

Why old duplicated genes are not thrown away in (paleo)polyploids? Example from the *petC* gene in *Brassica napus*

Plant Molecular Biology

Maud Facon¹, Gwenaëlle Deniot¹, Maryse Lodé-Taburel¹, Héloïse Archambeau¹, Emilie Montes¹, Younès Delloero¹, Loeiz Maillat¹, Anne-Marie Chèvre¹, Mathieu Rousseau-Gueutin^{1*}

¹IGEPP, INRAE, Institut Agro, Université de Rennes, 35650 Le Rheu, France

Mathieu Rousseau-Gueutin : mathieu.rousseau-gueutin@inrae.fr

Supplementary methods

Identification and evolution of the Brassicaceae *petC* genes

To identify the number of *petC* copies in Brassicaceae and more specifically in *B. napus*, we first retrieved the Arabidopsis Columbia *petC* sequence (in nucleotide and amino-acid) on TAIR 10 (<https://www.arabidopsis.org/>). To identify its homologous sequences in other Brassicaceae species, Blastn and Blastp analyses were performed against gene coding sequences (in nucleotide and amino-acid) of various Brassicaceae species available at the BRAD database (<http://brassicadb.cn/>), including *B. juncea* (cv ‘AU213’ v1 and ‘Tumida T84-66’ v2), *B. napus* (cv ‘ZS11’ v2 and ‘Darmor’ v10), *B. nigra* (‘Ni100’), *B. oleracea* (cv ‘HDEM’ v1 and ‘JZS’ v2), *B. rapa* (‘Chiifu’ v3.0 and ‘Z1’ v2), and *Sisymbrium irio*. We thereafter realized a multiple alignment using Geneious 11.0.14.1 (<https://www.geneious.com>) and performed phylogenetic analyses (Bootstrap: 1000; outgroup: *A. thaliana*) by either Maximum Parsimony or Maximum Likelihood (model Kimura 2 parameter) using MEGA 10 (Kumar et al, 2018) in order to identify the different *petC* copies and their origin in each species. The details of the gene accession numbers are given in Table S1. In this phylogenetic analysis, we also included the *B. napus* cv Drakkar *petC* sequences that we obtained via cloning and sequencing (sequences available in Table S4).

Selective pressure acting on the four *petC* nuclear genes

The selective pressure acting on each *petC* paleologous copy (A03, A09, C03, C09) was calculated by estimating the synonymous (Ks) and non-synonymous (Ka) nucleotide substitution rates between orthologous *Brassica* sequences using the yn00 method (Yang and Nielsen, 2000) from the PAML package (Yang, 2007). The results obtained from this analysis are given in Table S2.

Evaluation of the four *petC* copies transcription level using RNA Seq data

To determine the transcription level of the four *petC* gene copies, we used three Paired End RNA Seq libraries available at the EBI (<https://www.ebi.ac.uk/ena/browser/view/PRJEB5461?show=reads>) obtained from *B. napus* cv. Darmor leaves (Chalhoub et al. 2014): ERR431525; ERR431518; ERR431522. We then used a RNA-Seq snakePipes (v2.5.1; Bhardwaj et al. 2019) pipeline in “alignment” mode, which firstly performs the quality checks, trims and filters the reads, and then uses STAR (v2.7.4a; Dobin and Gingeras, 2015) with the following parameters (--alignEndsType EndToEnd --outFilterMismatchNmax 2) to map the cleaned RNA Seq reads against *B. napus* cv. Darmor v10 (Rousseau-Gueutin *et al.* 2021). Subsequently, the number of reads that mapped on each *petC* copy was determined using Feature Counts (-C -Q 10 -prima; Liao et al. 2014). Finally,

we calculated the normalization factors using the Trimmed Mean of M values (TMM) method and performed RPKM normalization (reads per kilobase of transcript per million reads mapped) using edgeR (v3.38.4; Robinson et al. 2010).

Amplification, cloning and sequencing of the *B. napus* cv. Drakkar *petC* gene copies

To obtain the full-length gene sequences of four *petC* gene copies in *B. napus* cv. Drakkar that we used for genetic transformation experiments, we first developed primer pairs (sup data S1) using the *B. napus* cv. Darmor *petC* gene sequences (on chromosomes A03, C03, A09, C09). To develop primer pairs specific to each copy, we designed forward and reverse primers before and after the 3' and 5'UTR of each *petC* copy using Netprimer (<http://www.premierbiosoft.com/netprimer/>): see Table S5. Then, we amplified by PCR each copy in *B. napus* cv. Drakkar. Each High-Fidelity PCR amplification was performed in a total volume of 20 μ l containing 2 μ l of 10 \times ExTaq (TaKaRa), 1.6 μ l of 2.5 mM deoxyribonucleotide mix, 2 μ l of each primer (10 mM), 0.1 μ l of ExTaq polymerase (5 U μ l⁻¹) and 20 ng of template DNA. The thermal cycling program corresponds to a first denaturation step at 98°C for 2 min, followed by 35 cycles, each with 10s denaturation at 98°C, 30s annealing at 60°C, and an extension of 1 min at 72°C. A final extension of 5 min at 72°C was performed. Thereafter, each PCR product was cloned using the Strataclone PCR cloning kit (Agilent). All procedures (PCR, ligation, and cloning) were performed twice. For each PCR product, a total of four clones were sequenced (Genoscreen). Different internal primers were also developed to sequence the full gene for all copies (Table S3). The sequences of these different cloned fragments were then aligned and phylogenetic analyses were performed as described previously to identify from which chromosome derived the different cloned sequences. The different *B. napus* cv Drakkar sequences are available in Table S4.

Design of a pair of guides simultaneously targeting all *B. napus* cv. Drakkar *petC* copies

To knock-out one, several or all four *B. napus* cv Drakkar *petC* gene copies, and to prevent the design of numerous guides but also of primer pairs specific of each copy and of each guide, we decided to design only two guides that targeted with the same efficiency the different genes copies (*i. e.* 100% of similarity between the four copies for the targeting region). Thereafter, we used the CRISPOR software using *B. napus* cv. Westar as the reference and identified two guides (guide 320 5'-3': ATCGTTTCCAAGTGCATCCT TGG; guide 502 5'-3': ATCTTGGATGTGTTGTGCCATGG) targeting all four *petC* gene copies (Haeussler et al. 2016; <http://crispor.tefor.net/>) and located in the third and fourth exon (corresponding to the “Rieske functional domain”), respectively. We also verified the absence of nucleotidic differences between *B. napus* cv. Westar and *B. napus* cv. Drakkar for these two guides. For these latter, there are no identified off-targets (apart from the 4 *petC* copies) with less than three mismatches.

Vector Construction

A fragment of 927 bp was firstly synthesized by IDT (Integrated DNA technologies, USA, see detail synthesized sequence in Table S6) and validated by Sanger sequencing. It contained an attB1 sequence, the first guide driven by the BnU3 promoter and followed by a scaffold, the second guide driven by a BnU6 promoter followed by a scaffold and an attB2 region. Taking advantage of the attB1 and attB2 regions and using the

Gateway cloning system, the synthesized fragment was inserted into the pDONR207 (Thermo Fisher Scientific) via a BP reaction. Thereafter, this vector was inserted by LR reaction into a pDE-Cas9 destination vector (Addgene # 61433) containing a spectinomycin bacterial resistance gene, an *Arabidopsis thaliana* codon optimized Cas9 (driven by PcUbi4-2 constitutive Ubiquitin 4-2 promoter from *Petroselinum crispum*) and a BlpR sequence conferring Basta resistance. This transformation vector is referred as pDE-Bn_petC_all.

Plant transformation

B. napus cv. Drakkar seeds obtained from the BrACySol Biological Resource Center (Rennes-Ploudaniel, France; https://www6.rennes.inrae.fr/igepp_eng/About-IGEPP/Platforms/BrACySol) were sterilized by treating them with 70% ethanol for two minutes, followed by 20 minutes with bleach (2,6%) and tween, and three rinses with sterile water (five minutes each). The sterilized seeds were thereafter left to dry overnight. Hundred seeds were sown on MS medium (pH 5,8) containing 2 g/l Gelrite and 20 g/l sucrose. They were then grown in a growth chamber (16h day/8h night, 20°C) for 10 days. From these plantlets, a total of 400 *B. napus* cv. Drakkar hypocotyl segments were pre-conditioned on MS (pH 5,8) containing 2 g/l Gelrite, 1 mg/L 2,4D and 30 g/l sucrose. They were left on this medium four days and in the same growth condition. Preconditioned hypocotyls were placed in *Agrobacterium tumefaciens* C58 GV3101 (pMp90) containing the pDE-Bn_petC_all vector suspension (at DO600 = 0,1) in liquid MinA medium (20X salt solution: 20g/l [NH₄]₂SO₄ and 10g/l sodium citrate; 20X buffer solution: 274 g/l K₂HPO₄ and 90 g/l KH₂PO₄; sucrose at 20% and MgSO₄ at 1M) for 30 minutes (slight shaking at 50 rpm). Infected hypocotyls were thereafter placed on the same medium that was used to pre-condition the hypocotyls. Selection and callus induction were initiated by placing the treated hypocotyls on MS-medium containing 2 g/l Gelrite, 1 mg/l 2,4D 30 g/l sucrose, 25 mg/l of the bialaphos plant herbicide and 400 mg/l timentin (Ticarcillin disodium/Clavulanate potassium 15:1; Duchefa). For this experiment, we also prepared two control plates: one with non-inoculated explants but antibiotic selection (to verify the antibiotic selection efficacy), as well as another one with non-inoculated explants and without antibiotic selection (control of organogenesis).

The T0 plants resistant to Bialaphos were regenerated as rooted plantlets in a growth chamber (16h day/8h night, 20°C). The plant regeneration required a transfer on a different or renewed medium every two weeks: firstly on an adapted MS-ORG medium favoring organogenesis (MS pH 5,8, Gelrite 2 g/l, BAP 4 mg/l, Zeatine 2 mg/l, sucrose 30 g/l, bialaphos 2,5 mg/L and timentin 400 mg/L), followed by MS-BOURG medium (twice) favoring plant regeneration (containing MS pH 5,8, Gelrite 2 g/l, BAP 3 mg/l, Zeatine 2 mg/l, sucrose 30 g/l bialaphos 2,5 mg/l and timentin 400 mg/l), thirdly a transfer on a MS-ELONG medium (often three times) favoring plant elongation (containing MS pH 5,8, Gelrite 3 g/l, BAP 0.05 mg/l, sucrose 30 g/l, bialaphos 2,5 mg/l and timentin 400 mg/t), and finally a MS-ENR medium (3-4 times) favoring plant rooting (containing MS pH 5,8, Gelrite 3 g/l, IBA 5 mg/l, sucrose 30 g/l, bialaphos 2,5 mg/l and timentin 400 mg/l). All these steps required in total about 6 months of culture before transferring the plantlets to greenhouse. They were left there for acclimation for three weeks before being vernalized for a month. The vernalized plants were thereafter selfed (on floral buds) to produce T1 seeds.

Identification of putative mutated copies in T0 transformed plants

DNA was extracted from the six T0 plants that survived on selection media. The presence of T DNA was determined by using PCR and primers amplifying the *Bar* and *Cas9* genes (primer sequences in supp. data S1). Thereafter we searched for the putative presence of mutations within one or several *petC* copies using primer pairs specific to each *petC* and surrounding the 320 or 502gRNA (Table S5). Copy specific PCR products were cloned (Strataclone cloning kit Agilent) and sequenced by Sanger (Genoscreen).

Production of homozygous mutant plants for one or several copies and Cas9 free

A few hundred T1 progenies deriving from each four T0 mutated plants (see details in S5) were grown on MS medium supplemented with sucrose 20g/l. Genomic DNA from each plant was extracted using the beadex maxi plant kit (LGC Genomics, Teddington Middlesex, UK) on the oKtopure robot at the GENTYANE platform (INRA Clermont-Ferrand, France). The absence of the *cas9* gene within the progenies was tested using the following primers in 5'-3': Cas9-F1-GD CGTTGATGAGGTGGCATAACCA / Cas9-R1-GD CAGCGTACTGATCTCCGATCTG. Within these latter plants, we searched for the presence of a fixed mutation at a given locus, by Sanger sequencing (and cloning when necessary) as described above. Within these latter, we notably obtained three mutant plants K.O. for one or two *petc* copies (16h, 37d, and 42a) that were *cas9* free and that we selfed to obtain T2 seeds.

Photosystem II fluorescence measurements

Photosystem II (PSII) fluorescence measurements were performed on leaves at similar phenological stages from two-month-old plants of either wild type or T2 single or double *petC* mutants (4 leaves per accession). PSII fluorescence was measured on an Open FluorCam FC 800-O/2020 (Photon System Instrument) using “Kautsky effect” protocol. “EI shutter” and “sensitivity” parameters were adjusted for each leaf prior to measurements to prevent a potential saturation of the fluorescence signal at high light intensities. For measurements, plants were subjected at a light intensity (PPFD) of 971 $\mu\text{mol photons. m}^{-2}.\text{s}^{-1}$ in order to saturate the PSII electron transfer rate and evaluate its maximal capacity at ambient CO_2 (See Figure S1 and Delloero *et al.*, 2021). This PPFD was obtained with a level intensity of 70% for actinic light 2, based on calibration curves performed with independent PPFD measurements at the leaf level. Measurements were recorded every 5 min and only the fourth measurement was used to calculate saturating PSII Electron Transport Rate (ETR_{sat}) as follows: $\text{ETR}_{\text{sat}} = \phi\text{PSII} \times 0.5 \times \text{PPFD} \times \alpha_{\text{leaf}}$ (Maxwell and Johnson, 2000), with ϕPSII corresponding to the steady-state quantum yield of PSII ($= (\text{FmLss} - \text{FtLss}) / \text{FmLss}$ with the FluorCam equations), α_{leaf} corresponding to the light absorption coefficient of a leaf ($\alpha_{\text{leaf}} = 0.85$ for C_3 plant species; Peterson and Havir, 2001) and the factor 0.5 reflecting the partitioning of light between PSI and PSII.

Since ETR_{sat} may vary among plants independently of the *petC* mutation (intrinsic differences of chlorophyll levels per leaf area for the tested leaves), the results (in $\mu\text{mol electrons.m}^{-2}.\text{s}^{-1}$) were normalized with respect to chlorophyll contents per leaf area (g Chl.m^{-2}) leading to ETR_{sat} values expressed in $\mu\text{mol electrons.s}^{-1}.\text{g}^{-1}$ Chl. Chlorophyll content per leaf area was determined from five leaf discs of 0.8 cm^2 (4 cm^2) harvested in liquid nitrogen at the end of PSII measurements for each analyzed leaf, exactly as previously described in Delloero *et al.* (2021). Briefly, chlorophylls a and b were extracted by depletion (white pellet) with ice-cold acetone and

quantified by spectrophotometry using dedicated coefficients from Lichtenthaler (1987). Total soluble proteins per leaf area and Rubisco content were also analyzed using Bradford reagent and 15% SDS-PAGE gel (Dellero *et al.*, 2021) and no differences were observed between wild-type and mutant plants.

References

- Bhardwaj V, Heyne S, Sikora K, Rabbani L, Rauer M, Kilpert F, Richter AS, Ryan DP, Manke T. (2019) snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics*. **35**, 4757-4759.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., et al. (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed. *Science*. **345**, 950-953.
- Dellero, Y., Jossier, M., Bouchereau, A., Hodges, M., Lepout L. (2021) Leaf Phenological Stages of Winter Oilseed Rape (*Brassica napus* L.) Have Conserved Photosynthetic Efficiencies but Contrasted Intrinsic Water Use Efficiencies at High Light Intensities. *Front. Plant Sci.* **12**.
- Dobin, A., Gingeras, T. R., Spring, C., Flores, R., Sampson, J., et al. (2016) Mapping RNA-seq with STAR. *Curr Protoc Bioinformatics*. **51**, 586-597.
- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., et al. (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 1-12.
- Kumar, S., Stecher, G., Li, M., Nnyaz, C., and Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 6.
- Liao, Y., Smyth, G. K., and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923-930.
- Lichtenthaler, H. K. (1987) Chlorophylls and Carotenoids: Pigments of Photosynthetic Biomembranes. *Methods in Enzymol.* **148**, 350-382.
- Maxwell, K., and Johnson, G. N. (2000) Chlorophyll fluorescence—a practical guide. *J. Exp. Bot.* **51**, 659-668.
- Peterson, R. B., and Havir, E. A. (2001) Photosynthetic properties of an *Arabidopsis thaliana* mutant possessing a defective PsbS gene. *Planta* **214**, 142-152.
- Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139-40.
- Rousseau-Gueutin, M., Belser, C., Da Silva, C. Richard, G., Istace, B., et al. (2021) Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*. **9**, gaa137.
- 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.