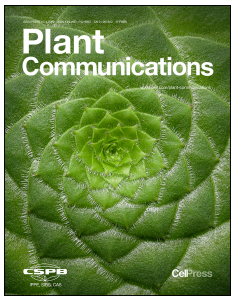
Journal Pre-proof



Creation of rapeseed germplasm with high polyunsaturated fatty acid content by relative introgression of *Brassica carinata*

Yan Niu, Wenwen Li, Yinghui Yang, Hao Wang, Zhesi He, Han Qin, Yikai Zhang, Dandan Hu, Jing Wang, Chunyu Zhang, Guangsheng Yang, Ian Bancroft, Jun Zou

PII: S2590-3462(24)00614-X

DOI: https://doi.org/10.1016/j.xplc.2024.101193

Reference: XPLC 101193

To appear in: *PLANT COMMUNICATIONS*

Received Date: 14 July 2023

Revised Date: 30 August 2024

Accepted Date: 15 November 2024

Please cite this article as: Niu, Y., Li, W., Yang, Y., Wang, H., He, Z., Qin, H., Zhang, Y., Hu, D., Wang, J., Zhang, C., Yang, G., Bancroft, I., Zou, J., Creation of rapeseed germplasm with high polyunsaturated fatty acid content by relative introgression of *Brassica carinata*, *PLANT COMMUNICATIONS* (2024), doi: https://doi.org/10.1016/j.xplc.2024.101193.

Creation of rapeseed germplasm with high polyunsaturated fatty acid content by relative introgression of *Brassica carinata*

Dear Editor,

Brassica napus L., commonly known as rapeseed, canola or oilseed rape, is the world's third oilseed crop and accounts for ~12% of major worldwide oil production (FAO., 2022). Rapeseed provides not only healthy and nutritionally balanced edible oil for humans, but also protein-rich fodder for animals, and renewable materials for biodiesel and industrial applications. Polyunsaturated fatty acids (PUFAs), including omega-3 fatty acids and omega-6 fatty acids, are mainly essential fatty acid and play a crucial role in the human diet, and humans obtain these fatty acids through dietary sources. Most crop seeds and vegetable oils, including canola, soybean, corn, and sunflower oils, are major sources of omega-6 fatty acids in the form of linoleic acid with low proportions of omega-3 fatty acids (Saini and Keum, 2018). There are several key genes responsible for PUFAs, such as *FAD2*, which converts oleic acid to linoleic acid and *FAD3*, which converts linoleic acid to linolenic acid (Okuley et al., 1994; Yadav et al., 1993).

As a young allotetraploid species formed by interspecific crosses between the diploid ancestors of *B. rapa* (2n = 20, AA) and *B. oleracea* (2n = 18, CC), *B. napus* (2n=38, AACC) has a limited genetic diversity due to the short history of cultivation and domestication and long-term double-low breeding (Allender and King, 2010; Friedt and Snowdon, 2010; Prakash et al., 2011). Wild relatives and other related genetic resources represent a favorable genetic variation that can be exploited to develop climate-resilient modern crops (Wang et al., 2020; Zhou et al., 2021). During the process of interspecific hybridization, exotic introgression occurs as a natural gene flow and it not only introduces new genetic variations but also increases the

complexity of crop genomes. However, how to mine the elite or favorable genetic variation with mixed genomic variations and introgression is a challenge. Therefore, we aim to explore the novel germplasm harboring rich exotic introgression with more efficiency and provide insights for explaining the regulatory mechanisms for exotic introgression and genomics-assisted wild relative introgression.

To broaden the genetic basis of rapeseed, through interspecific crosses between 75 *B. carinata* accessions and 122 *B. rapa* accessions, we obtained more than 20,000 lines of new-type *B. napus* gene pool (Hu et al., 2019; Zou et al., 2018). From the gene pool of new-type *B. napus*, we select one line called HLL45 with five large-scale homoeologous exchange (HE) including C5/B1, A2/C2, A3/C3, A4/C4 and A9/C8 (Figure 1A, Supplemental Figure 1). Interestingly, HE between C5/B1 resulted in an introgression fragment with more than 5Mb on C5 from B1 chromosome of *B. carinata* (Supplemental Figure 1). We investigated agronomic traits and quality traits for HLL45 and found that it is a novel germplasm of high polyunsaturated fatty acid (PUFA) content at 40%~45%, including linoleic acid (LEI) and linolenic acid (LEN) content. Compared to conventional *B. napus* varieties, the PUFA content of HLL45 was 15 percentage higher, mainly contributed by LEI, but its seed oil content decreased (Figure 1B and Supplemental Table 1). Our lines should therefore be considered pre-breeding material, from which new elite cultivars can be developed. Using a genetic map constructed by HG DH population (Hu et al., 2024), we identified a major QTL, *qFA-C5-4,* that explained 63.2%-64.6%, 37.3%-73.3%, and 7.4% of the phenotypic variation for oleic acid content (OLE), LEI and LEN, respectively (Figure 1B, Supplemental Figure 2 and Supplemental Table 2-5). Notably, the markers defining QTL *qFA-C5-4* were located in the *B. carinata* B1 segment introgression region on chromosome C5, which contains an orthologue of the *FAD2* gene, which encodes the enzyme that converts OLE to LEI (Supplemental Figure 2). Therefore, PUFA content in rapeseed appears to be associated with the B1

introgression from *B. carinata*.

To comprehensively study the characteristics and effect of exotic introgression, we assembled four high-quality genomes including HLL45 and its representational parents (Supplemental Figure 3 and 4) by a combination of Illumina short reads, Oxford Nanopore Technology (ONT) long reads, Hi-C reads and genetic map. The assembled genome for HLL45 comprised 344 contigs (N50, 18.0 Mb) and had a total size of 927 Mb, with high BUSCO completeness at 99.0% and a mapping rate of 99.7% from short reads, leading to a highly accurate and nearly complete assembly (Supplemental Figure 4 and Supplemental Table 6). We successfully anchored 99.0% of the contigs to 19 pseudochromosomes exhibiting good collinearity with HG genetic map (Hu *et al.*, 2024), and annotated 101,495 protein-coding genes for HLL45 (Supplemental Figure 5 and Supplemental Table 6). We identified a total of 11,611 exotic genes with mainly on A05 and C07 in HLL45, including 4,538 genes from *B. rapa and* 7,073 genes from *B. carinata.* Among exotic genes from *B. carinata*, 1,043 genes and 5,919 genes were exotic genes from B and C sub-genome respectively, with the rest genes coming from unanchored sequences. 80% of exotic genes from B sub-genome of *B. carinata* were located on chromosome C5 and were mainly enriched in fatty acid metabolism, unsaturated fatty acids biosynthesis and linoleic acid metabolism (Figure 1D). For the major QTL *qFA-C5-4* located in the 1.34 Mb region (50.0 Mb ~ 51.4 Mb) of C5 in HLL45 genome, there were 217 genes in the region and 131 genes were different expressed genes in all of 10 different seed development stages between HLL45 and ZS11 (Figure 1E and Supplemental Figure 6). *BcaFAD2.B1* (BnaHLLC05G045860) had the highest expression in seeds, and co-expression network analysis indicated that it was the hub gene of PUFA (Supplemental Figure 7). We identified a six-Mb exotic introgression on C5 (from 49.9 Mb to 55.9 Mb) in HLL45 from B1 of *B. Carinata*, showing the *FAD2* on C5 was replaced by *BcaFAD2.B1* (Figure 1F). Lines in HG population carrying the C5 SV contained 8.56% higher LEI than

those lines with no SV (Figure 1G and Supplemental Table 7). We further confirmed SV on C5 was stably inheritable (Supplemental Figure 8). In the HG population, the high-LEI lines carried the *BcaFAD2.B1* and C5 SV while the low-LEI lines did not (Figure 1H and Supplemental Table 8). These results implicate the six-Mb exotic introgression on C5 play an important role in increasing PUFA.

We further surveyed the expression level of *FAD2* orthologues of six species in *Brassica* and found that *BcaFAD2.B1* had the highest expression, which had 89% of total orthologs expression in *B. carinata* (Supplemental Figure 9). In HLL45, three *FAD2* orthologues was expressed and mainly contributed by *FAD2.B1* with reaching its highest expression at 36 DAF, while *FAD2.C5* was not expressed in HLL45 (Figure 1I). *FAD2.A5* and *FAD2.C1* in HLL45 were both expressed at about the similar level as them in traditional *B. napus* ZS11, while *FAD2.B1* was expressed more highly than *FAD2.*C5 in ZS11, which was confirmed by RT-PCR (Figure 1J). Except in seeds, *FAD2.B1* was also expressed in root, leaf and stem (Supplemental Figure 10). During the seed development, the PUFA content in HLL45 was higher than that of ZS11, especially LEI. After 18 DAF, the LEI content in HLL45 was higher than in ZS11, and the difference increased as the developmental time extends (Figure 1J and Supplemental Figure 11). Therefore, the increased PUFA content of HLL45 can be explained that *FAD2.C5* was replaced by *BcaFAD2.B1* with highly expressed level.

FAD2.B1 was the nearest clade to *FAD2* in *A. thaliana* and had many sequences in promoter and enhancer within the 800bp upstream of gene, while *FAD2.C5* did not have (Supplemental Figure 12 A, B and Supplemental Table 9). Compared to *BcaFAD2.B1* and *BniFAD2.B1* with one exon, *BnaFAD2.*C5 and *BolFAD2.C5* had one additional exon encoding 73 AA and same conserved domain (Supplemental Figure 12C and 13). Besides, *FAD2.C5* in HLL45 (BnaHLLC05G045860) and *BcaFAD2.B1* (Bca4012B1G035280) had the same structure, but they both had almost

identical C-termini but distinct N-termini with *BnaFAD2.C5* (BnaC05G0480500ZS) (Supplemental Figure 12D).

High PUFA content of rapeseed oil is a desirable trait which has been unachievable using conventional germplasm. We have demonstrated the use of wild relative introgression to achieve this, implicating high expression of one of the *FAD2* orthologues from the B genome, which is not normally present in oilseed rape, as causative. Our study provides both evidence for how PUFA content in seed oil above the normally occurring range is controlled, and an example for others to follow of how to use genomics to accelerate use of wild relative introgression for crop improvement.

Data availability

The genome assembly have been deposited in the Genome Warehouse at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession number PRJCA025800. The genomic sequencing data and transcriptome sequencing data have been deposited in National Genomics Data Center (GSA: CRA016277) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (32171982, 31970564), the National Key Research and Development Program of China (2021YFF1000100), the Fundamental Research Funds for the Central Universities (2662023PY004) and the UK Department for Environment, Food and Rural Affairs Oilseed Rape Genetic Improvement Network (OREGIN) (CH0110).

Author Contributions

Y.N. performed the research, analyzed the data and wrote the paper. W.L. performed field trials, genetic mapping and QTL analysis. Y.Y. performed phenotyping, sampling and molecular marker analysis. H. W. and Y. Z. contributed to the genome assembly. J.Z and H.Q contributed to plant material development, cultivation, phenotyping, and QTL analysis. Z.H. performed GDTPs analysis. D.H. contributed to QTL and genome resequencing analysis. J. W. and C.Z. contributed to the design for RT-PCR and fatty acid composition analysis. G.Y. contributed to the design of field trials and revise the manuscript. J.Z. and I.B. designed the project, analyzed the data, and wrote the manuscript. All authors edited and approved the final manuscript.

Declaration of competing interest

The authors declare no competing interests.

REFERENCES

Allender, C.J., and King, G.J. (2010). Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. BMC Plant Biol 10:54. 10.1186/1471-2229-10-54.

FAO. (2022). WORLD FOOD AND AGRICULTURE STATISTICAL YEARBOOK 2022 (FAO).

Friedt, W., and Snowdon, R. (2010). Oilseed rape. Oil crops:91-126.

Hu, D., Zhang, W., Zhang, Y., Chang, S., Chen, L., Chen, Y., Shi, Y., Shen, J., Meng, J., and Zou, J. (2019). Reconstituting the genome of a young allopolyploid crop, *Brassica napus*, with its related species. Plant Biotechnol J 17:1106-1118. 10.1111/pbi.13041.

Hu, D., Lu, J., Li, W., Yang, Y., Xu, J., Qin, H., Wang, H., Niu, Y., Zhang, H., Liu, Q., et al. (2024). The occurrence, inheritance, and segregation of complex genomic structural variation in synthetic *Brassica napus*. The Crop Journal https://doi.org/10.1016/j.cj.2024.01.002.

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., and Browse, J. (1994). *Arabidopsis FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6:147-158. 10.1105/tpc.6.1.147.

Prakash, S., Wu, X.i., and Bhat, S.R. (2011). History, Evolution, and Domestication of *Brassica* Crops (Plant Breeding Reviews).

Saini, R.K., and Keum, Y.S. (2018). Omega-3 and omega-6 polyunsaturated fatty acids: Dietary sources, metabolism, and significance — A review. Life Sciences 203:255-267.

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., Lyu, Z., Chen, L., Xu, S., Guo, J., et al. (2020). Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat. Science 36810.1126/science.aba5435.

Yadav, N.S., Wierzbicki, A., Aegerter, M., Caster, C.S., Pérez-Grau, L., Kinney, A.J., Hitz, W.D., Booth, J.R., Jr., Schweiger, B., Stecca, K.L., et al. (1993).

Cloning of higher plant omega-3 fatty acid desaturases. Plant Physiol 103:467-476. 10.1104/pp.103.2.467.

Zhou, Y., Bai, S., Li, H., Sun, G., Zhang, D., Ma, F., Zhao, X., Nie, F., Li, J., Chen, L., et al. (2021). Introgressing the *Aegilops tauschii* genome into wheat as a basis for cereal improvement. Nat Plants 7:774-786. 10.1038/s41477-021-00934-w.

Zou, J., Hu, D., Mason, A.S., Shen, X., Wang, X., Wang, N., Grandke, F., Wang, M., Chang, S., Snowdon, R.J., et al. (2018). Genetic changes in a novel breeding population of *Brassica napus* synthesized from hundreds of crosses between *B. rapa* and *B. carinata*. Plant Biotechnol J 16:507-519. 10.1111/pbi.12791.

Figure 1 The exotic introgression on C5 in HLL45 from *B. carinata* increasing PUFA content.

(A) Large-scale homoeologous exchanges (HEs) of new-type *B. napus* HLL45 identified by Genome Display Tile Plots. Red, black, orange, purple and green arrows represent HEs of A2/C2, A3/C3, A4/C4, C5/B1 and A9/C8, respectively.

(B) The upper figure represents the fatty acid component distribution for five *B. napus* accessions. LEI, LEN, OLE and EA mean linoleic acid content, linolenic acid content, oleic acid content and erucic acid content, respectively. The bottom figure represents the QTL mapping for fatty acid traits using HG DH populations (280 lines).

(C) The genomic features of exotic introgression in HLL45 (100 kb window). I-V means gene density, TE density, exotic gene density from BraA (*B. rapa*), BcaB and BcaC (B and C sub-genome of *B. carinata*).

(D) Pathway enrichment plot of exotic genes from BcaB in HLL45 (top 15).

(E) The volcano plot of 217 genes within the major QTL on C5 in seeds (40 days after pollination). Compared to ZS11, most candidate genes in HLL45

were significantly up-regulated, especially BnaHLLC05G045860 (*BcaFAD2.B1*).

(F) The collinearity of chromosome C5 among HLL45, *B. napus* Darmor-*bzh* and *B. carinata* C4012.

(G) The generation, transmission process and effect of SV on C5. Difference on linoleic acid (LEI) and oleic acid (OLE) content between lines with C5 SV and no SV from HG population. Statistical significance was determined by two-tailed unpaired Student's T-test.

(H) PCR validation for C5 SV and exotic gene *BcaFAD2.B1*. The left figure represents PCR genotyping assay for lines with different linoleic acid (LEI) content from HG populations on C5 SV markers. PCR amplicons of 211 bp and 577 bp represent the presence and absence of the SV, respectively. The left figure represents PCR genotyping assay for HLL45, HS3 and lines with different linoleic acid (LEI) content from HG populations on *BcaFAD2.B1* markers. PCR amplicons of 424 bp and 222 bp represent the presence and absence of *BcaFAD2.B1*, respectively.

(I) The gene expression pattern of three *FAD2* orthologues in different tissues and seed development in HLL45 and conventional *B. napus* ZS11 based on FPKM.

(J) The upper figure means LEI and LEN content in different seed development of new-type *B. napus* HLL45 and conventional *B. napus* ZS11. LEI means linoleic acid content and LEN means linolenic acid content. The bottom figure means the gene expression pattern of three *FAD2* orthologues in different seed development of HLL45 and conventional *B. napus* ZS11 based on the relative expression level from RT-PCR.

Yan Niu[1], Wenwen Li[1], Yinghui Yang[1], Hao Wang[1], Zhesi He[2], Han Qin[1], Yikai Zhang[1], Dandan Hu[1], Jing Wang[1], Chunyu Zhang[1], Guangsheng Yang[1], Ian Bancroft[2*], Jun Zou[1*]

[1]National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

[2]Department of Biology, University of York, Heslington, York, YO10 5DD, UK

***Correspondence: Jun Zou (zoujun@mail.hzau.edu.cn) and Ian Bancroft (ian.bancroft@york.ac.uk)**