PLANT GENOMICS

The coffee genome provides insight into the convergent evolution of caffeine biosynthesis

France Denoeud,[1,2,3] Lorenzo Carretero-Paulet,[4] Alexis Dereeper,[5] Gaëtan Droc,[6] Romain Guyot,[7] Marco Pietrella,[8] Chunfang Zheng,[9] Adriana Alberti,[1] François Anthony,[5] Giuseppe Aprea,[8] Jean-Marc Aury,[1] Pascal Bento,[1] Maria Bernard,[1] Stéphanie Bocs,[6] Claudine Campa,[7] Alberto Cenci,[5,10] Marie-Christine Combes,[5] Dominique Crouzillat,[11] Corinne Da Silva,[1] Loretta Daddiego,[12] Fabien De Bellis,[6] Stéphane Dussert,[7] Olivier Garsmeur,[6] Thomas Gayraud,[7] Valentin Guignon,[10] Katharina Jahn,[9,13,14] Véronique Jamilloux,[15] Thierry Joët,[7] Karine Labadie,[1] Tianying Lan,[4,16] Julie Leclercq,[6] Maud Lepelley,[11] Thierry Leroy,[6] Lei-Ting Li,[17] Pablo Librado,[18] Loredana Lopez,[12] Adriana Muñoz,[19,20] Benjamin Noel,[1] Alberto Pallavicini,[21] Gaetano Perrotta,[12] Valérie Poncet,[7] David Pot,[6] Priyono,[22] Michel Rigoreau,[11] Mathieu Rouard,[10] Julio Rozas,[18] Christine Tranchant-Dubreuil,[7] Robert VanBuren,[17] Qiong Zhang,[17] Alan C. Andrade,[23] Xavier Argout,[6] Benoît Bertrand,[24] Alexandre de Kochko,[7] Giorgio Graziosi,[21,25] Robert J Henry,[26] Jayarama,[27] Ray Ming,[17] Chifumi Nagai,[28] Steve Rounsley,[29] David Sankoff,[9] Giovanni Giuliano,[8] Victor A. Albert,[4]* Patrick Wincker,[1,2,3]* Philippe Lashermes[5]*

Coffee is a valuable beverage crop due to its characteristic flavor, aroma, and the stimulating effects of caffeine. We generated a high-quality draft genome of the species *Coffea canephora*, which displays a conserved chromosomal gene order among asterid angiosperms. Although it shows no sign of the whole-genome triplication identified in Solanaceae species such as tomato, the genome includes several species-specific gene family expansions, among them *N*-methyltransferases (NMTs) involved in caffeine production, defense-related genes, and alkaloid and flavonoid enzymes involved in secondary compound synthesis. Comparative analyses of caffeine NMTs demonstrate that these genes expanded through sequential tandem duplications independently of genes from cacao and tea, suggesting that caffeine in eudicots is of polyphyletic origin.

With more than 2.25 billion cups consumed every day, coffee is one of the most important crops on Earth, cultivated across more than 11 million hectares. Coffee belongs to the Rubiaceae family, which is part of the Euasterid I clade and the fourth largest family of angiosperms, consisting of more than 11,000 species in 660 genera (*1*). We sequenced *Coffea canephora* ($2n = 2x = 22$ chromosomes), an outcrossing, highly heterozygous diploid, and one of the parents of *C. arabica* ($2n = 4x = 44$ chromosomes), which was derived from hybridization between *C. canephora* and *C. eugenioides* (*2*). A total of 54.4 million Roche 454 single and mate-pair reads and 143,605 Sanger bacterial artificial chromosome–end reads were generated from a doubled haploid accession, representing ~30× coverage of the 710-Mb genome (*3*). Additional Illumina sequencing data (60×) were used to improve the assembly (table S1) (*4*). The resulting assembly consists of 25,216 contigs and 13,345 scaffolds with a total length of 568.6 Mb (80% of 710 Mb), including 97 Mb (17%) of intercontig gaps. Eighty percent of the assembly is in 635 scaffolds, and the scaffold N50 (the scaffold size above which 50% of the total length of the sequence assembly can be found) is 1.26 Mb (table S2). A high-density genetic map covering 349 scaffolds and comprising ~64% of the assembly (364 Mb) and 86% of the annotated genes was anchored to the 11 *C. canephora* chromosomes (*4*). More than 96% of the scaffolds larger than 1 Mb were anchored (Fig. 1A).

We annotated 25,574 protein-coding genes (*4*) (table S6), 92 microRNA precursors, and 2573 organellar-to-nuclear genome transfers (*4*). Transposable elements account for ~50% of the genome (*4*), of which ~85% are long terminal repeat (LTR) retrotransposons. Large-scale comparison between *C. canephora* LTR retrotransposons and those of reference plant genomes shows outstanding conservation of several *Copia* groups across distantly related genomes, suggesting that horizontal mobile element transfers may be more frequent than generally recognized (*5*–*8*).

Structurally, the coffee genome shows no sign of a whole-genome polyploidization in its lineage since the γ triplication at the origin of the core eudicots (*9*) (Fig. 1B). Coffee contains exactly three paralogous regions for each of the seven pre-γ ancestral chromosomes (Fig. 1B). Coffee chromosomal regions show unique one-to-one correspondences with grapevine chromosomes (Fig. 1C and fig. S12) and a one-to-three correspondence with the tomato genome, which underwent a second lineage-specific triplication during its evolutionary history (*10*). Although grapevine, a rosid, is the most conservative core eudicot in terms of integrity of gross chromosomal structure, coffee displays less gene-order divergence to all other rosids, despite being an asterid itself (*9*). Coffee also shows little syntenic divergence relative to other sequenced asterids (Fig. 1D, table S17, and supplementary text).

To classify gene families in the *C. canephora* genome, we ran OrthoMCL on inferred protein sequences from coffee, grapevine, tomato, and *Arabidopsis* (*4*), generating 16,917 groups of orthologous genes (fig. S5). To examine coffee-specific gene family expansions with potential adaptive value, we fit different branch models implemented in BadiRate (*11*) to these orthogroups (*4*). In the coffee lineage, 202 orthogroups

[1]Commissariat à l'Energie Atomique, Genoscope, Institut de Génomique, BP5706, 91057 Evry, France. [2]CNRS, UMR 8030, CP5706, Evry, France. [3]Université d'Evry, UMR 8030, CP5706, Evry, France. [4]Department of Biological Sciences, 109 Cooke Hall, University at Buffalo (State University of New York), Buffalo, NY 14260, USA. [5]Institut de Recherche pour le Développement (IRD), UMR Résistance des Plantes aux Bioagresseurs (RPB) [Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), IRD, UM2)], BP 64501, 34394 Montpellier Cedex 5, France. [6]CIRAD, UMR Amélioration Génétique et Adaptation des Plantes Méditerranéennes et Tropicales (AGAP), F-34398 Montpellier, France. [7]IRD, UMR Diversité Adaptation et Développement des Plantes (CIRAD, IRD, UM2), BP 64501, 34394 Montpellier Cedex 5, France. [8]Italian National Agency for New Technologies, Energy and Sustainable Development (ENEA) Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy. [9]Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa, Ontario K1N 6N5, Canada. [10]Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier Cedex 5, France. [11]Nestlé Research and Development Centre, 101 Avenue Gustave Eiffel, Notre-Dame-d'Oé, BP 49716, 37097 Tours Cedex 2, France. [12]ENEA Trisaia Research Center, 75026 Rotondella, Italy. [13]Center for Biotechnology, Universität Bielefeld, Universitätsstraße 27, D-33615 Bielefeld, Germany. [14]AG Genominformatik, Technische Fakultät, Universität Bielefeld, 33594 Bielefeld, Germany. [15]Institut National de la Recherche Agronomique (INRA), Unité de Recherches en Génomique-Info (UR INRA 1164), Centre de Recherche de Versailles, 78026 Versailles Cedex, France. [16]Department of Biology, Chongqing University of Science and Technology, 4000042 Chongqing, China. [17]Department of Plant Biology, 148 Edward R. Madigan Laboratory, MC-051, 1201 West Gregory Drive, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [18]Departament de Genètica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Diagonal 643, Barcelona 08028, Spain. [19]Department of Mathematics, University of Maryland, Mathematics Building 084, University of Maryland, College Park, MD 20742, USA. [20]School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Avenue, Ottawa, Ontario K1N 6N5, Canada. [21]Department of Life Sciences, University of Trieste, Via Licio Giorgieri 5, 34127 Trieste, Italy. [22]Indonesian Coffee and Cocoa Institute, Jember, East Java, Indonesia. [23]Laboratório de Genética Molecular, Núcleo de Biotecnologia (NTBio), Embrapa Recursos Genéticos e Biotecnologia, Final Av. W/5 Norte, Parque Estação Biológia, Brasília-DF 70770-917, Brazil. [24]CIRAD, UMR RPB (CIRAD, IRD, UM2), BP 64501, 34394 Montpellier Cedex 5, France. [25]DNA Analytica Srl, Via Licio Giorgieri 5, 34127 Trieste, Italy. [26]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia 4072, Australia. [27]Central Coffee Research Institute, Coffee Board, Coffee Research Station (Post) - 577 117 Chikmagalur District, Karnataka State, India. [28]Hawaii Agriculture Research Center, Post Office Box 100, Kunia, HI 96759-0100, USA. [29]BIO5 Institute, University of Arizona, 1657 Helen Street, Tucson, AZ 85721, USA.
*Corresponding author. E-mail: vaalbert@buffalo.edu (V.A.A.); pwincker@genoscope.cns.fr (P.W.); philippe.lashermes@ird.fr (P.L.)

Corrected 8 October, 2014; see full text.

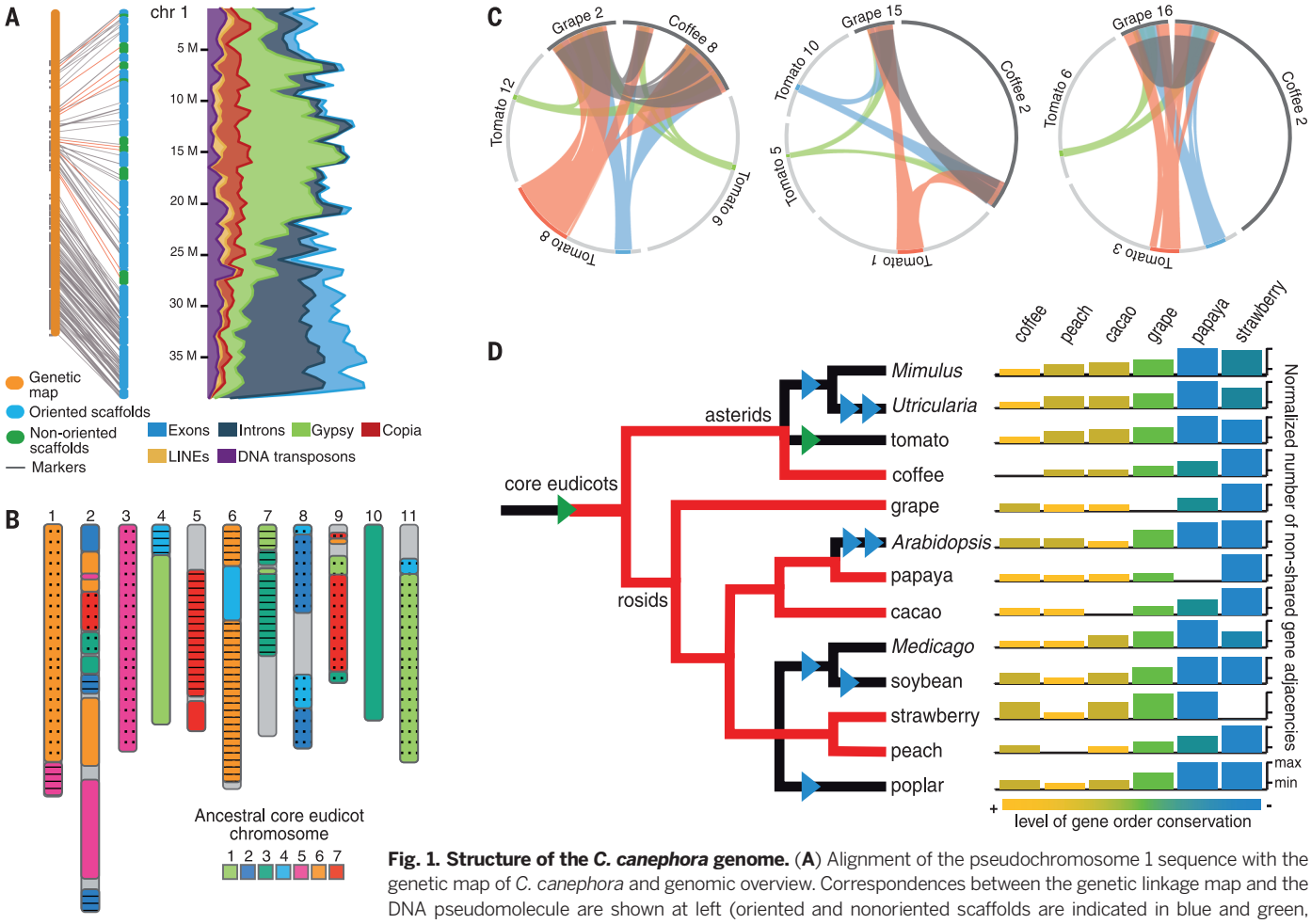

Fig. 1. Structure of the *C. canephora* genome. (**A**) Alignment of the pseudochromosome 1 sequence with the genetic map of *C. canephora* and genomic overview. Correspondences between the genetic linkage map and the DNA pseudomolecule are shown at left (oriented and nonoriented scaffolds are indicated in blue and green, respectively; gray lines denote consistent data; orange lines indicate markers with an approximate genetic location). The relative proportions (percentage of nucleotides) in sliding windows (1-Mb size, 500-kb step) of transposable elements (*Copia* in red, *Gypsy* in green) and genes (exons in blue, introns in dark blue) are shown at right. (**B**) Coffee chromosomal blocks descending from the seven ancestral core eudicot chromosomes. The three paralogous descendants of the seven ancestral chromosomes are shown in shared colors but different textures. (**C**) Comparison of three grapevine chromosomes (descendants of the prehexaploidization core eudicot chromosome) mapped to a single coffee chromosome and three regions in the tomato genome. (**D**) Phylogeny and genome duplication history of core eudicots. Arrowheads indicate tetraploidization (blue) or hexaploidization (green) events. Red lines trace lineages of six species that have not undergone further polyploidization. Bar graphs and colors reflect gene-order differences (table S17) between each of the six species (column labels) and the entire set, showing the gene order conservatism of coffee, especially among asterids, and of peach and cacao among rosids.

clustering 1270 genes were supported as expanded (Akaike information criterion > 2.7). Among gene ontology (GO) terms annotating these, 98 out of 4300 generic terms were significantly over- or underrepresented (table S14). Most GOs enriched in *C. canephora* ($P < 0.05$) belonged to two main functional categories: defense response and metabolic process, the later including different catalytic activities (table S15).

Among defense response functions, there is a clear expansion of nucleotide binding site disease-resistance genes (*12*, *13*) in the *C. canephora* genome (*4*). Most genes that grouped together within single orthogroups were tandemly arrayed, suggesting that *R* genes evolved by tandem duplication and divergence of linked gene families (supplementary text). Several gene functions involved in secondary metabolite biosynthesis are significantly expanded in the *C. canephora* genome, including enzymes associated with the production of phenylpropanoids such as flavonoids and isoflavones (naringenin 3-dioxygenase, isoflavone 2′-hydroxylase), alkaloids (strictosidine synthase, tropine dehydrogenase), monoterpenes (e.g., menthol dehydrogenase), and caffeine [*N*-methyltransferases (NMTs)] (Fig. 2). For example, indole alkaloids such as the monoamine oxidase inhibitor yohimbine and antimalaria drug quinine are prominent secondary compounds of the coffee family and its parent order, Gentianales (*14*), and the GO term indole biosynthetic process was highly enriched ($P < 0.001$) in coffee relative to tomato, grapevine, and *Arabidopsis*.

Caffeine is a purine alkaloid synthesized by several eudicot plants, including coffee, cacao (*Theobroma cacao*), and tea (*Camellia sinensis*) (Fig. 2). Caffeine is synthesized in both coffee leaves, where it has insecticidal properties (*15*), and fruits and seeds, where it inhibits seed germination of competing species (*16*). The late steps in caffeine biosynthesis are mediated by a series of NMTs (Fig. 2A) (*17*).

Among coffee-expanded genes, NMT activity is one of the more highly enriched GO terms (table S15). A single gene family (ORTHOMCL170) clusters 23 genes in coffee, but none in grapevine, tomato, or *Arabidopsis* (table S12), and this cluster contains genes encoding known enzymes of the caffeine biosynthetic pathway (*18*, *19*). Maximum likelihood (ML) phylogenetic analysis of ORTHOMCL170 with tea and cacao NMTs that have similar activities reveals species-specific gene clades (Fig. 2C). We analyzed these relationships in a broader evolutionary context by including genome-wide samples of NMTs from coffee, cacao, and other eudicot species. ML trees show that the genes encoding the closest *Arabidopsis* NMT relatives of coffee caffeine biosynthetic enzymes are involved in benzoic, salicylic, and nicotinic functions (*4*) (supplementary text). Caffeine biosynthetic NMTs from coffee nested within a gene clade distinct from those of cacao or tea, which group together as sister lineages. Thus, a

Corrected 8 October, 2014; see full text.

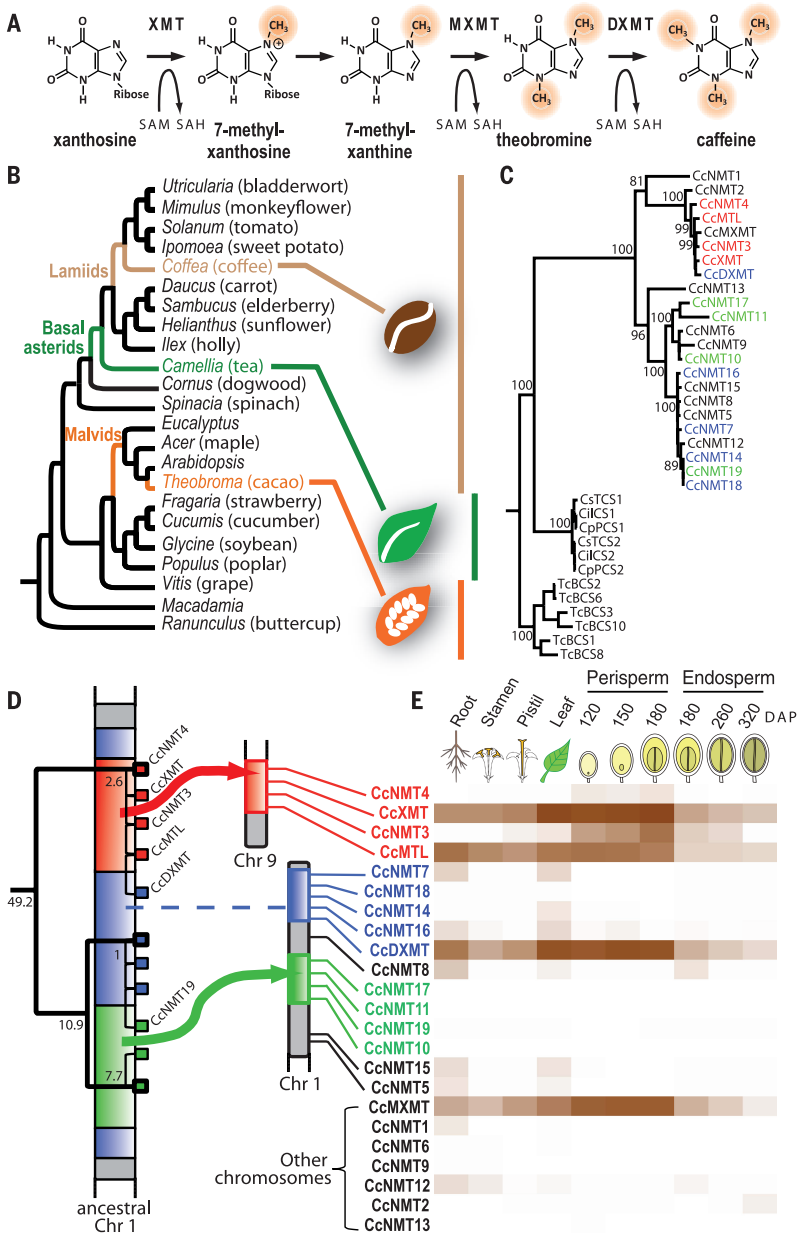

Fig. 2. Evolution of caffeine biosynthesis. (**A**) The principal caffeine biosynthetic pathway. Three methylation steps are necessary to produce caffeine from xanthosine, involving the successive action of three NMTs: xanthosine methyltransferase (XMT), theobromine synthase [7-methylxanthine methyltransferase (MXMT)], and caffeine synthase [3,7-dimethylxanthine methyltransferase (DXMT)]. SAM, *S*-adenosylmethionine; SAH, *S*-adenosylhomocysteine. (**B**) Evolutionary position of caffeine-producing plants with respect to other eudicots (phylogeny adapted from www.mobot.org/MOBOT/research/APweb/). (**C**) ML phylogeny of coffee, tea, and cacao NMTs. Bootstrap support values (percentages) from 1000 replicates are shown next to relevant clades. Branch lengths are proportional to expected numbers of nucleotide substitutions per site. Colors identify genes assignable to the genomic blocks denoted in (D). (**D**) (Left) A model summarizing the duplication history of coffee NMT genes, following the phylogeny in (C). Three distinct tandem gene arrays evolved in situ on chromosome 1 from nearby gene duplicates (bold squares). The red and green blocks, colored as in (C), translocated (to chromosome 9) or rearranged (to elsewhere on chromosome 1) from their ancestral locus (blue region), respectively. (Right) Gene orders on modern chromosomes. Translocation of the red block, containing the putative caffeine NMT metabolic cluster, left the phylogenetically derived *CcDXMT* gene behind. Similarly, *CcNMT19* is a derived gene within its own NMT clade that remained in place following movement of the green block. Numbers at branches indicate relative times since major duplication events or diversification times of the tandem arrays, calculated from approximately neutral synonymous substitution rates. (**E**) Expression profiles (reads per kilobase per million reads mapped) of known *Coffea canephora* NMTs. The genes in the putative metabolic cluster (along with *CcDXMT* and *CcMXMT*) exhibit similar expression patterns, higher in perisperm than endosperm. Data are plotted as log2 values. DAP, days after pollination.

minimum of two independent origins of caffeine biosynthetic NMT activity can be inferred, as proposed previously (*20*).

Microsynteny analyses of ORTHOMCL170, which includes three tandem arrays, show that some known and putative coffee caffeine synthase genes—*CcXMT* (encoding xanthosine *N*-methyltransferase), *CcMTL*, and *CcNMT3*—form a tight assemblage of coexpressed tandem duplicates (Fig. 2D) reminiscent of a metabolic gene cluster (*21*, *22*). Given that some plant metabolic gene clusters are of relatively recent origin (*23*), we sought to further unravel the role of gene duplication in the expansion of the coffee NMT gene family (Fig. 2D) (supplementary text). The three main coffee NMT clades in ORTHOMCL170 are distributed among a minimum of three genomic blocks; however, some phylogenetically recent tandem duplicates have moved away from their original positions via block rearrangements (Fig. 2D). One such movement involving the putative metabolic cluster appears to have left the *CcDXMT* gene (encoding 3,7-dimethylxanthine methyltransferase) behind, physically separated from its ancestral tandem array. In cacao, the functionally characterized *TcBCS1* gene has a tandem duplicate, but this pair of genes evolved independently from the NMT tandem arrays found in *C. canephora* (fig. S29). We also examined the role of positive selection (PS) in the evolution of caffeine biosynthesis among coffee, tea, and cacao (*4*) (supplementary text). We found significant evidence for PS [likelihood ratio test for PAML (Phylogenetic Analysis by Maximum Likelihood) branch-site test, $P = 5.78 \times 10^{-3}$ (*24*)] only for the coffee NMT lineage, indicating that the independent evolution of caffeine biosynthesis in coffee was adaptive and probably involved specific amino acid changes fixed by PS. These results highlight the distinct acquisition of caffeine biosynthesis in the coffee plant, providing an example of convergent evolution of secondary metabolic pathways encoded by tandemly duplicated genes.

Genomic functional diversification via tandem duplication may have helped shape other aspects of coffee bean chemical composition. Linoleic acid, which is produced by the oleate desaturase FAD2, is the major polyunsaturated fatty acid in the coffee bean (*25*, *26*), where it contributes to aroma composition and flavor retention after roasting (*4*). Coffee has six *FAD2* genes compared with one in *Arabidopsis*, and most of these have arisen from tandem duplications on chromosome 1 (fig. S33). RNA sequencing data suggest transcriptional specialization for two of the six *FAD2* copies, with *CcFAD2.3* being actively

Corrected 8 October, 2014; see full text.

transcribed in developing endosperm (supplementary text). Peak transcript abundance coincides with the dramatic increase in linoleic acid content that occurs during seed development at the perisperm-endosperm transition (*27*).

Our analysis of the adaptive genomic landscape of *C. canephora* identifies the convergent evolution of caffeine biosynthesis among plant lineages and establishes coffee as a reference species for understanding the evolution of genome structure in asterid angiosperms.

REFERENCES AND NOTES

1. E. Robbrecht, J. F. Manen, *Syst. Geogr. Plants* **76**, 85–146 (2006).
2. P. Lashermes *et al*., *Mol. Gen. Genet.* **261**, 259–266 (1999).
3. M. Noirot *et al*., *Ann. Bot. (London)* **92**, 709–714 (2003).
4. Materials and methods are available as supplementary materials on *Science* Online.
5. S. Schaack, C. Gilbert, C. Feschotte, *Trends Ecol. Evol.* **25**, 537–546 (2010).
6. A. Roulin *et al*., *BMC Evol. Biol.* **9**, 58 (2009).
7. M. El Baidouri *et al*., *Genome Res.* **24**, 831–838 (2014).
8. C. Moisy, A. H. Schulman, R. Kalendar, J. P. Buchmann, F. Pelsy, *Theor. Appl. Genet.* **127**, 1223–1235 (2014).
9. O. Jaillon *et al*., *Nature* **449**, 463–467 (2007).
10. S. Sato *et al*., *Nature* **485**, 635–641 (2012).
11. P. Librado, F. G. Vieira, J. Rozas, *Bioinformatics* **28**, 279–281 (2012).
12. S. H. Hulbert, C. A. Webb, S. M. Smith, Q. Sun, *Annu. Rev. Phytopathol.* **39**, 285–312 (2001).
13. L. McHale, X. Tan, P. Koehl, R. W. Michelmore, *Genome Biol.* **7**, 212 (2006).
14. F. Gleason, R. Chollet, *Plant Biochemistry* (Jones and Bartlett, Sudbury, MA, 2011).
15. J. A. Nathanson, *Science* **226**, 184–187 (1984).
16. A. Pacheco, J. Pohlan, M. Schulz, *Allelopathy J.* **21**, 39–56 (2008).
17. H. Ashihara, H. Sano, A. Crozier, *Phytochemistry* **69**, 841–856 (2008).
18. A. A. McCarthy, J. G. McCarthy, *Plant Physiol.* **144**, 879–889 (2007).
19. M. Ogawa, Y. Herai, N. Koizumi, T. Kusano, H. Sano, *J. Biol. Chem.* **276**, 8213–8218 (2001).
20. E. Pichersky, E. Lewinsohn, *Annu. Rev. Plant Biol.* **62**, 549–566 (2011).
21. B. Field, A. E. Osbourn, *Science* **320**, 543–547 (2008).
22. M. Matsuno *et al*., *Science* **325**, 1688–1692 (2009).
23. B. Field *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16116–16121 (2011).
24. J. Zhang, R. Nielsen, Z. Yang, *Mol. Biol. Evol.* **22**, 2472–2479 (2005).
25. D. Villarreal *et al*., *J. Agric. Food Chem.* **57**, 11321–11327 (2009).
26. S. Dussert, A. Laffargue, A. de Kochko, T. Joët, *Phytochemistry* **69**, 2950–2960 (2008).
27. T. Joët *et al*., *New Phytol.* **182**, 146–162 (2009).

ACKNOWLEDGMENTS

We acknowledge the following sources for funding: ANR-08-GENM-022-001 (to P.L.); ANR-09-GENM-014-002 (to P.W.); Australian Research Council (to R.J.H.); Natural Sciences and Engineering Research Council of Canada (to D.S.); CNR-ENEA Agrifood Project A2 C44 L191 (to G.Gi.); FINEP-Qualicafé, INCT-CAFÉ (to A.C.A.); NSF grants 0922742 (to V.A.A.) and 0922545 (to R.M.); and the College of Arts and Sciences, University at Buffalo (to V.A.A.). We thank P. Facella (ENEA) for Roche 454 sequencing and Instituto Agronômico do Paraná (Paraná, Brazil) for fruit RNA. This work was supported by the high-performance cluster of the SouthGreen Bioinformatics platform (UMR AGAP) CIRAD (www.southgreen.fr). The *C. canephora* genome assembly and gene models are available on the Coffee Genome Hub (http://coffee-genome.org) and the CoGe platform (www.genomevolution.org). Sequencing data are deposited in the European Nucleotide Archive under the accession numbers CBUE020000001 to CBUE020025216 (contigs), HG739085 to HG752429 (scaffolds), and HG974428 to HG974439 (chromosomes). Gene family alignments and phylogenetic trees for BAHD acyltransferases and NMTs are available in the GreenPhylDB (www.greenphyl.org/cgi-bin/index.cgi) under the gene family IDs CF158535 and CF158539 to CF158545, respectively. We declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6201/1181/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S33
Tables S1 to S27
References (*28*–*175*)

28 April 2014; accepted 29 July 2014
10.1126/science.1255274

GENOME EDITING

Prevention of muscular dystrophy in mice by CRISPR/Cas9–mediated editing of germline DNA

Chengzu Long,[1]* John R. McAnally,[1]* John M. Shelton,[2] Alex A. Mireault,[1] Rhonda Bassel-Duby,[1] Eric N. Olson[1]†

Duchenne muscular dystrophy (DMD) is an inherited X-linked disease caused by mutations in the gene encoding dystrophin, a protein required for muscle fiber integrity. DMD is characterized by progressive muscle weakness and a shortened life span, and there is no effective treatment. We used clustered regularly interspaced short palindromic repeat/Cas9 (CRISPR/Cas9)–mediated genome editing to correct the dystrophin gene (*Dmd*) mutation in the germ line of *mdx* mice, a model for DMD, and then monitored muscle structure and function. Genome editing produced genetically mosaic animals containing 2 to 100% correction of the *Dmd* gene. The degree of muscle phenotypic rescue in mosaic mice exceeded the efficiency of gene correction, likely reflecting an advantage of the corrected cells and their contribution to regenerating muscle. With the anticipated technological advances that will facilitate genome editing of postnatal somatic cells, this strategy may one day allow correction of disease-causing mutations in the muscle tissue of patients with DMD.

Duchenne muscular dystrophy (DMD) is caused by mutations in the gene for dystrophin on the X chromosome and affects approximately 1 in 3500 boys. Dystrophin is a large cytoskeletal structural protein essential for muscle cell membrane integrity. Without it, muscles degenerate, causing weakness and myopathy (*1*). Death of DMD patients usually occurs by age 25, typically from breathing complications and cardiomyopathy. Hence, therapy for DMD necessitates sustained rescue of skeletal, respiratory, and cardiac muscle structure and function. Although the genetic cause of DMD was identified nearly three decades ago (*2*), and several gene- and cell-based therapies have been developed to deliver functional *Dmd* alleles or dystrophin-like protein to diseased muscle tissue, numerous therapeutic challenges have been encountered, and no curative treatment exists (*3*).

RNA-guided, nuclease-mediated genome editing, based on type II CRISPR (clustered regularly interspaced short palindromic repeat)/Cas (CRISPR-associated) systems, offers a new approach to alter the genome (*4*–*6*). In brief, Cas9, a nuclease guided by single-guide RNA (sgRNA), binds to a targeted genomic locus next to the protospacer adjacent motif (PAM) and generates a double-strand break (DSB). The DSB is then repaired either by nonhomologous end-joining (NHEJ), which leads to insertion/deletion (indel) mutations, or by homology-directed repair (HDR), which requires an exogenous template and can generate a precise modification at a target locus (*7*). Unlike other gene therapy methods, which add a functional, or partially functional, copy of a gene to a patient's cells but retain the original dysfunctional copy of the gene, this system can remove the defect. Genetic correction using engineered nucleases (*8*–*12*) has been demonstrated in immortalized myoblasts derived from DMD patients in vitro (*9*), and rodent models of rare diseases (*13*), but not yet in animal models of relatively common and currently incurable diseases, such as DMD.

The objective of this study was to correct the genetic defect in the *Dmd* gene of *mdx* mice by CRISPR/Cas9–mediated genome editing in vivo. The *mdx* mouse (C57BL/10ScSn-*Dmd*^{mdx}/J) contains a nonsense mutation in exon 23 of the *Dmd* gene (*14*, *15*) (Fig. 1A). We injected Cas9, sgRNA, and HDR template into mouse zygotes to correct the disease-causing gene mutation in the germ line (*16*, *17*), a strategy that has the potential to correct the mutation in all cells of the body, including myogenic progenitors. Safety and efficacy of CRISPR/Cas9–based gene therapy was also evaluated.

[1]Department of Molecular Biology and Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. *These authors contributed equally to this work. †To whom correspondence should be addressed. E-mail: eric.olson@utsouthwestern.edu

Corrected 8 October, 2014; see full text.

The coffee genome provides insight into the convergent evolution of caffeine biosynthesis

France Denoeud *et al.*

Science **345**, 1181 (2014);
DOI: 10.1126/science.1255274

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.