

Genes gained and lost during the evolution and domestication of *Brassica napus*

Hong Zhou

Triticeae research institute, Sichuan Agricultural University

Xueling Ye

Triticeae Research Institute, Sichuan Agricultural University

Yanling Ma

Institute of crop sciences, Chinese Academy of Agricultural Sciences

Haiyan Hu

School of life science and technology, Henan Institute of science and technology

Yuming Wei

Triticeae Research Institute, Sichuan Agricultural University

Zheng Youliang

Triticeae Research Institute, Sichuan Agricultural University

Yaxi Liu

Triticeae Research Institute, Sichuan Agricultural University

Chunji Liu (✉ chunji.liu@csiro.au)

Research article

Keywords: Polyploidization, gene ratio, gene function, diploids, allopolyploids

Posted Date: January 9th, 2020

DOI: <https://doi.org/10.21203/rs.2.20421/v1>

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

Abstract

Background: *Brassica napus* is one of the most important sources of vegetable oil for human nutrition and biofuel. It is an allotetraploid formed about 7500 years ago by hybridization between *B. rapa* and *B. oleracea*. Results from earlier studies show that the allopolyploidization process was accompanied by rapid and intensive changes and abundant homoeologous exchanges between the subgenomes have been accumulated during its short history of evolution.

Results: By comparing differences between 19 artificially synthesized and 30 natural genotypes, we assessed possible changes in gene ratio, diversity and functional groups during the evolution and domestication of this species. This comparison revealed that gene ratio and diversity between the two subgenomes have hardly changed. However, large numbers of genes have been lost and many new genes gained. Compared with the artificial genotypes, the natural ones contain much lower proportions of genes conferring resistance and tolerance to biotic and abiotic stresses but much higher proportions of genes associated with seeds development and metabolic processes. The diploid donor for the A subgenome of *B. napus* contributed more genes involved in agronomic traits and the C subgenome donor contributed more genes related to cellular development and metabolic process.

Conclusions: Our results show that genes conditioning resistance and tolerance to both biotic and abiotic stresses have suffered stronger selection during the evolution and domestication of *B. napus*, and that changes in different aspects including gene content and genome size in the allotetraploid are not random but dictated by its two diploid donors.

Background

Polyploidization, a process of enveloping two or more genomes within a single nucleus, is a common feature in the evolution of many species [1, 2]. Depending on whether a similar or two or more different genomes are involved, they can be divided into auto- or allopolyploids. Several important crop species are allopolyploids. They include common wheat (*Triticum aestivum* L, genome AABBDD, $2n = 6x = 42$), durum wheat (*T. durum* L, genome AABB, $2n = 4x = 28$), cotton (*Gossypium hirsutum*, genome AACC, $2n = 4x = 52$), and *Brassica napus* (genome AACC, $2n = 4x = 36$).

Results from previous studies show that different species react differently to allopolyploidization events. Wheat [3] and *Brassica* [4] are among the species for which rapid and massive changes in DNA sequences were detected in newly synthesized allopolyploid genotypes. However, such changes were not detected in either cotton [5] or *Arabidopsis* [6]. Further, asymmetric evolution, measured with different characteristics including gene number, genetic diversity or methylation, has been widely reported for various species of allopolyploids. It is widely believed that different subgenomes of a polyploid species may respond differently to internal as well as external stimuli [7]. Possible evolutionary advantages of subgenome asymmetry have been proposed in various species [8–13].

Brassica napus is one of the most important oil crops with a world annual production of over 60 million tonnes [14, 15] and it has also been widely used as a model species in studying allopolyploid evolution [15, 16].

It is a recent allopolyploid evolved from the spontaneous hybridization between *B. rapa* ($2n = 20$; genome AA) and *B. oleracea* ($2n = 18$; genome CC) occurred some 7,500 years ago [15, 17–20].

The initial report on rapid and extensive changes in DNA sequences in the early generations after polyploidization were based on a study of analysing artificially synthesized allopolyploid genotypes of *B. napus* and *B. juncea* with the use of several DNA probes [15]. Sequencing and assembling the genome of the diploid *B. oleracea* [7] and comparing it with that of *B. rapa* (The Brassical *rapa* genome sequencing project consortium 2011) detected asymmetrical evolution in *B. napus* [7]. Sequencing the genome of the allotetraploid *B. napus* revealed that abundant homeologous exchanges may have occurred between its two subgenomes [20].

To enrich the genetic diversity depleted from interspecific hybridization and intensive selection, artificially synthesized genotypes analogous to the natural allotetraploid *B. napus* were obtained and used in breeding programs [21–24]. Different from those natural genotypes, the artificial genotypes have hardly been exposed to evolution and domestication. Thus, differences between the artificial and the natural genotypes reflect to a large degree changes accumulated during the evolution and domestication of this species. Further, recent studies have shown the significance of dispensable genome components [25–27] thus the potential issues with the use of a single genotype to represent any species in different studies. The sequences for multiple genotypes for both the natural and artificial allotetraploids genotypes of *B. napus* offer a unique opportunity to investigate possible changes in genes accumulated during the evolution and selection of this important crop species. Further, the availability of sequences from multiple genotypes for each of the two diploid donor species [7, 28–30] makes it possible to differentiate genes contributed by either of them in the allopolyploids.

Results

Genes lost and gained during the evolution and domestication of the allotetraploid

Hits for about 90% of the reads from the 30 natural and 19 artificial *B. napus* genotypes were found in the reference genome of *B. napus* and they contain a total of 46,941 genes (Fig. 1). Hits not found in the reference genome accounted for about 9.2% of the reads from the natural genotypes and 11.7% of those from the artificial genotypes. Assembling these unmapped reads based on SOAP denovo identified 24,505 unique genes from the natural genotypes, and 24,879 unique genes from the artificial genotypes (Additional file 1: Table S1, Fig. 1). Thus, the ratio of genes shared between the artificial and natural genotypes is about 48.7%, and the unique genes detected in both the artificial and natural genotypes accounted for about 51.3% of all genes detected in this study.

Difference in functional groupings of genes between natural and artificial genotypes

GO annotation assessments against the 24,505 unique genes from the natural genotypes identified 80 enriched GO terms (Fig. 2, Additional file 2: Table S2, S3). They include those involved in plant development including 'transferase activity', 'macromolecule metabolic process', 'cellular macromolecule metabolic process', 'phosphorus metabolic process', 'seed development', and 'fruit development'. A similar assessment against the 24,879 unique genes from the artificial genotypes identified 48 enriched GO terms. Different from those from the natural genotypes, most of the enriched GO terms are related to responses to stresses.

Specifically, they include 'response to stress', 'chromosome organization', 'cell cycle', 'cell cycle process', 'cell division' and 'mitotic cell cycle process'.

KEGG pathway analyses found that, compared with those in the artificial genotypes, significantly enriched pathways in the natural genotypes are involved in 'Phosphatidylinositol signaling system', 'RNA degradation', and 'Inositol phosphate metabolism'. These pathways have been shown to regulate pollen tube growth, root hair tip growth, flowering and maturation [31–34]. In contrast, pathways of 'Aminoacyl-tRNA biosynthesis', 'Propanoate metabolism' and 'Ribosome biogenesis in eukaryotes' were significantly enriched among the unique genes in the artificial genotypes ($P < 0.05$) (Additional file 3: Table S4, Fig. 3). Previous studies show that these pathways are predominantly involved in drought stress and disease response [35–38]. These results indicate that many genes related to stress responses have been lost and that genes related to seeds development and metabolic processes have been significantly increased during the evolution and domestication of this species.

Difference in *B. napus* genes derived from the two different diploid donors

To minimize possible influences of dispensable genome components (DGCs), transcriptomic data from 30 *B. rapa* and 26 *B. oleracea* were used to assess genes derived from either of these two diploid donors in the formation and domestication of the allotetraploid *B. napus*. With the use of an identity threshold of 95%, 39,583 non-redundant CDS (coding sequences) were identified from the *B. rapa* genotypes and 42,521 from the *B. oleracea* genotypes. Aligning these CDS against the genome of the allotetraploid *B. napus* found that the C subgenome shared 52.8% of its genes with its diploid donor *B. oleracea* and that the A subgenome shared 47.2% of its genes with its diploid donor *B. rapa*. Clearly, the ratio of shared genes between the A subgenome and its diploid progenitor *B. rapa* was substantially lower than that between the C subgenome and its diploid progenitor *B. oleracea* (Table 1).

Table 1

Genes shared between subgenomes and their respective progenitors and gene ratios between the subgenome.

Species	Genome	Gene number	Retained from AA		Retained from CC		A/(A + C)	C/(A + C)
			No. CDS	percentage	No. CDS	percentage		
B. rapa	AA	39583						
B. oleracea	CC	42521						
Lost genes	AACC	24783	8659	21.88%	9612	22.61%	47.39%	52.61%
Retained genes	AACC	46941	14821	37.44%	17314	40.72%	46.12%	53.88%
Natural allotetraploid	AACC	71377	18569	46.91%	20183	47.47%	47.92%	52.08%
Artificial allotetraploid	AACC	71724	19653	49.65%	21897	51.50%	47.30%	52.70%

Lost genes represent genes detected in the artificial allotetraploid genotypes but not in the natural genotypes of *B. napus*. Retained genes represent genes detected in both the artificial and natural allotetraploid genotypes of *B. napus*.

GO term analysis showed significant differences in functional groups of genes derived from the two diploid progenitors. Of the genes shared between the natural and artificial genotypes, the numbers of enriched GO terms for two ('molecular function' and 'biological process') of the three functional classes were significantly higher for genes derived from *B. rapa* than those from *B. oleracea*. Those genes derived from *B. rapa* were enriched in transcription factor, plant-type development and negative regulation of several biological process. Of the genes unique in the artificial allotetraploid genotypes, the numbers of enriched GO terms from *B. oleracea* were significantly larger than those from *B. rapa* (Table 2). The most significantly enriched GO terms included cellular development, response to abiotic stimulus, metabolic process and positive regulation of several biological processes (Fig. 4; Additional file 4: Table S5,6,7).

Table 2

Numbers of enriched GO terms among “retained” and “lost” genes derived from two diploid donors.

GO Term	Genes shared between natural and artificial genotypes		Unique genes in the artificial genotypes	
	genes derived from <i>B. rapa</i>	genes derived from <i>B. oleracea</i>	genes derived from <i>B. rapa</i>	genes derived from <i>B. oleracea</i>
cellular component	3	19	0	46
	15	1	7	16
molecular function				
biological process	73	15	5	77
“retained” genes represent genes shared between natural and artificial <i>B. napus</i> , “lost” genes represent unique genes in the artificial allotetraploid genotypes				

Discussion

B. napus is an important oil crop and is also a model species used widely in studying polyploid evolution. Previous studies detected rapid and extensive changes in DNA sequences immediately following the polyploidization events [15] and asymmetric evolution with abundant homeologous exchanges has been reported from sequencing genomes of the allotetraploid [20] and its diploid donor species [7, 28]. Thus, difference in gene ratio between the two subgenomes between the artificial and natural genotypes was expected. Similar to those reported in previous studies, we also detected subgenome asymmetry in this study and found that the A subgenome contains a smaller number of genes than the C subgenome in each of the allopolyploid genotypes assessed. Our results also show that the genes contributed to the allotetraploid *B. napus* differ in functional categories and they also suffered different selection pressures during evolution and domestication. However, little change in gene ratios between the artificial and natural genotypes was detected, although the two groups of genotypes differ dramatically in time exposed to evolution and domestication. The most likely explanation to this lack of difference in gene ratio between these two groups of genotypes is that changes, including homeologous exchanges and gene conversion events in the evolution and domestication [19, 20], may not be random but dictated by the founding genomes. This likelihood is further supported by the facts that the differences in both genome size and relative diversity between the two subgenomes of *B. napus* [20] also exist between its two diploid donors [7, 14, 39, 40].

Although sequences from multiple genotypes were used for each of the species investigated, they unlikely captured all genes present in any of them. Clearly, only expressed genes can be detected from the transcriptomic sequences, and the expression of many genes can be tissue-specific [41–43]. Importantly, obtaining all genes from any of the genotypes used was not the intention of this study and it is also not required for estimating either gene ratios between subgenomes or relative enrichment of genes in a given categories.

Differences in gene functional groupings between the artificial and natural genotypes showed that the genes conferring resistance and tolerance to both biotic and abiotic stresses were reduced and that genes involved in vegetative and productive growth were increased during the evolution and domestication of *B. napus*. The change for genes conferring resistance to biotic stresses detected during evolution and domestication was similar to that detected in an earlier study by assessing a selected panel of resistance genes [27]. The likelihood that genes conferring disease resistance tend to suffer stronger selection than genes in other categories during domestication has also been reported in other species including common bean (*Phaseolus vulgaris*) [44] and barley [45]. The reduced numbers of genes conferring tolerance to biotic and abiotic stresses during evolution and domestication provide further evidence suggesting that investment in defence requires allocation of limiting resources and hence trade-offs with other traits [46] or that defence strategies could have harmful pleiotropic effects [47–49].

It is of note that as many as 51% of the genes detected in this study were unique to either the artificial or natural genotypes of *B. napus*. Considering from the pan-genome point of view, they represent the dispensable genome component (DGC). This percentage is more than doubled that reported for this species in an earlier study [25]. However, it is comparable with those reported in maize [50, 51] and rice [52]. Clearly, DGCs would vary depending on not only the number of genotypes but also the relative diversity of the genotypes assessed. The size of the DGC would increase with the increase in the numbers of individuals assessed for a given species and assessing genotypes with a wider diversity range should lead to larger DGCs. An earlier study showed that the artificial genotypes are more diverse than the natural ones [25]. Thus, the inclusion of both artificial and natural genotypes in this study must have contributed to the large DGC obtained.

Interestingly, genes in the DGC are similar to those ‘lost’ during the evolution and domestication in that they all contain higher proportions of genes conferring disease resistance and lower proportions of genes related to vegetative and productive growth [44, 45]. The results from this study indicate that the similarity is unlikely due to the possibility that most of the genes lost during evolution belong to DGCs. The results from comparing the artificial and natural genotypes of *B. napus* showed that the numbers of ‘lost’ and ‘gained’ genes during evolution and domestication are similar (Fig. 1; Additional file 1: Table S1). It may not be difficult to understand why wild genotypes contain higher percentages of genes conferring tolerance to biotic and abiotic stresses as traits selected for in domestication and breeding, such as kernel sizes in cereals and oil content in Brassica, may not align with those favoured by a plant species under natural selection. However, possible reasons for the larger proportions of genes related to tolerance to biotic and abiotic stresses in the DGCs are not clear. Understanding the underlying reasons why many more genes conferring tolerance to biotic and abiotic stresses in DGCs could be highly values in designing plant breeding.

Conclusions:

By comparing artificial and natural genotypes of *B. napus*, we found that, although gene ratios between the two subgenomes having hardly changed, genes conditioning resistance and tolerance to both biotic and abiotic stresses have suffered stronger selection and the proportions of genes associated with seeds development and metabolic processes have increased during the evolution and domestication of this allotetraploid species. Our results also show that the functional groups of genes contributed by the two diploid donors seem to differ, with the A subgenome donor contributed more genes involved in agronomic traits and the C subgenome donor contributed more genes related to cellular development and metabolic process. Available evidences all point to the likelihood that changes accumulated during the evolution of *B. napus* are not random but dramatically affected by the genomes of its two diploid donors.

Methods

Genome and transcriptome sequences used

Sequences used in this study were all in the public domains. They included the genome sequences from 30 natural genotypes of *B. napus* and 19 artificially synthetic genotypes analogous to the natural allotetraploid *B. napus* (artificial *B. napus*) [16]. The artificial genotypes were obtained from independent hybridizations between the A-subgenome donor *B. rapa* and the C-subgenome donors *B. oleracea*. These genomic sequences were paired-end 100-bp reads generated on an Illumina HiSeq2000 (Illumina Inc., San Diego, CA) (Additional file 5: Table S8). Transcriptomic sequences from 30 datasets of *B. rapa* (diploid donor for the A subgenome) and 26 datasets of *B. oleracea* (the diploid donor for the C subgenome of the allotetraploid *B. napus*) were also used (Additional file 5: Table S9). The transcriptomic sequences were obtained from a wide range of studies on different environmental factors and treatments [29, 30]. All of the sequences were downloaded from the National Centre for Biotechnology Information (NCBI) Short Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>).

Genomic sequence analyses

The genomic sequences were assessed using the software FastQC (version 0.11.5) [53], and then trimmed and filtered using the software SolexaQA++ [54]. Low quality reads were removed with the criteria of $Q < 30$ and length < 50 bp. Following quality controls, the remaining reads from the 30 natural and 19 artificial genotypes were assessed against the reference genome of *B. napus* (GCF_000686985.2_Bra_napus_v2.0) using Bowtie2 [55] v2.2.6 (–end-to-end –sensitive). Unmapped reads were assembled using SOAPdenovo2-r240 [56] with the default settings based on k-mer size 37. The GapCloser (including in SOAPdenovo2 package) was used to reduce gaps among assembled scaffolds.

To remove contaminants, the assembled scaffolds were checked using the stand-alone DeconSeq toolkit (version 0.4.3) [57] with the default parameters. Genome sequences from bacterial, virus and human were downloaded from NCBI and used to build the ‘remove’ databases which were used to identify contaminant transcripts. Genome sequences of maize, rice, wheat, soybean (*Glycine max*), sorghum (*Sorghum Bicolor*), *Arabidopsis thaliana*, *Brachypodium distachyon* and *Medicago truncatula* were downloaded from EnsemblPlants database and used to build the ‘retain’ database. Gene prediction was conducted using

Augustus with options -species = arabidopsis –genemodel = complete (version 3.2.3) [58] with the default settings and a minimum length of 300nt (100 amino acids). Detailed information of reads used for mapping, assembly and annotation can be found in Table S1.

Transcriptomic sequence analyses

Following quality controls same as the procedure for genomic sequences, transcriptomic sequences from each of the two diploid species were assembled using Trinity (version 2.5.1) [59] with K-mers = 25 and de novo assembled with a minimum length of 200 bp. Redundant sequences were then removed using the cd-hit package (version 4.6.4) [60] with a sequence similarity threshold of 95%. CDS from the assembled sequences were then predicted using TransDecoder (version 5.0.2)

(<https://github.com/TransDecoder/TransDecoder/releases>).

Identification of shared genes and their functional prediction

In estimating genes shared between the subgenomes with their respective diploid donors, BLAST+ (version 2.7.1) [61] was used to identify syntenic genes with a minimum E-value of 1e-5, a minimum coverage size of 100 bp and a minimum identity threshold of 85%. The sequences from the diploid genotypes were first used as a query to BLAST against the sequences of the allotetraploid genotypes and a reciprocal analysis was then carried out by using the polyploid sequences as a query to BLAST against the diploid sequences. A custom script [62] was used to retrieve all those sequences where the polyploid or diploid sequences gave the best hits. These sequences were used to calculate the percentages of genes shared between a given subgenome and its diploid donors. GO (gene ontology) classification and KEGG (Kyoto Encyclopedia of Genes and Genomes) internal annotation were carried out using the KOBAS online service (<http://kobas.cbi.pku.edu.cn/index.php>) [63, 64].

Declarations

Availability of data and materials

All data analyzed during this study are included in the Additional files.

Acknowledgments

H.Z. and X.Y. are grateful to the Sichuan Agricultural University and the China Scholarship Council for funding his visit to CSIRO. H.H. wishes to acknowledge Henan Institute of Science and Technology and CSC for supporting her visit to CSIRO.

Funding

This work was supported by the Commonwealth Scientific and Industrial Organization (CSIRO), Australia (Project code: R-10191-01).

Authors' contributions

C.L., Y.-L.Z., Y.W. and Y.L. conceived the study; H.Z., X.Y., M.Y. and H.H. collected and analyzed data; H.Z. and C.L. prepared the manuscript with contribution from other authors.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Wendel JF: **Genome evolution in polyploids**. In: *Plant molecular evolution*. Springer; 2000: 225-249.
2. Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X: **Gene retention, fractionation and subgenome differences in polyploid plants**. *Nature plants* 2018, **4**(5):258-268.
3. Feldman M, Levy AA: **Genome evolution due to allopolyploidization in wheat**. *Genetics* 2012, **192**(3):763-774.
4. Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC: **Genomic changes in resynthesized Brassica napus and their effect on gene expression and phenotype**. *The Plant Cell* 2007, **19**(11):3403-3417.
5. Liu B, Brubaker C, Mergeai G, Cronn R, Wendel J: **Polyploid formation in cotton is not accompanied by rapid genomic changes**. *Genome* 2001, **44**(3):321-330.
6. Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge R, Comai L: **Genomewide nonadditive gene regulation in Arabidopsis allotetraploids**. *Genetics* 2006, **172**(1):507-517.
7. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S: **The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes**. *Nature communications* 2014, **5**:3930.
8. Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM: **Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes**. *Genetics* 1997, **147**(3):1381-1387.
9. Grover C, Gallagher J, Szadkowski E, Yoo M, Flagel L, Wendel J: **Homoeolog expression bias and expression level dominance in allopolyploids**. *New Phytologist* 2012, **196**(4):966-971.
10. Pont C, Murat F, Guizard S, Flores R, Foucrier S, Bidet Y, Quraishi UM, Alaux M, Doležel J, Fahima T: **Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo-and neoduplicated subgenomes**. *The Plant Journal* 2013, **76**(6):1030-1044.
11. Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L: **Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication**. *Nature Genetics* 2017, **49**(4):579.

12. Fan K, Li F, Chen J, Li Z, Lin W, Cai S, Liu J, Lin W: **Asymmetric evolution and expansion of the NAC transcription factor in polyploidized cotton.** *Frontiers in plant science* 2018, **9**:47.
13. Song Z, Wu X, Gao Y, Cui X, Jiao F, Chen X, Li Y: **Genome-wide analysis of the HAK potassium transporter gene family reveals asymmetrical evolution in tobacco (*Nicotiana tabacum*).** *Genome* 2019, **62**(4):267-278.
14. Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng J, Wang X: **Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing.** *Nature biotechnology* 2011, **29**(8):762.
15. Song K, Lu P, Tang K, Osborn TC: **Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution.** *Proceedings of the National Academy of Sciences* 1995, **92**(17):7719-7723.
16. Schmutzer T, Samans B, Dyrszka E, Ulpinnis C, Weise S, Stengel D, Colmsee C, Lespinasse D, Micic Z, Abel S: **Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*.** *Scientific data* 2015, **2**:150072.
17. Nagaharu U: **Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization.** *Jpn J Bot* 1935, **7**(7):389-452.
18. Ziolkowski PA, Kaczmarek M, Babula D, Sadowski J: **Genome evolution in Arabidopsis/Brassica: conservation and divergence of ancient rearranged segments and their breakpoints.** *The Plant Journal* 2006, **47**(1):63-74.
19. Gaeta RT, Pires JC: **Homoeologous recombination in allopolyploids: the polyploid ratchet.** *New Phytologist* 2010, **186**(1):18-28.
20. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B: **Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome.** *science* 2014, **345**(6199):950-953.
21. Kråling K: **Utilization of genetic variability of resynthesized rapeseed.** *Plant Breeding* 1987, **99**(3):209-217.
22. Becker H, Engqvist G, Karlsson B: **Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers.** *Theoretical and Applied Genetics* 1995, **91**(1):62-67.
23. Gómez-Campo C: **Biology of Brassica coenospecies**, vol. 4: Elsevier; 1999.
24. Zou J, Hu D, Mason AS, Shen X, Wang X, Wang N, Grandke F, Wang M, Chang S, Snowdon RJ: **Genetic changes in a novel breeding population of *Brassica napus* synthesized from hundreds of crosses between *B. rapa* and *B. carinata*.** *Plant biotechnology journal* 2018, **16**(2):507-519.
25. Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IA: **The pangenome of an agronomically important crop plant *Brassica oleracea*.** *Nature communications* 2016, **7**:13390.
26. Bayer PE, Hurgobin B, Golicz AA, Chan CKK, Yuan Y, Lee H, Renton M, Meng J, Li R, Long Y: **Assembly and comparison of two closely related *Brassica napus* genomes.** *Plant biotechnology journal* 2017, **15**(12):1602-1610.
27. Hurgobin B, Golicz AA, Bayer PE, Chan CKK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IA: **Homoeologous exchange is a major cause of gene presence/absence variation**

- in the amphidiploid *Brassica napus*. *Plant biotechnology journal* 2018, **16**(7):1265-1274.
28. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F: **The genome of the mesopolyploid crop species *Brassica rapa***. *Nature genetics* 2011, **43**(10):1035.
 29. Jin S-W, Rahim MA, Afrin KS, Park J-I, Kang J-G, Nou I-S: **Transcriptome profiling of two contrasting ornamental cabbage (*Brassica oleracea* var. *acephala*) lines provides insights into purple and white inner leaf pigmentation**. *BMC genomics* 2018, **19**(1):797.
 30. Wu X, Zhu ZB, Chen JH, Huang YF, Liu ZL, Zou JW, Chen YH, Su NN, Cui J: **Transcriptome analysis revealed pivotal transporters involved in the reduction of cadmium accumulation in pak choi (*Brassica chinensis* L.) by exogenous hydrogen-rich water**. *Chemosphere* 2019, **216**:684-697.
 31. Kost B, Lemichez E, Spielhofer P, Hong Y, Toliaas K, Carpenter C, Chua N-H: **Rac homologues and compartmentalized phosphatidylinositol 4, 5-bisphosphate act in a common pathway to regulate polar pollen tube growth**. *The Journal of cell biology* 1999, **145**(2):317-330.
 32. Liu K, Li L, Luan S: **An essential function of phosphatidylinositol phosphates in activation of plant shaker-type K⁺ channels**. *The Plant Journal* 2005, **42**(3):433-443.
 33. Kusano H, Testerink C, Vermeer JE, Tsuge T, Shimada H, Oka A, Munnik T, Aoyama T: **The Arabidopsis phosphatidylinositol phosphate 5-kinase PIP5K3 is a key regulator of root hair tip growth**. *The Plant Cell* 2008, **20**(2):367-380.
 34. Georges F, Das S, Ray H, Bock C, Nokhrina K, Kolla VA, Keller W: **Over-expression of *Brassica napus* phosphatidylinositol-phospholipase C2 in canola induces significant changes in gene expression and phytohormone distribution patterns, enhances drought tolerance and promotes early flowering and maturation**. *Plant, cell & environment* 2009, **32**(12):1664-1681.
 35. Chen J, Song Y, Zhang H, Zhang D: **Genome-wide analysis of gene expression in response to drought stress in *Populus simonii***. *Plant molecular biology reporter* 2013, **31**(4):946-962.
 36. Wang Z-X, Zhou X-Z, Meng H-M, Liu Y-J, Zhou Q, Huang B: **Comparative transcriptomic analysis of the heat stress response in the filamentous fungus *Metarhizium anisopliae* using RNA-Seq**. *Applied microbiology and biotechnology* 2014, **98**(12):5589-5597.
 37. Sloan KE, Warda AS, Sharma S, Entian K-D, Lafontaine DL, Bohnsack MT: **Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function**. *RNA biology* 2017, **14**(9):1138-1152.
 38. Wei H, Wang L, Hassan M, Xie B: **Succession of the functional microbial communities and the metabolic functions in maize straw composting process**. *Bioresource technology* 2018, **256**:333-341.
 39. Bus A, Körber N, Snowdon RJ, Stich B: **Patterns of molecular variation in a species-wide germplasm set of *Brassica napus***. *Theoretical and Applied Genetics* 2011, **123**(8):1413-1423.
 40. Delourme R, Falentin C, Fomeju BF, Boillot M, Lassalle G, André I, Duarte J, Gauthier V, Lucante N, Marty A: **High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L.** *BMC genomics* 2013, **14**(1):120.
 41. Coram TE, Brown-Guedira G, Chen X: **Using transcriptomics to understand the wheat genome**. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 2008, **3**(83):1-9.

42. Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, Herms DA: **Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer (*Agrilus planipennis*)**. *PLoS one* 2010, **5**(10):e13708.
43. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpour S, Danielsson A, Edlund K: **Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics**. *Molecular & Cellular Proteomics* 2014, **13**(2):397-406.
44. Córdova-Campos O, Adame-Álvarez RM, Acosta-Gallegos JA, Heil M: **Domestication affected the basal and induced disease resistance in common bean (*Phaseolus vulgaris*)**. *European Journal of Plant Pathology* 2012, **134**(2):367-379.
45. Ma Y, Liu M, Stiller J, Liu C: **A pan-transcriptome analysis shows that disease resistance genes have undergone more selection pressure during barley domestication**. *BMC genomics* 2019, **20**(1):12.
46. Tian D, Traw M, Chen J, Kreitman M, Bergelson J: **Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana***. *Nature* 2003, **423**(6935):74.
47. Parker MA: **The pleiotropy theory for polymorphism of disease resistance genes in plants**. *Evolution* 1990, **44**(7):1872-1875.
48. Horns F, Hood ME: **The evolution of disease resistance and tolerance in spatially structured populations**. *Ecology and evolution* 2012, **2**(7):1705-1711.
49. Preece C, Livarda A, Christin PA, Wallace M, Martin G, Charles M, Jones G, Rees M, Osborne CP: **How did the domestication of Fertile Crescent grain crops increase their yields?** *Functional ecology* 2017, **31**(2):387-397.
50. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K: **Insights into the maize pan-genome and pan-transcriptome**. *The Plant Cell* 2014, **26**(1):121-135.
51. Jin M, Liu H, He C, Fu J, Xiao Y, Wang Y, Xie W, Wang G, Yan J: **Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation**. *Scientific reports* 2016, **6**:18936.
52. Sun C, Hu Z, Zheng T, Lu K, Zhao Y, Wang W, Shi J, Wang C, Lu J, Zhang D: **RPAN: rice pan-genome browser for ~ 3000 rice genomes**. *Nucleic acids research* 2016, **45**(2):597-605.
53. Andrews S: **FastQC: a quality control tool for high throughput sequence data**. In.: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.
54. Cox MP, Peterson DA, Biggs PJ: **SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data**. *BMC bioinformatics* 2010, **11**(1):485.
55. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nature methods* 2012, **9**(4):357.
56. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y: **SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler**. *Gigascience* 2012, **1**(1):18.
57. Schmieder R, Edwards R: **Fast identification and removal of sequence contamination from genomic and metagenomic datasets**. *PloS one* 2011, **6**(3):e17288.
58. Stanke M, Morgenstern B: **AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints**. *Nucleic acids research* 2005, **33**(suppl_2):W465-W467.
59. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: **Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data**. *Nature*

biotechnology 2011, **29**(7):644.

60. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22**(13):1658-1659.
61. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications.** *BMC bioinformatics* 2009, **10**(1):421.
62. Powell JJ, Fitzgerald TL, Stiller J, Berkman PJ, Gardiner DM, Manners JM, Henry RJ, Kazan K: **The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias.** *Plant biotechnology journal* 2017, **15**(4):533-543.
63. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L: **KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.** *Nucleic acids research* 2011, **39**(suppl_2):W316-W322.
64. Ai C, Kong L: **CGPS: A machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways.** *Journal of genetics and genomics* 2018, **45**(9):489-504.

Supplementary Information

Additional file 1:

Table S1: Genome sequences from the 49 genotypes of *B. napus* used in this study.

Additional file 2:

Table S2: GO enrichment analysis of genes from the natural and artificial genotypes of *B. napus*.

Table S3: The 42 GO terms for genes shared between the natural and artificial genotypes of *B. napus*.

Additional file 3:

Table S4: KEGG enrichment analysis of genes from natural and artificial genotypes of *B. napus*.

Additional file 4:

Table S5: GO terms for genes shared between the natural and artificial genotypes of *B. napus* and derived from either of its diploid donors

Table S6: GO terms for genes found in the artificial genotypes of *B. napus* and derived from either of its diploid donors.

Table S7: GO enrichment analysis of lost and retained genes derived from its two diploid donors.

Additional file 5:

Table S8: Genomic sequences of the 49 *B. napus* genotypes used in this study

Table S9: Transcriptomic sequences used in this study.

Figures

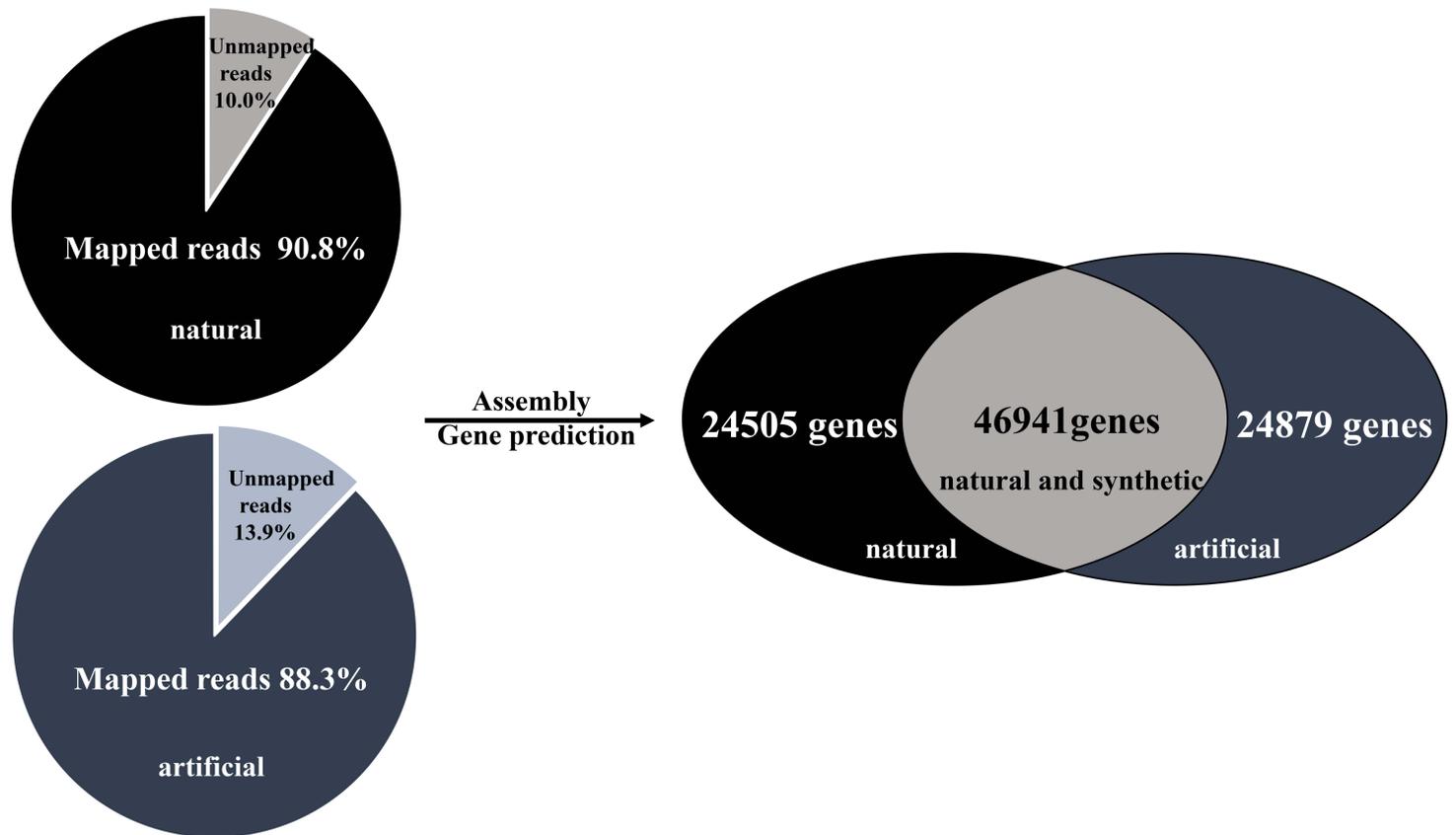


Figure 1

Sequence reads and predicted genes from 30 natural and 19 artificial genotypes of *B. napus*.

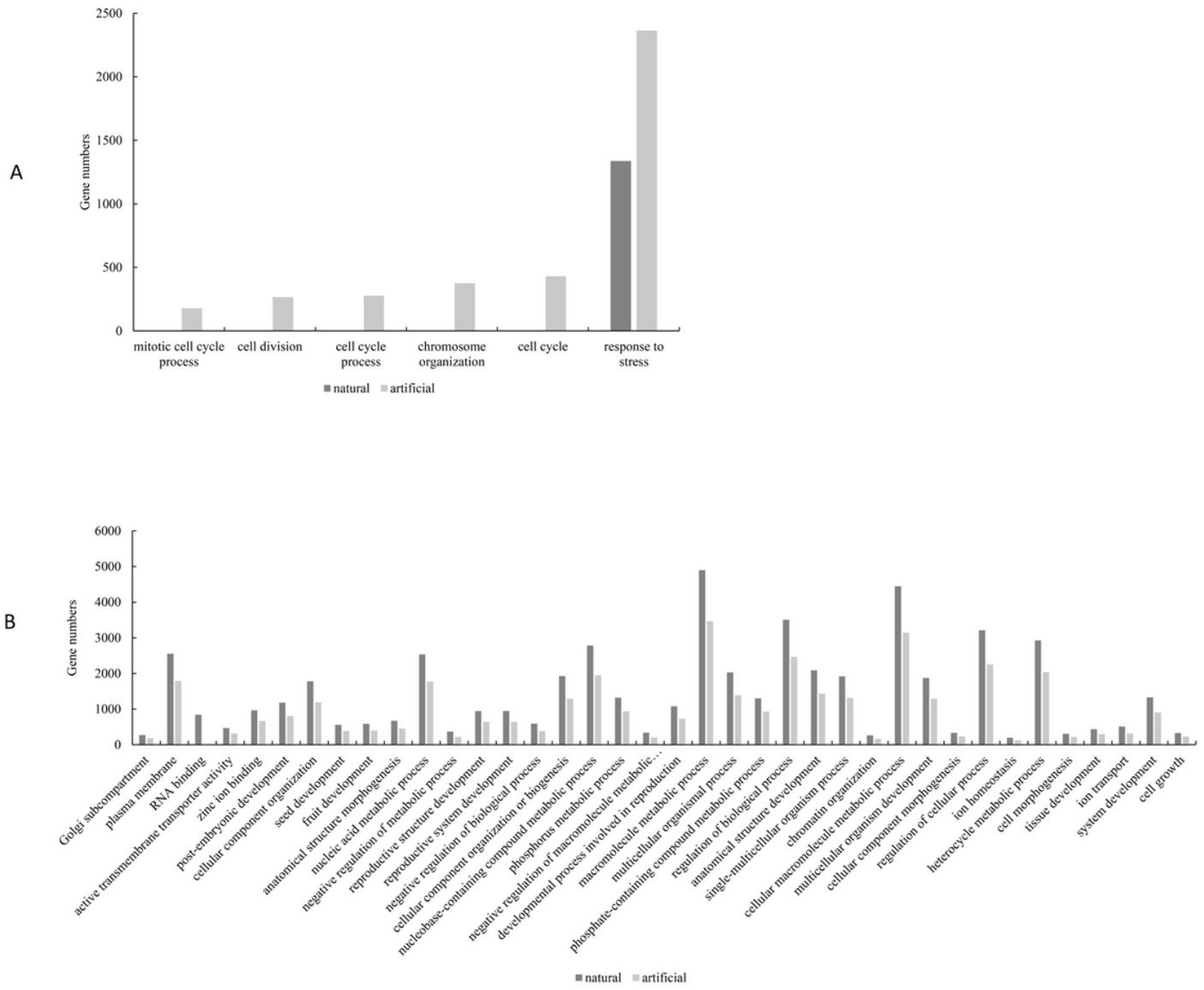


Figure 2

Differences in GO terms of genes between artificial and natural genotypes of *B. napus*. 2A, GO terms significantly enriched in only the artificial genotypes; 2B, GO terms significantly enriched in the natural genotypes. All GO terms were significantly different between input genes and backgrounds.

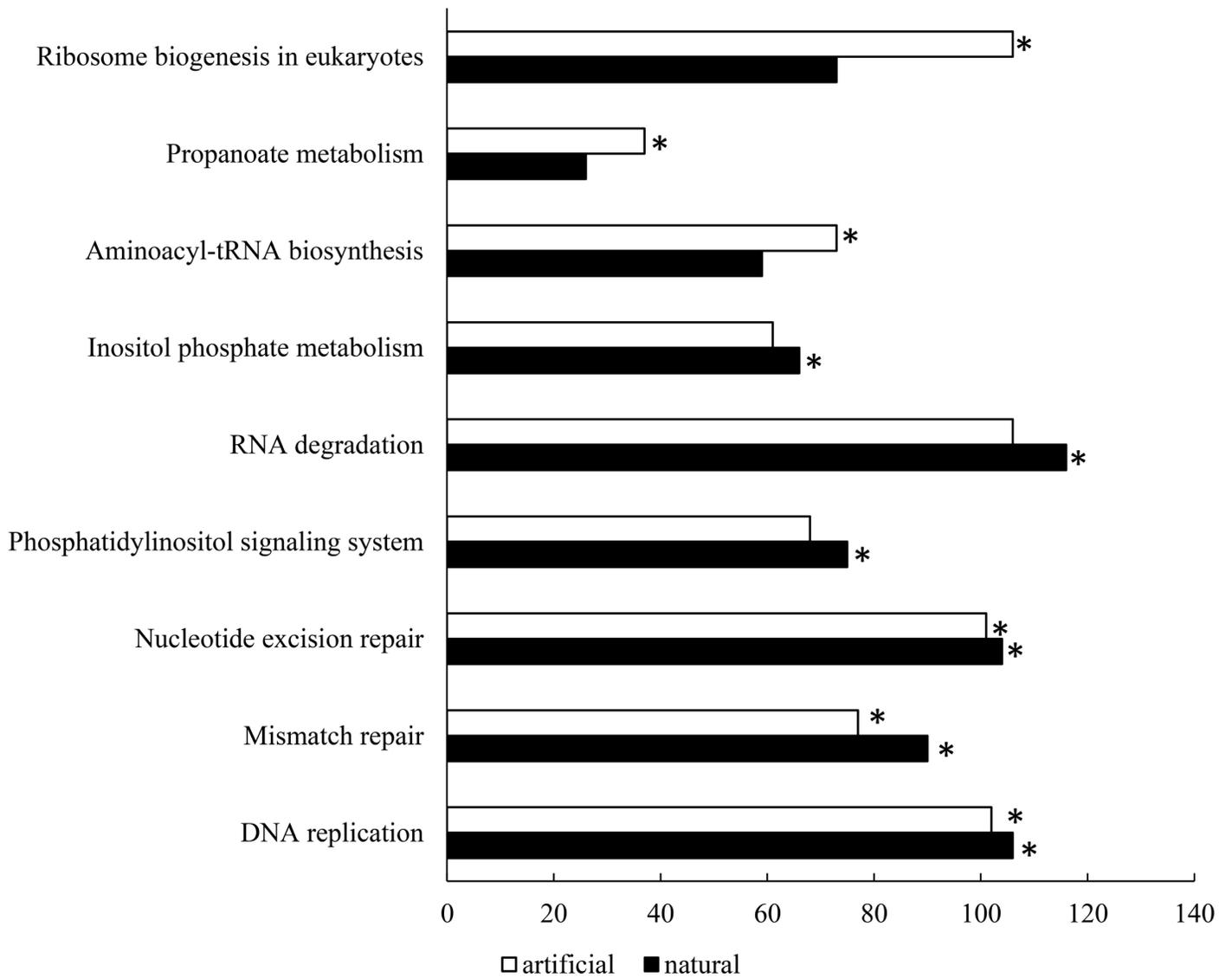


Figure 3

Enriched KEGG pathways which are significantly different between the natural and artificial genotypes of *B. napus*. The numbers of KEGG pathways are given on the X-axis, and those marked with '*' indicate significant difference between input genes and backgrounds.

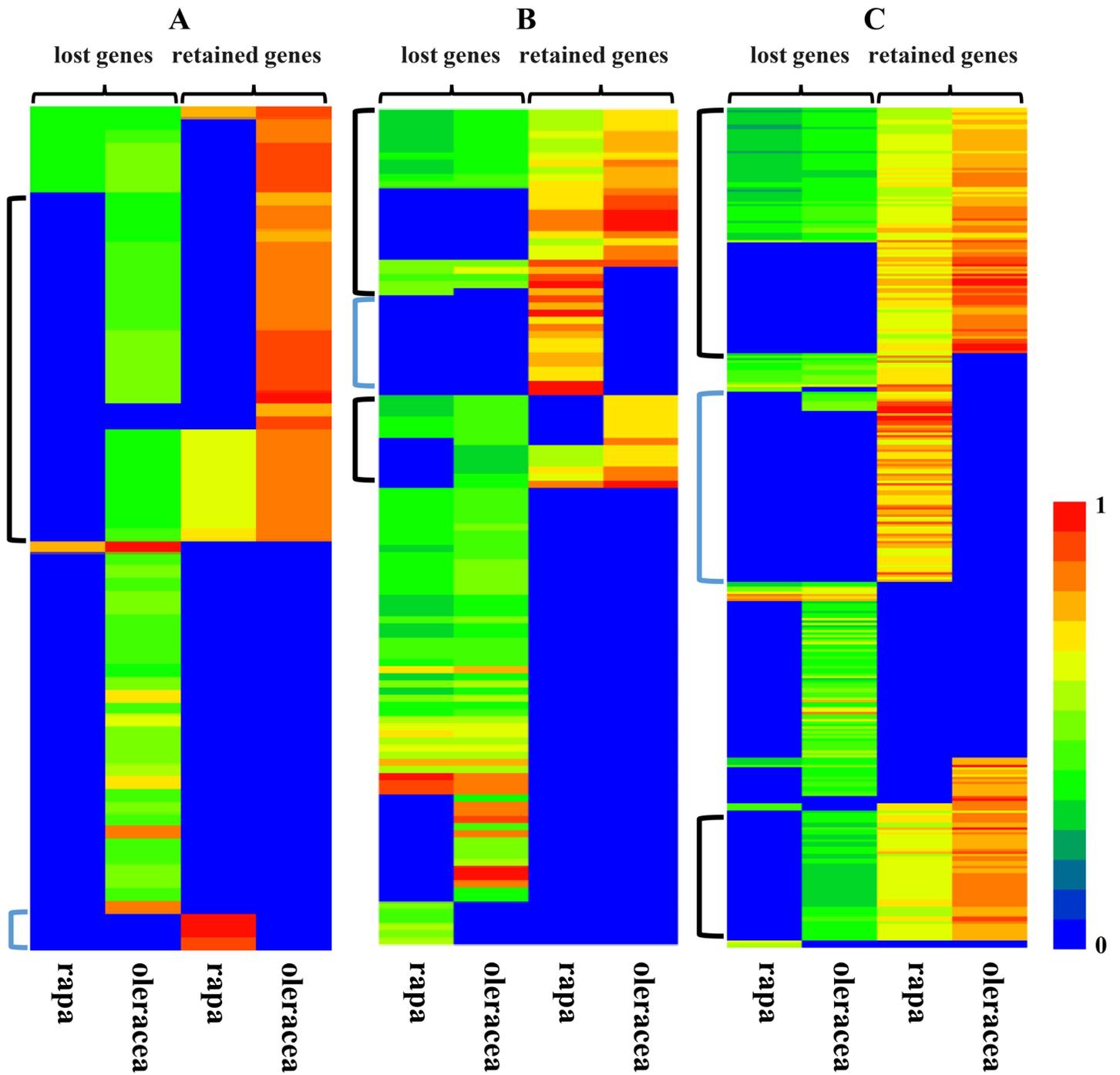


Figure 4

Heatmaps for allotetraploid genes derived from either of the diploid donor species. ‘Lost genes’ are those present in only the artificial genotypes and ‘retained genes’ are those found in both the natural and artificial allotetraploid genotypes. The three heatmaps represent the three different GO categories, ‘A’ for cellular component; ‘B’ for molecular function, and ‘C’ for biological process. Rows represent different GO terms, and different colors represent the ratio of input gene numbers and background gene number for each of the GO terms. The blue and black brackets at the left side represent GO terms for genes from *B. rapa* and *B. oleracea*, respectively. All GO terms showed significant differences between input genes and backgrounds.

Supplementary Files

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

- [Additionalfile3.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile1.xlsx](#)
- [Additionalfile2.xlsx](#)