Supplementary Note 5

Incomplete gene fractionation after paleopolyploidy: the first study case in flowering plants revealed by comparison of the *Petunia axillaris N*. and *Solanum lycopersicum* genomes

Laurie Grandont[1 +], Haibao Tang[2, 3 +], Mitrick Johns[4], Eric Lyons[3]*, M. Eric Schranz[1]*

[1]Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
[2] Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China.
[3] School of Plant Sciences, iPlant Collaborative, University of Arizona, Tucson, AZ 85721, USA.
[4]Department of Biological Sciences, Northern Illinois University, DeKalb IL 60115, USA.

[+] These authors contributed equally to this work

*for correspondence: eric.schranz@wur.nl; ericlyons@email.arizona.edu

Running title: Comparative genomic analysis of *Petunia* and tomato

ABSTRACT

Polyploidy events, or whole genome duplications, have had a strong impact on land plant diversification, adaptation and speciation. Genomic investigations have found that polyploidy is ubiquitous among angiosperms and have identified independent lineage-specific ancient polyploidizations. Traces of these ancient polyploidy events, or paleopolyploidy events, can still be identified although duplication events are followed by massive gene loss (fractionation) and chromosomal structural rearrangements. The sequencing of the genome of *Petunia axillaris N* offers an ideal opportunity to study paleopolyploidy and gene fractionation in the evolutionary context of radiation due to the unique phylogenetic location of *Petunia* in the Solanaceae family. Our study confirms the previously inferred Solanaceae paleohexaploidy event. We also demonstrate that the *Petunia* lineage has experienced at least two rounds of paleohexapolyploidization, the older gamma hexaploidy event, which is shared with other Eudicots, and the more recent Solanaceae paleohexaploidy event that is shared with tomato and other Solanaceae species. Despite the shared paleohexaploidy event, we found that the process of gene fractionation is less profound in Petunia compared to tomato. This indicates that fractionation of gene content was not complete when these lineages diverged and independent gene loss events may have contributed to the speciation of the lineages, similar to what has been observed in *Saccharomyces* yeasts but so far not shown in flowering plants.

INTRODUCTION

Genomic analysis of numerous plant species has found that polyploidy is ubiquitous among angiosperms, with shared and independent lineage-specific ancient polyploidy (or paleopolyploidy) events (Soltis et al., 2008; Jiao et al., 2011). While most paleopolyploidy events are ancient genome doublings (paleotetraploidies), there are a few important examples of ancient triplications (paleohexaploidies). For example, the gamma polyploidy event near the origin of Eudicots (e.g. Rosids and Asterids) is a genome triplication most clearly seen by the analysis of the grape genome (Jaillon et al., 2007). Ancient genome triplication events have also been detected by analysis of Brassica and Cleomaceae species (Barker et al., 2009; Wang et al., 2011; Cheng et al., 2013). Additionally, the genome analysis of tomato suggested that there was an ancient polyploidy event somewhere during the evolution of the plant family Solanaceae (Tomato Genome Consortium, 2012), however the exact timing and nature of this triplication is not yet established.

Sub-genomes created by paleo-polyploidization can differentiate in terms of gene density due to uneven or biased gene fractionation (Thomas et al., 2006), and levels of gene expression due to genome dominance (Schnable et al., 2011). Biased gene fractionation and genome dominance have been found between subgenomes/duplicated regions in some plants and yeast (Schnable et al., 2011; Wang et al., 2011). In other plants, a second pattern of gene fractionation involving no bias either in gene fractionation or in genome dominance has been observed (Garsmeur et al., 2013; Chalhoub et al., 2014). Interestingly, patterns of genome fractionation from a shared polyploidy event seem to be similar within plant clades (for example, across grasses and across crucifers). This contrasts with observations from *Saccharomyces* yeasts where early-branching clades show an independent pattern of genome fractionation than later branching species (Scannell et al., 2006). Paleopolyploidy events may contribute to the radiation between crown-groups (large phylogenetic clades) and sister-groups (smaller early-branching clades) by providing genetic material for the evolution of novel trait(s) (reviewed in Schranz et al., 2012; Tank et al., 2015). However, it is not yet clear why there appears to be a lag between the timing of the polyploidy event and the eventual radiation of species (referred to as the WGD radiation lag-time). One possible cause and/or mechanism could be due to the time to establish differential genome fractionation between crown-group species and early-branching sister-species. However, to date very few sister-lineages have been sequenced (*Aethionema* of the Brassicaceae being a rare example of this) to address this question.

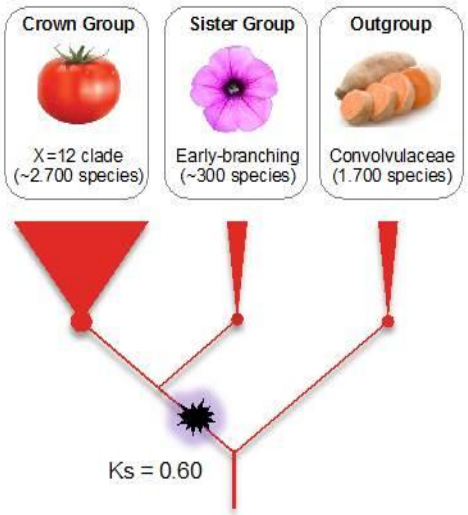


Figure 1: Simplified phylogeny of crown-group Solanaceae (x=12 clade) which includes most important crop species in relation to the early-branching sister-group including Petunia and the out-group family Convovulaceae including sweet potato. The Crown- and Sister-groups of the Solanaceae share an ancient genome triplication (paleohexaploidy) (shown by star-burst) with duplicate gene pairs having an average Ks divergence of 0.60.

In this study we focus on the paleopolyploid history and gene fractionation in two Solanaceae species, tomato (a member of the crown-group) and *Petunia axillaris N* (a member of the smaller-sister group). The Solanaceae plant family contains more than 3000 species, with most of the species found in the crown-group defined as the large "x=12" clade (Särkinen et al., 2013) and the Convolulaceae, including sweet potato being the outgroup family (Figure 1). The Solanaceae crown-group contains a variety of important crops such as tomato, potato, tobacco, and eggplant. Genome collinearity analysis between tomato and grape has been used in establishing a genome triplication in the Solanaceae (Tomato Genome Consortium, 2012). Genomic dot plots revealed that much of the tomato genome was covered in regions either duplicated or triplicated in comparison to grape (Tomato Genome Consortium, 2012). While the existence of triplicated segments suggested that a genome triplication was highly probable in the Solanaceae lineage, they could also have arisen through genome doubling followed by segmental duplications. These ambiguous genomic patterns have led to alternative interpretations. Analysis of the *Mimulus guttatus* genome, which is another related lineage to the Solanaceae in the family Phrymaceae, does not share the Solanaceae paleohexaploidy event, but has an independent paleotetraploidy event (Ibarra-Laclette et al., 2013). Analysis of *Mimulus* and tomato did not yield additional insight into the structure of the Solanaceae paleohexplaoidy event due to the independence of their respective WGD events.

Here we validate previous findings and investigate the role of whole genome triplication and gene fractionation in the evolutionary context of radiation by taking advantage of the phylogenetic location of *Petunia*, which is part of the Solanaceae "first-branching" or "sister-group". We showed that *Petunia* has experienced a paleohexaploidy event subsequent to its divergence from the grape lineage and that this event is dated to have occurred before tomato-*Petunia* split (and thus predated the other x=12 crown-group species as well). Additionally, we showed that following this shared genome triplication; the tomato genome has retained fewer genes than the *Petunia* genome. The high fractionation level in the tomato lineage may have contributed to the initial difficulties in clearly identifying the three triplicated regions within the tomato genome. We finally show that gene fractionation occurred in "two steps" with a first shared fractionation process in the *Petunia* and tomato lineages consecutive/subsequent to their common Solanaceae paleohexaploidy event. Gene fractionation then continued independently following the divergence of these two lineages. This is evidence that fractionation of gene content was not complete when these lineages diverged and may have contributed to the diversification of the lineages, similar to what has been observed in *Saccharomyces* yeasts but until now not yet described in flowering plants.

RESULTS

Petunia shares the Eudicot paleohexaploidy event

We first addressed if the *Petunia* genome shares the gamma genome triplication with other Eudicots. We performed whole genome collinearity analysis of grape and *Petunia axillaris N* (Figure 2). The genome hexaploidy event, or "gamma", could be clearly highlighted in a self-self grape comparison, as the grape genome has not undergone a more recent polyploidy event since gamma (Jaillon et al., 2007). Indeed, by analyzing a given genomic region by traversing the grape/grape dot plot horizontally we identified the three syntenic regions generated by the Eudicot paleohexaploidy gamma event (Figure 2a; see red dashed lines for example). Similarly, in the grape vs. *Petunia* dot plot we observed three syntenic regions (in green), called out-paralogs, which also originated from the gamma event (Figure 2b; see red dashed lines). In the grape-*Petunia* dot plot, the syntenic blocks on the diagonal correspond to the orthologous regions between grape and *Petunia*. The orthologous regions show a smaller Ks value (~1.58) than the "gamma" regions (~2.69), confirming that the gamma triplication predates the divergence of grape and *Petunia* (Figure 2c).

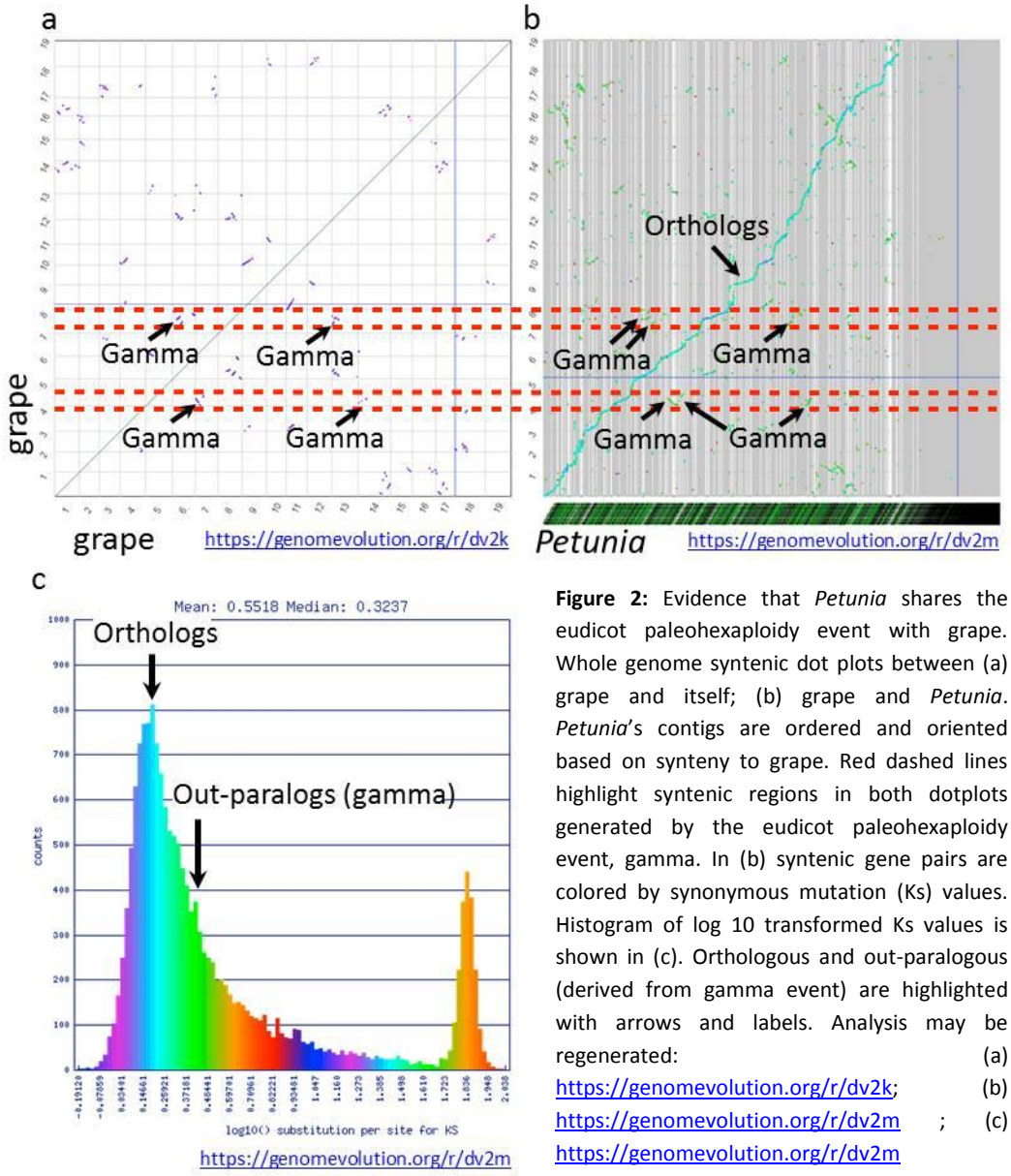


Figure 2: Evidence that *Petunia* shares the eudicot paleohexaploidy event with grape. Whole genome syntenic dot plots between (a) grape and itself; (b) grape and *Petunia*. *Petunia*'s contigs are ordered and oriented based on synteny to grape. Red dashed lines highlight syntenic regions in both dotplots generated by the eudicot paleohexaploidy event, gamma. In (b) syntenic gene pairs are colored by synonymous mutation (Ks) values. Histogram of log 10 transformed Ks values is shown in (c). Orthologous and out-paralogous (derived from gamma event) are highlighted with arrows and labels. Analysis may be regenerated: (a) https://genomevolution.org/r/dv2k; (b) https://genomevolution.org/r/dv2m ; (c) https://genomevolution.org/r/dv2m

***Petunia* shares the more recent hexaploidy event with crown-group Solanaceae**

Refined analyses of orthologous regions from our grape-*Petunia* comparison also reveal that the *Petunia* lineage underwent one additional independent hexaploidy event after its divergence from the grape lineage (Figure 3). By zooming in the grape-*Petunia* dot plot along the orthologous regions, we observed that cluster of orthologous regions appear to be either duplicated or triplicated in *Petunia* (Figure 3b), providing evidence for a subsequent polyploidy event (either tetraploidy or hexaploidy) in the *Petunia* lineage. These duplicated or triplicated segments were isolated as "boxed" areas along the diagonal, as a result of the Syntenic Path Assembly (SPA) algorithm in CoGe SynMap (Lyons et al., 2011). We then scrutinized manually each boxed region for microsynteny and duplicated regions were subsequently searched in order to capture additional syntenic regions, which can be missed or misplaced by the automated syntenic path assembly algorithm used to order *Petunia*'s contigs. In all cases of syntenic regions that appeared duplicated in the grape-*Petunia* dot plot, an additional syntenic region was identified in an exhaustive search. More precisely, we carefully analyzed the microsynteny for each box regions identified in the Figure 3b (labeled A-I) as shown in Figures 4-6. Each figure identifies three syntenic orthologous regions of *Petunia* to one region of grape, consistent with a paleohexaploidy event in *Petunia*.

The *Petunia* regions show consistent patterns of fractionation following polyploidy when compared to an unduplicated outgroup region from grape. Of the nine regions from grape analyzed, all but one had three syntenic regions in *Petunia*, thus confirming that the triplication is genome-wide although with varying degrees of clarity. The genomic regions A, C, F, G, H and I showed similar patterns of retention/loss (fractionation) across each of the triplicated regions. Note that in region F, it is possible to see a genomic insertion in grape. By contrast, the third (small) syntenic region found by SynFind for the B region is highly fractionated with relatively fewer retained duplicated genes. The D and E regions appear to be more complicated and are impacted by other genomic rearrangement events. For the D region, a number of tandem duplications/inversions are present. In addition, the third region shows weak synteny with the grape genome indicating a higher degree of fractionation or poor assembly. For the E region, the first and the third *Petunia* regions do not have overlapping synteny with grape . These results provide unequivocal evidence that *Petunia* lineage has an independent hexaploidy event after its divergence with grape lineage.

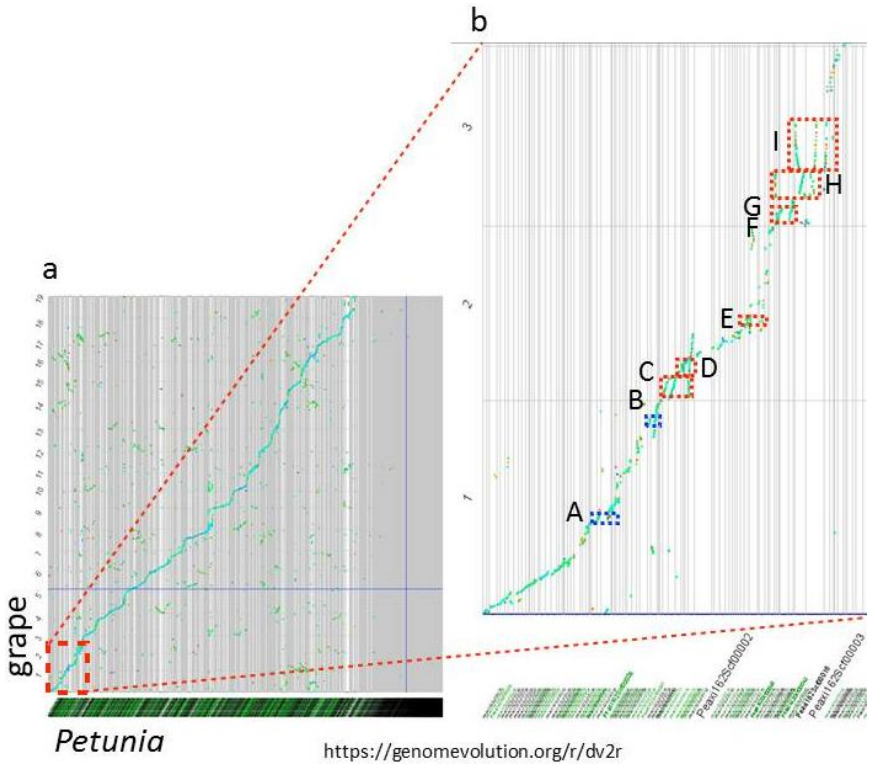


Figure 3: Evidence that the *Petunia* lineage had a subsequent hexaploidy event following its divergence with the grape lineage. Whole genome syntenic dot plot between grape and *Petunia* showing their entire genomes (a) or a zoomed in region of the dot plot (b). In the zoomed in region (b) multiple syntenic orthologous regions of *Petunia* are seen. These were given boxes to identify regions that appear to be triplicated (red) or duplicated (blue). Analysis may be regenerated: https://genomevolution.org/r/dv2r

Figure 4

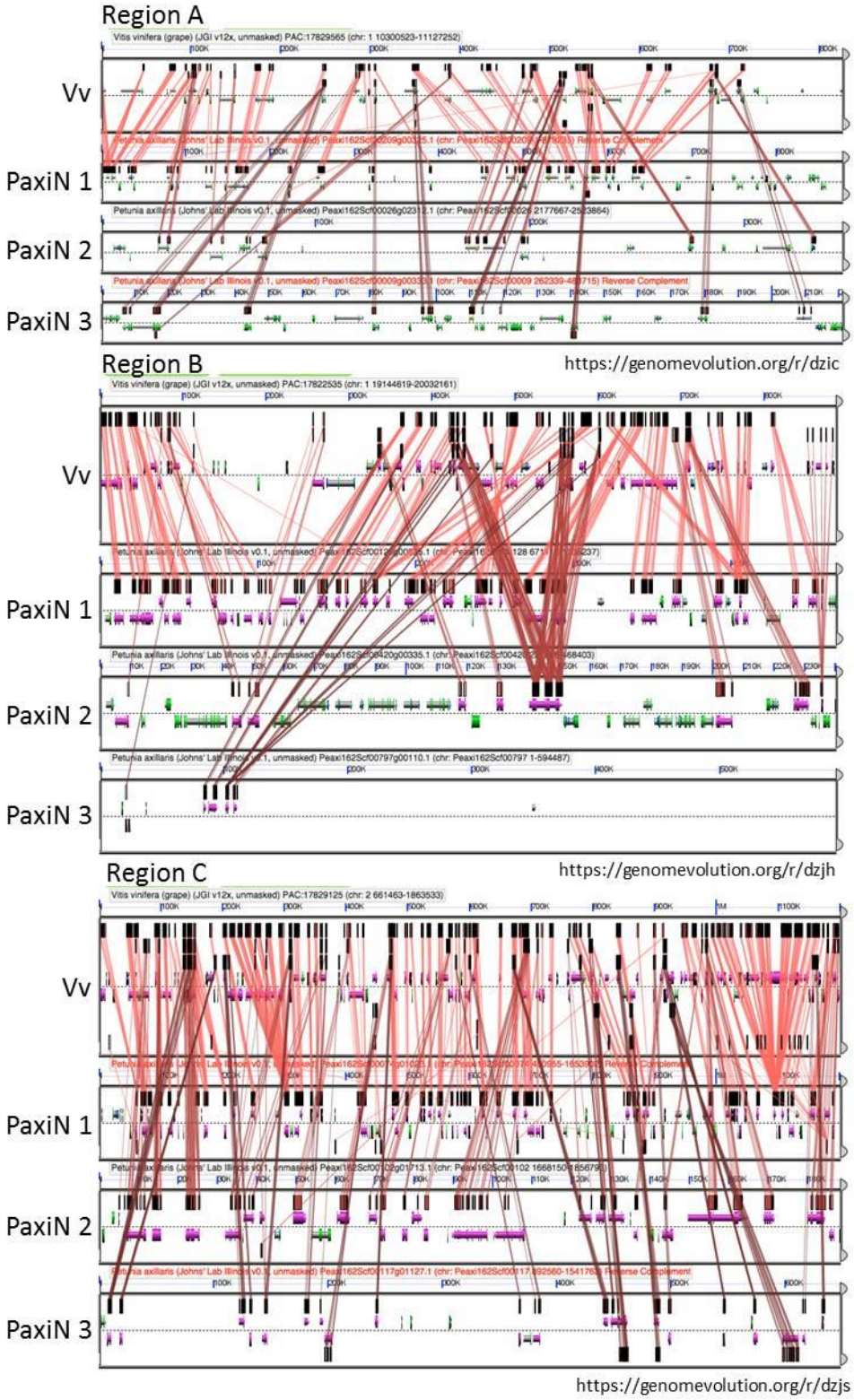
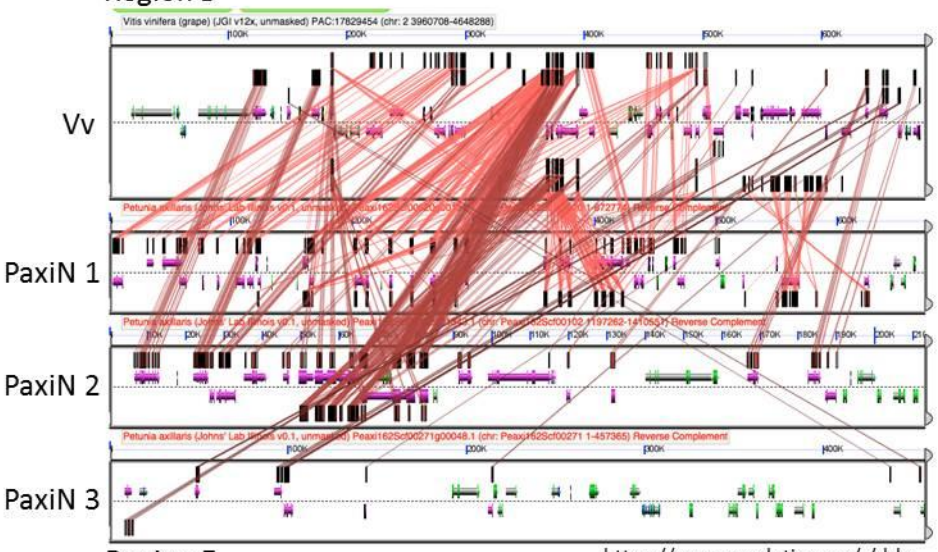

Region A
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/dzic
Region B
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/dzjh
Region C
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/dzjs

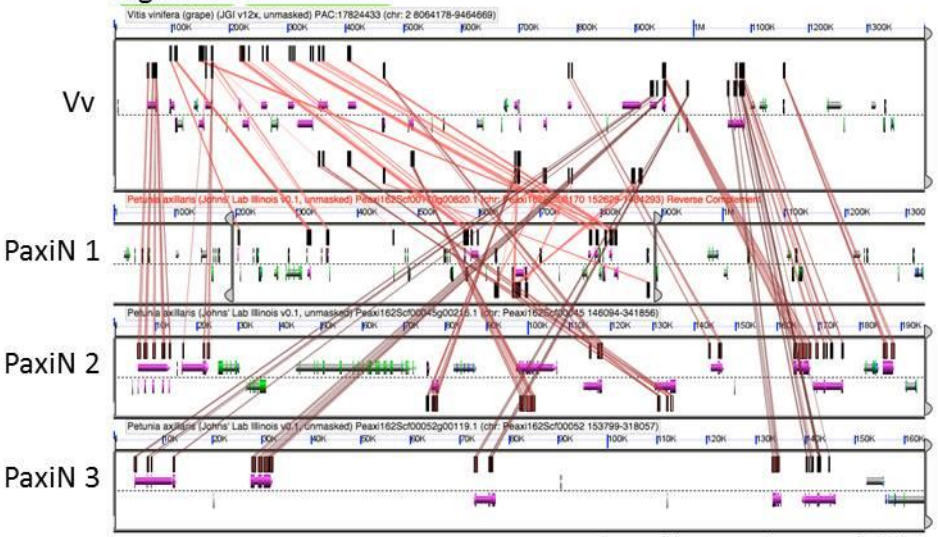
Figure 5

Region D



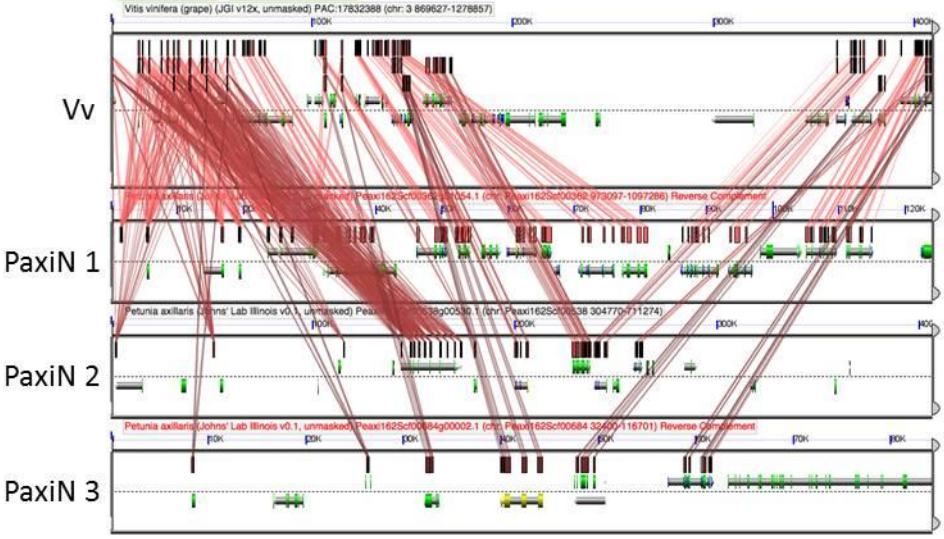
https://genomevolution.org/r/dzkg

Region E



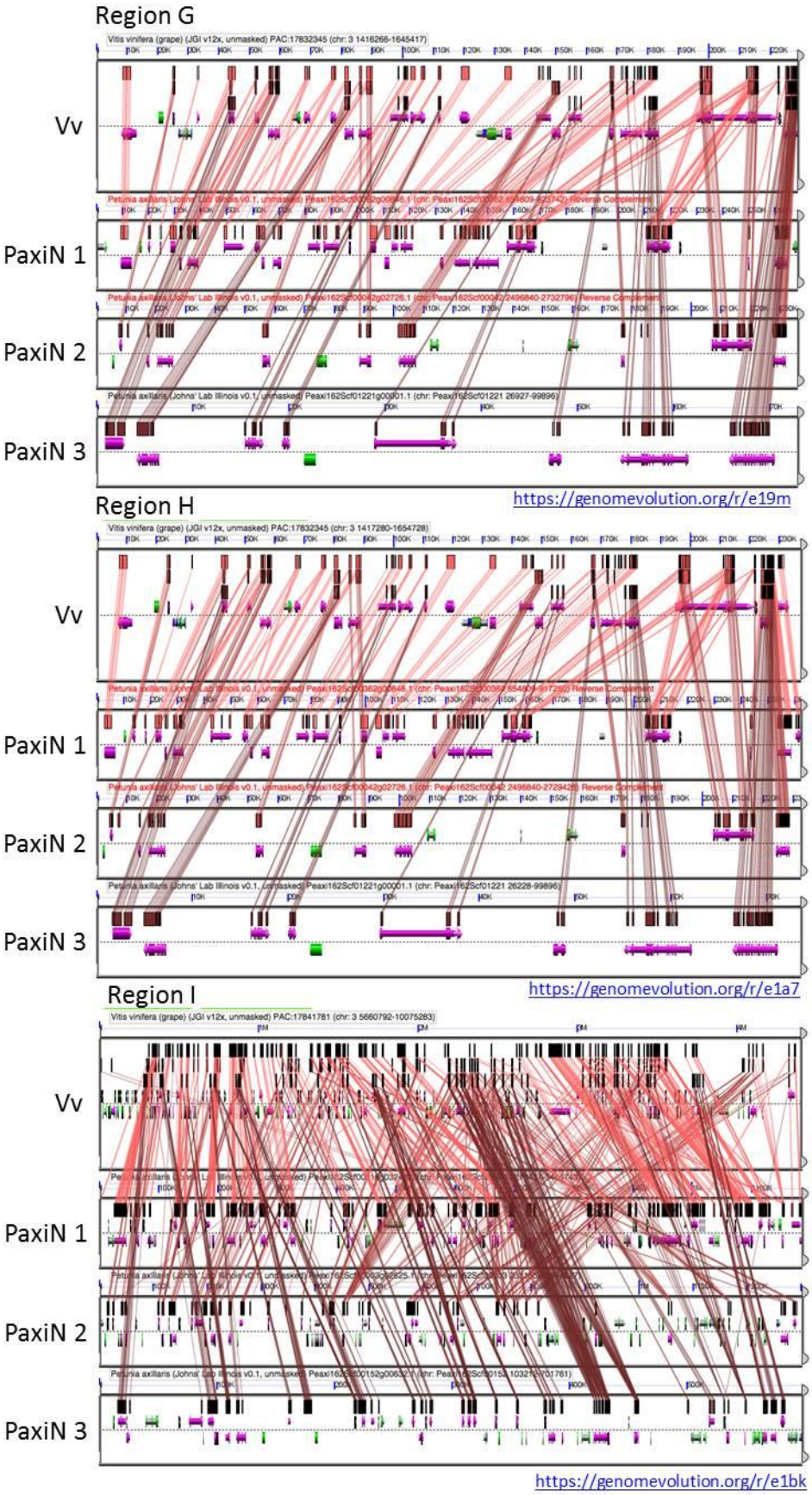
https://genomevolution.org/r/dzkn

Region F



https://genomevolution.org/r/e19j

Figure 6


Region G
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/e19m
Region H
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/e1a7
Region I
Vv
PaxiN 1
PaxiN 2
PaxiN 3
https://genomevolution.org/r/e1bk

Figures 4-6: Evidence that *Petunia* is triplicated compared to grape. Microsynteny analysis of selected regions identified in Figure 3. Each figure identifies three syntenic orthologous regions of *Petunia axillaris N* (PaxiN) to one region of grape (*Vitis vinifera*, Vv). Each figure has a link to generate the analysis along with notes about the quality of the pattern of synteny. Each genomic region is represented by a horizontal panel with a dashed line separating the top and bottom strands of DNA. Gene models are represented as colored arrows located immediately above and below the dashed line. Regions of sequence similarity are shown as colored blocks with lines connecting them between genomic panels. Synteny is inferred as a colinear arrangement of homologous genes forming parallel connecting lines.

We then examined if the subsequent hexaploidy event we found in the *Petunia* genome corresponds to the *Solanum* hexaploidy event previously found by analyzing the tomato, potato and tobacco genomes (Tomato Genome Consortium, 2012).

We analyzed the synteny between tomato and *Petunia* genomes (Figure 7). The corresponding dot plot highlights three types of genomic regions colored according to the *Ks* values; the orthologous regions (in purple) and two different out-paralogous regions (in blue and green). Much of the two genomes were covered in one-to-one matching orthologous regions, which suggests that the paleopolyploidy level is similar between the two genomes. The out-paralogous (II), in green, represent regions that are originated by the old Eurosid hexaploidy gamma event as indicated by the "weak" peak in the *Ks* value histogram (Figure 7b). While the out-paralogs (I), in blue, are originated from the youngest *Solanum* hexaploidy event. This result provides strong evidence that *Petunia* shares the *Solanum* hexaploidy event. This is further confirmed by the similar syntenic patterns seen between both *Solanum* and *Petunia* to grape (Figure 8).

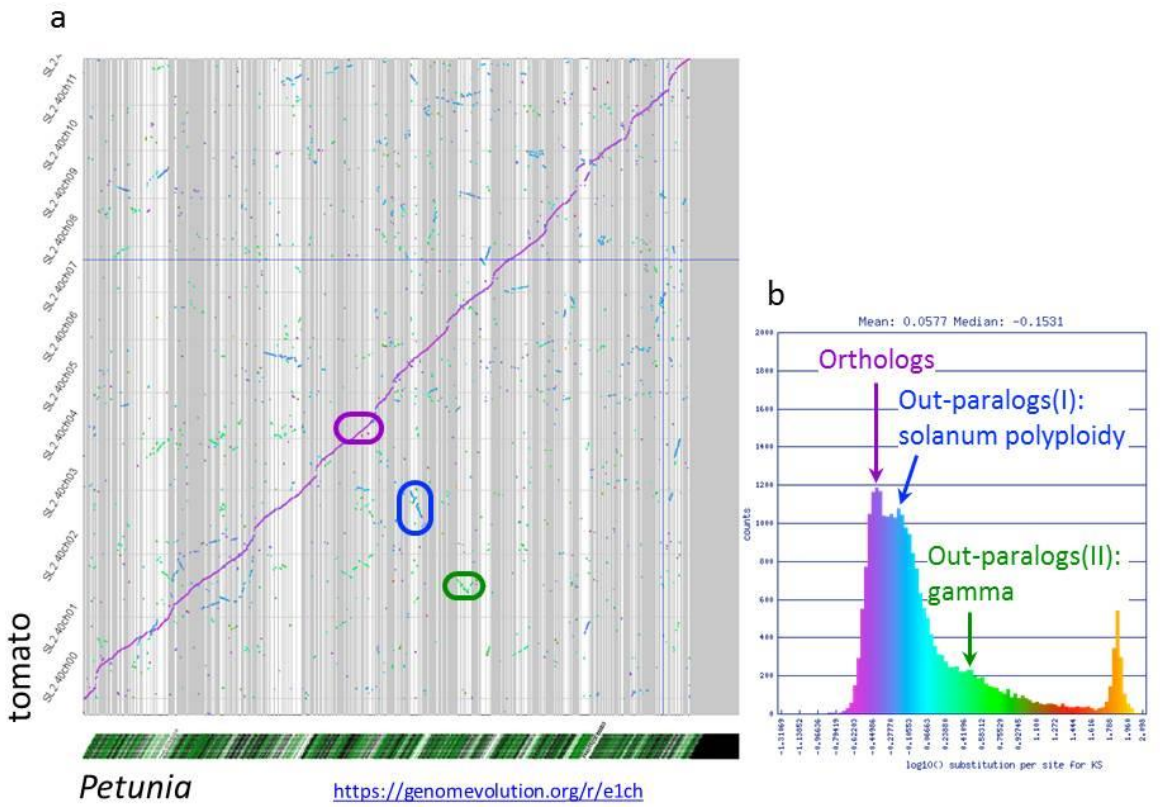


Figure 7: Evidence that *Petunia* and *Solanum* share a paleohexaploidy event. (a) Whole genome syntenic dot plot between tomato and *Petunia* with (b) histogram of log 10 transformed *Ks* values for syntenic gene pairs. Orthologous regions are colored purple; those derived from their shared *Solanum*-specific polyploidy event are colored blue; and those derived from the Eurosid/Eudicot paleohexaploidy (gamma) event are colored in green. Analysis may be regenerated: https://genomevolution.org/r/e1ch

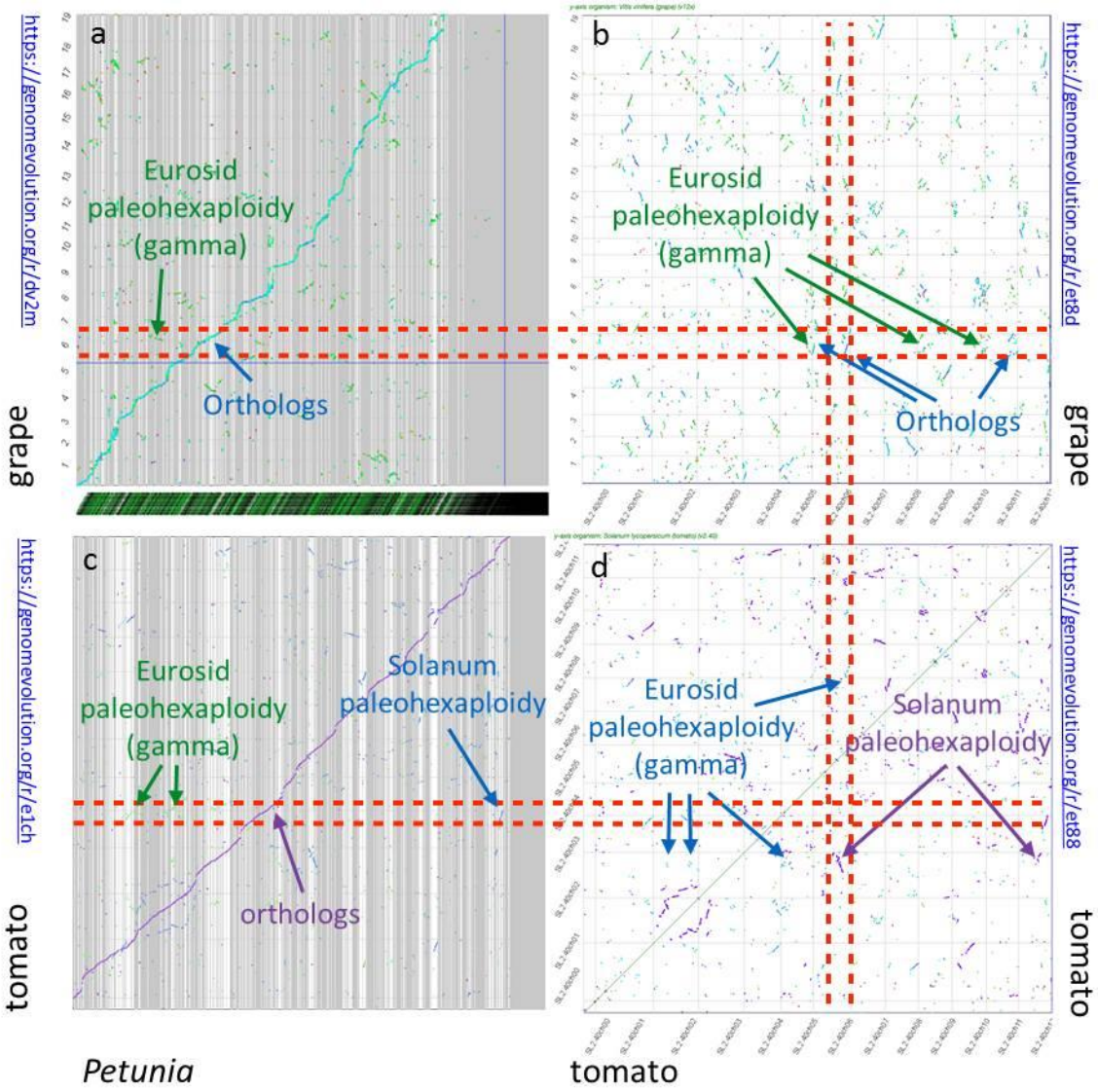


Figure 8: Evidence that *Petunia* shares the *Solanum* paleohexaploidy event. Whole genome syntenic dot plots between (a) grape and *Petunia*; (b) grape and tomato; (c) tomato and *Petunia*; (d) tomato and tomato. Syntenic regions are labeled in each dot plot as to their evolutionary origins and correlated patterns are highlighted with dashed red lines. Note, dashed red lines are not drawn vertically for *Petunia* across tomato and grape because the syntenic path ordering of *Petunia*'s contigs is not necessarily the same for those two genomes. Analyses may be regenerated: (a) https://genomevolution.org/r/dv2m; (b) https://genomevolution.org/r/et8d; (c) https://genomevolution.org/r/e1ch; (d) https://genomevolution.org/r/et88

An in-depth depiction of a set of syntenic regions

To go a step further we performed a detailed examination of one grape syntenic region, the region between bases 1,452,300 and 1,645,417 on chromosome 3 of grape (see Figure 6, region G, Table 1). This region provides evidence for the ancient triplication of the Solanaceae, followed by a later separation of the *Petunia* lineage from the *Solanum* lineage, and then by numerous independent gene loss and tandem duplication events. Here, a single grape region matches three distinct regions in both *Petunia axillaris* and *Solanum lycopersicum*. The region contains 22 grape genes, of which 20 have at least one match in a syntenic region of *Petunia* or tomato. All genes in all 7 syntenic segments are found in the same order and on the same strand relative to each other, allowing for

reverse complementation of the entire region as necessary. One region in each species is clearly dominant, which is the best match to the grape region. The dominant syntenic regions in *Petunia* (on scaffold Scf00362) and tomato (on chromosome 1) share 19 of the 20 matching genes, respectively; the one shared gene found in grape but not in either of the dominant regions is found on one of the other petunia syntenic regions.

The other four secondary syntenic regions share 6-9 genes with the grape region. The two secondary regions within each species show little similarity with each other. However, the syntenic regions on *Petunia* Scf00042 and tomato chromosome 10 appear to be derived from a common ancestor after the triplication event. They share 2 genes at the left end of the region not found in the other regions, as well as 7 other genes also found in the dominant regions. The only differences between the *Petunia* Scf00042 region and the tomato chromosome 10 region are a tandem duplication of a tomato gene, and one syntenic gene found in *Petunia* but not tomato. This similarity demonstrates that the triplication event predates the *Petunia-Solanum* split. The third pair of syntenic regions, found on *Petunia* Scf01221 and tomato chromosome 2, shows little evidence of shared common ancestry more recent than the triplication; at most one gene is shared by these regions that is not found in the other syntenic regions. These regions appear to either have been derived independently from the dominant region or subjected to enough fractionation to completely obscure their common ancestry.

The syntenic regions show evidence of many independent gene loss and tandem duplication events presumably since divergence from a common ancestor. Each of the six regions contains 1-14 genes with no homologues in the other regions. There are also 4 tandem duplications of shared genes (1 in *Petunia*, 3 in tomato) that are unique to a single syntenic region. In contrast, the tandem duplication of grape genes 17832342 and 17832343 must pre-date split between these species, as it is shared by one Petunia and two tomato syntenic regions. The length of the syntenic regions are also quite variable, ranging from 74,000 to 334,000 bp, highlighting different evolutionary trajectory among those regions following the shared event.

Another important feature of this region is the variation concerning which *Petunia* or tomato region contains the closest match to the syntenic grape gene in a tblastn search. For 10 of the 20 genes, the dominant syntenic region contains the best hit. However, for five *Petunia* genes and four tomato genes, the closest match is a gene not found in any syntenic region. The secondary regions have the best match for the remaining genes. The evolutionary forces that affect the subfunctionalization of duplicated genes have apparently affected individual genes in the *Petunia* and tomato genomes differently, and independently of the fractionation process.

Table 1. Description of grape syntenic region G, compared to *P. axillaris* and tomato.

		Matching genes						best tblastn hit to grape	
		petunia			tomato				
Grape gene	position on chr 3	Scf 00362	Scf 00042	Scf 01221	ch 01	ch 10	ch 02	petunia	tomato
--		--	g00274.1	--	--	g055230	--		

--		--	g02718.1	--	--	g055240	--		
PAC:17832347	1,459,600-1,461,579	g00940.1	--	g00001.1	g111270	--	g083820	other	other
PAC:17832346	1,468,355-1,469,527	g00091.1	g00273.1	--	g111280, g111300	g055250	g083800	Scf 00042	Chr 10
PAC:17832345	1,472,790-1,476,160	g00846.1	g02726.1	--	g111310	g055260	--	other	other
PAC:17832344	1,482,091-1,483,122	--	--	--	--	--	--		
PAC:17832343	1,484,329-1,485,994	g00852.1	--	--	g111320	--	g083790	Scf 00362	Chr 01
PAC:17832342	1,493,153-1,494,510	g00841.1	--	--	g111330	--	g083760	Scf 00362	Chr 01
PAC:17832341	1,497,503-1,500,305	g00845.1	g02623.1	g00031.1	g111340	g055340, g055370	--	Scf 01221	other
PAC:17832340	1,502,460-1,506,275	g00832.1	g02616.1	--	g111350	g055390	--	Scf 00362	Chr 10
PAC:17832339	1,512,511-1,518,020	g00840.1	--	g00026.1	g111360	--	--	Scf 00362	other
PAC:17832338	1,519,145-1,523,441	g00839.1	--	--	g111370	--	--	Scf 00362	Chr 01
PAC:17832337	1,527,977-1,528,708	g00838.1	--	g00027.1	g111380	--	--	Scf 01221	Chr 01
PAC:17832336	1,534,987-1,537,909	g00088.1	--	--	g111400	--	--	Scf 00362	Chr 01
PAC:17832335	1,543,989-1,546,820	--	--	--	--	--	--		Chr 01
PAC:17832334	1,549,919-1,551,865	g00082.1	--	--	g111430	--	--	Scf 00362	Chr 01
PAC:17832333	1,560,187-1,564,616	g00848.1	--	--	g111440	--	g083690	other	Chr 01
PAC:17832332	1,565,251-1,567,724	--	g02619.1	--	--	--	--	other	--
PAC:17832331	1,570,696-1,574,690	g00836.1	--	g00028.1	g111450, g111460	--	--	Scf 00362	Chr 01
PAC:17832330	1,584,232-1,585,073	g00831.1	g02519.1	g00042.1	g111500	g055410	--	Scf 00362	Chr 10
PAC:17832329	1,587,828-1,594,273	g00830.1	--	--	g111510	--	g083620, g083630	other	Chr 02
PAC:17832328	1,595,399-1,600,077	g00720.1	--	g00043.1	g111520	--	--	Scf 01221	Chr 01
PAC:17832327	1,613,566-1,639,214	g00725.1, g00716.1	g02521.1	--	g111530	g055450	--	Scf 00362	Chr 01
PAC:17832325	1,641,239-1,645,417	g00645.1	g02513.1	g00034.1	g111540	g055470	--	Scf 00042	Chr 10

The tomato genome is more fractionated than the *Petunia* genome.

We finally investigated gene fractionation in *Petunia* in comparison with tomato. Gene fractionation corresponds to the loss of duplicate genes after whole genome duplication (Langham et al., 2004; Thomas et al., 2006). We compared gene fractionation in Region H in detail (see Figure 6) by analyzing microsynteny between *Petunia* and tomato genomes with grape as reference (Figure 9 and Figure 10). While we selected H region for illustratrative purposes, patterns of other regions offer similar trends of fractionation.

As expected, due to the *Solanum*-hexaploidy event, we can see that both the *Petunia* and tomato genomes are triplicated in comparison with the grape genome (Figure 9). It appears that PaxiN 3 and Sl 3 are under-fractionated or dominant (i.e., retain more genes than other regions) while PaxiN 1, PaxiN 2, Sl 1 and Sl 2 are over-fractionated. Note that for the under-fractionated regions PaxiN 3 and Sl 3, gene fractionation is identical. To better identify regions that are differentially fractionated in the over-fractionated regions, synteny lines between genomic regions of *Petunia* and tomato vs. grape in PaxiN 3 and Sl 3 are not shown. Gene fractionation is significantly less complete in *Petunia* in comparison with tomato. In addition to shared fractionation events, we also observed independent gene fractionation in the two lineages with more fractionation events in tomato. Of the 13 regions differently fractionated between *Petunia* and tomato, 9 regions are independently retained in *Petunia* while only 3 regions are independently retained in tomato. In addition, *Petunia* retains more genes in over-fractionated regions than tomato (compare PaxiN 1 and PaxiN 2 with Sl 1 and Sl 2). These results are also validated by comparing the same genomic regions but with a different order (Figure 10). This representation highlights synteny between orthologs and allows us to observe independent fractionation in *Petunia* and tomato as well as shared fractionation events between these two lineages.

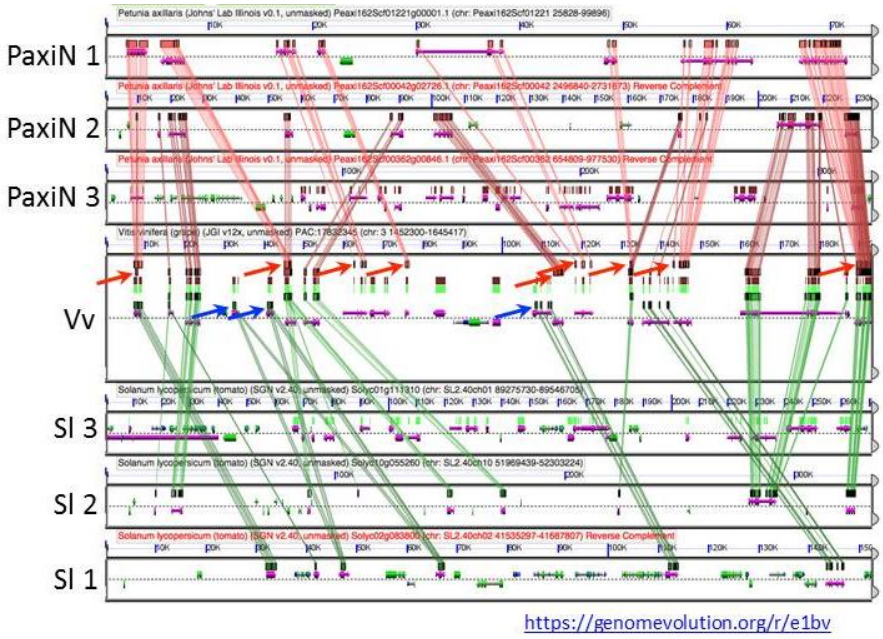


Figure 9: Evidence that divergence of tomato and *Petunia* lineages happened prior to the completion of fractionation. Microsynteny analysis of one region of grape (Vv) to three syntenic orthologous regions of *Petunia* (PaxiN) and tomato (Sl). Description of how to read this figure is described in summary for figures 4-6. Red arrows indicate genes retained in *Petunia* and fractionated in tomato. Blue arrows indicate genes retained in tomato and fractionated in *Petunia*. Analysis may be regenerated: https://genomevolution.org/r/e1bv

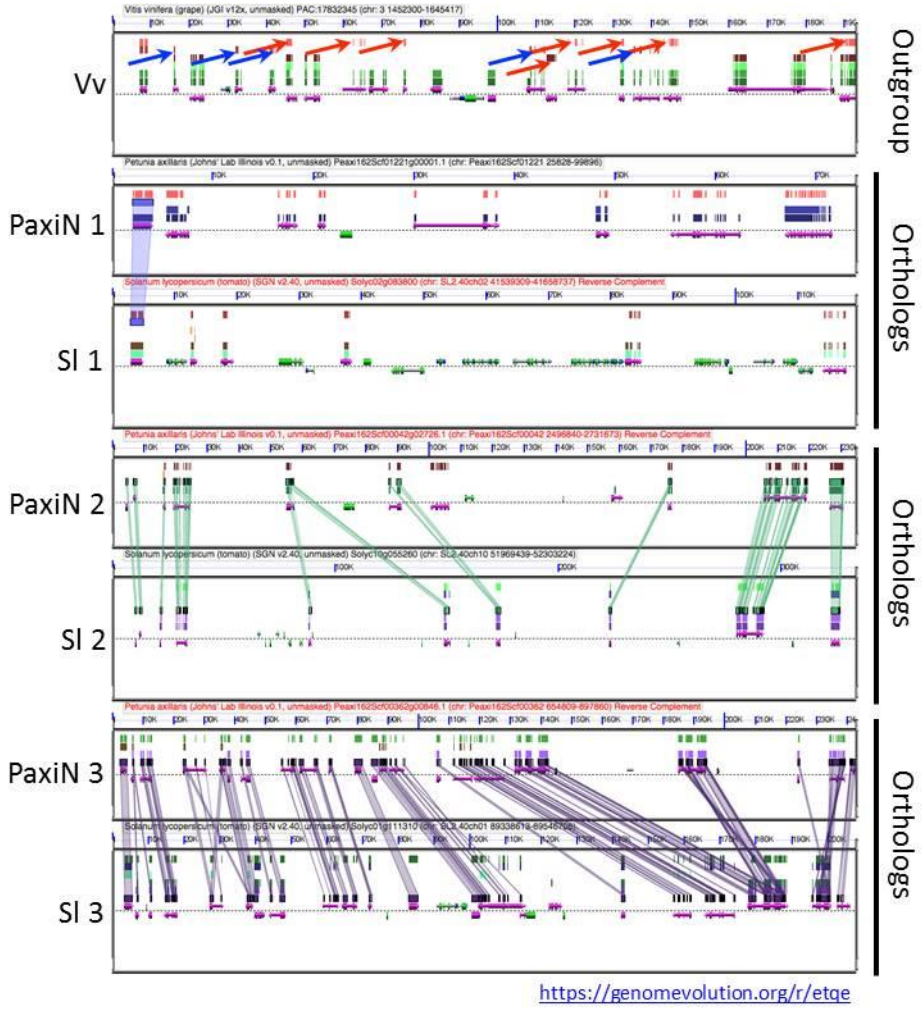


Figure 10: Fractionation partially occurred independently in each lineage, with more fractionation events in the tomato lineage. Microsynteny analysis description is similar to figure 9, but with different order to the regions. Red arrows indicate genes retained in *Petunia* and fractionated in tomato. Blue arrows indicate genes retained in tomato and fractionated in *Petunia*. Note that there is only a single syntenic gene set shared between orthologous regions PaxiN 1 and Sl 1. Analysis may be regenerated: https://genomevolution.org/r/etge

We finally included *Mimulus guttatus* in the analysis of microsynteny between *Petunia* and grape (Figure 11). *Mimulus guttatus* was shown to lack the *Solanum* paleohexaploidy event and to have an independent paleotetraploidy event (Ibarra-Laclette et al., 2013). Our analysis is consistent with this previous study as we could observe an independent fractionation in *Mimulus* in comparison with *Petunia* and tomato.

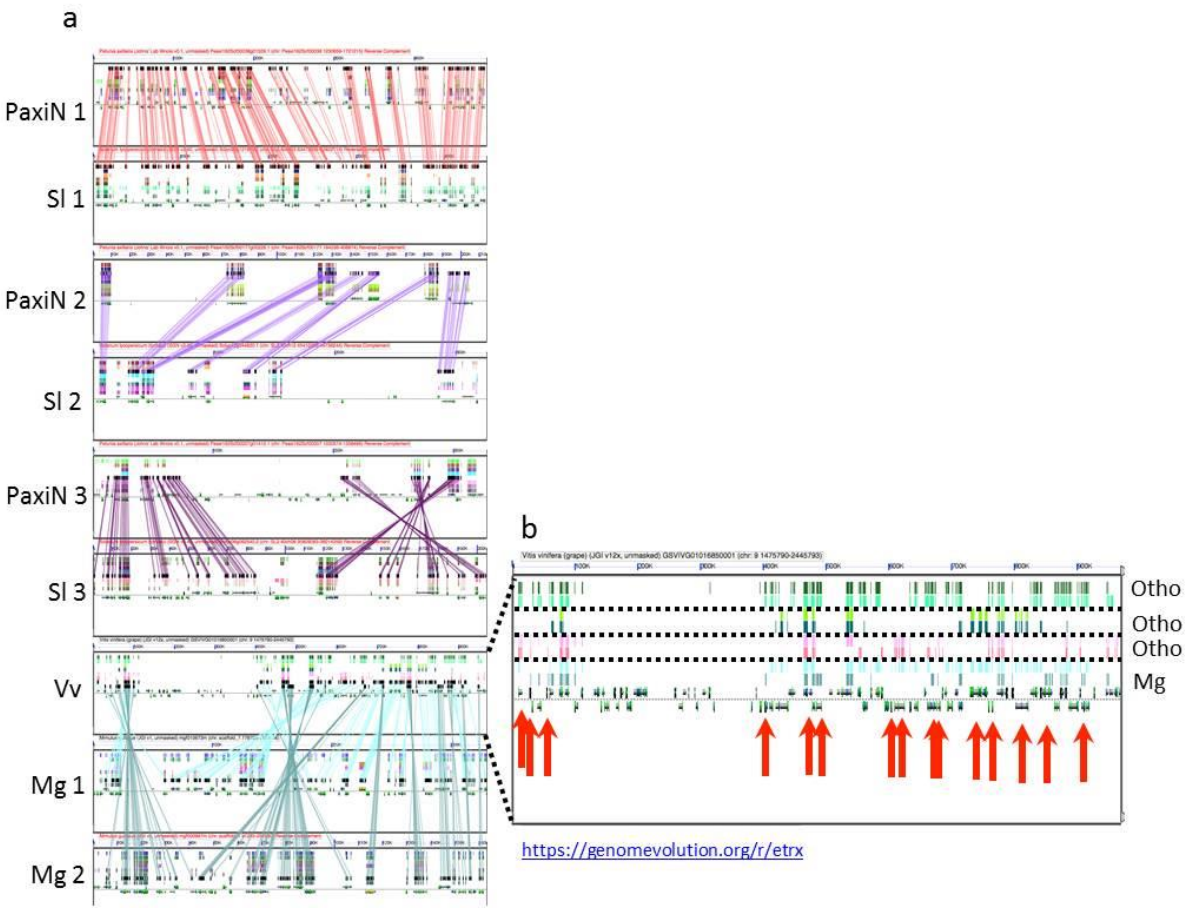


Figure 11: Independent gene fractionation in *Mimulus* in comparison with *Petunia* and tomato.
(a) Microsynteny analysis of *Petunia* (PaxiN) and tomato (Sl), *Mimulus guttatus* (Mg) and grape (Vv). (b) Close up of microsynteny panel for grape with red arrows marking independent fractionation of the *Mimulus* lineage and tomato/*Petunia* lineages. Analysis may be regenerated: https://genomevolution.org/r/etrx

DISCUSSION

Here we present the first analysis of paleopolyploidy for *Petunia*, which represents the sister-group to the larger and more diverse x=12 crown-group clade of the Solanaceae family (Särkinen et al., 2013) (Figure 1). Our study confirms the until now ambiguous Solanaceae paleohexaploidy event initially inferred in the analysis of the tomato genome. Further, we demonstrate that the *Petunia* lineage has experienced (at least) two rounds of paleopolyploidization, the older gamma hexaploidy event (Figure 2), which is shared with other Eudicots (Jaillon et al., 2007), and the more recent Solanaceae paleohexaploidy event which is shared with tomato and other x=12 species (Figures 3-8) (Tomato Genome Consortium, 2012). We have shown that the process of gene fractionation that facilitates the return to a diploid state (also known as diploidization), occurred, in part, independently in *Petunia* and tomato despite the shared Solanaceae paleohexaploidy event, similar to what has been observed in *Saccharomyces* yeasts but until now not yet described in flowering plants.

Validation of the Solanaceae paleohexaploidy event in the genome of tomato.

Genome collinearity analysis between tomato and grape is ambiguous in establishing a complete independent genome triplication in the Solanaceae or, potentially evidence for segmental duplication (i.e. some regions duplicated and others triplicated) (Ibarra-Laclette et al., 2013; Figure 8). Our study notably showed that the *Petunia* genome is less fractionated than the tomato genome (Figure 9). This would explain the difficulties in identifying a clear 3:1 orthologous syntenic relationship between tomato and grape by performing genome colinearity analysis (Ibarra-Laclette et al., 2013; Figure 8). In view of our results, we could hypothesize that an independent genome triplication occurred in the Solanaceae family but the high degree of gene fractionation in tomato makes it difficult to identify the three orthologous regions. A similarly high degree of gene fractionation could also be explain the interpretation of paleopolyploidy in the potato genome. Both tomato and potato genomes are relatively complete, suggesting that the lack of triplicated regions is not due to incomplete genome assembly. Indeed, genomic analysis showed that one genomic region of grape is (only) syntenic to two regions in the potato genome, leading to an inaccurate inference of only paleotetraploidy (The Potato Genome Consortium, 2011).

Gene fractionation occurred independently in *Petunia* and tomato genomes following their divergence despite the shared paleopolyploidy history

Our analysis is the first detailed description of an independent gene fractionation from a common polyploidy event in plants. While some fractionation of gene content probably occurred prior to the divergence of these lineages, as indicated by the shared gene fractionation events in *Petunia* and tomato genomes, our results demonstrate that gene fractionation also continued independently in these two lineages (Figure 9 and Figure 10) following their divergence, even though they shared a paleohexaploidy event (Figure 2 and Figure 3).

First, we observed in *Petunia* and tomato the presence of two classes of genomic regions with distinct levels of gene fractionation, the over-fractionated (i.e., more genes are lost) and the under-fractionated (i.e. fewer genes are lost) regions (Figure 9, Figure 10 and Figure 11), supporting a biased fractionation process following polyploidy. In each microsynteny analyses, one genomic region is still under-fractionated (PaxiN 3 and Sl 3 in Figures 9 and 10, PaxiN 1 and Sl 1 in Figure 11) while the two other are over-fractionated. This indicates that 3 sub-genomes, with contrasting gene contents, coexist in *Petunia* and tomato genomes. The same situation is found in other

paleopolyploids such as in *Brassica rapa* (Wang et al., 2011; Tang et al., 2012), a diploid species with three subgenomes originating from a whole genome triplication (Wang et al., 2011). A two-step theory implicating a differential subgenome evolution was proposed to explain the genome triplication event in *B. rapa* and also more broadly to explain fractionation after a paleohexaploidy event (Lyons et al., 2008). The first step of this model involves the formation of a tetraploid with two subgenomes. The two new subgenomes then experience loss of duplicated genes resulting in two fractionated subgenomes. The second step consists of the formation of the new tetraploid between the previous fractionated diploid genome and a new diploid genome, which experienced another round of gene fractionation. The final diploid genome will thus contain three subgenomes, two subgenomes that experienced two rounds of fractionation (and are thus over-fractionated plus one subgenome with only one round of fractionation (under-fractionated).

Second, *Petunia* genome is less fractionated than the tomato genome (Figure 9). Indeed, Petunia retains more genes in over-fractionated regions than tomato while difference in the dominant region is less profound. Analysis of gene fractionation between tomato and potato as well as between *Petunia* and potato would be helpful to decipher whether gene fractionation is similar in tomato and in potato and to confirm if fractionation in Petunia and potato is also independent. This observation is similar to what happens in yeast where early-branching clades, in our study *Petunia*, show a different pattern of fractionation than latter branching species, represented here by tomato and eventually by potato (Scannell et al., 2006). It would be interesting to compare gene fractionation between *Petunia* and other "later branching" Solanaceae species such as tobacco in order to confirm this similarity. *Petunia* is well situated to be a better genome comparator for Solanaceae genomes than the tomato genome, due to its low degree of gene fractionation. Nevertheless, some aspects of the *Petunia* genome structure and evolution still need to be elucidated in order to represent an ideal reference for genomics, specifically, a higher quality assembly into pseudomolecules.

METHODS

Whole genome collinearity analysis

Whole genome dot plots were performed using SynMap in the comparative genomics platform, CoGe (Lyons et al., 2008). Each analysis generates a tiny-url that links back to SynMap configured to regenerated the analysis exactly as shown. URLs include embedded options for configuring SynMap and for these analyses, the syntenic path assembly and coloration of syntenic gene pairs by synonymous mutation rate (Ks; CodeML; http://abacus.gene.ucl.ac.uk/software/paml.html) were extensively used. For details on how to use SynMap, please see Tang and Lyons (2012).

Microsynteny analysis

Microsynteny analysis was performed using GEvo in the comparative genomics platform, CoGe (Lyons and Freeling, 2008). As with SynMap, each GEvo analysis generates a tiny-url to regenerate the analysis exactly as configured; these links are provided in the figure legends.

ACKNOWLEDGEMENTS

LG is supported by a H2020 Marie Curie fellowship (H2020-MSCA-IF-2014). EL and HT are supported by USDA 2013-00984 and NSF IOS – 1339156. HT is supported by the "100 Talent Plan" award by the Fujian provincial government. CoGe is part of the Powered by iPlant program (NSF DBI – 1265383). MES was funded by a NWO-VIDI grant.

REFERENCES

Barker, M.S., Vogel, H., and Schranz, M.E. (2009). Paleopolyploidy in the Brassicales: analyses of the Cleome transcriptome elucidate the history of genome duplications in Arabidopsis and other Brassicales. Genome biology and evolution **1,** 391-399.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., and Wincker, P. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345,** 950-953.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Brautigam, A., Weber, A.P., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.G., Xu, X., and Schranz, M.E. (2013). The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. The Plant cell **25,** 2813-2830.

Garsmeur, O., Schnable, J.C., Almeida, A., Jourda, C., D'Hont, A., and Freeling, M. (2014). Two evolutionarily distinct classes of paleopolyploidy. Molecular biology and evolution **31,** 448-454.

Ibarra-Laclette, E., Lyons, E., Hernandez-Guzman, G., Perez-Torres, C.A., Carretero-Paulet, L., Chang, T.H., Lan, T., Welch, A.J., Juarez, M.J., Simpson, J., Fernandez-Cortes, A., Arteaga-Vazquez, M., Gongora-Castillo, E., Acevedo-Hernandez, G., Schuster, S.C., Himmelbauer, H., Minoche, A.E., Xu, S., Lynch, M., Oropeza-Aburto, A., Cervantes-Perez, S.A., de Jesus Ortega-Estrada, M., Cervantes-Luevano, J.I., Michael, T.P., Mockler, T., Bryant, D., Herrera-Estrella, A., Albert, V.A., and Herrera-Estrella, L. (2013). Architecture and evolution of a minute plant genome. Nature **498,** 94-98.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., Wincker, P., and French-Italian Public Consortium for Grapevine Genome, C. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449,** 463-467.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., and dePamphilis, C.W. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473,** 97-100.

Langham, R.J., Walsh, J., Dunn, M., Ko, C., Goff, S.A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. Genetics **166,** 935-945.

Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. The Plant journal : for cell and molecular biology **53,** 661-673.

Lyons, E., Freeling, M., Kustu, S., and Inwood, W. (2011). Using genomic sequencing for classical genetics in E. coli K12. PloS one **6,** e16717.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D., and Freeling, M. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant physiology **148,** 1772-1781.

Sarkinen, T., Bohs, L., Olmstead, R.G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC evolutionary biology **13,** 214.

Scannell, D.R., Byrne, K.P., Gordon, J.L., Wong, S., and Wolfe, K.H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature **440,** 341-345.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America **108,** 4069-4074.

Schranz, M.E., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Current opinion in plant biology **15,** 147-153.

Soltis, D.E., Bell, C.D., Kim, S., and Soltis, P.S. (2008). Origin and early evolution of angiosperms. Annals of the New York Academy of Sciences **1133,** 3-25.

Tang, H., and Lyons, E. (2012). Unleashing the genome of brassica rapa. Frontiers in plant science **3,** 172.

Tang, H., Woodhouse, M.R., Cheng, F., Schnable, J.C., Pedersen, B.S., Conant, G., Wang, X., Freeling, M., and Pires, J.C. (2012). Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics **190,** 1563-1574.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., and Harmon, L.J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. The New phytologist **207,** 454-467.

Thomas, B.C., Pedersen, B., and Freeling, M. (2006). Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome research **16,** 934-946.

Tomato Genome, C. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485,** 635-641.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weisshaar, B., Liu, B., Li, B.,

Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I.A., Batley, J., Kim, J.S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S.J., Choi, S.R., Lee, T.H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong, Z., Zhang, Z., and Brassica rapa Genome Sequencing Project, C. (2011). The genome of the mesopolyploid crop species Brassica rapa. Nature genetics **43,** 1035-1039.