Development and Application of Novel InDel Markers in Flax (*Linum usitatissimum* L.) Through Whole-Genome Re-Sequencing

Hui Jiang

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Gen Pan

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Touming Liu

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Li Chang

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Siqi Huang

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Huijuan Tang

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Yuan Guo

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Yena Wu

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Jie Tao

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Anguo Chen (✉ cagibfc@126.com)

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Research Article

Keywords: Flax, In Del, Re-sequence, Oil content, Genetic diversity

Posted Date: July 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-700179/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Flax is an important oil and fibre crop grown in Northern Europe, Canada, India, and China. The development of molecular markers has accelerated the process of flax molecular breeding and has improved yield and quality. Presently, simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers in the whole genome have been developed for flax. However, the development of flax insertion/deletion (InDel) markers has not been reported. A total of 17,110 InDel markers were identified by comparing whole-genome re-sequencing data of two accessions (87 – 3 and 84 – 3) with the flax reference genome. The length of InDels ranged from 1–277 bp, with 1–15 bp accounting for the highest rate (95.55%). The most common InDels were in the form of single nucleotide (8840), dinucleotide (3700), and trinucleotide (1349), and chromosome 2 (1505) showed the highest number of InDels among flax chromosomes, while chromosome 10 (913) presented with the lowest number. From 17,110 InDel markers, 90 primers that were evenly distributed in the flax genome were selected. Thirty-two pairs of polymorphic primers were detected in two flax accessions, and the polymorphism rate was 40.70%. Furthermore, genetic diversity analysis, population structure and principal component analyse (PCA) divided 69 flax accessions into two categories, namely oilseed flax and fibre flax using 32 pairs of polymorphic primers. Additionally, correlation analysis showed that InDel-26 and InDel-81 were associated with oil content traits, and two candidate genes (*lus10031535* and *lus10025284*) tightly linked to InDel-26 or InDel-81, might be involved in flax lipid biosynthesis and lipid metabolism. This study is the first to develop InDel markers based on re-sequencing in flax and clustered the markers into two well-separated groups for oil and fibre. The results demonstrated that InDel markers developed herein could be used for flax germplasm identification, genetic diversity analysis, and molecular marker-assisted breeding.

Introduction

Flax (*Linum usitatissimum* L., $2n = 30$) is an important economic crop that is mainly cultivated in Canada, China, America, and India. Flax is a multipurpose plant with a high comprehensive utilisation value and is used in food, medicine, and feed, while fibre is widely used in manufacturing industries and in the development of composite materials and fireproof materials [1, 2]. Therefore, cultivation of high-yield and high-quality flax varieties has emerged as one of the current breeding goals. Molecular marker-assisted breeding (MAS) is a new type of breeding technology that can help achieve the breeding goal of increasing crop yield and quality in a more direct, rapid, and efficient manner. In recent years, with the development of sequencing technology, development of high-quality molecular markers in the whole genome has been realised. Random amplified polymorphic DNA (RAPD) [3–5], amplified fragment length polymorphism (AFLP) [6, 7], simple sequence repeats (SSR) [8–14], and single nucleotide polymorphism (SNP) [15–21] molecular markers have been identified in flax. However, there no research has been conducted on flax InDel molecular markers. Therefore, the development of InDel for the expansion of molecular markers can accelerate the process of flax molecular breeding, thereby aiding the cultivation of high-quality flax varieties.

AFLP, RAPD, SSR, and SNP markers in flax have been studied and used in germplasm identification, genetic diversity analysis, genetic map construction, QTL mapping, and genome-wide association analysis. Presently, SNPs are mainly used for genetic map construction [16, 22], QTL mapping [22, 23], and genome-wide association analysis [17–19, 24]. Based on principles of the SLAF technology, a genetic map with the highest density using the best quality of flax was constructed [16, 23], and the QTL loci and candidate genes related to the agronomic traits of flax were screened through association analysis [17]. The SSR markers were widely used to analyse the genetic diversity of flax, and flax was divided into oil, fibre, and oil-fibre flax according to its use [13, 14]. However, few studies have

shown that the classification of flax varieties differs depending on the type of marker used [13]. Although studies have been conducted on different types of molecular markers in flax, the continuous development of new molecular markers can help provide technical support for the study of molecular marker-assisted breeding of flax. With the development of re-sequencing technology in recent years, mining of high-quality molecular markers based on the whole genome can accelerate the research on flax molecular breeding.

With the development of high-throughput sequencing technology, the development of SNP and InDel markers based on the whole genome has been widely considered in crop molecular marker-assisted breeding (MAS). InDel refers to the insertion or deletion of nucleotides in the plant genome, and such nucleotides are widely distributed, dense, and abundant in the plant genome [25–27]. InDel characteristics are similar to those of SSR, and InDel markers are composed of short fragments of different lengths. Owing to the low cost and high efficiency of sequencing, InDel can be developed easily using high-throughput sequencing [27, 28]. InDel markers have been used in rice [29–31], jute [32], quinoa [33], cabbage [25], cucumber [34], tomato [35] and other crops [26, 28, 36–38], and are mainly used for germplasm resource identification, genetic diversity analysis, and genetic map construction. However, there are no reports available on the development and application of InDel markers in flax.

To accelerate the application of more molecular markers in flax breeding, for the first time, re-sequencing was used to identify and develop flax InDel markers. In this study, based on the two materials with different oil contents developed in our laboratory in the early stage, our aims were to: 1) re-sequence the two materials (87 – 3 and 84 – 3) and compare them with the reference genome sequence to develop the InDel marker; 2) design the InDel primers and evaluate the quality of InDel markers use the two accessions (87 – 3 and 84 – 3); 3) use polymorphic primers to analyse the genetic diversity of 69 flax accessions; and 4) conduct correlation analysis between InDel markers and oil content traits, as well as perform the screening of candidate genes related to oil content traits.

Methods

Plant materials

In the early stage, we prepared 87 – 3 (male, 35.94%) and 84 – 3 (female, 45.77%) samples with different oil contents for re-sequencing, and a total of 69 flax accessions were collected for genetic analysis. Among them, 10 accessions were provided by the Gansu Academy of Agricultural Sciences, 20 accessions were provided by the Heilongjiang Academy of Agricultural Sciences, and 37 private accessions were obtained (Table S1). These flax accessions were derived from regions across the globe and were divided into oil and fibre according to their use. All materials were planted in the western centre of the Chinese Academy of Agricultural Sciences, and the oil content of flaxseed was determined by conducting near-infrared spectral analysis (NIRS) after maturity was reached.

DNA sample preparation and re-sequencing

Young leaf tissues of 10 individuals of 84 – 3 and 87 – 3 were collected for performing *de novo* genomic re-sequencing. Fresh and young leaf tissues of 69 flax accessions at the first branching stage were collected, and DNA extraction was performed using the CTAB method [39]. DNA quality and quantity were determined using an Eppendorf BioSpectrometer (Eppendorf, Hamburg, Germany). DNA quantification was further conducted using a fluorometer and DNA samples were diluted to a 10-ng/L working solution.

InDel detection and primer design

The flax reference genome sequence was obtained from NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCA_000224295.2) with the re-sequencing data of 87 – 3 and 84 – 3 to detect the InDel loci. The Primer 3.0 software (<http://pgrc.ipk-gatersleben.de/misa/primer3.html>) was used to design primers for the InDel loci. One pair of primers with the highest scoring was selected from the experimental design results.

InDel genotyping

Primers were designed based on the detected InDel loci. According to the distance of every 4 Mb of the whole genome, InDel primers evenly distributed on the flax chromosome were selected for polymorphism analysis. InDel-primed polymerase chain reactions (PCRs) were performed in 10 µL reaction volumes, which included 8 µL of the PCR mix solution (CW BIO, Beijing, China), 0.5 µL of the forward primer (10 nmol/L), 0.5 µL of the reverse primer (10 nmol/L), and 1 µL of the DNA template. PCR amplification was performed under the following conditions: (1) pre-denaturation at 95°C for 3 min; (2) 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min; and (3) 72°C for 5 min.

Genetic diversity assay and population structure analyse

The two materials (87 – 3 and 84 – 3) were used for InDel primer polymorphism detected, and band types were detected by performing analysis using polypropylene gel. The genetic diversity of 69 flax accessions was analysed with polymorphic primer. Each polymorphic band detected using the same primer represented an allelic mutation. To generate molecular data matrices, clear bands for each fragment were scored for each accession for each primer pair and were recorded as 1 (presence of a fragment), 0 (absence of a fragment), and 9 (complete absence of band). The PowerMarker software (version 3.25) was used to provide data on basic summary statistics [40], which included the polymorphism information content (PIC), the number of alleles (N_a), major allele frequency (MAF), and gene diversity for each InDel. A clustering map was constructed based on the genetic distances and the unweighted pair-group method with arithmetic means (UPGMA) using MEGA V6.0.

The population structure of 69 flax accessions was estimated using the software STRUCTURE v2.3.4. The number of presumed populations (K) was set from 1 to 10 and running 10 times per round, burn in time and MCMC (Markov Chain Monte Carlo) replication number both to were set to 100,000. The STRUCTURE HARVESTER [41] was used to calculate the optimal K value to determine the most suitable number of sub-populations. In order to further analyse genetic relationships of flax accessions, a principal component analyse (PCA) was performed based on variability trends of InDel using the NTSYSpc version 2.10.

Correlation analysis between InDel markers and oil content traits

Correlation analysis was conducted using the SPSS software (version 13.0) to analyse the correlation between the InDel markers that could clearly be used to distinguish 69 flax accessions and the oil content trait data. Pearson correlation analysis was used to analyse the correlation coefficient between the flax oil content traits and InDel markers. A multiple linear stepwise regression method [42] was used to fit the InDel markers that showed significant correlation with the oil content traits of flax. The X-axis indicates the genotype generated with the presence of each InDel marker, and the Y-axis indicates the trait of flax oil content.

Results

Distribution of InDel markers

The clean base quantity generated was 3.69 G for 84 – 3 and was 1.01 G for 87 – 3 with an average of 2.35 G. Using the Burrows-Wheeler Alignment (BWA), 3.55 G and 0.96 G of 84 – 3 and 87 – 3 were respectively mapped to the flax reference genome sequence. The mapping read depth was 18.93× for 84 – 3 and 6.02× for 87 – 3, respectively (Table 1). Genome-wide InDel polymorphism led to the generation of 17,110 InDels between 84 – 3 and 87 – 3 (Table S2), exhibiting an InDels density of 54.12 InDel/Mb. InDel sites varied from 1–277 bp, and the number of InDel sites decreased as the length of InDel increased (Fig. 1). The number of InDels of 1–3 bp showed the highest value (81.17%), followed by that of InDels of 4–15 bp (14.37%); the number of InDels of > 15 bp showed the least value (4.45%). Among all InDels, the most common InDel sites were single nucleotide (8840) accounting for 51.67%, dinucleotide (3700) accounting for 21.62%, and trinucleotide (1349) accounting for 7.88%. Furthermore, the repeats of A (2524) and T (2551) were more commonly present in single nucleotide sites, and there were four repeat types of dinucleotide (CA, CT, TA, and AT) and trinucleotide (TTC, AAT, CTT, and TAA) (Fig. S1). Additionally, the distribution of InDels on each chromosome was also different (Fig. 2A). Chromosome 2 exhibited the highest number of InDels (1505), and chromosome 10 demonstrated the lowest number (913). The average density of InDels in the whole genome of flax was 54.12 InDels/Mb, and the InDel density of each chromosome was different. The highest density of chromosome 5 was 64.97 InDels/Mb, while the lowest density of chromosome 8 was 40.35 InDels/Mb (Fig. 2B).

InDel marker polymorphism analysis

Based on data comprising 17,110 InDels, 17,110 primer pairs were designed. Ninety pairs of InDel primers were selected according to the distance of each 4 MB and a difference of more than three bases in the whole genome, as shown in Table S3. The length of all primers was 24 bp, and the size of the product was 80–250bp. The effectiveness of 90 pairs of InDel primers showed that 86 primers led to the successful amplification of bands, 32 pairs of polymorphic primers, and the polymorphism rate was 40.70% (Fig. 3; Fig. S2).

Genetic diversity analyse of 69 flax accessions

Data on 32 InDel primers were used to analyse the genetic diversity of the 69 flax accessions (Fig. 4). The number of alleles (N_a) ranged from 2–3 (average of 2.25) (Table S4). Twenty-four InDels exhibited two alleles, and eight InDels presented with three alleles. The polymorphism information content (PIC) varied from 0.0674–0.3958, with an average of 0.3097, and the gene diversity ranged from 0.0698–0.5138, with an average of 0.3899. Additionally, the average major allele frequency (MAF) was 0.6941, which varied from 0.5000–0.9638. InDel-51 presented with the highest PIC (0.3958) and gene diversity (0.5138). InDel-13 presented with the lowest PIC (0.0674), gene diversity (0.0698), and the highest MAF (0.9638).

The genetic relationships between the 69 flax accessions were analysed. As shown in Fig. 5, the 69 accessions were unambiguously classified into two clusters according to oil and fibre content. Group I included 35 accessions of oil. Group II included 34 accessions of fibre. However, the two accessions for oil (Baya 7, Kuerle) were classified into Group II. Moreover, five InDel markers (InDel-14, InDel-26, InDel-40, InDel-74, and InDel-81) could be used to largely classify 69 flax accessions for oil and fibre among the 32 polymorphic primers, and the five markers could be further analysed (Fig. 4; Fig. S3).

Population structure and principal component analyse

In order to further determine the population structure of 69 flax germplasm, STRCUTURE 2.3.4 was used to analyse the population structure. The results of Delta K values of different assumed (K) populations show that a sharp peak

was given at $K = 2$, which indicated that there were two sub-populations in 69 flax accessions (Fig. 6A). As shown in Fig. 6B, the population structure analysis classified 69 flax genotypes into two sub-populations, with 33 fibre genotypes in population 1 and 36 oil genotypes in population 2. In addition, PCA analysis divided all accessions into two major groups (Fig. 7). The results of the two-dimensional scatter plot showed that principal component 1 and principal component 2 account for 63.7% and 21.0% of the variation, respectively.

Correlation analysis and candidate gene prediction

To further explore InDel markers related to oil content traits of flax, correlation analysis between five InDel markers and oil content traits was conducted (Table 2). The results showed that InDel-26 exhibited a significant positive correlation with oil content traits, whereas InDel-81 exhibited a significant negative correlation ($p < 0.05$). Meanwhile, the correlation analysis between InDel markers and flax type InDel-26 showed the existence of an extremely significant negative correlation, while InDel-81 showed the existence of an extremely significant positive correlation ($p < 0.01$) (Table 3).

Further, the marker loci information was compared with the flax reference genome data to identify the gene locus related to the two markers. InDel-26 and InDel-81 were located on chromosomes 12 and 8, respectively. The genes close to InDel-26 were *lus10031534* and *lus100315535*, and functional annotations were found to be RNA-binding (RRM/RBD/RNP motifs) family protein and serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B prime gamma, respectively (Table 4). The genes close to InDel-81 were *lus10025284* and *lus10010835*, and their functional annotations included Lipin family protein and protein kinase superfamily protein, respectively.

Discussion

In this study, re-sequencing was performed for the first time to develop InDels related to the oil content of flax. Based on the re-sequencing data, 17,110 InDels were detected with a density of 54.12 InDel/Mb in the whole genome. Compared with jute (*Corchorus* spp.) (51,172) [32], quinoa (*Chenopodium quinoa* Willd.) (842,783) [33], apple (*Malus* spp.) (25,924) [38], tea plant (*Camellia sinensis* (L.) O. Kuntze) (255,218) [36], and pepper (*Capsicum* spp. L.) (1,651,856) [28], the number of InDels detected in this study was low but was higher than that of cucumber (*Cucumis sativus*) (10,470) [34] and sesame (*Sesamum indicum* L.) (7,477) [37]. Compared with other markers in flax, the number of InDels was less than that of SNPs [16, 19] and more than most SSR markers [8, 10, 13,

43]. The main reason for obtaining such results may be attributed to the fact that compared with the complicated operations of RAPD, AFLP, and SSR, the use of a re-sequencing technology can aid an easy detection and obtainment of a larger number and higher quality of SNPs and InDels in the whole genome. Moreover, the length of InDel obtained in this study was 1–277 bp, with the largest number of single nucleotides (51.67%), followed by dinucleotides (21.62%) and trinucleotides (7.88%). The small InDel fragments (1–15bp) accounted for a high rate (95.55%). Additionally, A and T were the most commonly occurring single nucleotides, and these results were consistent with reported by previous studies [26, 32, 44]. The InDel marker refers to the difference in nucleotides in the plant genome. In this study, using 84 – 3 and 87 – 3 with different oil contents constructed in the early stage of experiments, the InDel marker was developed for the first time in flax by re-sequencing, which was helpful for conducting flax molecular marker-assisted breeding.

Analysis of the genetic relationship between species has important statistical significance for guiding crop genetic breeding. To verify the developed InDel markers, 32 pairs of polymorphic primers were screened from 90 pairs of InDel markers evenly distributed on the chromosomes in 37 flax accessions. The polymorphism rate was 40.70%,

which was lower than that of pepper (67.4%) [28], cucumber (54.81%) [34], and *Brassica rapa* (77%) [26]. Additionally, the PIC obtained was low (0.3097), which was consistent with the results reported by previous studies [10, 13, 14]. This indicated that InDel was more suitable for the identification of flax germplasm resources, since the degree of variation of flax genes was not substantial and the genetic background was narrow.

Genetic diversity analysis can help identify germplasm resources that play a key role in crop breeding through the genetic evolution of crops. Based on the results of the previous analysis, 32 polymorphic InDel markers could be used to divide 69 flax accessions into two groups, namely oil and fibre. Group I was established for oil, and Group II was established for fibre. This classification was consistent with that reported in previous studies [7, 13, 14]. In addition, population structure analysis showed that when Delta K had a peak at K = 2, 69 flax genotypes were divided into two sub-populations, indicating that 69 flax accessions has a relatively simple population structure and a single genetic source. The PCA results also divided all accessions into two categories. The population structure and PCA classification results were consistent with the UPGMA clustering results, indicated that the classification results obtained in this study were reliable.

In previous studies, high-throughput sequencing techniques have been used for the development of SSR and SNP markers for QTL mapping and GWAS analysis of flax-related quality traits [12, 17, 19, 24]. In this study, two InDel markers (InDel-26 and InDel-81) related to flax oil content traits were identified through correlation analysis of the developed InDel markers. Traditional breeding is the major method for the development of flax varieties, and molecular breeding remains in the exploratory stage. In this study, the InDel marker (InDel-26, InDel-81) could be used for the selection of traits with high oil content during flax breeding, and might further accelerate the process of flax molecular breeding.

Furthermore, through the comparison of gene loci information, four gene loci (*lus10031534*, *lus10031535*, *lus10025284*, and *lus10010835*) near InDel-26 and InDel-81 were found. *lus10031535* is functionally annotated as serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B prime gamma, and the homologous gene in *Arabidopsis thaliana* is *PUX5* (UniProtKB: Q7Y175; GeneID: 827210), which encodes a protein containing the ubiquitin regulatory domain. In *Pichia pastoris*, Ubx1 and Ubx2 were reported to play a key role in the synthesis of unsaturated fatty acids by exerting influence on Spt23, and both are reportedly involved in the formation of lipid droplets and protein degradation [45]. *lus10025284* is functionally annotated as a lipid family protein, and its homologous gene in *Arabidopsis thaliana* is *PAH1* (UniProtKB: Q9SF47; GeneID: 820113), which catalyses the dephosphorylation of phosphates to generate diacylglycerols [46–48]. Therefore, it is speculated that *lus10031535* and *lus10025284* may function in flax lipid biosynthesis and lipid metabolism; however, the function of these genes in flaxseed warrants further verification using transgenic experiments.

Conclusion

In this study, 17,110 InDel markers were developed based on re-sequencing, and thirty-two pairs of polymorphic primers were obtained among accessions 84 – 3 and 87 – 3. Thereafter, 69 accessions were divided into two groups (oil and fibre) using 32 polymorphic primers based on population structure analysis, PCA analysis and the UPGMA clustering. Furthermore, the correlation analysis showed that InDel-26 and InDel-81 were related to the oil content of flax, and two candidate genes (*lus10031535* and *lus10025284*) may relate to lipid biosynthesis and lipid metabolism. This is the first study conducted on the development of InDel markers based on re-sequencing in flax, which may provide a theoretical basis for the identification of flax germplasm resources, genetic map construction, QTL mapping, and GWAS.

Declarations

Author contribution Anguo Chen designed experiment and provided resources. Hui Jiang performed experiment, analysed data and wrote original draft. Gen Pan designed experiment, analysed data and edited article. Touming Liu designed experiment. Li Chang and Siqi Huang edited article. Huijuan Tang investigated data of traits. Yuan Guo provide resources. Yena Wu and Jie Tao performed part of experiment.

Founding This research was funded by National Natural Science Foundation of China (31771852 and 31771853) and Special college-level overall planning project for basic scientific research business expenses of Chinese Academy of Agricultural Sciences (Y2019XK15-06).

Data availability All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethics approval Not applicable.

Informed consent Informed consent was obtained from all individual participants include in the study prior to sample collection.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

References

1. Westcott ND, Muir AD (2003) Flax seed lignan in disease prevention and health promotion. *Phytochem Rev* 2(3):401–417. <https://doi.org/10.1023/B:PHYT.0000046174.97809.b6>
2. Zuk M, Richter D, Matuła J, Szopa J (2015) Linseed, the multipurpose plant. *Ind Crop Prod* 75:165–177. <https://doi.org/10.1016/j.indcrop.2015.05.005>
3. Fu Y, Diederichsen A, Richards KW, Peterson G (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPD. *Genet Resour Crop Evol* 49(2):167–174. <https://doi.org/10.1023/A:1014716031095>
4. Fu Y, Guerin S, Peterson GW, Diederichsen A, Rowland GG, Richards KW (2003) RAPD analysis of genetic variability of regenerated seeds in the canadian flax cultivar CDC normandy. *Seed Sci Technol* 31(1):207–211. <https://doi.org/10.15258/sst.2003.31.1.22>
5. Kumari A, Paul S, Sharma V (2017) Genetic diversity analysis using RAPD and ISSR markers revealed discrete genetic makeup in relation to fibre and oil content in *Linum usitatissimum* L. genotypes. *Nucleus* 61(1):45–53. <https://doi.org/10.1007/s13237-017-0206-7>
6. Spielmeyer W, Green AG, Bittisnich D, Mendham N, Lagudah ES (1998) Identification of quantitative trait loci contributing to fusarium wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). *Theor App Genet* 97(4):633–641. <https://doi.org/10.1007/s001220050939>
7. Chandrawati, Maurya R, Singh PK, Ranade SA, Yadav HK (2014) Diversity analysis in Indian genotypes of linseed (*Linum usitatissimum* L.) using AFLP markers. *Gene* 549(1):171–178.

<https://doi.org/10.1016/j.gene.2014.07.067>

8. Cloutier S, Niu Z, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet 119(1):53–63. <https://doi.org/10.1007/s00122-009-1016-3>
9. Cloutier S, Ragupathy R, Niu Z, Duguid S (2011) SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. Mol Breeding 28(4):437–451. <https://doi.org/10.1007/s11032-010-9494-1>
10. Soto-Cerda BJ, Carrasco RA, Aravena GA, Urbina HA, Navarro CS (2011) Identifying novel polymorphic microsatellites from cultivated flax (*Linum usitatissimum* L.) following data mining. Plant Mol Biol Report 29(3):753–759. <https://doi.org/10.1007/s11105-010-0270-5>
11. Asgarinia P, Cloutier S, Duguid S, Rashid K, Mirlohi AF, Banik M, Saeidi G (2013) Mapping quantitative trait loci for powdery mildew resistance in flax (*Linum usitatissimum* L.). Crop Sci 53:2462–2472. <https://doi.org/10.2135/cropsci2013.05.0298>
12. Chandrawati, Yadav HK (2017) Development of linkage map and mapping of QTLs for oil content and yield attributes in linseed (*Linum usitatissimum* L.). Euphytica 213(11):258. <https://doi.org/10.1007/s10681-017-2050-x>
13. Wu J, Zhao Q, Wu G, Zhang S, Jiang T (2017) Development of Novel SSR Markers for Flax (*Linum usitatissimum* L.) Using Reduced-Representation Genome Sequencing. Front Plant Sci 7:2018. <https://doi.org/10.3389/fpls.2016.02018>
14. Pan G, Chen A, Li J, Huang S, Tang H, Chang L, Zhao L, Li D (2020) Genome-wide development of simple sequence repeats database for flax (*Linum usitatissimum* L.) and its use for genetic diversity assessment. Genet Resour Crop Evol 67(4):865–874. <https://doi.org/10.1007/s10722-020-00882-y>
15. Kumar S, You FM, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genom 13(1):684. <https://doi.org/10.1186/1471-2164-13-684>
16. Yi L, Gao F, Siqin B, Zhou Y, Li Q, Zhao X, Jia X, Zhang H (2017) Construction of an SNP-based high-density linkage map for flax (*Linum usitatissimum* L.) using specific length amplified fragment sequencing (SLAF-seq) technology. PLoS One 12(12):e0189785. <https://doi.org/10.1371/journal.pone.0189785>
17. Xie D, Dai Z, Yang Z, Sun J, Zhao D, Yang X, Zhang L, Tang Q, Su J (2018) Genome-Wide Association Study Identifying Candidate Genes Influencing Important Agronomic Traits of Flax (*Linum usitatissimum* L.) Using SLAF-sEq. Front Plant Sci 8:2232. <https://doi.org/10.3389/fpls.2017.02232>
18. Soto-Cerda BJ, Cloutier S, Quian R, Gajardo HA, Olivos M, You FM (2018) Genome-Wide Association Analysis of Mucilage and Hull Content in Flax (*Linum usitatissimum* L.) Seeds. Int J Mol Sci 19(10):2870. <https://doi.org/10.3390/ijms19102870>
19. You FM, Xiao J, Li P, Yao Z, Jia G, He L, Kumar S, Soto-Cerda B, Duguid SD, Booker HM, Rashid KY, Cloutier S (2018) Genome-Wide Association Study and Selection Signatures Detect Genomic Regions Associated with Seed Yield and Oil Quality in Flax. Int J Mol Sci 19(8):2303. <https://doi.org/10.3390/ijms19082303>
20. Singh N, Agarwal N, Yadav HK (2019) Genome-wide SNP-based diversity analysis and association mapping in linseed (*Linum usitatissimum* L.). Euphytica 215(8). <https://doi.org/10.1007/s10681-019-2462-x>
21. Hoque A, Fiedler JD, Rahman M (2020) Genetic diversity analysis of a flax (*Linum usitatissimum* L.) global collection. BMC Genom 21(1):557. <https://doi.org/10.1186/s12864-020-06922-2>
22. Zhang J, Long Y, Wang L, Dang Z, Zhang T, Song X, Dang Z, Pei X (2018) Consensus genetic linkage map construction and QTL mapping for plant height-related traits in linseed flax (*Linum usitatissimum* L.). BMC

Plant Biol 18(1):160. <https://doi.org/10.1186/s12870-018-1366-6>

23. Wu J, Zhao Q, Zhang L, Li S, Ma Y, Pan L, Lin H, Wu G, Yuan H, Yu Y, Wang X, Yang X, Li Z, Jiang T, Sun D (2018) QTL Mapping of Fiber-Related Traits Based on a High-Density Genetic Map in Flax (*Linum usitatissimum* L.). Front Plant Sci 9:885. <https://doi.org/10.3389/fpls.2018.00885>
24. Xie D, Dai Z, Yang Z, Tang Q, Deng C, Xu Y, Wang J, Chen J, Zhao D, Zhang S, Zhang S, Su J (2019) Combined genome-wide association analysis and transcriptome sequencing to identify candidate genes for flax seed fatty acid metabolism. Plant Sci 286:98–107. <https://doi.org/10.1016/j.plantsci.2019.06.004>
25. Lv H, Yang L, Kang J, Wang Q, Wang X, Fang Z, Liu Y, Zhuang M, Zhang Y, Lin Y, Yang Y, Xie B, Liu B, Liu J (2013) Development of InDel markers linked to Fusarium wilt resistance in cabbage. Mol Breeding 32(4):961–967. <https://doi.org/10.1007/s11032-013-9925-x>
26. Liu B, Wang Y, Zhai W, Deng J, Wang H, Cui Y, Cheng F, Wang X, Wu J (2013) Development of InDel markers for *Brassica rapa* based on whole-genome re-sequencing. Theor Appl Genet 126(1):231–239. <https://doi.org/10.1007/s00122-012-1976-6>
27. Liu G, Xu J, Zhang M, Li P, Yang X (2016) Exploiting Illumina sequencing for the development of InDel markers in watermelon (*Citrullus lanatus*). J Hortic Sci Biotech 91(3):220–226. <https://doi.org/10.1080/14620316.2015.1133609>
28. Guo G, Zhang G, Pan B, Diao W, Liu J, Ge W, Gao C, Zhang Y, Jiang C, Wang S (2019) Development and Application of InDel Markers for *Capsicum spp.* Based on Whole-Genome Re-Sequencing. Sci Rep 9(1):3691. <https://doi.org/10.1038/s41598-019-40244-y>
29. Hayashi K, Yoshida H, Ashikawa I (2006) Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. Theor Appl Genet 113(2):251–260. <https://doi.org/10.1007/s00122-006-0290-6>
30. Wu D, Wu H, Wang C, Tseng H, Hwu K (2013) Genome-wide InDel marker system for application in rice breeding and mapping studies. Euphytica 192(1):131–143. <https://doi.org/10.1007/s10681-013-0925-z>
31. Sahu PK, Mondal S, Sharma D, Vishwakarma G, Kumar V, Das BK (2017) InDel marker based genetic differentiation and genetic diversity in traditional rice (*Oryza sativa* L.) landraces of Chhattisgarh, India. PLoS One 12(11):e0188864. <https://doi.org/10.1371/journal.pone.0188864>
32. Yang Z, Dai Z, Xie D, Chen J, Tang Q, Cheng C, Xu Y, Wang T, Su J (2018) Development of an InDel polymorphism database for jute via comparative transcriptome analysis. Genome 61(5):323–327. <https://doi.org/10.1139/gen-2017-0191>
33. Zhang T, Gu M, Liu Y, Lv Y, Zhou L, Lu H, Liang S, Bao H, Zhao H (2017) Development of novel InDel markers and genetic diversity in *Chenopodium quinoa* through whole-genome re-sequencing. BMC Genom 18(1):685. <https://doi.org/10.1186/s12864-017-4093-8>
34. Adedze YMN, Lu X, Xia Y, Sun Q, Nchongboh CG, Alam MA, Liu M, Yang X, Zhang W, Deng Z, Li W, Si L (2021) Agarose-resolvable InDel markers based on whole genome re-sequencing in cucumber. Sci Rep 11:3872. <https://doi.org/10.1038/s41598-021-83313-x>
35. Ngan TP, Kim MK, Sim SC (2016) Genetic variations of F1 tomato cultivars revealed by a core set of SSR and InDel markers. Sci Hortic 212:155–161. <https://doi.org/10.1016/j.scienta.2016.09.043>
36. Liu S, An Y, Tong W, Qin X, Samarina L, Guo R, Xia X, Wei C (2019) Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of InDel markers for genetic research. BMC Genom 20(1):935. <https://doi.org/10.1186/s12864-019-6347-0>

37. Kizil S, Basak M, Guden B, Tosun HS, Uzun B, Yol E (2020) Genome-Wide Discovery of InDel Markers in Sesame (*Sesamum indicum* L.) Using ddRADSeq. *Plants (Basel)* 9(10):1262. <https://doi.org/10.3390/plants9101262>
38. Wang X, Shen F, Gao Y, Wang K, Chen R, Luo J, Yang L, Zhang X, Qiu C, Li W, Wu T, Xu X, Wang Y, Cong P, Han Z, Zhang X (2020) Application of genome-wide insertion/deletion markers on genetic structure analysis and identity signature of *Malus* accessions. *BMC Plant Biol* 20(1):540. <https://doi.org/10.1186/s12870-020-02744-2>
39. Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8(19):4321–4325. <https://doi.org/10.1093/nar/8.19.4321>
40. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129. <https://doi.org/10.1093/bioinformatics/bti282>
41. Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour* 4(2):359–361. <https://doi.org/10.1007/s12686-011-9548-7>
42. Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446. <https://doi.org/10.1534/genetics.104.026831>
43. Cloutier S, Miranda E, Ward K, Radovanovic N, Reimer E, Walichnowski A, Datla R, Rowland G, Duguid S, Ragupathy R (2012) Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (*Linum usitatissimum* L.). *Theor Appl Genet* 125(4):685–694. <https://doi.org/10.1007/s00122-012-1860-4>
44. Wei L, Miao H, Li C, Duan Y, Niu J, Zhang T, Zhao Q, Zhang H (2014) Development of SNP and InDel markers via de novo transcriptome assembly in *Sesamum indicum* L. *Mol. Breeding* 34(4):2205–2217. <https://doi.org/10.1007/s11032-014-0174-4>
45. Zhang M, Yu Q, Liu Z, Liang C, Zhang B, Li M (2017) UBX domain-containing proteins are involved in lipid homeostasis and stress responses in *Pichia pastoris*. *Int J Biochem Cell Biol* 90:136–144. <https://doi.org/10.1016/j.biocel.2017.08.006>
46. Nakamura Y, Koizumi R, Shui G, Shimojima M, Wenk MR, Ito T, Ohta H (2009) *Arabidopsis* lipins mediate eukaryotic pathway of lipid metabolism and cope critically with phosphate starvation. *Proc Natl Acad Sci U S A* 106(49):20978–20983. <https://doi.org/10.1073/pnas.0907173106>
47. Eastmond PJ, Quettier AL, Kroon JT, Craddock C, Adams N, Slabas AR (2010) Phosphatidic acid phosphohydrolase 1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in *Arabidopsis*. *Plant Cell* 22(8):2796–2811. <https://doi.org/10.1105/tpc.109.071423>
48. Mietkiewska E, Siloto RM, Dewald J, Shah S, Brindley DN, Weselake RJ (2011) Lipins from plants are phosphatidate phosphatases that restore lipid synthesis in a *pah1Δ* mutant strain of *Saccharomyces cerevisiae*. *FEBS J* 278(5):764–775. <https://doi.org/10.1111/j.1742-4658.2010.07995.x>

Tables

Table 1

Summary of the original sequencing data of 84-3 and 87-3

Samples	Clean base (G)	Map ratio (%)	Q20 (%)	Depth
84 – 3	3.69	96.34	98.11	18.93
87 – 3	1.01	95.1	98.01	6.02
Average	2.35	95.72	98.06	12.475

Table 2

Correlation coefficients between InDel markers and oil content traits in flax

Marker	Oil content	InDel-14	InDel-26	InDel-40	InDel-74
InDel-14	-0.015				
InDel-26	0.278*	0.011			
InDel-40	-0.186	0.023	0.106		
InDel-74	-0.070	0.743**	0.145	0.123	
InDel-81	-0.543**	0.060	-0.118	0.480**	0.228
* P < 0.05 indicates a significant correlation.					
** P < 0.01 indicates an extremely significant correlation.					

Table 3

Correlation coefficients between InDel markers and type of flax

Marker	Type	InDel-14	InDel-26	InDel-40	InDel-74
InDel-14	0.083				
InDel-26	-0.475**	0.011			
InDel-40	0.250*	0.023	0.106		
InDel-74	0.175	0.743**	0.145	0.123	
InDel-81	0.639**	0.060	-0.118	0.480**	0.228
* P < 0.05 indicates a significant correlation.					
** P < 0.01 indicates an extremely significant correlation.					

Table 4

Candidate gene information of InDel markers related to oil content traits in flax

Marker	Pos	Chr	ID	scaffold	Start	End	Functional annotation
InDel-26	20001016	12	Lus10031534	scaffold863	20003169	20003873	RNA-binding (RRM/RBD/RNP motifs) family protein
			Lus10031535	scaffold863	19995796	19998319	Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B prime gamma
InDel-81	12546472	8	Lus10025284	scaffold339	12554843	12558982	Lipin family protein
			Lus10010835	scaffold440	12531041	12533379	Protein kinase superfamily protein

Figures

Figure 1

Frequency of InDel type of flax

Figure 2

The number of InDel (A) and density of InDel (B) on each chromosome in the flax genome

Figure 3

Amplification products in 87-3 and 84-3 with InDel-25–InDel-48. M: 2000bp marker; 1: 87-3, 2: 84-3

Figure 4

Amplification products of 69 flax accessions using InDel-26 (A) and InDel-81 (B). a: InDel-26 amplified band; b: InDel-81 amplified band. M: 2000bp marker; 1 and 2 represent male and female respectively; 3-38: flax accessions for oil; 39-69, flax accessions for fibre.

Figure 5

Cluster analysis of 69 flax accessions based on 32 InDel markers; red numbers represent oil flax, blue represent fibre flax

Figure 6

(A) Delta K values for different assumed (K) populations and (B) population structure analyse divided 69 flax accessions into two populations based on 32 InDel markers (K=2)

Figure 7

Principal component analysis plot of 69 flax accessions based on 32 InDel markers

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata.docx](#)