figure S12. Venn diagram comparing the number of genes shared within selective sweeps from three different comparisons. Related to Figure 4.
**Figure S13.** Amino acid substitutions resulting from the large-effect SNP mutations for *LuXTH* and *LuROPGAP3* involved in the selective sweeps of landrace-to-fiber. Red indicates amino acid substitutions between landrace and fiber flax, from the top to the bottom. Related to Figure 3.
**Figure S14.** Expression analysis of candidate genes in selective sweeps. (a) Expression of genes with respect to oil content and seed size in boll. (b) Expression of genes with respect to secondary cell wall biosynthesis, stem length, and flowering time. Related to Figures 3 and 4.
Figure S15. Overlaps between selective sweeps and QTL loci. (a) Selective sweeps of landrace-to-oil. (b) Selective sweeps of oil-to-fiber. (c) Selective sweeps of landrace-to-fiber. Selection signals were defined by the top 5% $\pi_{cis}$ and $F_{ST}$ values (the genomic regions below and above the horizontal lines, respectively). The arrows indicate the QTL loci associated with several important agronomic traits. Related to Figures 3 and 4.
Figure S16. Phylogenetic tree of MYB46/83 genes from 11 species. Related to Figures 2 and 4.
Figure S17. InDel identified in *LuMYB46-2*. (a) InDel in *LuMYB46-2*. Longya-10 gene structure is shown at the top (exons in orange), nucleotide and amino acid sequence are shown at the bottom. At the bottom of the figure, the upper to the lower layers indicate pale flax, Longya-10, and Heiya-14. (b) Verification of InDel identified between Longya-10, Heiya-14 and pale flax genomes using Sanger sequencing. Red indicates InDel between Longya-10, Heiya-14, and pale flax. Related to Figure 2.
Figure S18. The collinear block relationships between flax and grape. Related to Figures 2 and 4.
Figure S19-1. Expression analysis of genes associated with secondary cell wall biosynthesis by qRT-PCR between Longya-10, Heiya-14 and pale flax. Data are represented as mean ± SEM. Related to Figure 2.
Figure S19-2. Expression analysis of genes associated with secondary cell wall biosynthesis by qRT-PCR between Longya-10, Heiya-14 and pale flax. Data are represented as mean ± SEM. Related to Figure 2.
Figure S20. Differential expressions of genes associated with secondary cell wall biosynthesis in stem. (a) Differential expressions of genes between landrace and oil-use flax. (b) Differential expressions of genes between fiber-use and oil-use flax. (c) Differential expressions of genes between landrace and fiber-use flax. Related to Figure 3.
### Supplemental Tables

**Table S1.** Trait performance of Longya-10, Heiya-14, and pale flax. Related to Figure 2.

| Accession | Plant height (cm) | Branch number | Thousand seed weight (g) | Flowering time |
|-----------|-------------------|---------------|--------------------------|----------------|
| Longya-10 | 71.6              | 5.6           | 7.509                    | 60d            |
| Heiya-14  | 93.7              | 3.5           | 5.011                    | 67d            |
| pale flax | 42.6              | 72.4          | 1.232                    | 300d           |
**Table S2.** Summary of genomic sequencing for Longya-10, Heiya-14, and pale flax. Related to Table 1.

| Accession | Insert size | Number | Data (Gb) | Depth (X) |
|-----------|-------------|--------|-----------|-----------|
| Longya-10 | 180bp       | 3      | 21.50     | 41.81     |
|           | 500bp       | 1      | 13.60     | 26.46     |
|           | 3kb         | 1      | 7.28      | 14.16     |
|           | 4kb         | 1      | 10.85     | 21.11     |
|           | 5kb         | 1      | 3.78      | 7.45      |
|           | 8kb         | 1      | 3.43      | 6.68      |
|           | 10kb        | 1      | 3.49      | 6.79      |
|           | 15kb        | 1      | 3.19      | 6.21      |
|           | 17kb        | 1      | 1.04      | 2.02      |
|           | Total       | 11     | 68.16     | 132.69    |
| Heiya-14  | 220bp       | 1      | 27.95     | 53.98     |
|           | 500bp       | 1      | 20.21     | 39.02     |
|           | 3kb         | 1      | 6.60      | 12.74     |
|           | 4kb         | 1      | 6.61      | 12.75     |
|           | 5kb         | 1      | 7.40      | 14.29     |
|           | 8kb         | 1      | 4.75      | 9.17      |
|           | Total       | 6      | 73.52     | 141.92    |
| pale flax | 220bp       | 1      | 22.34     | 42.26     |
|           | 500bp       | 1      | 7.08      | 13.39     |
|           | 3kb         | 1      | 8.13      | 15.38     |
|           | 5kb         | 1      | 6.20      | 11.72     |
|           | 8kb         | 1      | 5.35      | 10.13     |
|           | Total       | 6      | 49.10     | 92.88     |
Table S3. Evaluation of single-nucleotide error rate. Related to Table 1.

| Accession | Contig length(bp) | Correct base number (bp) | Error base number (bp) | Error rate (%) |
|-----------|-------------------|--------------------------|------------------------|----------------|
| Longya-10 | 287,985,064       | 287,985,040              | 24                     | 0.00           |
| Heiya-14  | 300,856,602       | 300,671,827              | 184,755                | 0.06           |
| pale flax | 287,903,089       | 287,901,288              | 1,801                  | 0.0006         |
### Table S4. Assessment of genome assembly completeness with CEGMA. Related to Table 1.

| Accession   | Number of 458 CEG* present in assembly | Percent of 458 CEGs present in assemblies | Number of 248 highly conserved CEGs present | % of 248 highly conserved CEGs present |
|-------------|----------------------------------------|------------------------------------------|-------------------------------------------|---------------------------------------|
| Longya-10   | 454                                    | 99.13%                                   | 243                                       | 97.98%                                |
| Heiya-14    | 453                                    | 98.91%                                   | 243                                       | 97.98%                                |
| pale flax   | 452                                    | 98.69%                                   | 245                                       | 98.79%                                |
Table S5. Assessment of genome assembly completeness with BUSCOs. Related to Table 1.

| Accession | Complete BUSCOs(C) | Complete and single-copy BUSCOs(S) | Complete and duplicated BUSCOs(D) | Fragmented BUSCOs(F) | Missing BUSCOs(M) | Total Lineage BUSCOs |
|-----------|-------------------|----------------------------------|-----------------------------------|----------------------|------------------|-----------------------|
| Longya-10 | 1318 (91.53%)     | 510 (35.42%)                     | 808 (56.11%)                      | 27 (1.88%)           | 95 (6.60%)       | 1440                  |
| Heiya-14  | 1308 (90.83%)     | 499 (34.65%)                     | 809 (56.18%)                      | 33 (2.29%)           | 99 (6.88%)       | 1440                  |
| pale flax | 1292 (89.72%)     | 606 (42.08%)                     | 686 (47.64%)                      | 33 (2.29%)           | 115 (7.99%)      | 1440                  |
Table S6. Assessment of genome assembly completeness with transcripts. Related to Table 1.

| Accession | Range of Length | Total Number | Aligned transcripts | Transcripts with coverage >=80% |
|-----------|-----------------|--------------|----------------------|-------------------------------|
|           |                 |              | Number | Percentage(%) | Number | Percentage(%) |
| Longya-10 | all             | 61,572       | 52,161 | 84.7          | 50,717 | 82.4          |
|           | >=500           | 20,732       | 20,576 | 99.3          | 19,842 | 95.7          |
|           | >=1000          | 11,808       | 11,792 | 99.9          | 11,317 | 95.8          |
|           | all             | 61,572       | 52,181 | 84.8          | 50,667 | 82.3          |
| Heiya-14  | >=500           | 20,732       | 20,584 | 99.3          | 19,829 | 95.6          |
|           | >=1000          | 11,808       | 11,792 | 99.9          | 11,310 | 95.8          |
|           | all             | 61,572       | 51,230 | 83.2          | 48,568 | 78.9          |
| pale flax | >=500           | 20,732       | 20,536 | 98.1          | 19,418 | 93.7          |
|           | >=1000          | 11,808       | 11,777 | 99.7          | 11,134 | 94.3          |
Table S7. Corrected Longya-10 assembly with Hi-C sequencing data. Related to Table 1.

| Scaffold number | Total Scaffold Length (bp) | Scaffold N50 (bp) | Scaffold N90 (bp) | Longest Scaffold (bp) | Total Gap Length (bp) |
|-----------------|-----------------------------|-------------------|-------------------|-----------------------|-----------------------|
| 2,006           | 305,958,589                 | 870,706           | 195,845           | 4,584,463             | 5,800,277             |

| Contig number   | Total Contig Length (bp)  | Contig N50 (bp)  | Contig N90 (bp)  | Longest Contig (bp)  | GC content (%)       |
|-----------------|---------------------------|------------------|------------------|----------------------|----------------------|
| 6,521           | 300,158,312               | 125,201          | 28,941           | 818,717              | 39.05                |
Table S8. Results of ordering and orienting the scaffolds on 15 groups for Longya-10. Related to Figure 1.

| Group            | Scaffold Number | Anchored Length (bp) |
|------------------|-----------------|----------------------|
| Lachesis Group0  | 109             | 25,013,800           |
| Lachesis Group1  | 101             | 22,850,753           |
| Lachesis Group2  | 84              | 22,716,348           |
| Lachesis Group3  | 79              | 22,492,499           |
| Lachesis Group4  | 74              | 21,429,037           |
| Lachesis Group5  | 75              | 21,895,496           |
| Lachesis Group6  | 109             | 21,978,438           |
| Lachesis Group7  | 84              | 18,495,440           |
| Lachesis Group8  | 75              | 21,823,055           |
| Lachesis Group9  | 59              | 19,127,934           |
| Lachesis Group10 | 99              | 16,188,687           |
| Lachesis Group11 | 91              | 17,796,027           |
| Lachesis Group12 | 66              | 15,877,710           |
| Lachesis Group13 | 72              | 18,869,614           |
| Lachesis Group14 | 97              | 15,888,048           |
| Total Sequences Clustered | 1,274 | 302,442,886 |
| Total Sequences Ordered and Oriented | 434 | 295,695,806 |
Table S9. Characteristics of protein-coding genes for Longya-10, Heiya-14, and pale flax. Related to Table 1 and Figure 1.

| Gene feature          | Longya-10 | Heiya-14 | pale flax |
|-----------------------|-----------|----------|-----------|
| Total gene number     | 43,668    | 43,826   | 43,424    |
| Total gene length (bp)| 109,376,018 | 109,600,288 | 101,797,390 |
| Average gene length (bp) | 2,505    | 2,501    | 2,344    |
| Total exon number     | 226,214   | 229,791  | 215,991  |
| Total exon length (bp)| 53,863,319 | 54,215,554 | 49,970,405 |
| Average exon length (bp) | 238      | 236      | 231      |
| Total intron number   | 226,213   | 229,790  | 215,990  |
| Total intron length (bp)| 55,512,699 | 55,384,734 | 51,826,985 |
| Average intron length (bp) | 245      | 241      | 240      |
Table S10. Annotation of protein-coding genes for Longya-10, Heiya-14, and pale flax. Related to Table 1.

| Annotation database | Longya-10 | Heiya-14 | pale flax |
|---------------------|-----------|----------|-----------|
| KOG                 | 25,055    | 15,775   | 21,540    |
| GO                  | 24,919    | 25,798   | 22,268    |
| KEGG                | 9,450     | 9,677    | 13,978    |
| SwissProt           | 33,005    | 34,147   | 27,472    |
| NR                  | 45,034    | 46,513   | 38,724    |
| All Annotated       | 46,044    | 47,559   | 39,567    |
Table S11. Prediction of non-coding RNAs for Longya-10, Heiya-14, and pale flax. Related to Figure 1.

| Accession   | rRNA | tRNA | miRNA | snRNA | snoRNA | Total |
|-------------|------|------|-------|-------|--------|-------|
| Longya-10   | 955  | 965  | 126   | 207   | 555    | 2808  |
| Heiya-14    | 722  | 986  | 115   | 202   | 543    | 2568  |
| pale flax   | 866  | 969  | 128   | 184   | 534    | 2681  |
## Table S12. Statistics of repeated sequences for Longya-10, Heiya-14, and pale flax. Related to Figure 1.

| Type                  | Number | Length (bp) | Percentage(%) |
|-----------------------|--------|-------------|---------------|
|                       | Longya-10 | Heiya-14 | pale flax | Longya-10 | Heiya-14 | pale flax | Longya-10 | Heiya-14 | pale flax |
| Class I/DIRS          | 3,025  | 3,259      | 5721      | 2,993,490 | 2,981,254 | 4557959 | 0.98      | 0.98      | 1.55      |
| Class I/LINE          | 16,134 | 14,093     | 10799     | 6,311,722 | 5,655,700 | 3495089 | 2.06      | 1.86      | 1.19      |
| Class I/LTR           | 556    | 1,964      | 884       | 157,115   | 677,495   | 151996  | 0.05      | 0.22      | 0.05      |
| Class I/LTR/Copia     | 32750  | 31,748     | 29661     | 24,275,676| 23,271,740| 22167895| 7.93      | 7.66      | 7.55      |
| Class I/LTR/Gypsy     | 27,918 | 23,952     | 1,964     | 18,737,539| 16,781,856| 17006063| 6.12      | 5.33      | 5.79      |
| Class I/PLE/LARD      | 37,372 | 32,506     | 46296     | 14,759,968| 13,643,677| 18267559| 4.82      | 4.49      | 6.22      |
| Class I/SINE          | 2,890  | 1,659      | 1215      | 637,127   | 324,134   | 260655  | 0.21      | 0.11      | 0.09      |
| Class I/TRIM          | 6,424  | 5,306      | 5473      | 4,511,135 | 3,849,307 | 4888713 | 1.47      | 1.27      | 1.67      |
| Class I/Unknown       | 1,855  | 2,000      | 1309      | 440,386   | 503,271   | 366263  | 0.14      | 0.17      | 0.12      |
| Class II/Crypton      | 7      | 10         | 16        | 416       | 638       | 991     | 0         | 0         | 0.00      |
| Class II/Helitron     | 5,008  | 6,247      | 2727      | 1,605,875 | 2,160,999 | 851914  | 0.52      | 0.71      | 0.29      |
| Class II/MITE         | 11,794 | 10,593     | 7235      | 2,533,023 | 2,510,195 | 1725056 | 0.83      | 0.83      | 0.59      |
| Class II/Maverick     | 563    | 263        | 129       | 172,289   | 141,370   | 103654  | 0.06      | 0.05      | 0.04      |
| Class II/TIR          | 15,564 | 14,814     | 15077     | 7,762,791 | 7,324,269 | 7678851 | 2.54      | 2.41      | 2.62      |
| Class II/Unknown      | 4,462  | 3,891      | 3434      | 2,708,376 | 2,396,024 | 1731831 | 0.89      | 0.79      | 0.59      |
| Potential Host Gene   | 3,553  | 3,680      | 1844      | 1,100,685 | 1,004,536 | 504930  | 0.36      | 0.33      | 0.17      |
| SSR                   | 17,434 | 17,463     | 4172      | 2,751,923 | 2,382,353 | 1100534 | 0.9       | 0.78      | 0.37      |
| Unknown               | 101,324| 102,348    | 83538     | 30,769,733| 29,809,961| 24541628| 10.06     | 9.82      | 8.36      |
| Total                 | 288,633| 275,796    | 244,460   | 122,229,269| 115,418,779| 109401581| 39.95     | 38.01     | 37.27     |
Table S13. Syntenic analysis between flax, grape and poplar genomes. Related to Figure 1.

| Ratio of orthologus regions | L. usitatissimum vs V.vinifera | L. usitatissimum vs P. trichocarpa |
|----------------------------|--------------------------------|-----------------------------------|
| 1:1                        | 1922(12.88M)                   | 2773(17.86M)                      |
| 2:1                        | 7443(48.09M)                   | 11352(71.73M)                     |
| 3:1                        | 6965(43.91M)                   | 10926(68.49M)                     |
| 4:1                        | 7883(49.09M)                   | 10892(64.35M)                     |
| 5:1                        | 301(2.03M)                     | 385(2.64M)                        |
| 6:1                        | 28(0.35M)                      | 42(0.27M)                         |

Note: The number of genes and the total length of genomic regions involved in syntenic blocks are shown.
Table S14. Comparison of SNVs and InDels between two cultivars and pale flax. Related to Figure 2.

|                                | Longya-10 vs pale flax | Heiya-14 vs pale flax |
|--------------------------------|------------------------|-----------------------|
| Total SNP number               | 3,623,057              | 3,686,366             |
| SNVs/kb                        | 11.37                  | 12.26                 |
| SNV number in intergenic region| 2,404,891              | 2,423,364             |
| SNV number in intron           | 722,871                | 738,135               |
| SNV number in CDS              | 495,295                | 524,867               |
| Nonsynonymous SNV number       | 251,564                | 268,516               |
| Gene number with nonsynonymous SNV | 31,385                 | 33,835                |
| Total InDel number             | 555,580                | 557,691               |
| InDel number/Kb                | 7.18                   | 7.57                  |
| InDel number in intergenic region | 372,368              | 371,744               |
| InDel number in intron         | 159,547                | 160,782               |
| InDel number in CDS            | 23,665                 | 25,165                |
| Gene number with InDel         | 10,749                 | 11,367                |
Table S19. Primer sequences for qRT-PCR. Related to Figure 2.

| Gene ID                  | Gene name | Primer sequence(5'→3')           | Predicted size of PCR products(bp) |
|--------------------------|-----------|----------------------------------|-----------------------------------|
| L.us.o.m.scaffold404.14 | LuFCA     | CAGGCTAAGCACAAGTAACGTGGACC       | 106                               |
| L.us.o.m.scaffold63.99  | LuALC     | CCCAATGGCTTTTCTCAATCTT           | 326                               |
| L.us.o.m.scaffold15.375 | LuLEC1    | AGACCACTCCAGCAGTGCCTTCT         | 237                               |
| L.us.o.m.scaffold196.102| LuMYB46-1 | CAATGGCAAGGTTGCTGGAGTGGTT       | 104                               |
| L.us.o.m.scaffold13.131 | LuMYB46-2-1| TCCAGGAAGGACACAGCAACGA          | 180                               |
| L.us.o.m.scaffold354.6  | LuMYB46-3 | AATGGAACAGGGTGGTGGAGTGG         | 158                               |
| L.us.o.m.scaffold69.1   | LuMYB83-1 | GGAATCTCTGCTTGCTGCTAAATCG       | 115                               |
| L.us.o.m.scaffold71.104 | LuMYB83-2 | GAGGATGAGGAGGACTCTGCTT           | 229                               |
| L.us.o.m.scaffold73.142 | LuMYB83-3 | TGGCTGGAAGAACACGACAGAGAGAG      | 246                               |
| L.us.o.m.scaffold100.96 | LuMYB83-4 | GGGAGGCAGTTAGTGTGGTGGGA         | 166                               |
Table S30. *Ks* values of gene pairs in flax *MYB46/MYB83* colinear blocks. Related to Figures 1, 2 and 4.

| Colinear blocks                      | No. of gene pairs | Average *Ks* value | Median *Ks* value |
|--------------------------------------|-------------------|--------------------|------------------|
| L.us.o.m.scaffold69.1 (LuMYB83-1)    | 13                | 0.1155             | 0.0899           |
| L.us.o.m.scaffold73.142 (LuMYB83-3)  | 92                | 0.1730             | 0.1487           |
| L.us.o.m.scaffold196.102 (LuMYB46-1) | 12                | 0.1381             | 0.1302           |
| L.us.o.m.scaffold13.131 (LuMYB46-2)  | 82                | 0.1670             | 0.1521           |
Transparent Methods

Genome sequencing and assembly

Genome of Longya-10 and Heiya-14, and wild pale flax were sequenced by whole genome shotgun sequencing strategy. A total of eleven, six and five libraries were constructed for Longya-10, Heiya-14, and pale flax, respectively. Paired-end sequencing was performed for these libraries using Illumina HiSeq2500 sequencing platform (Illumina, San Diego, CA, USA). After filtering low quality raw reads and removing adaptors and contaminated reads, the high-quality clean reads were used to de novo assemble the genomes. The whole genome was de novo assembled into longer contigs using ALLPATH-LG (Gnerre et al., 2011) with the default parameters; then the adjacent contigs connected by mate-pair information were linked to scaffolds using SSPACE v2.3 (Boetzer et al., 2011) and gaps were filled using GapCloser from the SOAPdenovo2 package (Luo et al., 2012).

Hi-C sequencing was used to improve the Longya-10 genome. In brief, fresh leaf samples were fixed with formaldehyde and lysed, and then the cross-linked DNA was digested with Hind III overnight. The sticky ends of these digested fragments were biotinylated and then ligated to each other to form chimeric circles. Biotinylated circles, which are chimeras of the physically associated DNA molecules from the original cross-linking, were enriched, sheared and processed into paired-end sequencing libraries. The paired-end reads were produced on the Illumina HiSeq2500 platform. The read pairs form Hi-C sequencing was mapped onto the genome
scaffolds of Longya-10 using Burrows-Wheeler Aligner (BWA) program (Li and Durbin, 2009) with default parameters. Only the unique mapped reads spanning two digested fragments which distally located but physically associated DNA molecules (defined as valid interaction pairs) were used for the next chromosome-level assembly. The scaffolds of Longya-10 genome were broken into fragments with a length of 50 Kb and were clustered by LACHESIS software (Burton et al., 2013) using valid interaction read pairs. The published genetic linkage map (Zhang et al., 2018) was used to validate the Hi-C assembly, by mapping the genetic markers of this map to the assembled Longya-10 genome with >99% coverage and >99% identity using BLAT (Kent, 2002), and then the congruence between the genetic map and the Longya-10 genome was constructed using ALLMAPS with default parameters (Tang et al., 2015).

**Genome evaluation**

To perform the transcriptome sequencing for genome evaluation, the cDNA library with fragment lengths of ~250 bp were constructed using total RNAs from mixed samples (root, stem, leaves, flower, and seed) of Longya-10. Thereafter, paired-end sequencing was performed using the Illumina HiSeq 2500 sequencing platform (Illumina, San Diego, CA, USA). After trimming the adaptor sequences and filtering low-quality reads, the remaining clean reads were *de novo* assembled into transcripts (unigenes) using Trinity (Grabherr et al., 2011).

Genome evaluation was carried out using several approaches as follows. The
single-nucleotide error rate was evaluated by mapping the reads to corresponding genome assembled using BWA program (Li and Durbin, 2009) with default parameter. The Core Eukaryotic Genes Mapping Approach (CEGMA) and Benchmarking Universal Single-Copy Orthologs (BUSCO) were used to evaluate the completeness of the assembled genomes using CEGMA v2.5 (Parra et al., 2007) and BUSCO v3.0.2b (Simao et al., 2015), respectively. In addition, the assembly quality of gene-coding region was evaluated by transcript alignment using BLAT (Kent, 2002), and the alignment of transcript to the genome with identity ≥ 98% and coverage ≥ 80% was requested.

**Genome annotation**

Protein-coding genes of three genomes were predicted based on de novo methods using Genscan v1.0 (Burge and Karlin, 1997), Augustus v2.5.5 (Stanke et al., 2006), GlimmerHMM v3.0.1 (Majoros et al., 2004), GeneID v1.3 (Blanco et al., 2007) and SNAP (Korf, 2004), with the default parameters. In addition, the transcriptome mentioned above were used to assist the annotation of these two genomes, by aligning the transcripts into genomes using PASA (Haas et al., 2003) and GMAP (Wu and Watanabe, 2005). Then, the consensus gene models were generated by integrating the results of two approaches using GLEAN (Elsik et al., 2007). For the genome of pale flax, besides the approaches mentioned above, the homologous peptides from the *Arabidopsis thaliana* (TAIR 10), *Populus trichocarpa* (http://ensemblgenomes.org, release-21) were aligned into genome assembled to identify homologous genes with GeMoMa v1.4.2 (Keilwagen et al., 2016). Thereafter, consensus gene models were
obtained by integrating all prediction methods using EVidenceModeler (EVM) (Haas et al., 2008). Finally, annotations of the predicted genes were performed by blasting their sequences against a number of nucleotide and protein sequence databases, including COG (Tatusov et al., 2003), KEGG (Kanehisa and Goto, 2000), NCBI-NR and Swiss-Prot (Boeckmann et al., 2003) with an E-value cutoff of 1e-5.

The non-coding RNAs were also predicted in three genomes. The rRNA fragments were identified by aligning the rRNA template sequences (Pfam database v22.0) using BLAST (Altschul et al., 1990) with E-value at 1e-10 and identity cutoff at 95%. The tRNAscan-SE v2.0 algorithms (Lowe and Eddy, 1997) with default parameters were applied to prediction of tRNA genes. The miRNA, snRNA and snoRNA genes were identified by mapping the genome sequences to the Rfam database v11.0 (Griffiths-Jones et al., 2003) using INFERNAL v1.1 software (Nawrocki and Eddy, 2013).

The repeat composition in three genomes assembled was estimated by building a repeat library employing the de novo prediction programs LTR-FINDER (Xu and Wang, 2007), MITE-Hunter (Han and Wessler, 2010), RepeatScout v1.0.5 (Price et al., 2005) and PILER-DF (Edgar and Myers, 2005). The database was classified using PASTEClassifier v1.0 (Wicker et al., 2007), and then, was combined with the Repbase database v20.01 (Bao et al., 2015) to create the final repeat library. Repeat sequences in the flax genomes were identified and classified using RepeatMasker program v4.0.6 (Tarailo-Graovac and Chen, 2009). The sequences that were BLAST against the LTR family with ≥ 80% identity and ≥ 80% coverage were deemed to be
Constructing phylogenetic tree of species and WGD analysis

Altogether OrthoMCL v3.1 (Li et al., 2003) clustering derived 212 shared single copy genes were extracted from *V. vinifera*, *L. biene* (pale flax), *L. usitatissimum* (Longya-10 and Heiya-14), *P. trichocarpa*, *R. communis* (Phytozome v12.1), *J. curcas* (GCA_000208675.2), *M. esculenta* (Phytozome v12.1), *A. thaliana*, *E. grandis* (Phytozome v12.1), *M. domestica* (Phytozome v12.1) and *M. truncatula* (Phytozome v12.1), aligned with MUSCLE v3.8.31 (Edgar, 2004) and phylogeny was constructed by PhyML software v3.0 (Guindon et al., 2009). The divergence time was estimated using MCMCtree program implemented in the PAML package v4.9 (Yang, 2007). Calibration times were obtained from the TimeTree database (http://www.timetree.org/).

To perform WGD analysis, the all-against-all BLASTP method was used to detect the paralogous genes in *L. usitatissimum* and *P. trichocarpa* and the orthologous genes in *L. usitatissimum-P. trichocarpa* with the E-value threshold of 1e-5. Homologous blocks were detected using MCScanX (Wang et al., 2012), and the synonymous substitution (*Ks*) values of the blocks were calculated using the HKY model (Hasegawa et al., 1985). The distribution of *Ks* value was used to determine the events of whole genome duplication (WGD). The WGD event was validated by performing a synteny search to compare the flax genome structure with that other related plant genomes. Synteny was searched for by performing comparisons of the
flax genome with *V. vinifera* (γ-WGD) (Jaillon et al. 2007), *P. trichocarpa* (γ-WGD and β-WGD) (Tuskan et al., 2006) genomes.

**Variation detection and positive selection analysis between the genomes of two cultivars and wild pale flax**

The software MUMmer v3.23 (Delcher et al., 2003) was used to align the genomes of Longya-10, Heiya-14 into pale flax genomes, respectively, using the parameters -maxmatch -c 90 -l 40; and then the program of one-to-one alignment block was used to filter the alignment results using the parameter delta -filter -1, and the program of show-snp were used to identify SNVs and InDels in the one-to-one alignment block (parameter -Clr TH). The annotation of the function for SNVs and InDels was performed by the snpEffv4.3 (Cingolani et al., 2012). Sliding window method (window size, 100 Kb; step, 100 Kb) was used to calculate the distribution of SNVs and InDels in each genome.

To identify positive selection genes (PSGs) in flax domestication, we searched the orthologous genes between cultivars (Longya-10 and Heiya-14) and pale flax, and performed CodeML plus a series of different likelihood ratio tests (LRTs) to the ratio of synonymous and non-synonymous changes at each codon on particular branch of the phylogeny (pale_flax, (Longya-10, Heiya-14)).

**Validation of InDels between the genomes of two cultivars and wild pale flax**

The InDel variations in ortholog in three flax genomes were validated by Sanger sequencing. First, we performed the PCR amplification for each InDel variation
from the Longya-10, Heiya-14 and pale flax, respectively, using the primer pairs spanning the entire InDels. Thereafter, these products were digested using 5 U *Exo*I (NEB) and 0.13 U shrimp alkaline phosphatase (Fermentas) and sequenced using a 3730xl DNA Analyzer (ABI, USA). Sequence contigs were assembled using SEQUENCHER 4.1.2 (Gene Codes Co.)

Quantitative real-time PCR

We collected bolls and stems from Longya-10, Heiya-14 and pale flax at 20 days post anthesis, all samples were immersed in liquid nitrogen and then stored at -80°C for RNA extraction. Total RNAs were extracted from the bolls, stems for pale flax, Longya-10 and Heiya-14 by using Plant Easy Spin RNA Miniprep Kit (BIOMIGA, USA). RNAs concentration and purity were determined by agarose gel electrophoresis and NanoDrop2000 spectrophotometer (Thermo, Wilmington, USA). Genomic DNA removing and cDNA synthesis were conducted with the PrimeScript™RT Reagent Kit with gDNA Eraser (Perfect Real Time; TaKaRa). cDNAs were diluted with RNase-free water and then used as the template for qRT-PCR.

qRT-PCR primers for candidate genes were designed using Primer Premier 5.0 (PREMIER Biosoft International, USA) with the following conditions: Tm around 63 °C, product size between 100 and 250 bp, primer length of 21-26 bp, and GC content of 40-60%. qRT-PCR was performed on the Eco Real-Time PCR System (Illumine). According to the manufacturer’s protocol, the PCR reaction volume was 20 μl containing 10 μl 2 × SYBR Mixture (BIOMIGA, USA), 0.5 μM each of forward
and reverse primers, 2 μl diluted cDNA and 6 μl RNase-Free Water. Reaction mixtures were incubated for 2 min at 50 °C, 10 min at 95 °C, followed by 40 amplification cycles of 15 s at 95 °C, 15 s at 60 °C and 15 s at 72°C, the final step melt curve was done for 10 s at 95 °C, 1 min at 65 °C, 1 s at 97 °C. All samples were amplified in triplicate times. GADPH was chosen for internal control (Huis et al., 2010). Data analysis was performed by transforming gene threshold cycle (Ct) into the relative expression level according to the delta CT method (Antonov et al., 2005).

**Analysis of MYB46/83 homologs**

To identify the homologs of the *Arabidopsis MYB46* and *MYB83* genes in other ten species, the 133 MYB genes in *Arabidopsis* provided by Stracke, et al (2001) were downloaded from the Arabidopsis Information Resource (https://www.arabidopsis.org/) and these genes were subsequently used as queries to blast against the ten genomes with an *E*-value cutoff of 1e-5. Then, the obtained MYB proteins between each species and *Arabidopsis* were aligned using MUSCLE (Edgar, 2004), and phylogenetic tree was constructed using the JTT+CAT model of FastTree v2.1 (Price et al., 2010). Finally, the phylogeny of all recognized MYB46/83 genes in eleven species was constructed. The *Ks* values of flax *MYB46/MYB83* gene pairs were calculated using the yn00 program of the PAML package.

**SNPs/InDels detection in flax populations**

To detect the population variation of flax, the DNA of 83 flax accessions was used to construct the library (~250 bp inserted fragment), and then paired-end sequencing was
performed for each library using Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA). After filtering, the clean reads were aligned against the Longya-10 genome assembled with the BWA (Li and Durbin, 2009), allowing no more than 4% mismatches and one gap. Thereafter, SAMtools (Li et al., 2009) was used to convert mapping results to bam format, and duplicated reads were filtered with the help of the Picard package. SNPs and small InDels discovery were performed using the GATK with the default parameters (McKenna et al., 2010). The GATK local realignment was performed to refine the read mapping in the presence of the variants. After realignment, SNP calling was carried out by the Haplotype Caller program of GATK (McKenna et al., 2010), with the following parameters: standard emit confidence (-stand_emit_conf), 10; standard call confidence (-stand_call_conf), 30. To reduce the false discovery rate of SNP/InDel, raw variant identified were filtered using Variant Filtration in GATK for the following parameters: QUAL, 30; call quality divided by depth (QD), 2.0; mapping quality (MQ), 40.0; Fisher’s exact text (FS), 60.0; minor allele frequency, 0.05; missing genotype rate, 0.2.

**Population genetic analysis**

SNPs identified from 83 accessions were used to estimate the genetic distance. The neighbor-joining tree was constructed under the p-distances model, with 1,000 replicates bootstrapping, and was visualized by MEGA5 (Tamura et al., 2011). Population structure was investigated using the ADMIXTURE program (Alexander et al., 2009), and each K value was run 100 times for obtaining it standard error. Principal component analysis was performed by the smartpca program of
EIGENSOFT 6.0 software (Price et al., 2006). To measure linkage disequilibrium (LD) levels in three flax groups, the correlation coefficient ($r^2$) of alleles was calculated using the PopLDdecay (Zhang et al., 2019), with the following parameters: -MAF 0.05 -Miss 0.2 -MaxDist 1000. The average $r^2$ value was calculated for each length of distance. To gain the insights into the genetic diversity and population differentiation, we calculated nucleotide diversity ($\pi$) and $F_{ST}$ values based on 100-Kb sliding windows in 10-Kb steps using the PopGen package of BioPerl (http://cran.r-project.org/web/packages/popgen/index.html).

Detection of selective sweeps

The nucleotide diversity ratio $\pi$ and the differentiation value $F_{ST}$ were used to detect the regions under selective sweeps during the improvement of oil and fiber flax from landrace. In the scanning procedure for identifying selective region, the sliding windows with a size of 100 Kb and a sliding step size of 10 Kb were performed., The $\pi$ and $F_{ST}$ value were estimated in each window, and the windows with the top 5% of the $\pi$ ratios and $F_{ST}$ values were selected and merged into candidate selective sweep regions. The SNP/InDel variations and allelic frequency of each mutant locus in the gene involved in the sweeps were estimated from the genetic group of fiber flax, oil flax and landrace using the SnpEff program (Cingolani et al., 2012).

Transcriptome sequencing

Stems and bolls for Tianshuixian (a landrace accession), Longya-10 and Heiya-14 at 20 days post anthesis were collected with two biological duplicates and immediately
frozen in liquid nitrogen. Total RNAs were isolated using the Trizol reagent (Invitrogen, USA) followed by treatment with RNase-free DNase I (Promega, USA) according to the manufacturers’ protocols. The quality of RNAs was then checked using an Agilent 2100 Bioanalyzer. Illumina RNA-Seq libraries were prepared and sequenced on a HiSeq 2500 system with a PE150 strategy following the manufacturer’s instructions (Illumina, USA). After trimmed based on their quality scores using the quality trimming program Btrim v0.2.0 (Kong, 2011), the clean reads were aligned to our Longya-10 genome assembled using TopHat (Trapnell et al., 2012). Differential expression of genes in the different tissues was calculated using Cuffdiff (Trapnell et al., 2012).

**Supplemental References**

Alexander D.H., Novembre J. and Lange K., Fast model-based estimation of ancestry in unrelated individuals, *Genome Res.* **19**, 2009, 1655–1664.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J., Basic local alignment search tool, *J. Mol. Biol.* **215**, 1990, 403–410.

Antonov J., Goldstein D.R., Oberli A., Baltzer A., Pirotta M., Fleischmann A., Altermatt H.J. and Jaggi R., Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization, *LabInvest* **85**, 2005, 1040–1050.

Bao W., Kojima K.K. and Kohany O., Repbase Update, a database of repetitive elements in eukaryotic genomes, *Mob. DNA* **6**, 2015, 11.
Blanco E., Parra G. and Guigó R., Using geneid to identify genes, *Curr. Protoc. Bioinformatics* **18**, 2007, 4.3.1–4.3.28.

Boeckmann B., Bairoch A., Apweiler R., Blatter M.C., Estreicher A., Gasteiger E., Martin M.J., Michoud K., Donovan C., Phan I., et al., The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, *Nucleic Acids Res.* **31**, 2003, 365–370.

Boetzer M., Henkel C.V., Jansen H.J., Butler D. and Pirovano W., Scaffolding pre-assembled contigs using SSPACE, *Bioinformatics* **27**, 2011, 578–579.

Burge C. and Karlin S., Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.* **268**, 1997, 78–94.

Burton J.N., Adey A., Patwardhan R.P., Qiu R., Kitzman J.O. and Shendure J., Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions, *Nat. Biotechnol.* **31**, 2013, 1119–1125.

Cingolani P., Platts A., Wang I.L., Coon M., Nquyen T., Wang L., Land S.J., Lu X. and Ruden D.M., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3, *Fly* **6**, 2012, 80–92.

Delcher A.L., Salzberg S.L. and Phillippy A.M., Using MUMmer to identify similar regions in large sequence sets, *Curr. Protoc. Bioinformatics* 2003, Chapter 10:Unit 10.3.
Edgar R.C. and Myers E.W., PILER: identification and classification of genomic repeats, *Bioinformatics* **21**, 2005, i152–i158.

Edgar R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* **32**, 2004, 1792–1797.

Elsik C.G., Mackey A.J., Reese J.T., Milshina N.V., Roos D.S. and Weinstock G.M., Creating a honey bee consensus gene set, *Genome Biol.* **8**, 2007, R13.

Gnerre S., Maccallum I., Przybylski D., Ribeiro F.J., Burton J.N., Walker B.J., Sharpe T., Hall G., Shea T.P., Sykes S., et al., High-quality draft assemblies of mammalian genomes from massively parallel sequence data, *Proc. Natl. Acad. Sci. U S A* **108**, 2011, 1513–1518.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* **29**, 2011, 644–652.

Griffiths-Jones S., Bateman A., Marshall M., Khanna A. and Eddy S.R., Rfam: an RNA family database, *Nucleic Acids Res.* **31**, 2003, 439–441.

Guindon S., Delsuc F., Dufayard J.F. and Gascuel O., Estimating maximum likelihood phylogenies with PhyML, *Methods Mol. Biol.* **537**, 2009, 113–137.

Haas B.J., Delcher A.L., Mount S.M., Wortman J.R., Smith R.K., Hannick L.I., Maiti R., Ronning C.M., Rusch D.B., Town C.D., et al., Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies, *Nucleic Acids Res.* **31**,
Haas B.J., Salzberg S.L., Zhu W., Pertea M., Allen J.E., Orvis J., White O., Buell C.R. and Wortman J.R., Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments, *Genome Biol.* **9**, 2008, R7.

Han Y. and Wessler S.R., MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences, *Nucleic Acids Res.* **38**, 2010, e199.

Hasegawa M., Kishino H. and Yano T., Dating of the human-ape splitting by a molecular clock of mitochondrial DNA, *J. Mol. Evol.* **22**, 1985, 160–174.

Huis R., Hawkins S. and Neutelings G., Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.), *BMC Plant Biol.* **10**, 2010, 71.

Jaillon O., Aury J.M., Noel B., Policriti A., Clepet C., Casaqrande A., Choisne N., Aubourg S., Vitulo N., Jubin C., et al., The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, *Nature* **449**, 2007, 463–467.

Kanehisa M. and Goto S., KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* **28**, 2000, 27–30.

Keilwagen J., Wenk M., Erickson J.L., Schattat M.H., Grau J. and Hartung F., Using intron position conservation for homology-based gene prediction, *Nucleic Acids Res.*
Kent W.J., BLAT—the BLAST-like alignment tool, *Genome Res.* **12**, 2002, 656–664.

Kong Y., Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies, *Genomics* **98**, 2011, 152–153.

Korf I., Gene finding in novel genomes, *BMC Bioinformatics* **5**, 2004, 59.

Li H. and Durbin R., Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* **25**, 2009, 1754–1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. and Durbin R., The sequence alignment/map format and SAMtools, *Bioinformatics* **25**, 2009, 2078–2079.

Li L., Stoeckert C.J. and Roos D.S., OrthoMCL: identification of ortholog groups for eukaryotic genomes, *Genome Res.* **13**, 2003, 2178–2189.

Lowe T.M. and Eddy S.R., tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.* **25**, 1997, 955–964.

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., et al., SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler, *GigaScience* **1**, 2012, 18.

Majoros W.H., Pertea M. and Salzberg S.L., TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders, *Bioinformatics* **20**, 2004, 2878–2879.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A.,
Garimella K., Altshuler D., Gabriel S., Daly M., et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res.* **20**, 2010, 1297–1303.

Nawrocki E.P. and Eddy S.R., Infernal 1.1: 100-fold faster RNA homology searches, *Bioinformatics* **29**, 2013, 2933–2935.

Parra G., Bradnam K. and Korf I., CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, *Bioinformatics* **23**, 2007, 1061–1067.

Price A., Patterson N., Plenge R., Weinblatt M., Shadick N. and Reich D., Principal components analysis corrects for stratification in genome-wide association studies, *Nat. Genet.* **38**, 2006, 904–909.

Price A.L., Jones N.C. and Pevzner P.A., De novo identification of repeat families in large genomes, *Bioinformatics* **21**, 2005, i351–i358.

Price M.N., Dehal P.S. and Arkin A.P., FastTree 2-approximately maximum-likelihood trees for large alignments, *PLoS One* **5**, 2010, e9490.

Simao F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V. and Zdobnov E.M., BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* **31**, 2015, 3210–3212.

Stanke M., Keller O., Gunduz I., Hayes A., Waack S. and Morgenstern B., AUGUSTUS: ab initio prediction of alternative transcripts, *Nucleic Acids Res.* **34**, 2006, W435–W439.
Stracke R., Werber M. and Weisshaar B., The \textit{R2R3-MYB} gene family in \textit{Arabidopsis thaliana}, \textit{Curr. Opin. Plant Biol.} \textbf{4}, 2001, 447–456.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S., MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, \textit{Mol. Biol. Evol.} \textbf{28}, 2011, 2731–2739.

Tang H., Zhang X., Miao C., Zhang J., Ming R., Schnable J.C., Schnable P.S., Lyons E. and Lu J., ALLMAPS: robust scaffold ordering based on multiple maps, \textit{Genome Biol.} \textbf{16}, 2015, 3.

Tarailo-Graovac M. and Chen N., Using RepeatMasker to identify repetitive elements in genomic sequences, \textit{Curr. Protoc. Bioinformatics} \textbf{25}, 2009, 4.10.1–4.10.14.

Tatusov R.L., Fedorova N.D., Jackson J.D., Jacobs A.R., Kiryutin B., Koonin E.V., Krylov D.M., Mazumder R., Mekhedov S.L., Nikolskaya A.N., et al., The COG database: an updated version includes eukaryotes, \textit{BMC Bioinformatics} \textbf{4}, 2003, 41.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L. and Pachter L., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, \textit{Nat. Protoc.} \textbf{7}, 2012, 562–578.

Tuskan G.A., Difazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A., et al., The genome of black cottonwood, \textit{Populus trichocarpa} (Torr. & Gray), \textit{Science} \textbf{313}, 2006, 1596–1604.
Wang Y., Tang H., Debarry J.D., Tan X., Li J., Wang X., Lee T.H., Jin H., Marler B., Guo H., et al., MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, *Nucleic Acids Res.* **40**, 2012a, e49.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., et al., A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.* **8**, 2007, 973–982.

Wu T.D. and Watanabe C.K., GMAP: a genomic mapping and alignment program for mRNA and EST sequences, *Bioinformatics* **21**, 2005, 1859–1875.

Xu Z. and Wang H., LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, *Nucleic Acids Res.* **35**, 2007, W265–W268.

Yang Z., PAML 4: phylogenetic analysis by maximum likelihood, *Mol. Biol. Evol.* **24**, 2007, 1586–1591.

Zhang C., Dong S.S., Xu J.Y., He W.M. and Yang T.L., PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files, *Bioinformatics* **35**, 2019, 1786–1788.

Zhang J., Long Y., Wang L., Dang Z., Zhang T., Song X., Dang Z. and Pei X., Consensus genetic linkage map construction and QTL mapping for plant height-related traits in linseed flax (*Linum usitatissimum* L.), *BMC Plant Biol.* **18**, 2018, 160.