

# Evolution of a hot genome

Victor A. Albert<sup>1</sup> and Tien-Hao Chang

Department of Biological Sciences, University at Buffalo, Buffalo, NY 14260

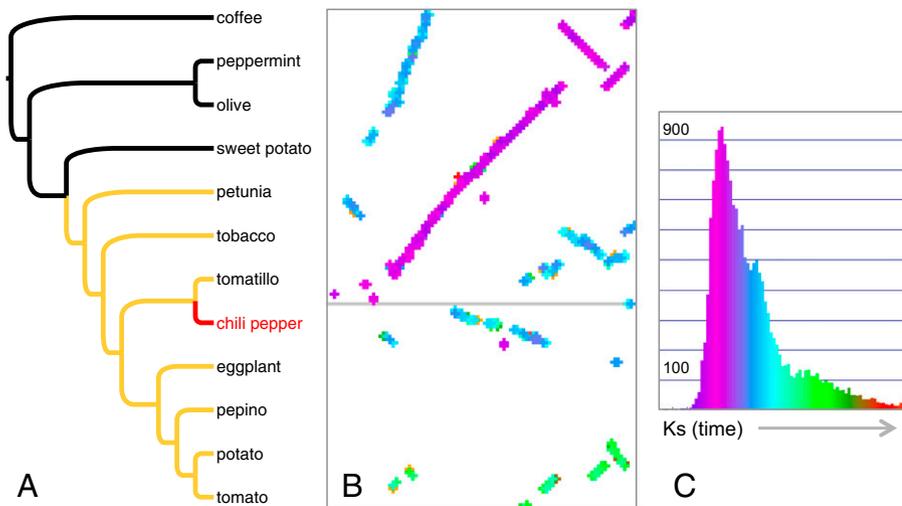
The chile pepper has a long and important human history. Chiles, members of the plant genus *Capsicum* in the Solanaceae family, evolved in the New World and were domesticated there several millennia ago (1). Although it is hard to imagine a Mexican, Szechuan, or Indian dinner without chiles, these characteristic cuisine elements are a relatively recent development, having only achieved their enormous Old World culinary impact in the centuries after Columbus. Indeed, several other Solanaceae crops are of entirely New World origin, with tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) of particular importance, but also pepino (*Solanum muricatum*) and tomatillo (*Physalis philadelphica*) (Fig. 1A). In testimony to the importance of this plant group, whole-genome sequences of tomato (2) and potato (3) were among the first 50 plant genomes available for study. Now, we can add chile pepper to the mix, as published in PNAS by Qin et al. (4) and independently by Kim et al. (5). Although both papers provide excellent resources to the plant science

community and share several similar findings (e.g., on the enormous expansion of the chile pepper genome by proliferation of mobile elements and on the biochemical genetics of the fruit's characteristic pungency), the work of Qin et al. (4), which includes considerable genomic resequencing of multiple cultivated varieties and a wild pepper, the chiltepin, provides important added data for evaluating the (in part artificially derived) adaptive landscape of cultivated peppers.

The recent proliferation of complete and high-quality nuclear genome sequences of flowering plants (angiosperms) has revolutionized our understanding of complex genome evolution. Despite the obvious biomedical and other human-related interest in genome evolution of mammals, their genomes are downright boring compared with those of angiosperms. Of course, this is in part a biased opinion, but not one terribly far removed from reality. For example, despite the fact that the mammalian lineage is approximately as old as that of

flowering plants, only in plants can we observe—over and over again—the evolutionary influence of whole-genome duplications (WGDs), whereby all genetic material is duplicated *en masse*, followed by an extended period of fractional gene loss among subgenomes, with those duplicate genes retained usually having partitioned their functions or evolved entirely new ones (6, 7). Gene copies derived from tandem or other small-scale duplication events often share similar fates. Furthermore, genetic linkages established in pre-WGD progenitors can be severed among daughter blocks from which gene duplicates have fractionated out.

Since the sequencing of the tomato genome, it has been known that both the tomato and potato genomes descend from an ancient whole-genome triplication event (2). As such, this triplication event must be older than the tomato–potato phylogenetic split at ~7.3 My. The tomato investigators used synonymous substitution rates in coding sequences (Ks) determined from duplicate gene pairs that descended from the triplication as proxies to calculate evolutionary time; from this comparison, researchers estimated that the triplication event occurred roughly an order of magnitude earlier than the tomato–potato split. Although Kim et al. (5) noted strong collinearity of the chile pepper genome to that of tomato, they did not explicitly address the triplication event. Qin et al. (4) convincingly provide this analysis, which we show here for the wild chiltepin genome against the tomato genome in our own Ks-colored syntenic dotplot (Fig. 1B and C) derived using CoGe (8). Thus, the triplication event clearly predates pepper–tomato divergence, which Qin et al. (4) date to ~36 Mya [but note that the dating game is tricky and that other authors have suggested a similar age for the entire Solanaceae (9)]. Regardless of the precise timing of the event, genome scientists can now use the chile pepper and *Solanum* genomes to examine how homeologous genes descending from the triplication have evolved



**Fig. 1.** The genome of chile pepper in an evolutionary context. (A) A phylogeny of lamiid angiosperms, the family Solanaceae in yellow and chile pepper in red. (B) Syntenic dotplot of tomato chromosome 2 (vertical) versus chile pepper pepper chromosomes 2 (Upper) and 4 (Lower), with syntenic gene duplicates colored according to their Ks values (synonymous substitutions per synonymous site), as plotted in the x axis of C against numbers of gene pairs on the y axis. Ks, increasing along the x axis, is a proxy for time since speciation or gene duplication. Purple, orthologous synteny (youngest); blue, paralogous synteny descending from the triplication observed in Solanaceae; green, paralogous synteny dating to the paleohexaploidy event at the base of core eudicots (oldest). Note corresponding purple, blue, and green peaks in the Ks plot, the latter two subtle.

Author contributions: V.A.A. and T.-H.C. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 5135.

<sup>1</sup>To whom correspondence should be addressed. E-mail: vaalbert@buffalo.edu.

over deeper time and among distinct phylogenetic groups (Fig. 1A). Adding even more piquancy is the fact that the common ancestor of all core eudicots experienced a whole-genome triplication of its own (10), meaning that the Solanaceae triplication has resulted in some triples of triples of homeologous genes, all of which can partition ancestral functions or evolve entirely new ones. In fact, we can make out what may be some triply triplicated blocks in the pepper–tomato syntenic dotplot (Fig. 1B and C). During domestication, any of this naturally evolved genomic complexity can in turn become a target of artificial selection.

Qin et al.'s (4) important new resequencing data for 20 different cultivated varieties provide a scaffold upon which to survey the genomic landscape of cultivated chile pepper domestication. Using SNPs among the different varieties compared with the reference genomes, highly homozygous genomic regions potentially subjected to “selective sweeps” were identified from a sliding-window analysis of sequence diversity. A selective sweep occurs when selection occurring at a particular locus sweeps the variation in closely linked genomic regions to fixation along with the naturally or artificially influenced genetic target (11). Closely linked loci can in turn form a coadapted complex that may have persisted through deep time (12), and given the frequency with which genome duplications and triplications have occurred during flowering plant evolutionary history, such complexes may have drastically different ages and compositions because of the rampant deletion process that postpolyploid genomes go through to produce highly fractionated syntenic blocks. Thus, some coadapted genetic linkages may become grist for the mill of domestication possibilities.

Qin et al. (4) identified 115 regions in the chile pepper genome, including 511 genes, that show the diversity-reducing signatures of selective sweeps. Among genes overlapping putatively swept regions, these authors highlighted several potentially domestication-related homologs, including relatives of rice *Rc* and a tomato xyloglucan endotransglucosylase/hydrolase (XTH). *Rc* is known to be related to pericarp color and seed dormancy; moreover, it is a homolog of the classic *ANTHOCYANIN 1 (AN1) myc* gene of petunia (13). *AN1* and orthologs are central players in the purple-red pigmentation of Solanaceae fruits, including chile peppers (14), so it is not difficult to imagine artificial selection occurring at this locus. Using this first-order hypothesis provided by Qin et al. (4), we as “downstream users” attempted to further study selection pressure

at pepper *AN1 (Capana09g001426)* using multispecies, sequence alignment-based molecular evolutionary analyses (contact V.A.A. for details). We found evidence for heterogeneous molecular evolutionary rates (HR) among codons in pepper *AN1*. HR can indicate either positive (protein-adaptive) selection or relaxation of purifying (sequence-conserving) selection (15), both of which may yield selectable traits. Given this finding, we surveyed the genes closely linked to pepper *AN1* (within ~3 Mb) for their possible domestication-related functions. Immediately upstream of pepper (and grape) *AN1* is a homolog (*Capana09g001419*) of an *Arabidopsis* gene involved in stomatal closure (*At4g17970*), positive alleles of which might be adaptive in the hot and arid environments where chiltepin grows and peppers were domesticated. Immediately downstream of pepper *AN1* (and close by in both tomato and grape) is a homolog of the Sec pathway gene *SCY2 (Capana09g001428)*, the mutant form of which causes embryo lethality at the globular stage in *Arabidopsis*. It seems possible that selection against negative pepper *SCY2* alleles could enhance seed yield, and seeds are borne on the most pungent part of the chile pepper fruit. Further downstream of pepper *SCY2* lies a gene (*Capana09g001429*) encoding a homolog of a heat-shock factor (*HsFA7a*) known to contribute to thermotolerance in *Arabidopsis*, positive alleles of which could also be helpful in desert climates. Similarly to pepper *AN1*, we observed significant HR in pepper *HsFA7a*. Finally, close downstream of *HsFA7a* is the pepper homolog (*Capana09g001431*) of an *Arabidopsis* gene (*At3g51820*) encoding an enzyme with prenyltransferase activity that may be in-

involved in tocopherol (vitamin E) biosynthesis. It is known that tocopherol production peaks in red, ripe chile pepper fruits (16), which may make sense given the close linkage of that gene to *AN1* in pepper, where anthocyanin pigmentation may have been artificially selected for. Neither the pepper *HsFA7a* nor the prenyltransferase have syntenic homologs in tomato, potato, and grape, suggesting that the complete linked gene set is unique to chile pepper.

Qin et al. (4) located another possible selective sweep containing a XTH gene (that we find most similar to *Arabidopsis ATXTH27*), which they argue may be involved in non-climacteric fruiting and delayed fruit softening. Only ~100 kb upstream of the pepper XTH lies a homolog of *ATMBD9 (At3g01460)*, the mutant form of which causes early flowering time and enhanced shoot branching, both traits that may have been selected for to increase yield. Similarly to pepper *AN1* and *HsFA7a*, we found these genes to be significant for HR in chile pepper. A close arrangement of pepper *ATXTH27* and *ATMBD9* homologs also occurs in tomato, but not potato.

To synthesize, the new chile pepper genomes produced by Qin et al. (4) provide a number of interesting hypotheses to follow up on for how genome duplications, and duplicate deletion or functional evolution, can “scramble” preexisting arrangements of genetic diversity to yield new combinations that might become targets of either natural or artificial selection. We can hope that many more studies in other species will include such complete genome assemblies and extensively resequenced individuals, both of which are vital components to uncovering genome evolutionary landscapes, both natural and artificial.

- 1 Perry L, et al. (2007) Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. *Science* 315(5814):986–988.
- 2 Consortium TTG; Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641.
- 3 Xu X, et al.; Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195.
- 4 Qin C, et al. (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc Natl Acad Sci USA* 111:5135–5140.
- 5 Kim S, et al. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet* 46(3):270–278.
- 6 Guo H, Lee T-H, Wang X, Paterson AH (2013) Function relaxation following by diversifying selection after whole-genome duplication in flowering plants. *Plant Physiol* 162(2):769–778.
- 7 Sankoff D, Zheng C, Zhu Q (2010) The collapse of gene complement following whole genome duplication. *BMC Genomics* 11:313.
- 8 Lyons E, Pedersen B, Kane J, Freeling M (2008) The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Tropical Plant Biology* 1(3–4):181–190.

- 9 Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-visited. *Am J Bot* 97(8):1296–1303.
- 10 Jiao Y, et al. (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biol* 13(1):R3.
- 11 Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457(7231):843–848.
- 12 Joron M, et al. (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol* 4(10):e303.
- 13 Gerats AGM, Farcy E, Wallroth M, Groot SPC, Schram A (1984) Control of anthocyanin synthesis in *Petunia hybrida* by multiple allelic series of the genes *An1* and *An2*. *Genetics* 106(3):501–508.
- 14 Stommel JR, Lightbourn GJ, Winkler BS, Griesbach RJ (2009) Transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. *J Am Soc Hortic Sci* 134(2):244–251.
- 15 Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155(1):431–449.
- 16 Osuna-Garcia JA, Wall MM, Waddell CA (1998) Endogenous levels of tocopherols and ascorbic acid during fruit ripening of New Mexican-type chile (*Capsicum annuum* L.) cultivars. *J Agric Food Chem* 46(12):5093–5096.