

## PLANT SCIENCE

# Mutation of a bHLH transcription factor allowed almond domestication

R. Sánchez-Pérez<sup>1,2,3\*</sup>†, S. Pavan<sup>4,5\*</sup>†, R. Mazzeo<sup>2,3,4</sup>, C. Moldovan<sup>2,3</sup>,  
R. Aiese Cigliano<sup>6</sup>, J. Del Cueto<sup>1,2,3,7</sup>, F. Ricciardi<sup>2,3,8</sup>, C. Lotti<sup>8</sup>, L. Ricciardi<sup>4</sup>,  
F. Dicenta<sup>1</sup>, R. L. López-Marqués<sup>9</sup>, B. Lindberg Møller<sup>2,3</sup>

Wild almond species accumulate the bitter and toxic cyanogenic diglucoside amygdalin. Almond domestication was enabled by the selection of genotypes harboring sweet kernels. We report the completion of the almond reference genome. Map-based cloning using an F<sub>1</sub> population segregating for kernel taste led to the identification of a 46-kilobase gene cluster encoding five basic helix-loop-helix transcription factors, bHLH1 to bHLH5. Functional characterization demonstrated that bHLH2 controls transcription of the P450 monooxygenase-encoding genes *PdCYP79D16* and *PdCYP71AN24*, which are involved in the amygdalin biosynthetic pathway. A nonsynonymous point mutation (Leu to Phe) in the dimerization domain of bHLH2 prevents transcription of the two cytochrome P450 genes, resulting in the sweet kernel trait.

Almond [*Prunus dulcis* Miller (D. A. Webb), syn. *Prunus amygdalus* L.] is the main tree nut species worldwide, with a cultivated area of about 1.9 million ha and an annual in-shell production exceeding 2.2 million metric tons (FAOSTAT 2017, [www.fao.org/faostat/en/#data/QC](http://www.fao.org/faostat/en/#data/QC)). The spatiotemporal framework of almond domestication is still controversial, although archaeobotany and genetic studies suggest it occurred in the Fertile Crescent—the “Cradle of Civilization”—during the first half of the Holocene (1–4). Cultivation in the Mediterranean Basin is documented by the presence of almonds among the artifacts found in historical places such as the tomb of Tutankhamen (5) and the Franchthi Cave in Greece (6). Most recently, almond was introduced in suitable habitats in America (primarily California) and the Southern Hemisphere. In addition to almond, the botanical family of Rosaceae encompasses an extraordinary variety of species of great economic importance in temperate regions, including apple (*Malus domestica*), apricot (*Prunus armeniaca*), Japanese apricot (*Prunus mume*), peach (*Prunus persica*), strawberry (*Fragaria vesca*), and sweet

cherry (*Prunus avium*) (7, 8). Among these, almond is unique in that the kernel (seed), rather than the flesh (mesocarp), represents the edible commercial product.

Kernels of wild almond species are bitter and highly toxic to humans and predators because the cyanogenic diglucoside amygdalin accumulates in the cotyledons (9–12). The key event enabling almond domestication was the selection of sweet kernel genotypes. Genetic studies showed that edible sweet almond kernels originate from a dominant mutation (13, 14) within the almond linkage group (LG) 5, at a locus referred to as *Sweet kernel* (*Sk*) (15, 16). The nature of the *Sk* gene has remained elusive.

Amygdalin is produced from the cyanogenic monoglucoside prunasin (9), which is biosynthesized in the seed coat (tegument). Recently, it was shown that the first two genes in the prunasin biosynthetic pathway, *PdCYP79D16* and *PdCYP71AN24*, are poorly expressed in the tegument of sweet genotypes relative to bitter

genotypes (12). The almond genome harbors an additional CYP79-encoding gene, named *PdCYP79A68* (12), a putative ortholog of the *P. mume* gene *PmCYP79A68* (17). *PmCYP79A68* was shown to have weak activity toward tryptophan, and it did not catalyze conversion of phenylalanine into the corresponding oxime, as required for a function in prunasin and amygdalin synthesis (17).

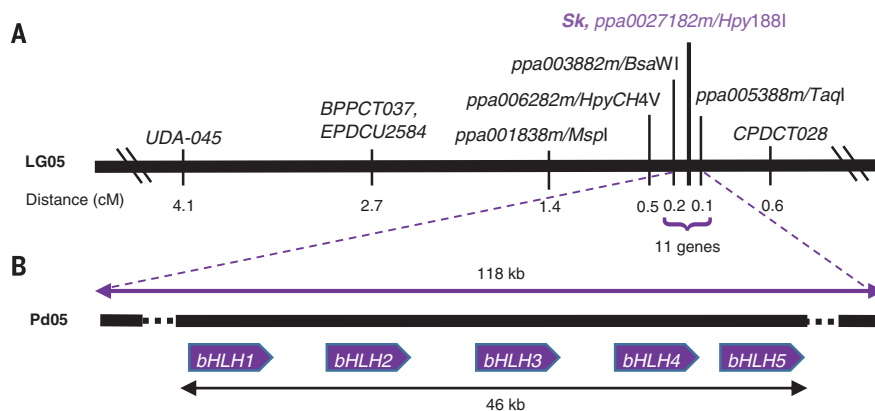
Despite the economic importance of almond, genomic resources in almond are limited relative to those available for other Rosaceae species (18–22). Here, we report the assembly of the draft genome of almond. The sequence information was used to unravel the genetic differences between bitter and sweet kernel genotypes. The results provide insights into almond domestication history, as well as a sequence inventory for studies of other almond agronomic traits such as flowering time, drought tolerance, and resistance to disease.

The genome ( $2n = 2x = 16$ , haploid genome size = 246 Mb) of the sweet homozygous almond cultivar Lauranne was sequenced with a combination of Illumina (paired-end and 5-kb mate pairs) and PacBio technologies. The long PacBio reads were used to perform the genome assembly, whereas the Illumina reads were used to perform gap filling and scaffolding. The scaffolds obtained were organized in pseudomolecules using the *P. persica* reference genome, as the genomes across the *Prunus* genus are essentially collinear (23–25). The final assembly constitutes 4078 scaffolds, of which 2572 are organized in eight pseudomolecules, with a final N50 of 21.8 Mb (where N50 is the minimum contig length needed to cover 50% of the genome) and a L90 of 306 (where L90 is the smallest number of contigs whose length sum makes up 90% of genome size) (table S1). A high level of collinearity between the almond genomic assembly and the recently reported single-nucleotide polymorphism (SNP)/short sequence repeat (SSR) genetic linkage maps (25–27) was demonstrated (fig. S1).

By combining ab initio, RNA sequencing (RNA-seq), and transcript sequences from *P. persica* and the Viridiplantae Swissprot database, the

<sup>1</sup>Department of Plant Breeding, CEBAS-CSIC, Campus Universitario de Espinardo, 30100 Espinardo, Spain. <sup>2</sup>Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark. <sup>3</sup>VILLUM Research Center for Plant Plasticity, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark. <sup>4</sup>Department of Soil, Plant and Food Science, University of Bari “Aldo Moro,” 70126 Bari, Italy. <sup>5</sup>Institute of Biomedical Technologies, National Research Council (CNR), 70126 Bari, Italy. <sup>6</sup>Sequentia Biotech SL, Campus de la UAB, 08193 Bellaterra, Barcelona, Spain. <sup>7</sup>Arboriculture Research Group, Agroscope, Conthey, Switzerland. <sup>8</sup>Department of the Sciences of Agriculture, Food and Environment, University of Foggia, 71100 Foggia, Italy. <sup>9</sup>Transport Biology Section, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark.

\*Corresponding author. Email: [rsanchez@cebas.csic.es](mailto:rsanchez@cebas.csic.es) (R.S.-P.); [stefano.pavan@uniba.it](mailto:stefano.pavan@uniba.it) (S.P.) †These authors contributed equally to this work.



**Fig. 1. Map-based cloning of the *Sk* gene.** (A and B) Comparison of linkage group 5 (LG05) (A) and pseudomolecule 5 (Pd05) (B) in the region where the *Sk* locus is mapped. Approximate positions of the five *bHLH* genes present within that region are shown.

*P. dulcis* genome was predicted to contain 27,817 genes (table S2), of which 16,747 had an annotation edit distance (AED)  $\leq 0.3$ . A total of 1080 conserved plant genes were identified among the annotated proteins with the BUSCO tool (28) (table S3). Of these, 95% were complete and present in a single copy, confirming the quality of the assembly and annotations. Functional annotation of the genes was then performed by searching for orthologous sequences in *P. persica*, *Arabidopsis thaliana*, and *P. mume*. In this way, a Gene Ontology term was assigned to 12,777 genes, whereas a description was given to 24,035 genes. About 34.6% of the genome was found to be repetitive, including 11.1% of long-term repeat elements (table S4).

Using a set of *Sk*-linked cleaved amplified polymorphic sequence (CAPS) and SSR markers within the almond LG5 (29), together with two *Sk*-linked CAPS markers developed in this study (table S5 and fig. S2), a large mapping population of 475 F<sub>1</sub> individuals segregating at the *Sk* locus was genotyped. The resulting map displayed the two markers ppa003882m/*Bsa*WI and ppa005388m/*Taq*I flanking the *Sk* locus, separated by 0.2 cM and 0.1 cM, respectively (Fig. 1A). Alignment of the two marker sequences against the assembled almond genome sequence allowed the definition of a physical interval of 118 kb, harboring 11 genes (Fig. 1B). A 46-kb cluster of five genes encoding putative basic helix-loop-helix (bHLH) transcription factors (*bHLH1* to *bHLH5*) was positioned within this interval. We investigated these *bHLH* genes for their ability to control amygdalin accumulation because (i) sweet and bitter almond genotypes are associated with markedly different transcriptional levels of the amygdalin biosynthetic genes *PdCYP79D16* and *PdCYP71AN24* (12), as confirmed by our RNA-seq experiment (Fig. 2); (ii) bHLH transcription factors are known to regulate the biosynthesis of several bioactive natural products (30); and (iii) we identified several bHLH-binding DNA matrices in the *PdCYP79D16* and *PdCYP71AN24* promoter regions (data S1).

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) time-course experiments on tegument tissues of the sweet cultivar Lauranne (*Sk/Sk*) and the bitter genotype S3067 (*sk/sk*) did not reveal differential expression for *bHLH1*, *bHLH2*, and *bHLH4* (fig. S3 and data S2); expression of *bHLH3* and *bHLH5* was not detected. A scan of a 20-kb genomic region associated with the *Sk* locus revealed higher heterozygosity in the bitter genome than in the sweet genome (fig. S4).

The expressed *bHLH* genes (*bHLH1*, 2, and 4) were sequenced in Lauranne and S3067 (fig. S5). In *bHLH1*, this highlighted an indel polymorphism of 161 bp leading to a truncated protein (262 versus 401 amino acids) in the sweet genotype. In *bHLH2*, a C/T SNP and a 3-bp indel were observed, resulting in a nonsynonymous Leu<sup>346</sup> → Phe (L346F) substitution and an Asn<sup>412</sup> insertion in the sweet genotype. In *bHLH4*, no polymorphism was found.

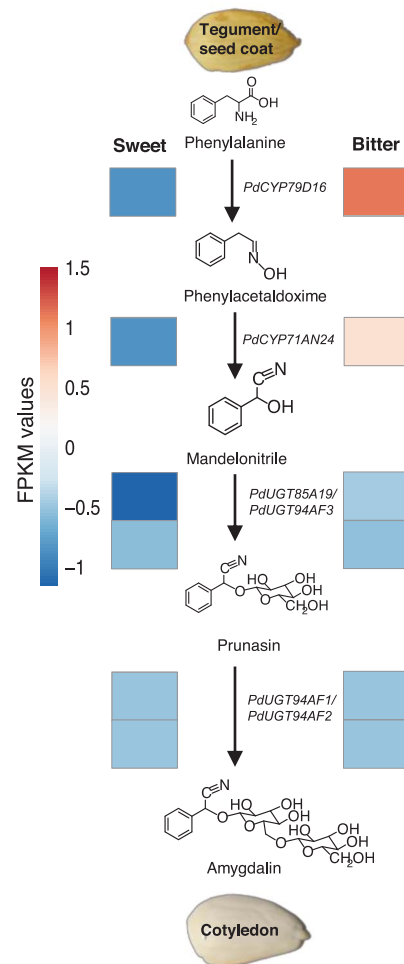
Functional characterization of *bHLH1* and *bHLH2* was carried out using a factorial transient  $\beta$ -glucuronidase (GUS) expression assay in *Nicotiana benthamiana*. The alleles were used in combination with the two *PdCYP79D16* promoter regions cloned from Lauranne and S3067 (henceforth referred to as *p79sweet* and *p79bitter*) and the three promoter regions cloned from *PdCYP71AN24* (henceforth referred to as *p71sweet1*, *p71sweet2*, and *p71bitter*) (Fig. 3). No significant differences were observed in GUS activity when expressing the two *bHLH1* alleles (Fig. 3A). In contrast, a significant induction of GUS activity was observed for the *bHLH2* bitter allele in combination with all versions of *p79* and *p71* (Fig. 3B). This indicated that *bHLH2* could be the *Sk* gene. DNA sequencing of *bHLH2* in 56 almond genotypes demonstrated cosegregation between the *Sk/sk* alleles and the C/T polymorphism in all cases except one (table S6), represented by the cultivar Atocha. However, Atocha displayed a different SNP positioned nearby, resulting in a Leu<sup>330</sup> → Arg (L330R) nonsynonymous change (fig. S6).

The L346F and L330R nonsynonymous mutations observed in sweet genotypes were both predicted to occur in the dimerization interphase typical of bHLH transcription factors (fig. S5). To study the possible functional consequences of the L346F mutation, we carried out direct mutagenesis of the coding sequence of the two *bHLH2* alleles, substituting the Leu<sup>346</sup> residue with Phe and vice versa. In addition, the last 151 amino acids of the two *bHLH2* alleles (henceforth called the “tail”) were also exchanged to investigate whether the extra Asn residue occurring at position 412 in the sweet genotype of bHLH2 influenced protein activity. Induction of GUS activity was observed only in association with the Leu<sup>346</sup> residue. Indel polymorphisms at the protein tail did not affect GUS activity (Fig. 3B). bHLH2 thus controls transcription of the P450 monooxygenase-encoding genes *PdCYP79D16* and *PdCYP71AN24* in the amygdalin biosynthetic pathway.

To understand how a change from Leu to Phe at position 346 affects activity of the bHLH2 transcription factor, we generated a structural model of the protein sequence surrounding this amino acid (Fig. 4, A to C). When dimerized, the corresponding Leu residues of each monomer are placed facing each other, allowing hydrophobic interactions between the amino acids. Change of this Leu into a Phe might have two consequences. The first is that steric hindrances would prevent dimer formation. This would be augmented by the presence of two additional phenylalanines in the near vicinity in the model. Alternatively, the dimer would be formed, but the ability of the aromatic rings to form  $\pi$ -stacks (31) would result in a dimer with an aberrant, nonfunctional conformation. To discriminate between these two possibilities, we expressed the *bHLH2* bitter and sweet transcription factor variants in *Escherichia coli* as a fusion to glutathione *S*-transferase (GST) or His, and performed pull-down assays (Fig. 4D) as well as

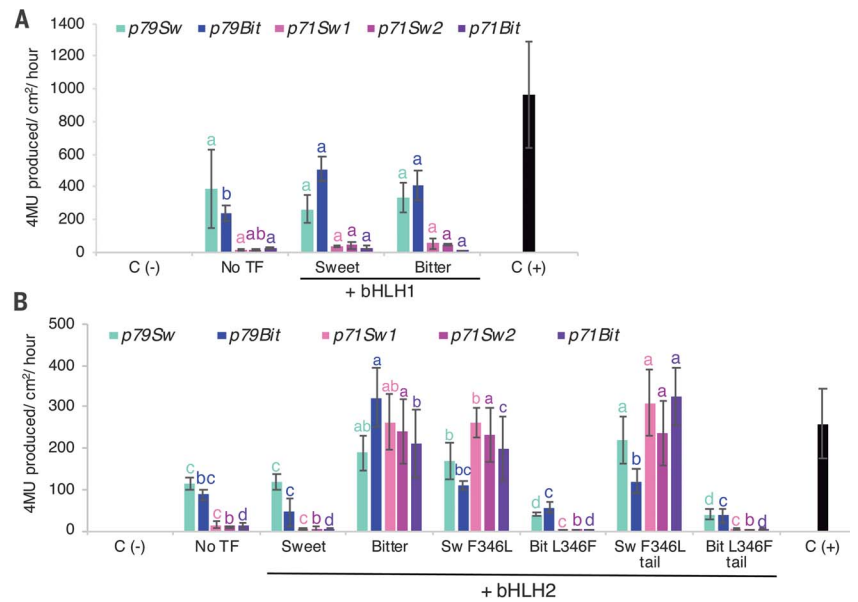
polyacrylamide gel electrophoresis under native conditions (fig. S7). Both protein versions migrated as a band with a molecular mass corresponding to the dimer, indicating that the L346F polymorphism does not affect the ability of bHLH2 proteins to form dimers. To test the DNA-binding ability of the bHLH2 dimers, we cloned a triple repeat of the six-nucleotide motif (CATGTG) present in all *p79* and *p71* promoter variants (data S1) upstream of a  $\beta$ -galactosidase gene in a yeast one-hybrid-compatible plasmid. Only the bHLH2 variants carrying Leu<sup>346</sup> were able to induce  $\beta$ -galactosidase activity (Fig. 4E), indicating that the L346F mutation generates a nonfunctional bHLH2 dimer.

The history of almond domestication is marked by the selection of sweet kernel genotypes, although other traits such as thinner endocarp and increased seed size might also have been objects of early selection (32). The knowledge of the

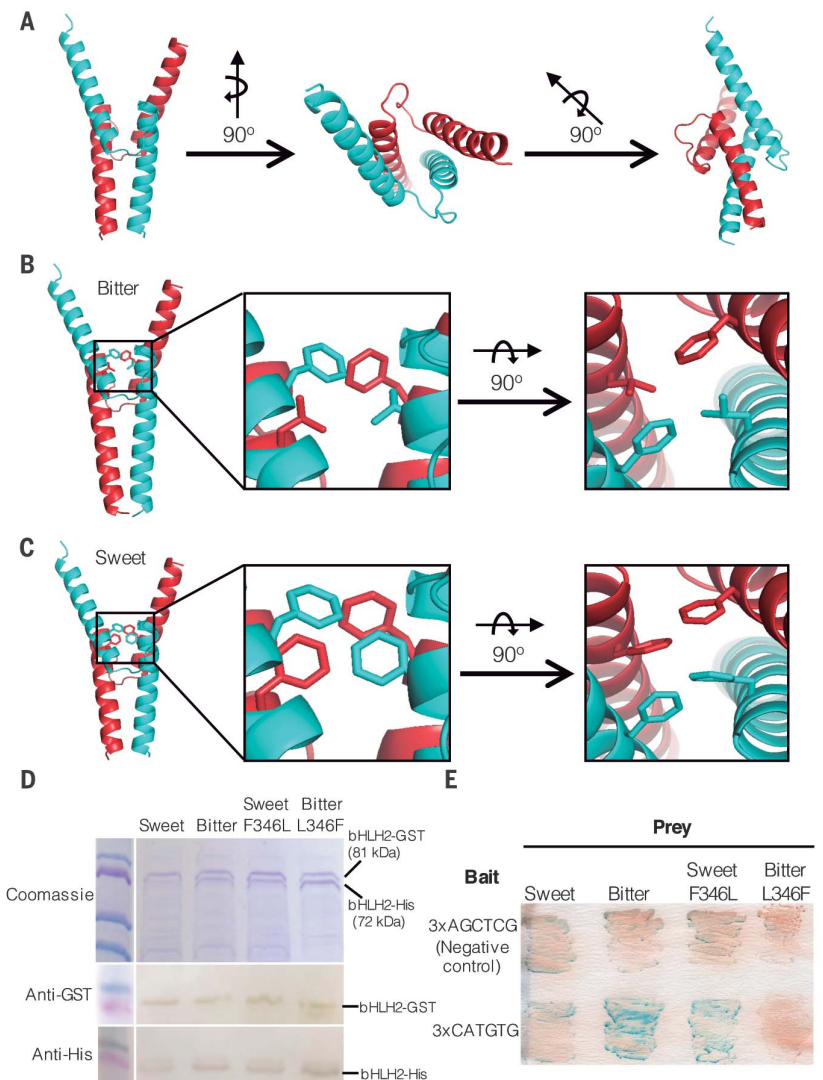


**Fig. 2. The amygdalin biosynthetic pathway and transcript levels of the encoding genes during fruit development.** Expression patterns (standardized FPKM values) of the genes of the amygdalin pathway are shown during tegument growth in the sweet genotype Lauranne and the bitter genotype S3067.

**Fig. 3. Trans-activation analysis of putative downstream gene promoters by the bHLH transcription factors localized in the *Sk* locus and expressed in the tegument during kernel development.** (A and B) The bar plots show the capacity of bHLH1 (A) and bHLH2 (B) variants of sweet (Sw) and bitter (Bit) genotypes to trans-activate the sweet and bitter promoters of *PdCYP79D16* (*p79*; light/dark blue colors) and *PdCYP71AN24* (*p71*; pink/lilac colors), respectively. The effects of mutations in bHLH2 bitter and sweet at position 346 as well of their C-terminal tails are also shown. Control experiments: p19 silencing suppressor [C(-)] and no transcription factor (no TF) as negative controls, and pCaMV35S: APL3:GUS as positive control [C(+)]. For each promoter, shown in different colors, data were compared using analyses of variance (ANOVA) least significant differences test with  $P < 0.05$ . Therefore, significant differences in the multiple combinations of promoters with or without TF are shown by different letters but with the same color.



**Fig. 4. A Leu → Phe substitution is responsible for formation of a nonfunctional bHLH2 dimer.** (A) Visualization of the relative position of two bHLH2 monomers (red and cyan) after dimer formation. (B and C) A single amino acid substitution (Leu to Phe) is present in the sweet variant, possibly interfering with formation of a functional dimer. (D) *E. coli*-expressed bHLH2 proteins were subjected to pull-down assays with a His-tag resin, transferred to polyvinylidene difluoride membranes, and detected with an antibody recognizing the His- and GST-tag engineered at the N-terminal end of the proteins. Precision Plus Protein Standard (Dual Color) was used as a protein standard, in which the pink-colored marker protein corresponds to 75 kDa. (E) The sweet and bitter variants of bHLH2 were tested in a yeast one-hybrid screening for their activation of a synthetic promoter including a triple repetition of a CATGTG motif, a typical binding sequence for the basic  $\beta$ -helix-loop-helix transcription factor family. As a negative control, a triple repetition of a random hexanucleotide sequence (AGCTCG) was used. The effects of the L346F mutation in bHLH2 bitter and sweet are also shown.



Downloaded from <http://science.sciencemag.org/> on June 21, 2019



toxicity of bitter almond and peach kernels was already realized in ancient Egypt in the second millennium BCE, where traitorous priests in the capital cities of Memphis and Thebes were poisoned to death with pits of peaches—a practice referenced in hieroglyphics as “death by peach” (33, 34). Caius Plinius Secundus (Pliny the Elder) wrote in his 37-volume encyclopedia *Naturalis Historia* that the Romans were proud of knowing how to remove the bitterness and toxicity from bitter almond kernels (35). In Basil of Caesaria’s *Hexameron* from the 4th century CE, it is stated that Greek agriculturists in Cappadocia had discovered that piercing of the trunk of a highly bitter almond tree near the root, so as to introduce a fat plug of pine into the middle of the pith, resulted in production of delicious sweet almonds (36).

Today, almond breeding often introduces desirable traits from genotypes heterozygous or homozygous for the bitter *sk* allele, thus requiring the selection of sweet kernel individuals in segregating populations. With the identification of the polymorphisms causally related to the kernel taste, such a selection could be efficiently performed early at the seedling stage.

Besides contributing to the advancement of almond genomics and breeding, the knowledge obtained in this study offers a route toward neodomestication of other plants (37), for controlling accumulation of not only other cyanogenic glucosides, such as linamarin and lotaustralin in cassava tubers (38), but also other specialized metabolites, such as anthocyanins in strawberry (39) saponins in the sweet/bitter quinoa lines (40) and formation of gossypol-containing glands in cotton (41). Finally, we envisage that DNA sequencing of the *bHLH* gene in almonds found at ancient archaeological sites may be used to delineate the roots of cultivation (42) or how ancient human populations moved across Asia, Europe (through the Mediterranean Basin), and the Americas (43).

## REFERENCES AND NOTES

1. G. Willcox, S. Fornite, L. Herveux, *Veg. Hist. Archaeobot.* **17**, 313–325 (2008).

2. D. Rivera Núñez, G. Matilla Séiquer, C. Obón de Castro, F. J. Alcaraz Ariza, in *Plants and Humans in the Near East and the Caucasus: Ancient and Traditional Uses of Plants as Food and Medicine, a Diachronic Ethnobotanical Review* (Armenia, Azerbaijan, Georgia, Iran, Iraq, Lebanon, Syria, and Turkey) (Edit.um, 2012), vol. 2, pp. 886–944.
3. M. Delplancke et al., *Evol. Appl.* **5**, 317–329 (2012).
4. M. Delplancke et al., *Mol. Ecol.* **22**, 1092–1104 (2013).
5. D. Zohary, M. Hopf, E. Weiss, *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford Univ. Press, 2012).
6. J. Hansen, J. M. Renfrew, *Nature* **271**, 349–352 (1978).
7. T. Yamamoto, S. Terakami, *Breed. Sci.* **66**, 148–159 (2016).
8. S. Farinati, A. Rasori, S. Varotto, C. Bonghi, *Front. Plant Sci.* **8**, 1247 (2017).
9. R. Sánchez-Pérez, K. Jørgensen, C. E. Olsen, F. Dicenta, B. L. Møller, *Plant Physiol.* **146**, 1040–1052 (2008).
10. R. Sánchez-Pérez, K. Jørgensen, M. S. Motawia, F. Dicenta, B. L. Møller, *Plant J.* **60**, 894–906 (2009).
11. R. Sánchez-Pérez et al., *Plant Physiol.* **158**, 1916–1932 (2012).
12. S. Thodberg et al., *Plant Physiol.* **178**, 1096–1111 (2018).
13. M. J. Heppner, *Genetics* **8**, 390–391 (1923).
14. F. Dicenta, J. E. García, *Heredity* **70**, 308–312 (1993).
15. R. Sánchez-Pérez et al., *Tree Genet. Genomes* **6**, 237–245 (2010).
16. T. Joobeur et al., *Theor. Appl. Genet.* **97**, 1034–1041 (1998).
17. T. Yamaguchi, K. Yamamoto, Y. Asano, *Plant Mol. Biol.* **86**, 215–223 (2014).
18. R. Velasco et al., *Nat. Genet.* **42**, 833–839 (2010).
19. V. Shulaev et al., *Nat. Genet.* **43**, 109–116 (2011).
20. Q. Zhang et al., *Nat. Commun.* **3**, 1318 (2012).
21. International Peach Genome Initiative et al., *Nat. Genet.* **45**, 487–494 (2013).
22. K. Shirasawa et al., *DNA Res.* **24**, 499–508 (2017).
23. E. Dirlwanger et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9891–9896 (2004).
24. S. Jung et al., *BMC Genomics* **10**, 76 (2009).
25. S. N. Goonetilleke et al., *G3* **8**, 161–172 (2018).
26. R. Sánchez-Pérez, F. Dicenta, P. Martínez-Gómez, *Tree Genet. Genomes* **8**, 379–389 (2012).
27. R. Sánchez-Pérez, W. Howad, F. Dicenta, P. Arús, P. Martínez-Gómez, *Plant Breed.* **126**, 310–318 (2007).
28. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, *Bioinformatics* **31**, 3210–3212 (2015).
29. F. Ricciardi et al., *Genes* **9**, 385 (2018).
30. W. R. Chezem, N. K. Clay, *Phytochemistry* **131**, 26–43 (2016).
31. G. B. McCaughey, M. Gagné, A. K. Rappé, *J. Biol. Chem.* **273**, 15458–15463 (1998).
32. D. Velasco, J. Hough, M. Aradhya, J. Ross-Ibarra, *G3* **6**, 3985–3993 (2016).
33. R. H. Davis, in *Toxic Substances in Crop Plants* (Royal Society of Chemistry, 1991), pp. 202–225.
34. D. Blum, *The Poisoner’s Handbook* (Penguin, 2010).
35. Pliny the Elder, in *Naturalis Historia. Liber XV (77 AD)*, p. 26.
36. Basil of Caesaria, in *Hexameron*, Homily 5, paragraph 7.

37. P. Gepts, in *Encyclopedia of Agriculture and Food Systems*, N. K. Van Alfen, Ed. (Elsevier, 2014), vol. 2, pp. 474–486.
38. K. Jørgensen et al., *Plant Physiol.* **155**, 282–292 (2011).
39. F. Zhao et al., *Sci. Rep.* **8**, 2721 (2018).
40. D. E. Jarvis et al., *Nature* **542**, 307–312 (2017).
41. D. Ma et al., *Nat. Commun.* **7**, 10456 (2016).
42. G. Willcox, *Science* **341**, 39–40 (2013).
43. C. L. Scheib et al., *Science* **360**, 1024–1027 (2018).

## ACKNOWLEDGMENTS

We thank T. Cremades Rosado and M. Dahl for technical help; the Biology of Stress and Plant Pathology Department and Biotechnology Fruit Trees Group at CEBAS-CSIC for use of their facilities; J. A. Egea for statistical analysis advice; and T. Sicheritz-Ponten, B. Petersen, and A. Dhingra for their initial contributions and efforts regarding the genome sequencing. L. M. De Luca, Johns Hopkins University, is acknowledged for providing the reference to Basil of Caesaria’s *Hexameron*. **Funding:** Supported by VILLUM Foundation grants to the VILLUM Center for Plant Plasticity (VKR023054) (B.L.M. and R.S.-P.) and VILLUM project 13234 (R.L.L.-M.); a University of Copenhagen Excellence Program for Interdisciplinary Research grant to the Center for Synthetic Biology (B.L.M.); European Research Council Advanced Grant ERC-2012-ADG\_20120314 (B.L.M.); VILLUM Young Investigator Program grant VKR023124 (R.S.-P.); and the projects “Mejora Genética del Almendro” (MINECO-Spain), Séneca Foundation grants for “Molecular Biology of Cyanogenesis in Almond” (11937-PI-09) (R.S.-P.), and “Breeding stone fruit species assisted by molecular tools” (19879/GERM/15) (F.D., J.D.C., and R.S.-P.). **Author contributions:** R.S.-P., J.D.C., and F.D. designed crosses, generated and evaluated plant material; R.S.-P., R.M., C.M., J.D.C., F.R., R.A.C., R.L.L.-M., and S.P. performed the research; R.S.-P., S.P., C.L., L.R., F.D., R.L.L.-M., and B.L.M. designed the research; and R.S.-P., S.P., R.A.C., R.L.L.-M., and B.L.M. wrote the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** The following sequences can be found at GenBank database with the following codes: pCYP71AN24sweet2, MK041085; pCYP71AN24sweet1, MK041086; pCYP71AN24bitter, MK041087; pCYP79D16sweet, MK041088; pCYP79D16bitter, MK041089; bHLH1sweetLauranne, MK041090; bHLH1bitterS3067, MK041091; bHLH2sweetLauranne, MK041092; bHLH2bitterS3067, MK041093; bHLH4, MK041094. The almond genome sequence is available at the DDBJ database: AP019297-AP019304. All data are available in the main text or the supplementary materials or in the databases previously named.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1095/suppl/DC1  
Materials and Methods  
Figs. S1 to S7  
Tables S1 to S9  
Data S1 and S2  
References (44–89)

30 November 2018; accepted 23 May 2019  
10.1126/science.aav8197

## Mutation of a bHLH transcription factor allowed almond domestication

R. Sánchez-Pérez, S. Pavan, R. Mazzeo, C. Moldovan, R. Aiese Cigliano, J. Del Cueto, F. Ricciardi, C. Lotti, L. Ricciardi, F. Dicenta, R. L. López-Marqués and B. Lindberg Møller

*Science* **364** (6445), 1095-1098.  
DOI: 10.1126/science.aav8197

### How to make almonds palatable

The domesticated almond tree has been feeding humans for millennia. Derivation from the wild, bitter, and toxic almond required loss of the cyanogenic diglucoside amygdalin. Sánchez-Pérez *et al.* sequenced the almond genome and analyzed the genomic region responsible for this shift. The key change turned out to be a point mutation in a transcription factor that regulates production of P450 monooxygenases in the biosynthetic pathway for the toxic compound.

*Science*, this issue p. 1095

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/364/6445/1095>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2019/06/12/364.6445.1095.DC1>

#### REFERENCES

This article cites 75 articles, 15 of which you can access for free  
<http://science.sciencemag.org/content/364/6445/1095#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)