
Article:
Ferreira de Carvalho, Julie, Lucas, Jeremy, Deniot, Gwenaelle et al. (17 more authors) (2019) Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in Brassica. The Plant journal. pp. 434-447. ISSN 1365-313X

https://doi.org/10.1111/tpj.14228

DR JULIE FERREIRA DE CARVALHO (Orcid ID : 0000-0001-6200-3344)

Article type : Original Article

Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in Brassica

Julie Ferreira de Carvalho[1*], Jérémy Lucas[1], Gwenaëlle Deniot[1], Cyril Falentin[1], Olivier Filangi[1], Marie Gilet[1], Fabrice Legeai[1], Maryse Lode[1], Jérôme Morice[1], Gwenn Trotoux[1], Jean-Marc Aury[2], Valérie Barbe[2], Jean Keller[3], Rod Snowdon[4], Zhesi He[5], France Denoeud[2,6,7], Patrick Wincker[2,6,7], Ian Bancroft[5], Anne-Marie Chèvre[1], Mathieu Rousseau-Gueutin[1*]

[1] IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, BP35327, 35653 Le Rheu Cedex, France

[2] Commissariat à l'Energie Atomique, Genoscope, Institut de biologie François-Jacob, BP5706, 91057 Evry, France

[3] UMR CNRS 6553 ECOBIO, OSUR, Université de Rennes 1, 35042 Rennes, France

[4] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

[5] Department of Biology, University of York, Heslington, York, YO10 5DD, United Kingdom

[6] UMR CNRS 8030, CP5706 Evry, France

[7] Université d'Evry-Val-d'Essonne, Université Paris-Saclay, 91000 Evry, France

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.14228

Corresponding authors:

Dr. Julie Ferreira de Carvalho; Address: IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, BP35327, 35653 Le Rheu Cedex, France
Phone : +33(0)2 232 359 97 ; E-mail : julie.ferreira-de-carvalho@inra.fr
Dr. Mathieu Rousseau-Gueutin ; Address: IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, BP35327, 35653 Le Rheu Cedex, France
Phone : +33(0)2 232 351 31 ; E-mail : mathieu.rousseau-gueutin@inra.fr

Running head: Evolution of cytonuclear interaction in polyploids

Keywords: interspecific hybridization, allopolyploidy, intergenomic conflicts, chloroplast, genome inheritance, duplicated genes, *Brassica rapa*, *Brassica oleracea*, *Brassica napus.*

SUMMARY

Several plastid macromolecular protein complexes are encoded by both nuclear and plastid genes. Therefore, cytonuclear interactions are held in place to prevent genomic conflicts that may lead to incompatibilities. Allopolyploidy resulting from hybridization and genome doubling of two divergent species, can disrupt these fine-tuned interactions, as newly formed allopolyploid species confront biparental nuclear chromosomes with uniparentally inherited plastid genome. To avoid any deleterious effects of unequal genome inheritance, preferential transcription of the plastid donor over the other one has been hypothesized to occur in allopolyploids. We used *Brassica* as a model to study the effects of paleopolyploidy in diploid parental species, as well as the effects of recent and ancient allopolyploidy in *Brassica napus* on genes implicated in plastid protein complexes. We first identified redundant nuclear copies involved in those complexes. Compared to cytosolic protein complexes and to genome-wide retention rates, genes involved in plastid protein complexes show a higher retention of genes in duplicated and triplicated copies. Those redundant copies are functional and undergoing strong purifying selection. We then compared transcription patterns and sequences of those redundant gene copies between resynthesized allopolyploids and their diploid parents. The neo-polyploids showed no biased subgenome expression or maternal homogenization via gene conversion despite presence of some non-synonymous substitutions between plastid genomes of parental progenitors. Instead, subgenome dominance was observed regardless of the maternal

progenitor. Our results provide new insights on the evolution of plastid protein complexes which could be tested and generalized in other allopolyploid species.

INTRODUCTION

Genome mergers, through horizontal transfer and interspecific hybridization, have played a central role in the origin, evolution and diversification of eukaryotes, providing some of the astonishing diversity among living organisms. One of the most important examples of this phenomenon was the formation of plants through the endosymbiotic engulfment of a cyanobacterium by a eukaryotic cell (Margulis, 1970; Deusch *et al.*, 2008). This cyanobacterial genome gradually evolved into the plastid genome, allowing plants to autonomously photosynthesize. Since this endosymbiotic event that occurred more than one billion years ago, plastid DNA and genes were transferred to the nucleus, creating genes with similar or new functions as well as novel regulatory elements (Martin *et al.*, 2002; Timmis *et al.*, 2004; Kleine *et al.*, 2009; Rousseau-Gueutin *et al.*, 2011; Rousseau-Gueutin *et al.*, 2013). Indeed, it has been shown that about 18% of *Arabidopsis thaliana* nuclear genes were in fact of cyanobacterial origin and result from the deluge of plastid DNA that was constantly transferred to the nucleus since this endosymbiotic event (Martin *et al.*, 2002). To date, the plant plastid genome is highly reduced compared to its cyanobacterial ancestor (Kleine *et al.*, 2009), with more than 90% of plant plastid proteins being nuclear-encoded. However, a number of essential macromolecular protein complexes, including Photosystem I, Photosystem II, Cytochrome b6f, Rubisco and ATP synthase, RNA polymerase PEP, Acetyl-CoA carboxylase, Caseinolytic protease and ribosomes are still encoded by interacting nuclear and plastid genes (Allen *et al.*, 2011; Lerbs-Mache, 2011; Rockenbach *et al.*, 2016; Weng *et al.*, 2016). Specifically, excess of non-synonymous mutations was found in genes encoding non-photosynthetic plastid protein complexes compared to cytosolic complexes (Sloan *et al.*, 2014). Therefore, as plastid-nuclear co-evolution seems to be important in those complexes, any genomic conflicts arising could potentially lead to cytonuclear incompatibilities.

Intergenomic conflict could arise in allopolyploids, which result from interspecific hybridization followed by genome doubling. This intriguing evolutionary process often entails profound structural alterations (including exchanges between homoeologous chromosomes, gene conversion events or activation of transposable elements) and functional modifications (at the epigenetic, transcriptional or translational levels) immediately after polyploid formation (Albertin *et al.*, 2006; Gaeta *et al.*, 2007; Parisod *et al.*, 2009; Tate *et al.*, 2009; Marmagne *et al.*,

2010; Szadkowski *et al.*, 2010; Xiong *et al.*, 2011; Lashermes *et al.*, 2014; Wendel *et al.*, 2016; Rousseau-Gueutin *et al.*, 2017). Additionally, duplicated genes may take different paths, including gene loss, pseudogenisation, neofunctionalization or subfunctionalization, where expression of the duplicated copies is partitioned between tissues or developmental stages (Grover *et al.*, 2012). A newly formed allopolyploid individual carries maternal and paternal nuclear chromosomes from both progenitor species. However, because maternal organelle inheritance is most common in higher plants only the maternal plastid genome is present. Plastid-nuclear incompatibilities and altered cytonuclear interactions have been reported in interspecific hybrids and allopolyploid species (reviewed in Sharbrough *et al.*, 2017). These alterations may result from several mechanisms including deletions in plastid intergenic sequences (Greiner *et al.*, 2008) and presence of non-synonymous substitutions putatively modifying regulatory networks and protein structure (Greiner *et al.*, 2011). The latter case has been investigated in allopolyploids. Authors have hypothesized occurrence of preferential transcription of maternal nuclear genes over the paternally inherited copies to avoid plastid-nuclear conflict. To date, this has been studied only in a few allopolyploid plant systems and exclusively on the Rubisco-encoded genes (reviewed in Sharbrough *et al.*, 2017). In cotton, for example, non-reciprocal homoeologous recombination events were found to occur between sub-genomes of natural allopolyploids (Gong *et al.*, 2012). Such events can potentially lead to homogenization via gene conversions occurring during allopolyploid evolution. Interestingly, maternal Rubisco homoeologs were preferentially expressed in both wild and cultivated allopolyploid cottons, as well as in resynthesized F_1 hybrids, compared to paternal homoeologs (Gong *et al.*, 2012). Differential evolution of maternal and paternal Rubisco genes was also reported between homoeologous genomes of *Arabidopsis suecica*, *Arachis hypogaea* and *Nicotiana tabacum* allopolyploids (Gong *et al.*, 2014). However, this scenario was not observed in the young allopolyploid *Tragopogon miscellus* and in resynthesized allotetraploid *Oryza* despite non-synonymous substitutions observed in the Rubisco plastid-encoded parental sequences of these species (Sehrish *et al.*, 2015; Wang *et al.*, 2017).

Here we describe the structural and functional (biased gene retention, gene conversion, selection, global and homoeo-specific gene expression) consequences of both ancient and recent allopolyploidy on nuclear genes implicated in different plastid protein complexes that are encoded both in the nucleus and plastid genomes (Allen *et al.*, 2011; Lerbs-Mache, 2011; Rockenbach *et al.*, 2016; Weng *et al.*, 2016). We used the *Brassica* polyploid complex, which represents an excellent system to study the impact of recurrent polyploidization events as well as immediate and long-term consequences of genome merger and doubling. The genomes of *B. rapa* (2n=20; AA) and *B. oleracea* (2n=18; CC) (Wang *et al.*, 2011; Parkin *et al.*, 2014) diverged

about 4 million years ago. Both arose from a triplication event, which occurred 22 million years ago (Beilstein *et al.*, 2010) and are thus considered as paleohexaploids. Thereafter, their triplicated genomes rapidly underwent extensive gene deletion through a fractionation process (Murat *et al.*, 2015). It has been shown in *B. rapa* that only 11.2% of genes retained their triplicated copies, while 36.4% have lost one copy and 52.3 % have lost two copies (Wang *et al.*, 2011). In other words, the majority of the genes are tending towards a return to single-copy status (Wang *et al.*, 2011), and overall gene content in *Brassica* diploid genomes has so far been reduced by about half. However, different cultigroups within *Brassica* species do not display the same number of gene copies, and may explain the astonishing and extensive number of morphotypes existing within each species (Cheng *et al.*, 2014; Cheng *et al.*, 2016; Golicz *et al.*, 2016). The allotetraploid *B. napus* (2n=38; AACC), which arose recently (less than 7500 years ago; Chalhoub *et al.*, 2014) via hybridization and genome doubling of *B. rapa* (2n=20; AA) as maternal genome donor and *B. oleracea* (2n=18; CC) as paternal genome donor (Allender and King, 2010), exemplifies the additional level of genome complexity that exists in *Brassica* allopolyploids. In both natural varieties and particularly in *de novo* resynthesized *B. napus*, rapid and extensive changes have been reported at the structural and gene expression levels (Gaeta *et al.*, 2007; Szadkowski *et al.*, 2010; Rousseau-Gueutin *et al.*, 2017; Hurgobin *et al.*, 2017). We used this valuable model system to address several questions regarding effects of allopolyploid speciation on genes implicated in cytonuclear interactions: (1) Are these genes preferentially retained in multicopies in paleopolyploids and allopolyploids? (2) Can we detect signatures of selection and/or functionalization in these duplicate copies? (3) Is the maternal subgenome preferentially expressed over the paternally inherited subgenome in allopolyploids? (4) Are paternal copies more prone to gene conversion than their corresponding maternal copies? And, finally, (5) is global gene expression in allopolyploids subject to transgression to compensate higher plastid number?

RESULTS

Nucleotide divergence in plastid genomes of diploid species

Plastid genomes of *B. rapa* var. Chiifu and *B. oleracea* var. TO1000 were retrieved from NCBI and aligned with *B. rapa* var. 'C1.3' and var. 'Z1' as well as *B. oleracea* var. 'RC34', var. 'HDEM' to study intra- and interspecific nucleotide conservation for the 59 plastid genes involved in cytonuclear interactions (Table S1). Comparison of their coding gene sequences revealed the presence of 11 non-synonymous substitutions (NSS) in six plastid genes between *B. rapa* and *B.*

oleracea (Table S2) whereas no NSS were identified within species. Specifically, NSS were found in the plastid genes *PetA* and *AtpA,* which each interacts at the macromolecular protein level with the nuclear genes *PetC* and *PetM* (within the Cytochrome b6f complex) and *AtpC*, *AtpD* and *AtpG* (within the ATP synthase complex), respectively. The other non-synonymous mutations were found in *rpoB* and *rpoC2* from the RNA polymerase PEP protein complex, *rpl22* encoding Ribosomal proteins and *accD* part of the Acetyl-CoA carboxylase complex. For comparison, other polyploidy complexes were analyzed for the same 59 plastid genes and NSS were identified between the diploid parents of the allotetraploid *Arabidopsis suecica* (Table S2, 45 NSS for 20 plastid genes) and *Nicotiana tabacum* (Table S2, 68 NSS for 21 plastid genes), as well as *Gossypium hirsutum* (Table S2, eight NSS for eight plastid genes).

Nuclear gene copy retention

Because of the triplication event in the ancestor of *Brassica* species (Figure S1), we first investigated the number of retained copies present in *B. rapa* var. Chiifu and *B. oleracea* var. TO1000 reference genomes for each of the 110 nuclear genes implicated in plastid protein complexes (Table S1). To that purpose, we performed phylogenetic analyses to determine gene copies retained after the triplication event rather from more small-scale duplications, as well as to determine the orthologous relationships between *B. rapa* and *B. oleracea* gene copies. We then, retained only duplicated copies originating from the triplication event specific of Brassicaceae. Overall, we found that 22 genes (20%) forming plastid complexes showed conservation of all three copies, 42 genes (38.2%) had lost one copy and 46 (41.8%) have returned to a single copy in *B. rapa* (Table 1). In *B. oleracea*, similar percentages were reported (Table 1). These proportions were significantly different from those observed genome-wide (Table 1, Chi^2, $P<1.10^{-12}$). Particularly, genes exhibiting retention of three and two copies were more frequent whereas genes in single copy were less abundant in our dataset compared to genome-wide retention rates (Chalhoub *et al.*, 2014). In order to determine if the conservation of genes in duplicate or triplicate was a general trend for genes involved in protein complexes, we determined the number of copies retained for genes involved in cytosolic complexes (Aryal *et al.*, 2014). We were able to assign retention rates for 86 and 84 genes implicated in cytosolic complexes for *B. rapa* and *B. oleracea*, respectively (Table S3). For *B. rapa*, 2.33% of genes were conserved in triplicate, 34.88% in duplicate and 62.79% in single copy (Table 1). Similar rates were observed in *B. oleracea* (Table 1). Thus, these retention rates observed in cytosolic complexes resembled those observed for overall gene data, but are in contradiction with our data on plastid protein complexes (Table 1).

Substitution rates

To examine the selective pressure acting on these redundant genes, previously transcribed Maximum Likelihood (ML) phylogenetic trees were used to investigate 197 gene copies representing 102 nuclear genes encoding plastid macromolecular proteins. All gene copies were found undergoing strong purifying selection with on average, dN/dS ratios around 0.1459 (Table S4, MIN=0.011; MAX=0.4724).

Gene conversion

To test for homogenization of homoeologous copies in our dataset, alignments were performed on all copies separately and include orthologous sequences of *B. rapa* and *B. oleracea* cultivars (i.e Chiifu, Z1, C1.3 and TO1000, HDEM, RC34) and homoeologs of *B. napus* var. Darmor reference genome (Chalhoub *et al.*, 2014). Only matrices without gaps were kept for further analyses. Thus, using 152 redundant copy alignments representing 97 genes, a total of 1,300 homoeo-SNPs were identified between *B. rapa* and *B. oleracea* (or 1.34 homoeo-SNP per 100bp). Among those homoeo-SNPs, 12.5% (or 161) were classified as converted SNPs (Figure 1, Table S5). Converted SNPs were then categorized into maternally homogenized SNP (in *B. napus* var. Darmor AA instead of AC) or paternally homogenized SNP (in *B. napus* var. Darmor: CC instead of AC). Following this classification, we identified 95 and 66 maternally and paternally converted SNPs, respectively (Table S5). At the SNP level, maternal homogenization was 1.44-fold higher than paternal homogenization. However, when counting the number of gene conversion events (spanning one or several SNPs), we detected 35 and 46 occurrences of maternal and paternal homogenization, respectively (Table S5). In that case, the ratio completely switch to 1.31-fold higher paternal homogenization.

Evolution of gene transcription in resynthesized polyploids and *B. napus* cultivars

i. Transcriptional dynamics within species. We first assessed the transcriptional status of the different copies previously identified in both diploid *Brassica* species used as progenitors in subsequent interspecific crosses. A large proportion (90%) was found expressed in both *B. rapa* and *B. oleracea* (Table S6). Additionally, we also investigated the differential expression of those copies within each species. We observed on average for *B. rapa* 47.7% and 56.8% of copies differentially expressed, for the one in duplicate and triplicate, respectively. For *B. oleracea*, those percentages were increased, with 66.67% and 80% of, respectively, duplicated and

triplicated copies differentially expressed (Table S7). As we could determine that those redundant copies all exhibited the same functional domains (from pfam analyses), we hypothesized that they also retained the same function within each species. Thus, to determine global level of transcription for each gene, we summed the expression of all redundant copies to explore their transcriptional dynamic following allopolyploidization.

ii. Transcriptional dynamic of each gene duplicated (i.e. orthologs) within and between progenitors. The two respective cultigroups of *B. oleracea* and *B. rapa* were experimentally crossed in two combinations to obtain two independent resynthesized polyploid crosses namely 'EMZ' and 'RCC'. First, the different cultivars of *B. oleracea* and *B. rapa* were compared within species to detect intraspecific variability and distinguish our two crosses. Between *B. oleracea* genotypes, the cultivar 'RC34' was preferentially expressed for 17 out of 110 genes (15.5%) genes whereas only 2 (1.8%) were found preferentially expressedin 'HDEM' (Table S8). Between *B. rapa* genotypes, the difference was less prominent with 6 and 10 genes preferentially expressed in 'C1.3' and 'Z1', respectively (Table S8). These contrasts were also visible when comparing differential expression patterns between species for both crosses. Progenitors of the 'EMZ' cross were compared and global gene expression was found different between Z1 and HDEM in 59.9% of the cases (Figure 2). Among those genes, 33.6% were preferentially expressed in *B. rapa* var. 'Z1' and thus, leaving 25.5% preferentially expressed in *B.* oleracea var. 'HDEM' (Figure 2). For the other cross, global gene expression between progenitors was found different in 65.5% of the genes. However, in contrast with the 'EMZ' cross, 'RCC' progenitors exhibited mostly preferentially expressed in *B. oleracea* var. 'RC34' with 38.2% of the genes impacted. In *B. rapa* var. 'C1.3', 27.3% of genes were found preferentially expressed (Figure 2).

iii. Transcriptional dynamic of A and C subgenomes (orthologous copies) following allopolyploidy. These progenitors were then used in independent crosses to form resynthesized polyploids to study the dynamic of transcription after interspecific hybridization and genome doubling in recent polyploids. To test our hypotheses of preferential utilization of maternal transcripts, we first considered subgenome dynamic during allopolyploid formation. More precisely, we identified subgenome expression profiles in the F1, S0 and resynthesized polyploids and compared it with each progenitor genome independently (Figure 3). In the interspecific hybrid 'RCC', the Ar subgenome revealed 5.5% of gene copies differentially expressed when compared to the Ar genome of *B. rapa* cv 'C1.3' (including 4.5 and 0.9% of genes down- and up-regulated in the F1, respectively) (Figure 3, Table S9). This ratio drastically increased after genome doubling with 17.3% of genes differentially expressed in S0 and 19.3% after a few meioses in resynthesized polyploids. The C subgenome revealed a different pattern. In the F1, 14.5% of

genes were classified differentially expressed between the C subgenome and *B. oleracea* cv 'RC34' whereas no differences could be detected in the S0 (Figure 3, Table S9). In the polyploids, 13.8% of genes were found differentially expressed (including 1.5 and 12.4% of genes showing up- and down-regulation, respectively).

In the EMZ cross, we found in the A subgenome of the interspecific hybrid 24.5% of the genes differentially expressed by comparison with *B. rapa* cv 'Z1' (including 12.7 and 11.8% of genes down- and up-regulated in the F1, respectively) (Figure 3, Table S10). This ratio drastically decreased after genome doubling and the first meiosis in resynthesized polyploids. On average only 8.6% of genes were differentially expressed between progenitor *B. rapa* and Ar subgenome of the polyploids (Figure 3, Table S10). The Co subgenome revealed a different pattern. In the F1, only 1.8% of genes were differentially expressed between the C subgenome and *B. oleracea* cv 'HDEM' (Figure 3, Table S10). In the polyploids, this ratio was increased to, on average, 11.3% of genes differentially expressed between the C subgenome and the maternal progenitor. Among those genes differentially expressed, 3.8 and 7.5% were found in the polyploids up- and down-regulation, respectively.

iv. Transcriptional dynamic of A vs C subgenomes (homoeologous copies) following allopolyploidy. Additionally, we also compared A vs C subgenome expression within a resynthesized polyploid to detect biased expressed genes (Figure 2, Figure 3, Table S11). In F1 'RCC' cross, we found half of the genes (50.0%) showing biased subgenome expression. The same trend was observed after genome doubling with half of the genes showing biased subgenome expression (Figure 2, Figure 3, Table S11). However, in the resynthesized 'RCC' polyploids an increased number of genes exhibiting biased subgenome expression was identified with on average 56.5% of genes showing genome dominance, including a larger portion (32.7%) towards the A genome (paternal progenitor). We also investigated subgenome bias expression in the 'EMZ' cross. In resynthesized polyploids, most of the genes were found biased towards the A subgenome (paternal progenitor) with on average 38.5% (Figure 3, Table S11).

To examine gene expression evolution in natural varieties of *B. napus,* we used the progenitor species from our crosses for comparisons as actual parental genotypes of *B. napus* varieties remain unknown. We observed in most cases, similar expression levels between the An subgenome of *B. napus* and the Ar genomes of 'Z1' (80.2%) and 'C1.3' (82.7%). Among genes differentially expressed, 10.9 and 11.8% are up-regulated in the polyploids compared to the progenitors 'C1.3' and 'Z1', respectively (Figure 3, Table S11). Regarding the Cn subgenome, genes showing no differential expression were encountered in 84.7 and 89.9.7% of the cases, compared to *B. oleracea* 'HDEM' and 'RC34', respectively. Among the differential patterns, the *B.*

oleracea progenitor exhibited mostly up-regulated genes with 9.4 and 7.8% in 'HDEM' and 'RC34', respectively (Figure 3, Table S11). Finally, we investigated biased subgenome dominance in *B. napus* varieties and found that 45.5% of genes exhibited expression biased, with predominantly An subgenome dominance (28.9% of genes).

v. Transcriptional dynamic of global gene expression following allopolyploidy. To identify possible patterns of transgressivity, we also investigated global gene expression dynamics during allopolyploidization. In both crosses, mostly additivity of parental gene expression levels was observed in the F1 hybrid, double haploid S0 and resynthesized polyploids. In F1 interspecific hybrids, only four genes were found up-regulated in F1 'RCC'. These up-regulated genes were not identified in subsequent S0 'RCC' (Figure 3, Table S12). In resynthesized polyploids, on average 2.6% and 2.9% of genes in EMZ and RCC respectively, were found differentially expressed (Figure 3, Table S12). Among those, a slightly higher number of genes were found down-regulated in polyploids, comparatively to up-regulated genes (ratios 1.3 and 1.7 in EMZ and RCC respectively). In *B. napus* varieties, similar patterns were observed, with on average additive global gene expression identified when compared to EMZ and RCC parental progenitors independently (Figure S2, Table S12). More specifically, between 5 and 21 genes were found differentially expressed in *B. napus* cultivars compared to EMZ and RCC respectively.

DISCUSSION

Nuclear genes encoding plastid macromolecular complexes retained in duplicated copies

Brassica species have been subjected to a WGT event that took place 22.5 mya (Beilstein *et al.*, 2010). Up to three and six duplicated copies can be found in the diploid and allotetraploid *Brassica* species, respectively. By performing BlastP analyses between *Brassica* species and subsequent phylogenetic analyses, we were able to unambiguously identify the evolutionary history of each nuclear gene involved in plastid protein complex and their orthologous relationships. More specifically, we observed that these genes were significantly kept in duplicates or triplicates. Interestingly, a similar preferential retention pattern of duplicated photosystem genes has been reported in *Glycine max* (Coate *et al.*, 2011). However, in wider studies encompassing various flowering plants, genes encoding organelle-related functions (including Photosynthesis) were described as preferentially returning to single gene status, except for *B. rapa* that was described as an outlier with comparatively higher number of duplicates than other species (de Smet *et al.*, 2013; Li *et al.*, 2016).

In Brassica, we tested if this retention of redundant genes may be due to their involvement in protein complexes and compared our dataset including nuclear genes involved in plastid protein complexes with genes encoding cytosolic protein complexes (Aryal *et al.*, 2014). Surprisingly, we found that genes implicated in these latter complexes have a strong tendency towards a return to single copy status and that the proportions of copy retention were identical to those found genome-wide (Wang *et al.*, 2011). We conclude that nuclear genes implicated in plastid protein complexes, in strong contrast to other plant protein complexes, tend towards long-term retention of duplicate and triplicate genes after ancient allopolyploidization. As mentioned previously by Coate and coauthors (2011), contrasted retention rates could be the consequence of both dosage sensitivity as well as protein topological structure. Indeed, high retention of multicopy genes supports the gene balance hypothesis, where genes whose products are involved in macromolecular complexes or networks are more sensitive to imbalance regarding concentration of protein subunits (Papp *et al.*, 2003). Consequently, any change in the stoichiometry of individual subunits would lead to impaired function and have deleterious consequences for the individual, promoting retention of multicopies for specific genes and biological functions (Birchler and Veitia, 2007; Birchler and Veitia, 2010; Aury *et al.*, 2006).

Conserved functions and contrasted evolutionary rates between copies duplicated by WGT

All redundant copies remained transcribed and shared the same functional domains. Globally, all nuclear gene copies involved in plastid protein complex were found experiencing strong purifying selection. However, some of these copies were found undergoing slight but significant increased mutations rates while still under purifying selection. This phenomenon was particularly evidenced in genes whose copies were retained in duplicates and triplicates. In addition, when looking at the differential transcription between those redundant copies in diploid species, contrasted expression levels were observed. These findings support the fractionation hypothesis previously observed in *Brassica rapa* (Wang *et al.*, 2011). Genome-wide, gene copies present in the Least Fractionated (LF) subgenome exhibit higher expression levels and accumulate fewer non-synonymous mutations than its paralogs in the MF1 and MF2 subgenomes. To explain this pattern, Freeling et al. (2012) proposed a model in which the most highly expressed gene copy produces more protein and thus contributes more to the required stoichiometric balance. Over a long evolutionary time, this model predicts that mutations impacting function would be more strongly selected in the more highly expressed copy than the

other down-regulated copy, explaining the different evolutionary rates observed between paralogous copies.

Stable interactions between plastid and nuclear genomes in hybrids and polyploids

One aspect of allopolyploid speciation that has been particularly overlooked is the consequences of cytonuclear stoichiometric disruption in hybrids and polyploids. Recent studies have addressed this issue on the *Rubisco* macromolecular complex and showed preferential utilization of maternal transcript to enhance protein conformation between nuclear and plastid-encoded genes. Most of the hypotheses are based upon the *a priori* that plastid genomes have strongly diverged in the diploid species. Therefore, the impact of cytonuclear disruption would depend on the time since divergence and plastid evolution rate in diploids.

Brassica rapa and *B. oleracea* that have diverged less than 4 mya exhibited 11 NSS in plastid coding regions. Number of NSS was actually similar to what was found between *Gossypium* parental species, which diverged 5 to 10 mya (Li *et al.*, 2015), but still lower compared to other species that have diverged between 5-7mya (*i.e Arabidopsis thaliana* and *A. arenosa* (Wang *et al.*, 2006), *Nicotiana sylvestris* and *N. tomentosiformis*; (Särkinen *et al.*, 2013). We tested the hypothesis of preferential maternal transcript utilization through gene conversion and transcriptional bias, previously reported in other models for the Rubisco complex (Gong *et al.*, 2012; Gong *et al.*, 2014). Therefore, since just a few non-synonymous mutations in other systems were sufficient to lead to strong functional changes, we tested in *B. napus* var. Darmor for gene conversion events and signs of homogenization towards the maternal homoeolog, as well as preferential transcription of maternally-derived transcripts over the paternal subgenomes in resynthesized *B. napus*. Regarding gene conversion, number of converted SNP from maternal-to-paternal homoelogs (i.e leading to maternal homogenization) exhibited higher ratio of 1.44 compared to SNPs converted from paternal-to-maternal homoeologs. This result is consistent with the genome-wide data (Chalhoub *et al.*, 2014) where maternal homogenization via SNP conversion of homoeolog copies were found to be 1.5 time more abundant. Therefore, our results did not support the hypothesis that maternal homogenization would be favored in genes implicated in plastid-nuclear macromolecular complexes.

Then, the fate of homoeolog transcripts in resynthesized allopolyploids was assessed and tested specifically for preferential utilization of maternal transcripts over paternally inherited ones. Differential expression towards the A subgenome (compared to C) was predominantly found in resynthesized polyploids and in natural varieties of *B. napus*. Although resynthesized *B. napus*

polyploids have *B. oleracea* as maternal genome donor and natural varieties have *B. rapa* as the maternal progenitor, the pattern of A subgenome dominance is conserved. Therefore, we can reject our first hypothesis that maternal transcripts would be preferentially transcribed over the paternal subgenome for nuclear genes involved in plastid protein complexes. Instead, we found evidence for the establishment of the A genome dominance in natural and resynthesized *B. napus* regardless of the maternal progenitor. This mechanism was previously proposed to occur genome-wide in *B. napus* polyploids (Wu *et al.*, 2018) but also in *Triticum* (Akhunova *et al.*, 2010) and *Gossypium* (Yoo *et al.*, 2012). It has been recently proposed that differences in Transposable Element density and preferential targeting by 24nt-siRNAs may explain this genome dominance (Cheng *et al.*, 2016; Li *et al.*, 2014; Shen *et al.*, 2015). Additionally, epigenetic differences were shown to play a major role in this phenomenon, as the lowly-expressed copies from duplicated genes were shown to present a higher methylation at CHH, CHG and CpG sites and have less open chromatin (Springer *et al.*, 2016; Edger *et al.*, 2017). Interestingly, this preferential expression of one subgenome over the other one occurs instantly following the hybridization of divergent genomes and significantly increases over generations (Edger *et al.*, 2017).

Thus, it is still unclear why nuclear genes encoding plastid proteins tend to retain redundant copies with the same function. One hypothesis is that a higher transcript level conferred by multiple gene copies may increase photosynthetic capacity, a frequently described feature of polyploid plants compared to their diploid progenitors (Warner and Edwards, 1993; Ilut *et al.*, 2012; Coate *et al.*, 2013). To test this assumption we compared global gene expression between diploid parents and resynthesized allopolyploids, as well as natural varieties of *B. napus*. However, we did not detect significant differences in global gene expression but mainly additivity in hybrids, resynthesized and natural varieties of *B. napus*. The few genes with up- or down-regulated global expression may possible be triggered by the presence of non-reciprocal homoeologous exchanges as previously observed in these resynthesized *B. napus* polyploids (Rousseau-Gueutin *et al.*, 2017).

Conclusion

Due to uniparental inheritance of the plastid genome, maternal preferential utilization of nuclear genes involved in plastid protein complexes has been proposed to occur in allopolyploid species. However, in *B. napus* no evidence of biased transcription or homogenization via gene conversion towards the maternal genome has been found. Instead, subgenome dominance was observed regardless of the maternal progenitor. Our conclusions differ from previous studies in allopolyploids whose parents diverged around the same time, but similar with other young and

resynthesized polyploids. Our results highlight the importance to take advantage of progenitor transcription levels to be able to distinguish the effect of plastid maternal inheritance on nuclear gene sequences from genome dominance, a common feature in polyploids.

EXPERIMENTAL PROCEDURES

Plant material

Two different resynthesized *B. napus* populations, hereafter designated 'RCC' and 'EMZ', were obtained by performing two crosses between different accessions of *B. oleracea* (female) and *B. rapa* (male), as reported previously (Rousseau-Gueutin *et al.*, 2017). The first cross was performed between a doubled haploid line ('RC34') of *B. oleracea* var. *alboglabra* (CC, 2n=2x=18) and a plant 'C1.3' belonging to the *B. rapa* var. *rapifera* (AA, 2n=2x=20) fodder variety 'Chicon'. The second cross was generated between a homozygous doubled haploid line ('HDEM') of *B. oleracea* var. *botrytis italica* (CC, 2n=2x=18) and the doubled haploid line ('Z1') of *B. rapa* var. *trilocularis* (AA, 2n=2x=20). The resulting amphiploid hybrids 'RCC-F1' and 'EMZ-F1' (AC, 2n=2x=19) were somatically doubled using colchicine (Chevre *et al.*, 1989), leading to S0 plants. By selfing the RCC-S0 and the EMZ-S0 plants, via hand pollination of floral buds before anthesis, we produced RCC-S1 and EMZ-S1 progenies. Thereafter, 110 S1 plants were grown under the same cage (one cage for 'RCC' and one for 'EMZ') in the presence of flies to ensure open-pollination and to facilitate outcrossing between plants. The following year, one seed per 110 G2 plants (from 'RCC' and 'EMZ', respectively) was grown in a cage (with flies) to obtain the G3 progenies. For each cross, five G3 individuals per population were retained for subsequent analyses and are referred in the manuscript as 'FG54, FG55, FG57, FG71, FG72' and 'FG78, FG84, FG90, FG92, FG108', for the RCC and EMZ crosses, respectively. In addition to these resynthesized *B. napus* lines, four natural *B. napus* varieties were also used: '*Bronowski*', '*Oro*', '*Petranova*' and '*Tanto*'.

Chloroplast genome assemblies and annotation

High molecular weight genomic DNA of the parental lines of the resynthesized *B. napus* populations (*B. rapa*: 'Z1' and 'C1.3'; *B. oleracea*: 'HDEM' and 'RC34') was obtained using the procedure detailed in Chalhoub et al. (2014). For subsequent plastome reconstruction, libraries were sequenced on an Illumina HiSeq2500 instrument (Illumina, San Diego, CA, USA) using 250 base-length read chemistry in a paired-end mode by the Genoscope (Evry, France). After the

Illumina sequencing, an in-house quality control process was applied to the reads that passed the Illumina quality filters. The first step discards low-quality nucleotides (Q < 20) from both ends of the reads. Next, Illumina sequencing adapters and primer sequences were removed from the reads. Then, reads shorter than 30 nucleotides after trimming were discarded. These trimming and removal steps were achieved using in-house-designed software based on the FastX package (http://www.genoscope.cns.fr/fastxtend/). Plastid genome assemblies were performed using the pipeline previously published in Keller et al. (Keller *et al.*, 2017). Briefly, *de novo* assemblies were built using "The organelle assembler" software (Coissac *et al.*, 2016). Each draft plastid genome was then verified and corrected by mapping the Illumina reads against each reconstructed genome using Bowtie 2 v2.0 (Langmead and Salzberg, 2012) (Parameters: Score-min G 52 8, Local alignment). Finally, annotation was performed using DOGMA (Wyman *et al.*, 2004) and verified visually after comparison with previously published *Brassica* plastid genomes. The four newly assembled *Brassica* plastid genomes were deposited in Genbank under the accession numbers MG717286, MG717287, MG717288 and MG717289. These different genomes were aligned using Geneious 10.1.3 (Kearse *et al.*, 2012). Coding sequences were then translated, extracted and aligned using Geneious 10.1.3 (Kearse *et al.*, 2012) to identify non-synonymous substitutions, either between cultivars of the same species (i.e between 'HDEM' and 'RC34' or between 'Z1' and 'C1.3') or between the parental species (i.e. *B. oleracea* 'HDEM' and 'RC34' against *B. rapa* 'Z1' and 'C1.3'). Non-synonymous substitutions were retrieved using a custom Perl script, which compares amino acid sequences and returns positions of each non-synonymous substitution within a gene. The script was also used to assess levels of non-synonymous mutations between parental plastid genomes of other polyploid systems (in *Arabidopsis suecica*, *Gossypium hirsutum* and *Nicotiana tabacum*). For that purpose, chloroplast genomes of parental species were retrieved from GenBank (NC_029334 and NC_000932 for *Arabidopsis thaliana* and *A. arenosa* ; JF317353 and NC_016668 for *Gossypium arboretum* and *G. raimondii* and NC_007500 and NC_007602 for *Nicotiana sylvestris* and *N. tomentosiformis*). Finally, non-synonymous nucleotide positions were checked for not being in a predicted RNA editing sites using prep-cp (Mower, 2009) online tool (http://prep.unl.edu/).

Evolutionary dynamics of nuclear genes involved in chloroplast protein complexes

In plants, some plastid protein complexes are encoded by both plastid and nuclear genes. We focused on 110 nuclear genes involved in these plastid protein complexes previously identified in *Arabidopsis* (Allen *et al.*, 2011; Lerbs-Mache, 2011; Rockenbach *et al.*, 2016; Weng *et al.*, 2016). The protein sequences of these 110 genes were retrieved from *Arabidopsis* genome in

TAIR 10 (https://www.arabidopsis.org/). Then, these different *Arabidopsis* amino-acid gene sequences were blasted (reciprocal BlastP) against *B. rapa* v 1.5 (Wang *et al.*, 2011), *B. oleracea* v.2.1.24 (Parkin *et al.*, 2014) and *B. napus* v. 1.5 (Chalhoub *et al.*, 2014) reference genomes. As a consequence, for each *Arabidopsis* gene, up to 3 redundant copies (genes triplicated by ancient whole duplication events (Glover *et al.*, 2016) can be found in the genomes of *B. rapa* and *B. oleracea*. Moreover, up to 6 copies can be found in the genome of *B. napus*, as only few genes were lost since its formation less than 7500 years ago (Chalhoub *et al.*, 2014). In our case, phylogenetic analyses for each gene were essential to comprehend the evolutionary history of each gene and differentiate between paralogous copies deriving from either polyploidy or non-polyploidy (i.e. tandem duplication, transposition or recombination) events. Additionally, this analysis enables assessment of which gene copies are orthologous between the respective *Brassica* species and which *B. napus* copies are homoeologs.

As *Arabidopsis* and *Brassica* species diverged about 27 million years ago (Murat *et al.*, 2015), only genes presenting at least 70% identity (CIP: Cumulative Identity Percentage) across 70% of the gene length of the query (CALP: Cumulative Alignment Length Percentage) were subsequently considered as putative homologs (Salse *et al.*, 2009). Each *Arabidopsis* gene was aligned with the different homologous gene copies identified in *B. rapa*, *B. oleracea* and *B. napus* using Geneious v 10.1.3 (Kearse *et al.*, 2012). The different data matrices produced were adjusted manually and subjected to molecular phylogeny using Maximum Parsimony (MP) and Maximum Likelihood (ML). Trees are available in Zenodo under DOI 10.5281/zenodo.1172694. MP and ML analyses were performed using MEGA 7 (Kumar *et al.*, 2016) with 1000 bootstrap replicates and using *Arabidopsis* as outgroup. Finally, amino acid sequences of all duplicated copies from the same gene function were checked to include the same functional domains using PFAM and the script pfam_scan.pl (Finn *et al.*, 2016). The different gene copies and their orthologous relationships in *B. rapa*, *B. oleracea* and *B. napus* are presented in Supplementary Table 1.

Also, in order to compare retention rates of nuclear genes involved in plastid protein complexes (both nuclear and plastid encoded) with those of other multi-protein complexes, we used gene encoding cytosolic proteins. For that purpose, we retrieved 270 *Arabidopsis* genes implicated in cytosolic protein complexes from Aryal et al. (Aryal *et al.*, 2014) Using identifiers, we recovered the amino-acid sequences from the TAIR website (https://www.arabidopsis.org/). Subsequently, 121 and 118 homologous sequences were identified using BLASTP in *B. rapa* and *B. oleracea* reference genomes, respectively. As performed for the plastid protein complexes, we only extracted BLAST results presenting a CIP/CALP of at least 70%. Our results were tested

statistically for their independence of frequencies with what was found genome-wide (Chalhoub *et al.*, 2014) using Chi^2.

Selective pressure acting on nuclear genes involved in plastid protein complexes

Selective pressure acting on the different paralogous copies for each nuclear gene was estimated using the codeml branch model (Yang and Nielsen, 2000) in the PAML 4 package (Yang, 2007). For each ML tree a file was generated in a newick format. Global dN/dS of the whole phylogenetic tree (Model M0) was determined and compared to dN/dS of specific branches (Model Branch Free). Specific branches were determined visually, nodes corresponding to one ancestral paralogous copies (including one copy of *B. rapa*, one copy of *B. oleracea* and two copies of *B. napus* if all copies were retained) were characterized with a star in the newick file to be recognized by the Branch Free Model. Then, significant dN/dS ratios across the tree were performed with a LRT against Model M0. Modified scripts are available in Github (github.com/juliefc/Brassica).

Identification of putative gene conversion events

To detect the presence of putative gene conversion in *B. napus*, we first reconstructed the sequence of each gene copy using DNAseq data obtained from the parents of the two resynthesized populations: *B. oleracea* ('RC34' and 'HDEM') and *B. rapa* ('C1.3' and 'Z1'). To that purpose, we first mapped DNAseq data obtained from *B. oleracea* against the 110 *B. oleracea* gene copies (including the different copies) identified from the *B. oleracea* TO1000 reference genome. Similarly, we mapped *B. rapa* DNAseq data against the 110 gene copies identified from the *B. rapa* Chiifu reference genome using Bowtie 2 (Langmead and Salzberg, 2012) (Parameters: Score-minG 52 8, Local alignment). The obtained *bam* files were converted into a *pileup* file using Samtools (Li *et al.*, 2009) in order to identify SNPs existing between the reference genome sequence and our genotypes. We then used a custom python script to reconstruct the sequence of each gene. Subsequently, we used Geneious to align the orthologous copies obtained from our *B. rapa* and *B. oleracea* genotypes with the reference genome sequences (*B. rapa* var. Chiifu, *B. oleracea* var. TO1000, and *B. napus* var. Darmor-*bzh*). Following the procedure detailed in Salmon *et al.* (2010) and using a custom-made script (available in github.com/juliefc/Brassica), we first detected the SNPs differentiating between the *B. rapa* and *B. oleracea* accessions (i.e. species-specific SNPs). A SNP was called if all *B. rapa* genotypes presented the same nucleotide whereas all *B. oleracea* genotypes showed the

alternative nucleotide. Following allopolyploidy, these SNPs are referred to as homoeo-SNP and should be retained in the genome of *B. napus*. In the case of gene conversion, homoeo-SNPs present in the polyploid genome may be homogenized. Therefore, only the nucleotide from one parent is present in both homoeologs. After identifying these homoeo-SNPs, we searched for the presence of converted SNPs and gene conversion events (involving several consecutive converted homoeo-SNPs). Finally, we determined which copy (*B. rapa* vs. *B. oleracea*) was used as a template to assess the direction of the conversion event.

RNAseq analyses

Overall, 21 samples were grown in triplicates in a phytotron for a period of one month with photoperiod 14/10 hours, thermoperiod 21/16°C and humidity 70/85%. These included the diploid parents of the two resynthesized *B. napus* lines, two F1 amphiploid hybrids ('RCC-F1' and 'EMZ-F1'), one resynthesized *B. napus* S0 (only 'RCC-S0' as the 'EMZ-S0' plant died), five resynthesized *B. napus* individuals in G3 (for both 'RCC' and 'EMZ' populations), and four *B. napus* varieties ('*Bronowski*', '*Oro*', '*Petranova*' and '*Tanto*). For the resynthesized *B. napus* individuals, triplicates from each genotype were obtained from cuttings, allowing avoiding genetic variations between biological replicates. For all samples, leaf tissues were collected the same day and directly ground in liquid nitrogen using a mortar and pestle. RNA was then extracted using the manufacturer's instructions for the Omega Biotek EZNA Plant RNA Kit and the libraries were prepared using the Illumina mRNAseq kit (RS-100-0801, Illumina Inc.) as described by Bancroft et al. (Bancroft *et al.*, 2011). In total, 50 libraries were prepared, with three libraries per genotype except for the natural *B. napus* varieties, for which only one library was created. Per library, more than 50Gb of raw paired-end reads of 100bp (insert size 400-500bp) were generated and deposited in NCBI SRA (BioProject ID PRJNA429568). Quality checks, adapters trimming and filtering were performed as described in Higgins *et al.* (2012). Cleaned RNAseq reads were mapped against concatenated *B. rapa* and *B. oleracea* genomes (Wang *et al.*, 2011; Parkin *et al.*, 2014) using STAR. For resynthesized *B. napus* polyploids, the concatenated genome of *B. rapa* and *B. oleracea* was preferred over the *B. napus* var. Darmor-*bzh* genome (Chalhoub *et al.*, 2014), since a resynthesized *B. napus* polyploid genome is genetically closer to the concatenation of both *B. rapa* and *B. oleracea* genomes than to the genome of a natural *B. napus* variety. RNAseq data obtained from *B. napus* varieties were also mapped against this concatenated genome to allow suitable comparisons of gene expression between resynthesized and natural *B. napus*. As the sequence similarity is very high between the *B. rapa* and *B. oleracea* genomes (about 98%) (Chalhoub *et al.*, 2014), very stringent

mapping parameters were used to ensure correct mapping of read data (only one mismatch accepted for the whole read length). In addition, only reads that mapped uniquely on the concatenated genome were considered, to avoid taking into account reads that mapped on both homoeologous genes (i.e. identical regions between *B. rapa* and *B. oleracea* coding sequences). The mapping parameters used for STAR are as follows: alignment end to end, no multimapping, one mismatch (Dobin *et al.*, 2013). Number of reads for each gene copy were counted using featureCount (Liao *et al.*, 2014). This software used mapping (*bam*) files coming from *B. rapa*, *B. oleracea* and concatenation of both genomes to determine the number of reads with default parameters. This way, we were able to distinguish subgenome contributions to global gene expression in the amphiploid hybrids and resynthesized allopolyploids. Differences in global gene expression between parents, F1, S0 and polyploids, as well as biases in subgenome expression, were tested using the package edgeR (Robinson and Oshlack, 2010). Briefly, raw read counts were first normalized using reads per kilobase per million mapped reads (RPKM), which takes into account library size for each sample as well as transcript size. Then, transcripts with low read count were trimmed from the analysis. Only transcripts with at least 1 RPKM in more than 10 samples (i.e. libraries) were retained. A second normalization step, Trimmed Mean of M-valued (TMM) (Robinson and Oshlack, 2010) allowed minimization of variance between samples and helped resolve statistical comparisons. Subsequently, data were tested with a Global Linear Model (McCarthy et al. 2012) with a likelihood ratio test, followed by a multiple testing correction using Benjamini-Hochberg (Benjamini and Hochberg, 1995). A transcript (or a gene from a subgenome) was considered differentially expressed (up- or down-regulated) if its corrected p-value was less than 0.05.

ACKNOWLEDGMENTS

This work was made possible by financial support from ERA-CAPS13.017_Evo-Genapus attributed to A.-M.C., R.S. and I.B. (UK Biotechnology and Biological Sciences Research Council BB/L027844/1) as well as an Agreenskills+ Fellowship to J.F.C. This work was additionally supported by the France Génomique National infrastructure, funded as part of the « Investissements d'Avenir » program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). We acknowledge the BrACySol BRC (INRA Ploudaniel, France) that provided us with the seeds of parental lines. We are grateful to H. Boulain for her help with the Branch specific analyses. We would also like to thank all the technical staff of the greenhouse for management of the plant material (especially L. Charlon, P. Rolland, J.-P. Constantin, J.-M. Lucas and F. Letertre). The authors have no conflict of interest to declare.

Accession numbers

The RNAseq datasets have been deposited to NCBI SRA under BioProject ID PRJNA429568. Assembled and annotated *Brassica sp* plastid genomes have also been deposited to NCBI under GenBank accession numbers MG717286, MG717287, MG717288, MG717289. Phylogenetic analyses are available in Zenodo under DOI 10.5281/zenodo.1172694 and custom made scripts in Github.com/juliefc/Brassica.

Short legends for Supporting Information

Figure S1. Evolutionary history of the genus Brassica since the divergence with Arabidopsis sp.

Figure S2. Differential gene expression in natural varieties of *Brassica napus* compared with A) RCC progenitors and B) EMZ progenitors.

Table S1. List of duplicated gene copies with their paleologous and orthologous relationships in Brassica species as well as the orthologs in *Arabidopsis thaliana* for all nuclear genes implicated in plastid protein complexes.

Table S2. Identification of non-synonymous mutations in plastid genes implicated in plastid protein complexes between progenitor species of four allopolyploid model species.

Table S3. Lists of nuclear genes implicated in cytosolic complexes. *Arabidopsis thaliana* identifiers were extracted from Aryal *et al*. 2014.

Table S4. For each duplicated gene copy, ratios of non-synonymous to synonymous rates are presented (omega).

Table S5. Number of homoeoSNPs, SNP conversion and gene conversion event for the different duplicated gene copies in *Brassica napus* var. Darmor.

Table S6. Levels of expression (RPKM) for all duplicated gene copies implicated in plastid protein complexes and identified in *Brassica oleracea* and *B. rapa* varieties.

Table S7. Significant gene expression comparisons between copies duplicated by ancient triplication event within *Brassica rapa* and *B. oleracea* varieties.

Table S8. Significant gene expression comparisons between varieties of *Brassica rapa* and *B. oleracea*.

Table S9. In the RCC cross, significant gene expression comparisons of each subgenome in a diploid context vs polyploid context.

Table S10. In the EMZ cross, significant gene expression comparisons of each subgenome in a diploid context vs polyploid context.

Table S11. Significant gene expression comparisons between subgenomes (A vs C) in diploids, hybrids and polyploids.

Table S12. Significant global gene expression comparisons between parents and hybrids or polyploids.

REFERENCES

Akhunova, A.R., Matniyazov, R.T., Liang, H. and Akhunov, E.D. (2010) Homoeolog-specific transcriptional bias in allopolyploid wheat. *BMC Genomics*, **11**, 505.

Albertin, W., Balliau, T., Brabant, P., Chevre, A.-M., Eber, F., Malosse, C. and Thiellement, H. (2006) Numerous and rapid nonstochastic modifications of gene products in newly synthesized Brassica napus allotetraploids. *GENETICS*, **173**, 1101–1113.

Allen, J.F., Paula, W.B.M. de, Puthiyaveetil, S. and Nield, J. (2011) A structural phylogenetic map for chloroplast photosynthesis. *Trends Plant Sci.*, **16**, 645–655.

Allender, C.J. and King, G.J. (2010) Origins of the amphiploid species Brassica napusL. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol.*, **10**, 54.

Aryal, U.K., Xiong, Y., McBride, Z., Kihara, D., Xie, J., Hall, M.C. and Szymanski, D.B. (2014) A proteomic strategy for global analysis of plant protein complexes. *Plant Cell*, **26**, 3867–3882.

Aury, J.-M., Jaillon, O., Duret, L., et al. (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**, 171.

Bancroft, I., Morgan, C., Fraser, F., et al. (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nat. Biotechnol.*, **29**, 762–766.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R. and Mathews, S. (2010) Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proc. Natl. Acad. Sci.*, **107**, 18724–18728.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat. Soc B*, **57**, 289–300.

Birchler, J.A. and Veitia, R.A. (2007) The Gene Balance Hypothesis: From Classical Genetics to Modern Genomics. *Plant Cell*, **19**, 395–402.

Birchler, J.A. and Veitia, R.A. (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.*, **186**, 54–62.

Chalhoub, B., Denoeud, F., Liu, S., et al. (2014) Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science*, **345**, 950–953.

Cheng, F., Wu, J., Cai, C., et al. (2016) Genome resequencing and comparative variome analysis in a *Brassica rapa* and *Brassica oleracea* collection. *Sci. Data*, **3**, 160119.

Cheng, F., Wu, J. and Wang, X. (2014) Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.*, **1**, 14024.

Chevre, A.M., Eber, F., Thomas, G. and Baron, F. (1989) Cytological studies of tetraploid kale (Brassica oleracea L. ssp acephala) obtained from diploid lines after colchicine treatment. Agronomie EDP Sciences, **9**(5), 521-525.

Coate, J.E., Powell, A.F., Owens, T.G. and Doyle, J.J. (2013) Transgressive physiological and transcriptomic responses to light stress in allopolyploid Glycine dolichocarpa (Leguminosae). *Heredity*, **110**, 160–170.

Coate, J.E., Schlueter, J.A., Whaley, A.M. and Doyle, J.J. (2011) Comparative evolution of photosynthetic genes in response to polyploid and nonpolyploid duplication. *Plant Physiol.*, **155**, 2081–2095.

Coissac, E., Hollingsworth, P.M., Lavergne, S. and Taberlet, P. (2016) From barcodes to genomes: extending the concept of DNA barcoding. *Mol. Ecol.*, **25**, 1423–1428.

Deusch, O., Landan, G., Roettger, M., Gruenheit, N., Kowallik, K.V., Allen, J.F., Martin, W. and Dagan, T. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol. Biol. Evol.*, **25**, 748–761.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.*, **29**, 15–21.

Edger, P.P., Smith, R., McKain, M.R., et al. (2017) Subgenome Dominance in an Interspecific Hybrid, Synthetic Allopolyploid, and a 140-Year-Old Naturally Established Neo-Allopolyploid Monkeyflower. *Plant Cell*, **29**, 2150–2167.

Finn, R.D., Coggill, P., Eberhardt, R.Y., et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.

Freeling, M., Woodhouse, M.R., Subramaniam, S., Turco, G., Lisch, D. and Schnable, J.C. (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.*, **15**, 131–139.

Gaeta, R.T., Pires, J.C., Iniguez-Luy, F., Leon, E. and Osborn, T.C. (2007) Genomic changes in resynthesized Brassica napus and their effect on gene expression and phenotype. *PLANT CELL*, **19**, 3403–3417.

Glover, N.M., Redestig, H. and Dessimoz, C. (2016) Homoeologs: What Are They and How Do We Infer Them? *Trends Plant Sci.*, **21**, 609–621.

Golicz, A.A., Bayer, P.E., Barker, G.C., et al. (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.*, **7**, 13390.

Gong, L., Olson, M. and Wendel, J.F. (2014) Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol. Biol. Evol.*, **31**, 2624–2636.

Gong, L., Salmon, A., Yoo, M.-J., Grupp, K.K., Wang, Z., Paterson, A.H. and Wendel, J.F. (2012) The Cytonuclear Dimension of Allopolyploid Evolution: An Example from Cotton Using Rubisco. *Mol. Biol. Evol.*, **29**, 3023–3036.

Greiner, S., Rauwolf, U., Meurer, J. and Herrmann, R.G. (2011) The role of plastids in plant speciation. *Mol. Ecol.*, **20**, 671–691.

Greiner, S., Wang, X., Herrmann, R.G., Rauwolf, U., Mayer, K., Haberer, G. and Meurer, J. (2008) The Complete Nucleotide Sequences of the 5 Genetically Distinct Plastid Genomes of Oenothera, Subsection Oenothera: II. A Microevolutionary View Using Bioinformatics and Formal Genetic Data. *Mol. Biol. Evol.*, **25**, 2019–2030.

Grover, C.E., Gallagher, J.P., Szadkowski, E.P., Yoo, M.J., Flagel, L.E. and Wendel, J.F. (2012) Homoeolog expression bias and expression level dominance in allopolyploids. *NEW Phytol.*, **196**, 966–971.

Higgins, J., Magusin, A., Trick, M., Fraser, F. and Bancroft, I. (2012) Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. *BMC Genomics*, **13**, 247.

Hurgobin, B., Golicz, A.A., Bayer, P.E., et al. (2017) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. *Plant Biotechnol. J.*

Ilut, D.C., Coate, J.E., Luciano, A.K., Owens, T.G., May, G.D., Farmer, A. and Doyle, J.J. (2012) A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *Am. J. Bot.*, **99**, 383–396.

Kearse, M., Moir, R., Wilson, A., et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinforma. Oxf. Engl.*, **28**, 1647–1649.

Keller, J., Rousseau-Gueutin, M., Martin, G.E., et al. (2017) The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from Lupinus. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes*, **24**, 343–358.

Kleine, T., Maier, U.G. and Leister, D. (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu. Rev. Plant Biol.*, **60**, 115–138.

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, **33**, 1870–1874.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Lashermes, P., Combes, M.-C., Hueber, Y., Severac, D. and Dereeper, A. (2014) Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. *Plant J.*, **78**, 674–685.

Lerbs-Mache, S. (2011) Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? *Plant Mol. Biol.*, **76**, 235–249.

Li, A., Liu, D., Wu, Jun, et al. (2014) mRNA and Small RNA Transcriptomes Reveal Insights into Dynamic Homoeolog Regulation of Allopolyploid Heterosis in Nascent Hexaploid Wheat. *Plant Cell*, **26**, 1878–1900.

Li, F., Fan, G., Lu, C., et al. (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.*, **33**, 524–530.

Li, H., Handsaker, B., Wysoker, A., et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y. and De Smet, R. (2016) Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. *Plant Cell*, **28**, 326–344.

Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.*, **30**, 923–930.

Margulis, L. (1970) Origin of Eukaryotic Cells. In New Haven, CT: Yale University Press.

Marmagne, A., Brabant, P., Thiellement, H. and Alix, K. (2010) Analysis of gene expression in resynthesized Brassica napus allotetraploids: transcriptional changes do not explain differential protein regulation. *New Phytol.*, **186**, 216–227.

Martin, W., Rujan, T., Richly, E., et al. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci.*, **99**, 12246–12251.

Mower, J.P. (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.*, **37**, W253–W259.

Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Crollius, H.R. and Salse, J. (2015) Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biol.*, **16**, 262.

Papp, B., Pál, C. and Hurst, L.D. (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature*, **424**, 194–197.

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M. and Ainouche, M. (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in Spartina. *New Phytol.*, **184**, 1003–1015.

Parkin, I.A., Koh, C., Tang, H., et al. (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol.*, **15**, R77.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, **11**, R25.

Rockenbach, K., Havird, J.C., Monroe, J.G., Triant, D.A., Taylor, D.R. and Sloan, D.B. (2016) Phylogenetic and Population Genetic Evidence for Positive Selection in Rapidly Evolving Plastid-Nuclear Enzyme Complexes. *Genetics*, genetics.116.188268.

Rousseau-Gueutin, M., Ayliffe, M.A. and Timmis, J.N. (2011) Conservation of Plastid Sequences in the Plant Nuclear Genome for Millions of Years Facilitates Endosymbiotic Evolution. *Plant Physiol.*, **157**, 2181–2193.

Rousseau-Gueutin, M., Huang, X., Higginson, E., Ayliffe, M., Day, A. and Timmis, J.N. (2013) Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiol.*, **161**, 1918–1929.

Rousseau-Gueutin, M., Morice, J., Coriton, O., et al. (2017) The Impact of Open Pollination on the Structural Evolutionary Dynamics, Meiotic Behavior, and Fertility of Resynthesized Allotetraploid Brassica napus L. *G3 Genes Genomes Genet.*, **7**, 705–717.

Salmon, A., Flagel, L., Ying, B., Udall, J.A. and Wendel, J.F. (2010) Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol.*, **186**, 123–134.

Salse, J., Abrouk, M., Bolot, S., et al. (2009) Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 14908–14913.

Särkinen, T., Bohs, L., Olmstead, R.G. and Knapp, S. (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.*, **13**, 214.

Sehrish, T., Symonds, V.V., Soltis, D.E., Soltis, P.S. and Tate, J.A. (2015) Cytonuclear Coordination Is Not Immediate upon Allopolyploid Formation in Tragopogon miscellus (Asteraceae) Allopolyploids. *PLoS ONE*, **10**, e0144339.

Sharbrough, J., Conover, J.L., Tate, J.A., Wendel, J.F. and Sloan, D.B. (2017) Cytonuclear responses to genome doubling. *Am. J. Bot.*, **104**, 1277–1280.

Shen, E., Zou, J., Hubertus Behrens, F., et al. (2015) Identification, evolution, and expression partitioning of miRNAs in allopolyploid Brassica napus. *J. Exp. Bot.*, **66**, 7241–7253.

Sloan, D.B., Triant, D.A., Wu, M. and Taylor, D.R. (2014) Cytonuclear Interactions and Relaxed Selection Accelerate Sequence Evolution in Organelle Ribosomes. *Mol. Biol. Evol.*, **31**, 673–682.

Smet, R.D., Adams, K.L., Vandepoele, K., Montagu, M.C.E.V., Maere, S. and Peer, Y.V. de (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci.*, **110**, 2898–2903.

Springer, N.M., Lisch, D. and Li, Q. (2016) Creating Order from Chaos: Epigenome Dynamics in Plants with Complex Genomes. *Plant Cell*, **28**, 314–325.

Szadkowski, E., Eber, F., Huteau, V., et al. (2010) The first meiosis of resynthesized Brassica napus, a genome blender. *New Phytol.*, **186**, 102–112.

Tate, J., Joshi, P., Soltis, K., Soltis, P. and Soltis, D. (2009) On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid Tragopogon miscellus (Asteraceae). *BMC Plant Biol.*, **9**, 80.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.*, **5**, 123–135.

Wang, J., Tian, L., Lee, H.S., et al. (2006) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. *Genetics*, **172**, 507–517.

Wang, X., Dong, Q., Li, X., Yuliang, A., Yu, Y., Li, N., Liu, B. and Gong, L. (2017) Cytonuclear Variation of Rubisco in Synthesized Rice Hybrids and Allotetraploids. *Plant Genome*, **10**(3).

Wang, X., Wang, H., Wang, J., et al. (2011) The genome of the mesopolyploid crop species Brassica rapa. *Nat. Genet.*, **43**, 1035–1039.

Warner, D.A. and Edwards, G.E. (1993) Effects of polyploidy on photosynthesis. *Photosynth. Res.*, **35**, 135–147.

Wendel, J.F., Jackson, S.A., Meyers, B.C. and Wing, R.A. (2016) Evolution of plant genome architecture. *Genome Biol.*, **17**, 37.

Weng, M.-L., Ruhlman, T.A. and Jansen, R.K. (2016) Plastid–Nuclear Interaction and Accelerated Coevolution in Plastid Ribosomal Genes in Geraniaceae. *Genome Biol. Evol.*, **8**, 1824–1838.

Wu, J., Lin, L., Xu, M., Chen, P., Liu, D., Sun, Q., Ran, L. and Wang, Y. (2018) Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. *BMC Genomics*, **19**, 586.

Wyman, S.K., Jansen, R.K. and Boore, J.L. (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.

Xiong, Z., Gaeta, R.T. and Pires, J.C. (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid Brassica napus. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 7908–7913.

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

Yang, Z. and Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.*, **17**, 32–43.

Yoo, M.J., Szadkowski, E. and Wendel, J.F. (2012) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity, **110**, 171-180*.

TABLE

	Plastid complexes				*Cytosolic complexes*		*Overall genes*	
	B. rapa	*B. napus A*	*B. oleracea*	*B. napus C*	*B. rapa*	*B. oleracea*	*B. rapa*	*B. oleracea*
tri% (nb)	20.0 (22)	17.3 (19)	19.1 (21)	18.2 (20)	2. (2)	3.6 (3)	4.9 (706)	4.7 (559)
di% (nb)	38.2 (42)	41.8 (46)	37.3 (41)	47.3 (52)	34.9 (30)	35.7 (30)	24.7 (3 521)	25.8 (3 053)
uni%(nb)	41.8 (46)	40.9 (45)	40.0 (44)	32.7 (36)	62.8 (54)	60.7 (51)	70.1 (10 003)	69.7 (8 172)
Obs Chi2	63.3	54.1	56.6	71.3	6.4	5.2	-/-	-/-
Exp Chi2 (1%. 1ddl)	6.6	6.6	6.6	6.6	6.6	6.6	-/-	-/-

Table 1. Copy retention for nuclear genes implicated in plastid and cytosolic complexes. Proportions of genes retained in triplicate, duplicate and single copy are given for *Brassica rapa, B. oleracea* and *B. napus* for two subsets of nuclear genes either implicated in plastid or cytosolic complexes. These proportions were then compared to retention rates for overall genes. Results of Chi^2 are presented for comparisons between plastid complexes and overall genes, and between cytosolic complexes and overall genes (NS=Non-significant).

FIGURES

Figure 1. Position of nuclear genes involved in nine plastid protein complexes and their homoeologous and paralogous relationships in *Brassica napus*. The outer circle represents the 19 chromosomes of *B. napus* var. Darmor (from Chalhoub et al. 2014). Genes are then positioned on the chromosomes and possible events of gene conversion events are illustrated with the black dashed lines. Homoeologous relationships between nuclear gene copies from the A and C subgenomes are presented by solid colored lines. The dashed colored lines represent the paralogous relationships between redundant copies within a subgenome.

Figure 2. Subgenome transcription dynamics in the progenitors, interspecific hybrids F1, resynthesized *B. napus* in S0 and in G3 (from both 'EMZ' and 'RCC' crosses) as well as in natural *B. napus* varieties. Each comparison is made between the individuals at the tips of the line. We compared either the transcription levels of *B. rapa* [Ar] vs *B. oleracea* [Co] subgenomes within polyploids, or between the subgenome of the polyploid *B. rapa* [Ar] or *B. oleracea* [Co] with the respective progenitor genome *B. rapa* [Ar] or *B. oleracea* [Co], respectively. For each comparison, the number of genes up-regulated in the individuals are presented at the tip of the line with percentage in brackets. Thickness of each tip is proportional to the percentage of up-regulated genes. Numbers in brackets in the middle of the line represent the proportions of genes non-differentially expressed for each comparison. By standard, *B. rapa* [A] genome is represented in blue and *B. oleracea* [C] genome in red.

Figure 3. Heatmap of RNAseq data for the two crosses 'RCC' and 'EMZ' studied. Global expression is presented as total disc surface for all 110 genes implicated in the nine plastid complexes. An asterisk was added for each significant differential global expression between resynthesized polyploids and the *in silico* S0, indicating either over- or lower-expression in polyploids. Biases in subgenome expression are also schematized as a gradient of color from red (*B. oleracea*) to blue (*B. rapa*) in hybrid, S0 and resynthesized polyploids. Individuals from the 'RCC' and 'EMZ' populations are depicted on the left and on the right of each panel, respectively.