

## PLANT GENETICS

# Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome

Boulos Chalhoub,<sup>1,†</sup> France Denoed,<sup>2,3,4,\*</sup> Shengyi Liu,<sup>5,\*</sup> Isobel A. P. Parkin,<sup>6,†</sup> Haibao Tang,<sup>7,8</sup> Xiyin Wang,<sup>9,10</sup> Julien Chiquet,<sup>11</sup> Harry Belcram,<sup>1</sup> Chaobo Tong,<sup>5</sup> Birgit Samans,<sup>12</sup> Margot Corr ea,<sup>2</sup> Corinne Da Silva,<sup>2</sup> J r my Just,<sup>1</sup> Cyril Falentin,<sup>13</sup> Chu Shin Koh,<sup>14</sup> Isabelle Le Clainche,<sup>1</sup> Maria Bernard,<sup>2</sup> Pascal Bento,<sup>2</sup> Benjamin Noel,<sup>2</sup> Karine Labadie,<sup>2</sup> Adriana Alberti,<sup>2</sup> Mathieu Charles,<sup>15</sup> Dominique Arnaud,<sup>1</sup> Hui Guo,<sup>9</sup> Christian Daviaud,<sup>16</sup> Salman Alamery,<sup>17</sup> Kamel Jabbari,<sup>1,18</sup> Meixia Zhao,<sup>19</sup> Patrick P. Edger,<sup>20</sup> Houada Chelaifa,<sup>1</sup> David Tack,<sup>21</sup> Gilles Lassalle,<sup>13</sup> Imen Mestiri,<sup>1</sup> Nicolas Schnel,<sup>13</sup> Marie-Christine Le Paslier,<sup>15</sup> Guangyi Fan,<sup>22</sup> Victor Renault,<sup>23</sup> Philippe E. Bayer,<sup>17</sup> Agnieszka A. Golicz,<sup>17</sup> Sahana Manoli,<sup>17</sup> Tae-Ho Lee,<sup>9</sup> Vinh Ha Dinh Thi,<sup>1</sup> Smahane Chalabi,<sup>1</sup> Qiong Hu,<sup>5</sup> Chuchuan Fan,<sup>24</sup> Reece Tollenaere,<sup>17</sup> Yunhai Lu,<sup>1</sup> Christophe Batail,<sup>2</sup> Jinxiong Shen,<sup>24</sup> Christine H. D. Sidebottom,<sup>14</sup> Xinfang Wang,<sup>5</sup> Aur lie Canaguier,<sup>1</sup> Aur lie Chauveau,<sup>15</sup> Aur lie B rard,<sup>15</sup> Gwena lle Deniot,<sup>13</sup> Mei Guan,<sup>25</sup> Zhongsong Liu,<sup>25</sup> Fengming Sun,<sup>22</sup> Yong Pyo Lim,<sup>26</sup> Eric Lyons,<sup>27</sup> Christopher D. Town,<sup>7</sup> Ian Bancroft,<sup>28</sup> Xiaowu Wang,<sup>29</sup> Jinling Meng,<sup>24</sup> Jianxin Ma,<sup>19</sup> J. Chris Pires,<sup>30</sup> Graham J. King,<sup>31</sup> Dominique Brunel,<sup>15</sup> R gine Delourme,<sup>13</sup> Michel Renard,<sup>13</sup> Jean-Marc Aury,<sup>2</sup> Keith L. Adams,<sup>21</sup> Jacqueline Batley,<sup>17,32</sup> Rod J. Snowdon,<sup>12</sup> Jorg Tost,<sup>16</sup> David Edwards,<sup>17,32,†</sup> Yongming Zhou,<sup>24,†</sup> Wei Hua,<sup>5,†</sup> Andrew G. Sharpe,<sup>14,†</sup> Andrew H. Paterson,<sup>9,†</sup> Chunyun Guan,<sup>25,†</sup> Patrick Wincker<sup>2,3,4,†</sup>

Oilseed rape (*Brassica napus* L.) was formed ~7500 years ago by hybridization between *B. rapa* and *B. oleracea*, followed by chromosome doubling, a process known as allopolyploidy. Together with more ancient polyploidizations, this conferred an aggregate 72× genome multiplication since the origin of angiosperms and high gene content. We examined the *B. napus* genome and the consequences of its recent duplication. The constituent A<sub>n</sub> and C<sub>n</sub> subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homeologous exchanges. Incipient gene loss and expression divergence have begun. Selection in *B. napus* oilseed types has accelerated the loss of glucosinolate genes, while preserving expansion of oil biosynthesis genes. These processes provide insights into allopolyploid evolution and its relationship with crop domestication and improvement.

The Brassicaceae are a large eudicot family (1) and include the model plant *Arabidopsis thaliana*. Brassicas have a propensity for genome duplications (Fig. 1) and genome mergers (2). They are major contributors to the human diet and were among the earliest cultigens (3).

*B. napus* (genome A<sub>n</sub>A<sub>n</sub>C<sub>n</sub>C<sub>n</sub>) was formed by recent allopolyploidy between ancestors of *B. oleracea* (Mediterranean cabbage, genome C<sub>0</sub>C<sub>0</sub>) and *B. rapa* (Asian cabbage or turnip, genome A<sub>r</sub>A<sub>r</sub>) and is polyphyletic (2, 4), with spontaneous formation regarded by Darwin as an example of unconscious selection (5). Cultivation began in Europe during the Middle Ages and spread worldwide. Diversifying selection gave rise to oilseed rape (canola), rutabaga, fodder rape, and kale morphotypes grown for oil, fodder, and food (4, 6).

The homozygous *B. napus* genome of European winter oilseed cultivar ‘Darmor-*bzh*’ was assembled with long-read [>700 base pairs (bp)] 454 GS-FLX+ Titanium (Roche, Basel, Switzerland) and Sanger sequence (tables S1 to S5 and figs. S1 to S3) (7). Correction and gap filling used 79 Gb of Illumina (San Diego, CA) HiSeq sequence. A final assembly of 849.7 Mb was obtained with SOAP (8) and Newbler (Roche), with 89% nongapped se-

quence (tables S2 and S3). Unique mapping of ~5× nonassembled 454 sequences from *B. rapa* (‘Chiifu’) or *B. oleracea* (‘TO1000’) assigned most of the 20,702 *B. napus* scaffolds to either the A<sub>n</sub> (8294) or the C<sub>n</sub> (9984) subgenomes (tables S4 and S5 and fig. S3). The assembly covers ~79% of the 1130-Mb genome and includes 95.6% of *Brassica* expressed sequence tags (ESTs) (7). A single-nucleotide polymorphism (SNP) map (tables S6 to S9 and figs. S4 to S8) genetically anchored 712.3 Mb (84%) of the genome assembly, yielding pseudomolecules for the 19 chromosomes (table S10).

The assembled C<sub>n</sub> subgenome (525.8 Mb) is larger than the A<sub>n</sub> subgenome (314.2 Mb), consistent with the relative sizes of the assembled C<sub>0</sub> genome of *B. oleracea* (540 Mb, 85% of the ~630-Mb genome) and the A<sub>r</sub> genome of *B. rapa* (312 Mb, 59% of the ~530-Mb genome) (9–11). The *B. napus* assembly contains 34.8% transposable elements (TEs), less than the 40% estimated from raw reads (tables S11 to S14) (7), with asymmetric distribution in the A<sub>n</sub> and C<sub>n</sub> subgenomes (table S12) as in the progenitor genomes (9–11). A small TE fraction has proliferated since *B. napus* separated from its progenitors (7), at lower rates in the *B. napus* subgenomes than the corresponding progenitor genomes (table S14 and figs. S9 and S10).

The *B. napus* genome contains 101,040 gene models estimated from 35.5 Gb of RNA sequencing (RNA-seq) data (tables S15 and S16) in combination with ab initio gene prediction, protein and EST alignments, and transposon masking (7). Of these, 91,167 were confirmed by matches with *B. rapa* and/or *B. oleracea* predicted proteomes. Genes are abundant in distal euchromatin but sparse near centromeric heterochromatin (Fig. 2). RNA-seq data revealed alternative splicing in 48% of genes, with frequent intron retention (62%) and rare exon skipping (3%) (tables S17 and S18 and fig. S11).

The *B. napus* A<sub>n</sub> and C<sub>n</sub> subgenomes are largely colinear to the corresponding diploid A<sub>r</sub> and C<sub>0</sub>

<sup>1</sup>Institut National de Recherche Agronomique (INRA)/ Universit  d’Evry Val d’Essonne, Unit  de Recherche en G n mique V g tale, UMR1165, Organization and Evolution of Plant Genomes, 2 rue Gaston Cr mieux, 91057 Evry, France. <sup>2</sup>Commissariat   l’Energie Atomique (CEA), Institut de G n mique (IG), Genoscope, BP5706, 91057 Evry, France. <sup>3</sup>Universit  d’Evry Val d’Essonne, UMR 8030, CP5706, Evry, France. <sup>4</sup>Centre National de Recherche Scientifique (CNRS), UMR 8030, CP5706, Evry, France. <sup>5</sup>Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture of People’s Republic of China, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China. <sup>6</sup>Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N 0X2, Canada. <sup>7</sup>J. Craig Venter Institute, Rockville, MD 20850, USA. <sup>8</sup>Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China. <sup>9</sup>Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA. <sup>10</sup>Center of Genomics and Computational Biology, School of Life Sciences, Hebei United University, Tangshan, Hebei 063000, China. <sup>11</sup>Laboratoire de Math matiques et Mod lisation d’Evry–UMR 8071 CNRS/Universit  d’Evry val d’Essonne–USC INRA, Evry, France. <sup>12</sup>Department of Plant Breeding, Research Center for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. <sup>13</sup>INRA, Institut de G n tique, Environnement et Protection des Plantes (IGEPP) UMR1349, BP35327, 35653 Le Rheu Cedex, France. <sup>14</sup>National Research Council Canada, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada. <sup>15</sup>INRA, Etude du Polymorphisme des G nomes V g taux, US1279, Centre National de G notypage, CEA–IG, 2 rue Gaston Cr mieux, 91057 Evry, France. <sup>16</sup>Laboratory for Epigenetics and Environment, Centre National de G notypage, CEA–IG, 2 rue Gaston Cr mieux, 91000 Evry, France. <sup>17</sup>Australian Centre for Plant Functional Genomics, School of Agriculture and Food Sciences, University of Queensland, St. Lucia, QLD 4072, Australia. <sup>18</sup>Cologne Center for Genomics, University of Cologne, Weyertal 115b, 50931 K ln, Germany. <sup>19</sup>Department of Agronomy, Purdue University, WSLR Building B018, West Lafayette, IN 47907, USA. <sup>20</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. <sup>21</sup>Department of Botany, University of British Columbia, Vancouver, BC, Canada. <sup>22</sup>Beijing Genome Institute–Shenzhen, Shenzhen 518083, China. <sup>23</sup>Fondation Jean Dausset–Centre d’Etude du Polymorphisme Humain, 27 rue Juliette Dodu, 75010 Paris, France. <sup>24</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. <sup>25</sup>College of Agronomy, Hunan Agricultural University, Changsha 410128, China. <sup>26</sup>Molecular Genetics and Genomics Laboratory, Department of Horticulture, Chungnam National University, Daejeon-305764, South Korea. <sup>27</sup>School of Plant Sciences, iPlant Collaborative, University of Arizona, Tucson, AZ, USA. <sup>28</sup>Department of Biology, University of York, Wentworth Way, Heslington, York YO10 5DD, UK. <sup>29</sup>Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China. <sup>30</sup>Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA. <sup>31</sup>Southern Cross Plant Science, Southern Cross University, Lismore, NSW 2480, Australia. <sup>32</sup>School of Plant Biology, University of Western Australia, WA 6009, Australia.

\*These authors contributed equally to this work. †Corresponding author. E-mail: chalhoub@evry.inra.fr (B.C.); other e-mail addressees are given in the supplementary materials.