

Evolution of Plant Genomes

Introduction

Modern plant genomes are quite variable

- ~150 megabase (Mb) *Arabidopsis thaliana* genome.
- 18,000 Mb hexaploidy wheat genome.

Why understanding the evolutionary history of genomes?

- Applied genetics perspective
- Application of comparative genomics for gene discovery.
 - *Arabidopsis terminal flower 1 (tfl1)*
 - Encodes a transcription factor
 - It controls indeterminacy/determinacy phenotype
 - *Arabidopsis tfl1* as a reference gene
 - Homolog of this gene also controls the phenotype in other
 - Dicot species
 - Snapdragon (*Antirrhinum*)
 - Pea (*Pisum sativum*)
 - Monocot
 - Rice (*Oryza sativum*)
- Mutations all results in a determinate phenotype

The relevant question

- To what degree are functional genes in one plant species conserved in another species?
 - Important to trace
 - Evolutionary events
 - Related to current organization of plant genomes

Polyploidy and the Construction of Plant Genomes

Whole genome duplication (WGD)

- Common event in the evolution of plant species
 - Entire genome doubles in size
 - Duplicates the same genome
- Two related diploid species merge
 - During mitosis
 - Chromatids migrate to separate daughter cells
 - If they move to only one cell
 - The cell will be a tetraploid
- If the 2x duplicate cell is involved in reproduction
 - Resulting gamete
 - 2x the normal number of cells
 - If 2x gamete unites
 - Offspring will be tetraploid

Polyploidy

- An organism that contains extra sets of chromosomes.
 - Tetraploids
 - Cultivated potato
 - Alfalfa
- For a success of any polyploidy
 - It must generate balanced gametes.
 - The same number of chromosomes as other gametes
- Embryos from gametes with the same number of gametes
 - Successfully survive

Other Polyploids

- Allopolyploids
 - Two species with very similar chromosomal structure and number intermate.
 - After chromosomal doubling organism, genome will have
 - Number of chromosomes equal to the sum of the number of chromosomes from each of the parent species.
- Examples of allopolyploid species
 - Tetraploid durum wheat ($x=14$)
 - Hexaploid bread wheat ($x=21$).
- Durum wheat arose from
 - Union of two diploid species ($x=7$) species
- Bread wheat arose from
 - Diploid wheat species with the tetraploid wheat species

Constructing the *A. thaliana* genome as a model for eudicot genome evolution

- With the whole genome sequence
 - Study the duplication history of the *A. thaliana* genome.
 - Ancestral duplication signatures could be inferred
 - Blastp analysis
 - Protein vs. protein comparison
 - Identifies gene pairs
 - E-value < -10 used in Fig. 1
 - ***Suggests genes are ancestrally related***
 - Duplicates are mapped relative position in the genome
 - Displayed using a dot blot
 - Blocks observed
 - Linear arrayed dots
 - Form a diagonal in the dot blot,
 - Signatures of a duplication event

Figure 1A

- Early comparison of the proteins in the *A. thaliana* genome
 - Red and green diagonals in the upper right panel
 - Block α_3
 - Chromosome 1 vs. chromosome 1 block
 - Signature of a duplicated block of genes
 - Genes that have the same conserved order
 - At two ends of the *A. thaliana* chromosome 1
 - Block α_5
 - Another pairs of duplicated genes on chromosome 1
 - Block α_8
 - Shared block on chromosomes 1 and 3
 - Block, α_{11}
 - Largest block
 - Ends of chromosomes 3 and 2
 - Total
 - 27 major duplicated blocks
 - Strong signals
 - *Signals of a recent duplication*

So how does this relate to the mechanism of genome construction?

- *A. thaliana* underwent a WGD
 - Chromosomes were broken
 - Rearranged into new chromosomes
 - New chromosomes developed
 - *Represent blocks of DNA from the progenitor species*

Progenitor *Arabidopsis* genome

- How it was modified by the duplication event
- Compare to species that is evolutionary close.
 - *A. lyrata*
 - 8 chromosomes
 - *A. thaliana*
 - 5 chromosomes
- Genetic maps developed using shared loci were

Fig. 2

- Five *A. thaliana* chromosomes
 - Constructed from ancestral genome with eight chromosomes
- At Chr 1
 - Blocks of AlyLG1 + AlyLG2
- At Chr II
 - Blocks of AlyLG3 + AlyLG4.
- Conclusion
 - ***Two species with different chromosome numbers consist of the same chromosomal blocks***

Fig. 1B – Early duplication events

- Shows evidence of more ancient duplications
 - 27 α duplications reoriented
 - Notice block $\alpha 5$
 - Two duplicates blocks in the same order
 - Two in an opposite orientation
 - Presumed ancestral order derived from these four blocks
 - Same procedure that uncovered the α blocks.
 - Two types of blocks discovered.
 - 22 β blocks
 - *Another duplication event in the A. thaliana lineage*

The 7 γ blocks

- Controversial
 - Hypothesis 1
 - Early duplication in the angiosperm lineage
 - Hypothesis 2
 - Duplication after the split of monocots and dicots
- Grapevine genome sequenced
 - Evidence from the genome appears to have resolved this question
 - Grape
 - Ancestor of the rosids
 - Group of species included *A. thaliana*.
 - Blast and dot blot analysis of grape genome

Figure 3

- Any genes shared with two other regions of the genome
 - Grape genome has a hexaploid history
- How about other species
 - Signal of hexaploidy is detected
 - **Figure 4**
 - Grape and poplar genomes were compared
 - Only triplicated regions in grape used
 - Triplicated regions
 - Two copies in poplar
 - *Hexaploid ancestry concept is supported*
 - *Poplar underwent an additional WGD after its divergence from the grape lineage*

Shared duplications in dicot and monocot analysed

- Grape and rice orthologs analyzed
 - Hypothesis 1
 - Rice shared the hexaploid ancestry
 - 3-to-3 relationship
 - Not observed
 - Hypothesis 2
 - Rice does not share the same hexaploid ancestry
 - 3-to-1 relationship observed
 - Conclusion
 - *Monocots and dicots do not share the same hexaploid history.*

(**Note:** See Tang et al. 2008. Genome Research 18:1944 for an alternative perspective.)

Summary of Eudicot Evolution

- Two diploid mate
 - Tetraploid species developed
- Tetraploid species mated to another diploid
 - Produce the ancestral hexaploid
 - All subsequent eudicots derived from this ancestor
 - *Signatures of the same duplications*
 - *Should be observed in their genome history*

Monocot genome evolution.

- Monocots also have a duplication history.
 - **Figure 5**
 - Compared rice and maize.
 - Maize chromosomes (y-axis) as the reference
 - Most rice genes found in two copies
 - Rice chromosomes (x-axis) as the reference
 - Blocks found three or four times in maize.
 - Conclusion
 - *WGD event in the history of monocots*
 - *An additional duplication occurred in the maize lineage.*

Unified model of grass evolution – developing the ancestor

- Based on sequences of genome sequences of
 - Rice
 - Sorghum
 - Brachypodium (a model grass species)
 - Maize
- 56-73 MYA
 - Ancestral grass species containing five chromosomes
 - Duplicated
 - Genome with ten chromosomes appeared
 - Then
 - A4 and A6 fractionated
 - Chromosomes A4, A6, and A2 appear
 - A7 and A10 fractionated
 - Chromosomes A7, A10, and A3 appear
 - Paleopolyploid developed
 - 12 chromosomes
 - *Progenitor of all of the modern grasses*

Unified model of grass evolution – developing the lineages

- Rice genome structure
 - Represents the ancient paleotetraploid.
 - Basic set of chromosomes
 - Building blocks for other genomes

Figure 6

- Breakage/translocation/fusion events
 - Involve chromosomal fragments from the n=12 ancestor.
 - Developed
 - Brachypodium
 - Poideae (representing the wheat lineage)
 - Panicoideae (representing the maize/sorghum lineage)
 - Panicoideae
 - Simplest history
 - Arose from only four breaks
 - Other lineages
 - More complex patterns of evolution
 - Maize genome
 - Underwent additional duplication
 - Additional breakage/translocation/fusion events
 - ***Constructed the modern maize chromosomes***

Summary

- Plant genomes
 - A long history of genome duplications
 - Unlike animal and fungal genomes,
- **Figure 7**
 - Illustrates the duplication history
 - (The γ event should be moved to the origin of the eudicot lineage.)
 - Significant role of WGD in development of plant species
 - Many duplications appear 55-70 MYA
 - Transition point
 - Cretaceous and Tertiary periods
 - Mass extinction of species
 - Hypothesis
 - Duplications gave plants the needed gene repertoire
 - *To survive this extinction*
 - *Flourish on earth*

(see Fawcett et al. 2009. PNAS USA 106:5737)

- **Figure 8**
 - Additional species were analyzed
 - Extended the analysis to deeper phylogeny
 - Additional duplication events determined
 - Ancestral seed plants
 - ζ at ~330 MYA
 - Ancestral angiosperms
 - ϵ at ~220 MYA

The Gene-based Evolution of Duplicated Genes

If duplications are a major signature of plant genomes

- Copy number of genes should equal the number of rounds of duplication.

Table 1

- Number of genes found within plant species
 - Complete genome sequence
 - If the hexoploidy concept is true for dicots, and
 - Grape only contains this hexaploid event
 - Estimate
 - Ancestral dicot contains ~10,000 genes (=30,000/30).

Similarly

- Poplar underwent an additional duplication,
 - Theoretically # of genes = 60,000 genes
- *A.thaliana* underwent two duplications
 - Theoretically # of genes = 120,000 genes
- ***Not observed***

Monocot calculations

- Rice, Brachypodium, and sorghum only contain a duplication event
 - Number of ancestral monocot genes
 - 15,000 (=30,000/2).
 - Maize
 - Additional duplication event
 - But has undergone a reduction to ~30,000 genes
- Conclusion
 - Necessary to reduce the number of genes to ensure the success of the species.

Diploidization.

The polyploid past history of plants

- Surprising result for Arabidopsis and rice genomes
 - Why??
 - Selected for sequencing because of their small genome sizes

Consequences of polyploidy?

- Doubling or tripling of the number of chromosomes
 - Evident for monocots.

Fate of the additional gene set from the WGD

- Concept
 - Species cannot maintain the entire set of duplicate chromosomes
 - New genes a problem
 - Generate deleterious mutations
 - Compromises the fitness of a genome
 - Genome must transition back to its original state.
 - Process is called
 - *Diploidization.*

To revert back to the diploid state

- Many duplicate genes must be eliminated from the gene set
 - But a recently duplicated genome
 - Soybean
 - Withstands the extra copies
 - Genome about 2X the basic set of 30,000 genes of hexaploid ancestral eudicot

Events associated with diploidization

- Duplicate genome must change its chromosome pairing pattern
 - After the duplications,
 - Four chromosomes pair
 - Form quadravalents
 - Chromosomal structure must be changed so
 - Bivalents must be formed
 - Result
 - Doubling of the chromosome number
 - Seen for the monocot lineage
 - Once bivalents are formed
 - Gene sets can evolve
 - Processes
 - Deletions and chromosomal rearrangements

Duplicate genes can undergo specific changes

- Common fate
 - Gene death of new copies
 - Loses associated with
 - Chromosomal breakage
 - Rearrangements.
 - Result
 - New basic set of chromosomes and genes will have appeared

Duplicate genes fate differs

- Some are retained as multicopy
 - Up to the ploidy level for that species
- Other reduced to only a single copy

“Deletion resistant” genes

- Not reduced to single copy
 - Dosage dependent
 - Mainly encode
 - Transcription factors
 - May lead to
 - Complex morphologies

“Duplication resistant” genes

- Must be maintained as single copy
 - Mainly encode
 - Enzymes or genes of unknown function

Table 1. The estimated number of genes in sequenced plant genomes.

Species	Estimated # of Genes (from www.phytozome.net)
<i>Eudicots</i>	
Cucumber	21,491
Cassava	47,164
Poplar	41,000
Medicago	50,692
Soybean	66,153
Arabidopsis	27,343
Papaya	27,332
Grape	30,434
Mimulus	25,530
<i>Monocots</i>	
Sorghum	34,496
Maize	32,540
Brachypodium	25,532
Rice	31,500

Developing new functions

Duplicate set of genes cannot be maintained

- Deleterious mutations can arise
- Duplicate genes are modified
 - Changes will provide
 - New functions
 - Altered altered functions
 - New functions may lead to the evolution of the species
 - Higher level of fitness
 - Evolutionary modifications of duplicate genes

Neofunctionalization.

- One duplicate gene maintains its original function
- Second gene evolves a function
 - May increase the adaptability of an individual

Subfunctionalization

- Modifies the duplicates
- Basic structure of both copies altered
 - Expression pattern of the gene changes
 - Results in a higher level of the protein production
- Alternately, the function of the original gene is maintained
 - Structure of both copies is significantly changed.
 - New copies retains
 - Part of the original function
 - Two genes work together
 - Function of the original gene maintained

Synteny: The Result of WGD and Reconstructing Plant Genomes

Synteny among plant species.

- Major result of the duplication history
 - Synteny
 - Maintenance of gene order between two species
 - Classic approach to synteny
 - Based on shared markers mapped onto two different species.
 - Macrosynteny is detected by
 - Large scale chromosomal blocks shared by two species.

Fig. 9

- Example of macrosynteny
 - Tomato and eggplant
 - Eggplant linkage group 4
 - Evolutionarily related to tomato
 - Linkage groups 10S and 4L.
 - Highly conserved marker order over many centimorgans of the two genomes

Genetic mapping of shared genes

- First method of comparing species
- Only way to compare species that have not been sequenced
- Many examples of synteny mapping in plants.
- The power of synteny mapping
 - Discovery of shared loci from two species
 - Control the same phenotype
 - Map to the same genetic location.

Fig. 8 again

- Major QTL for fruit striping
 - Eggplant linkage 4.
 - Previous work with tomato
 - Major QTL
 - Linkage group 10 of tomato
 - Syntenic marker and QTL observed here
- Hypothesis
 - Multiple loci are shared in the same macrosyntenic order
 - Same ancestral gene is controlling this trait in these two species.

Leveraging knowledge in one species for gene discovery in a second species

- Phenotypic traits mapped extensively in one species
 - Points a researcher working on a second species
 - Likely location of a similar gene in second species.
 - Leverage is
 - Great aid for genetic discovery
 - For species in where the discovery of important genetic factors are limited by a lack of funding

Figure 1. Dot blot display revealing duplication events. (from Bowers et al. 2003. Nature 422:433)

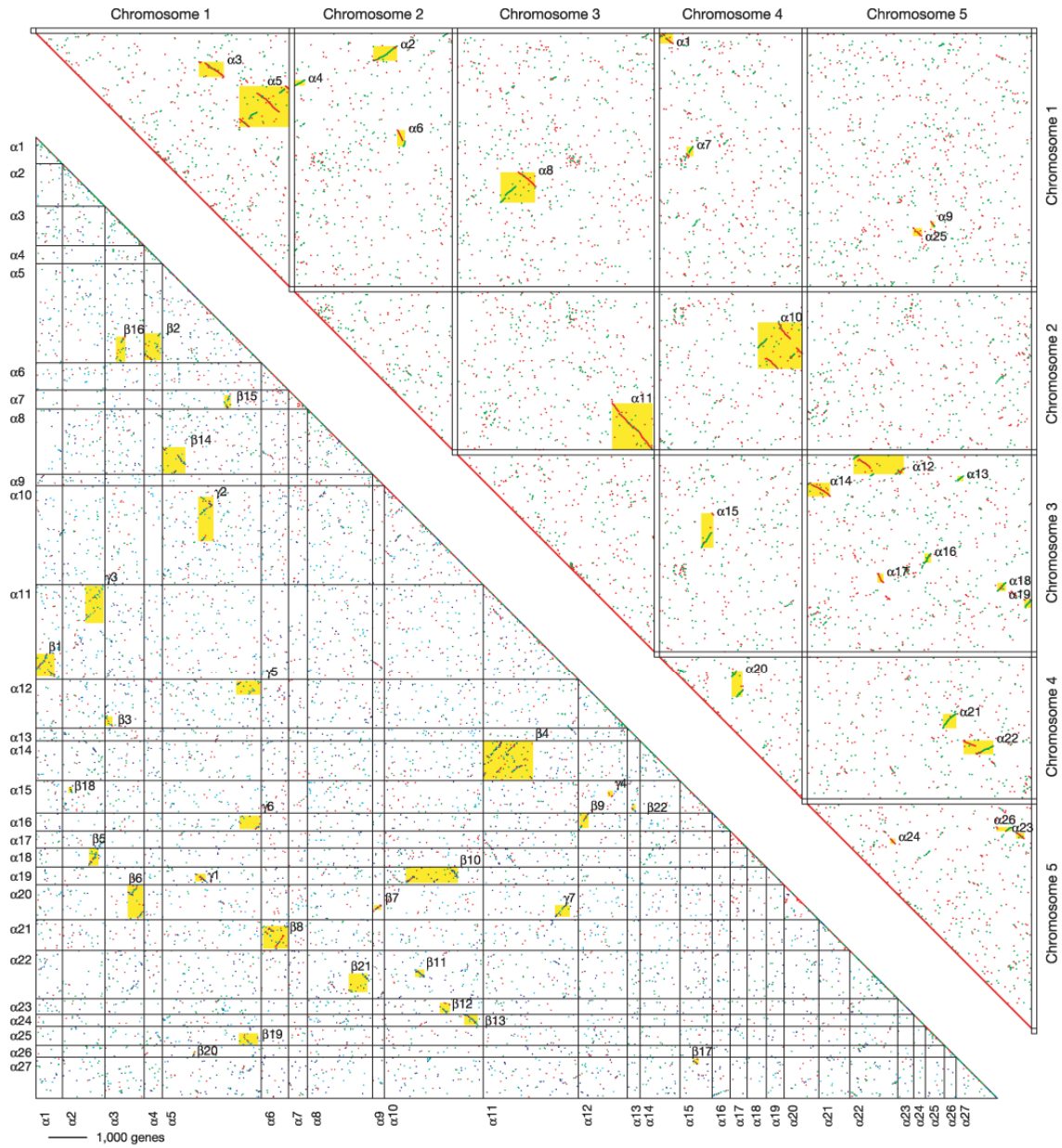


Figure 1 Arrangement of duplicated protein-encoding genes in *Arabidopsis thaliana*. Top left: the composition of the 26 large duplications (at left and bottom). Top right: 20 smaller duplications. Bottom axes represent 26,028 genes in their chromosomal duplications (see text) are highlighted. Colours show how the four Arabidopsis genomes contributed to duplications, distinguishing contributions to opposite (green) transcriptional orientations. For further analysis, 57 adjacent duplicated genes at left and bottom respectively from the: (1) lower-numbered chromosomes regions with opposite orientation and order explicable by localized inversions (red); (2) higher- and lower-numbered chromosomes (light blue); (3) lower- and higher-numbered chromosomes (dark blue); (4) higher-numbered chromosomes (green). Eight shorter duplications were plotted lower left and duplications. Higher-resolution versions of the figure and lists of gene orders are available (see Supplementary Information).

Figure 2. Comparative physical map of *A. thaliana* and the genetic map of *A. lyrata*. (from: Yogeewaran et al. Genome Research 15:505)

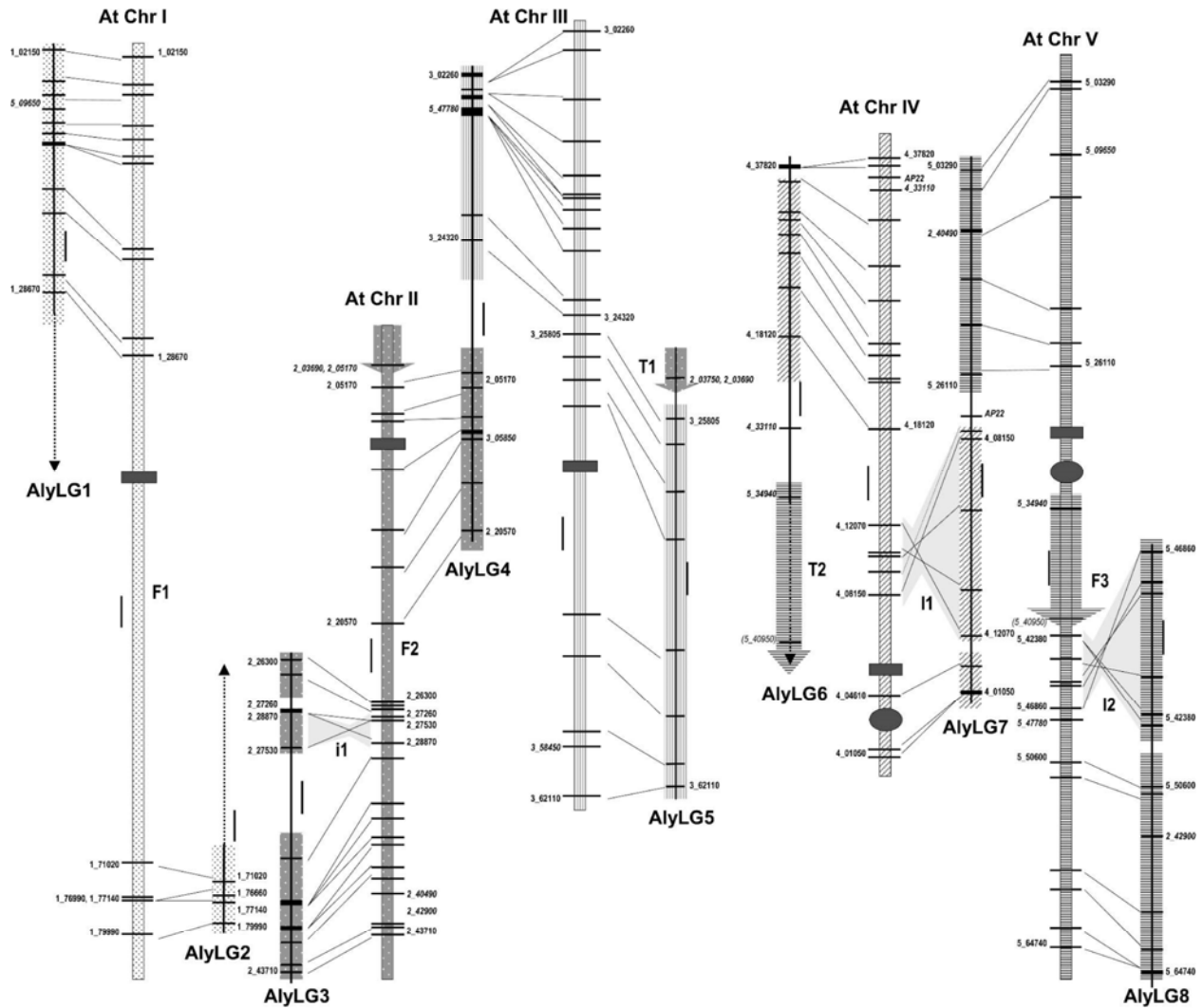


Figure 2. Colinearity of *A. lyrata* linkage map with the *A. thaliana* genome. *A. thaliana* chromosomes (At Chr I –V) are represented as patterned bars (drawn to scale, 1 unit = 1 Mbp; gray rectangles, centromeres; gray circles, heterochromatic knobs). *A. lyrata* linkage groups (Aly LG 1 –8) are shown in black (drawn to scale, 1 unit = 5cM). Sixteen colinear blocks are highlighted with the same pattern as the *At* chromosome to which they correspond. Markers defining the ends of each colinear block are shown on the map in black lettering. Markers mapping with LOD score less than 3.0 are featured in parentheses. Italicized markers map to translocated or nonsyntenic regions in *A. lyrata*. Translocations T1 and T2 are highlighted by arrows whose patterns correspond to the *At* chromosome where their colinear region lies. Major inversions I1 and I2 and minor inversion i1 are highlighted in light gray. Three chromosomal fusions are denoted as F1-F3.

Figure 3. Dot blot representation of duplicate regions of the grapevine genome. (from: Jaillon et al. 2007. Nature 449:463)

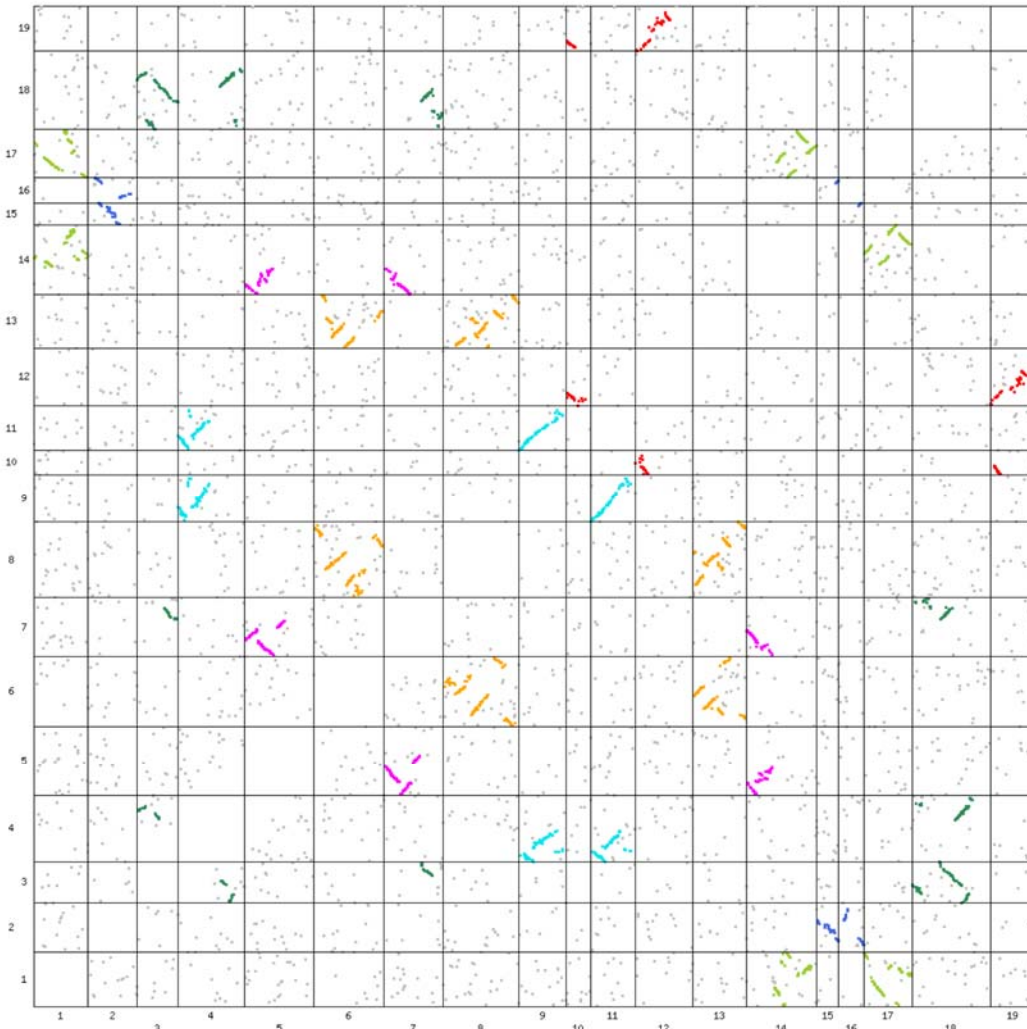


Figure S5. The grape genome originated from a polyploidy event that joined three ancestral genomes. The nineteen chromosomes of grape are represented on both the x and y axis. Dots represent the positions of paralogous pairs of genes. For clarity, intrachromosomal paralogs are not shown. Clusters of paralogs form a succession of dots, that indicate that the gene order of the ancestral genome was locally maintained. These clusters are painted in seven colours. Each colour marks paralogous blocks, that were colinear in the ancestors of the three constituents of the grape genome. Some regions are not painted in triplicate in this grid, either because a whole region is not visible in synteny with two others in the present-day grape genome (too many rearrangements or gene loss), or because one or two syntenic regions lie in supercontigs which are still not anchored.

Figure 4. Comparison of the triplicated blocks and the Poplar genome.
(from: Jaillon et al. 2007. Nature 449:463)

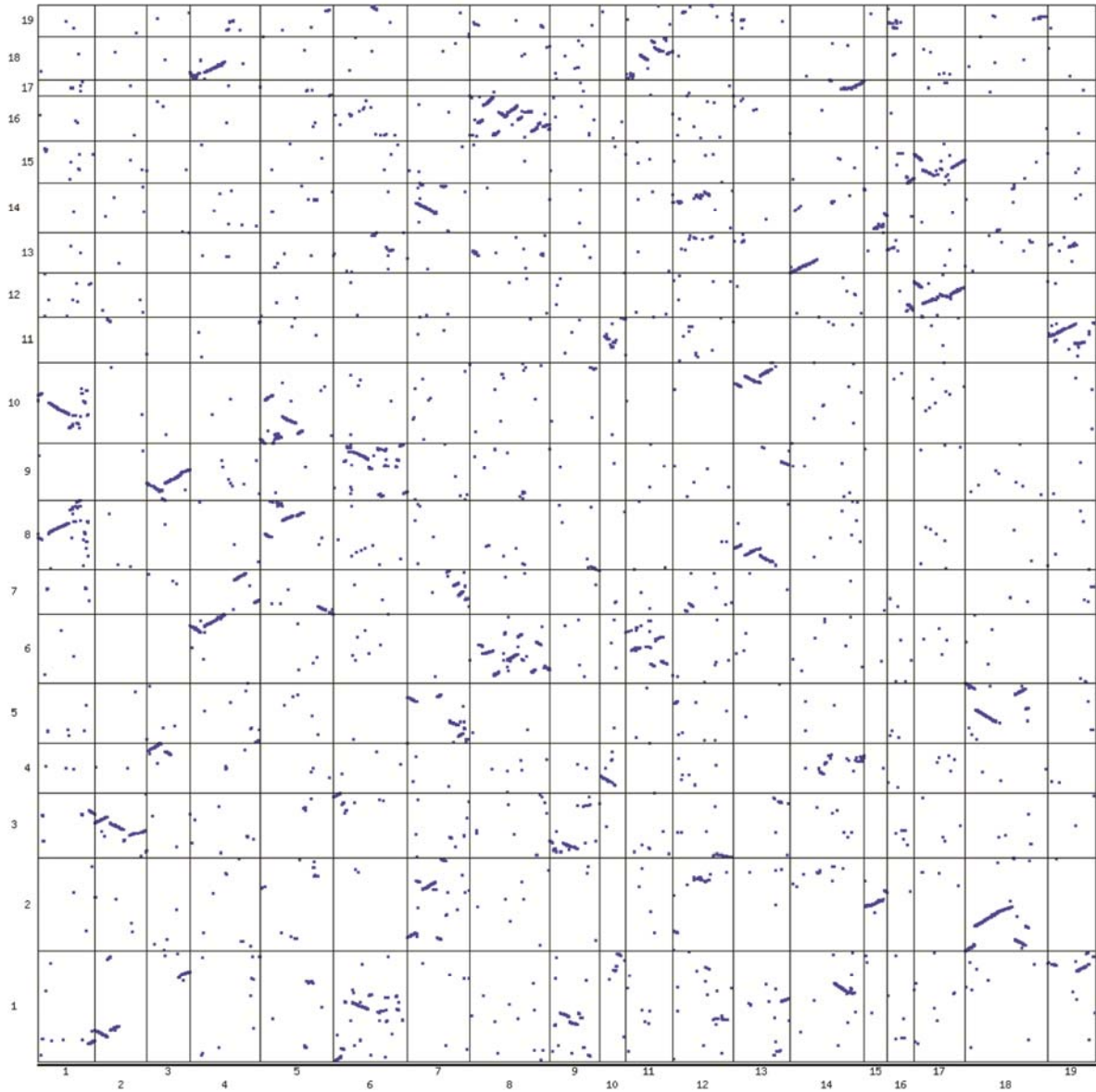


Figure S6. The distribution of 8,604 orthologous genes between *Vitis vinifera* (x axis) and *Populus trichocarpa* (y axis) chromosomes.

Figure 5. A comparison of maize and rice duplication events. (from: Wei et al. (2007) PLoS Genetics 3(7):e123, 1254)

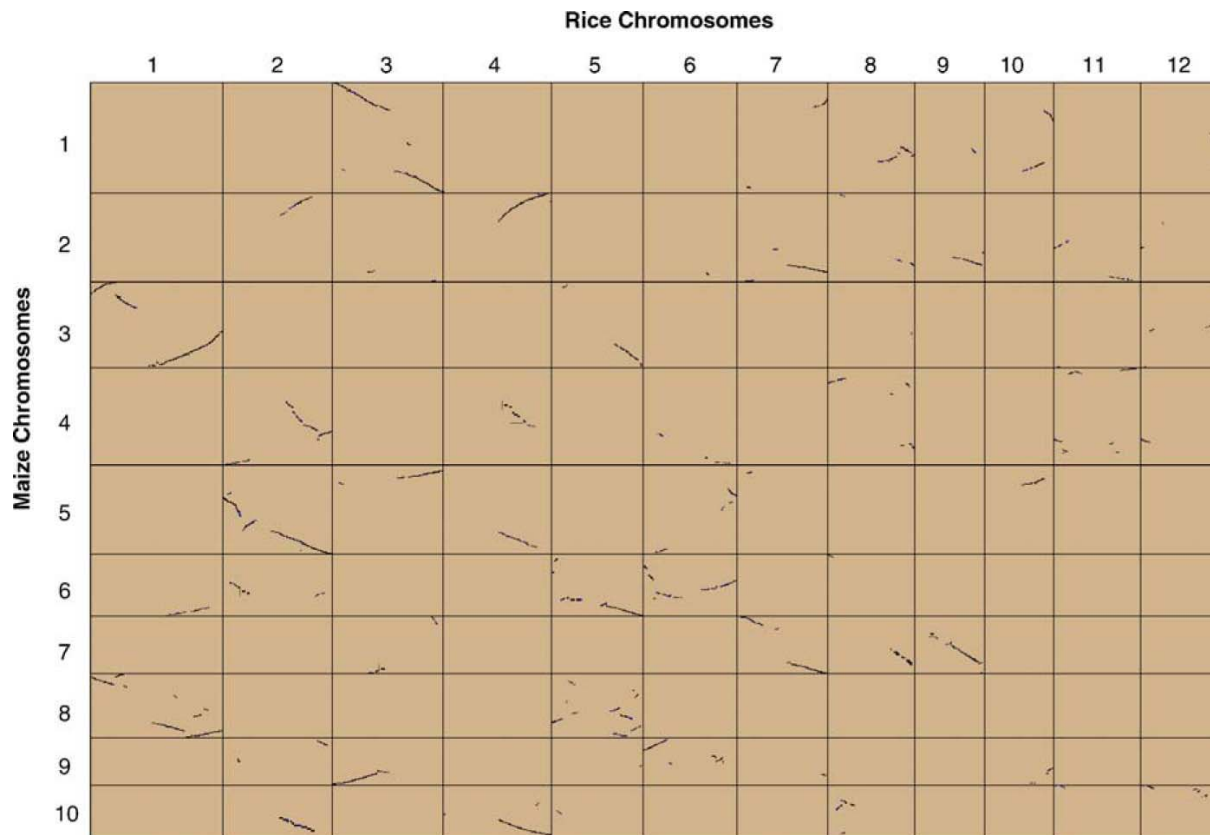
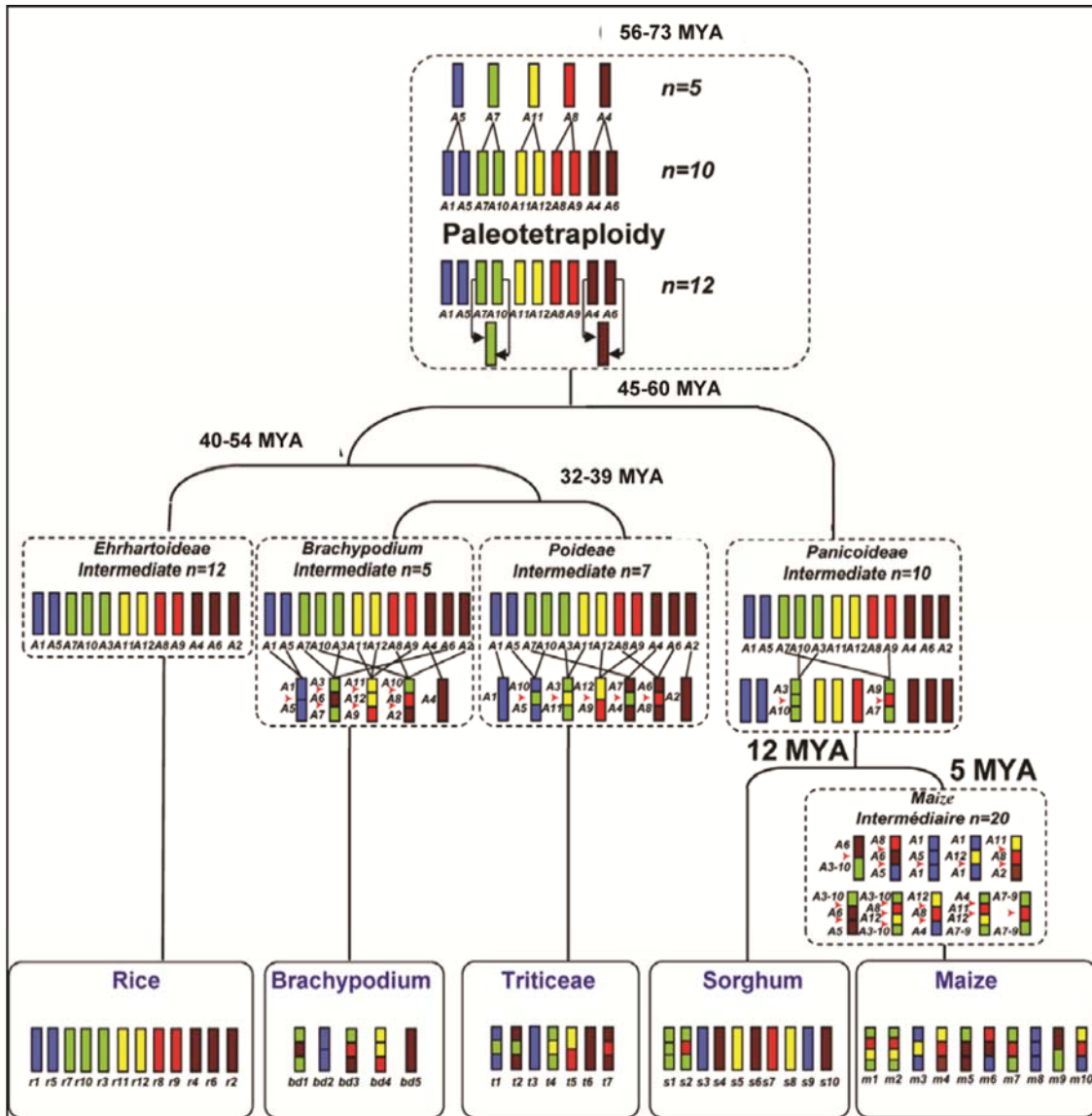


Figure 1. Dotplot Analysis of the Integrated Maize Map against Rice Pseudomolecules

Syntenic blocks were detected, and background noise was filtered with SyMAP [37]. The interactive dotplot can be viewed at <http://www.agcol.arizona.edu/sympap>. When clicking the related syntenic block, the detailed window with contig number will pop up. The viewer can select the preferred area and double click the selection, and then a graphic alignment is displayed.

doi:10.1371/journal.pgen.0030123.g001

Figure 6. A unified model of grass genome evolution. (from: Vogel et al. 2010. Nature 463:763.)



Supplementary Figure 18. Grass chromosome evolution model. The monocot chromosomes (r1-r12 for rice, t1-t7 for Triticeae, bd1-bd5 for *Brachypodium*, s1-s10 for sorghum, and m1-m10 for maize) are represented with a five colour code to illustrate the evolution of segments from a common ancestor with five proto-chromosomes and a $n=12$ intermediate as described in ⁶², and are named according to the rice nomenclature. The events that have shaped the structure of the 5 different grass genomes including the 7 *Brachypodium* chromosome nested insertion events during their evolution from the common ancestor are indicated as whole genome duplication, ancestral chromosome translocations and fusions, and lineage-specific nested chromosome insertions.

Figure 7. A summary of the duplication history of plants. (from Van de Peer et al. 2009. Trends in Plant Sciences 14:680)

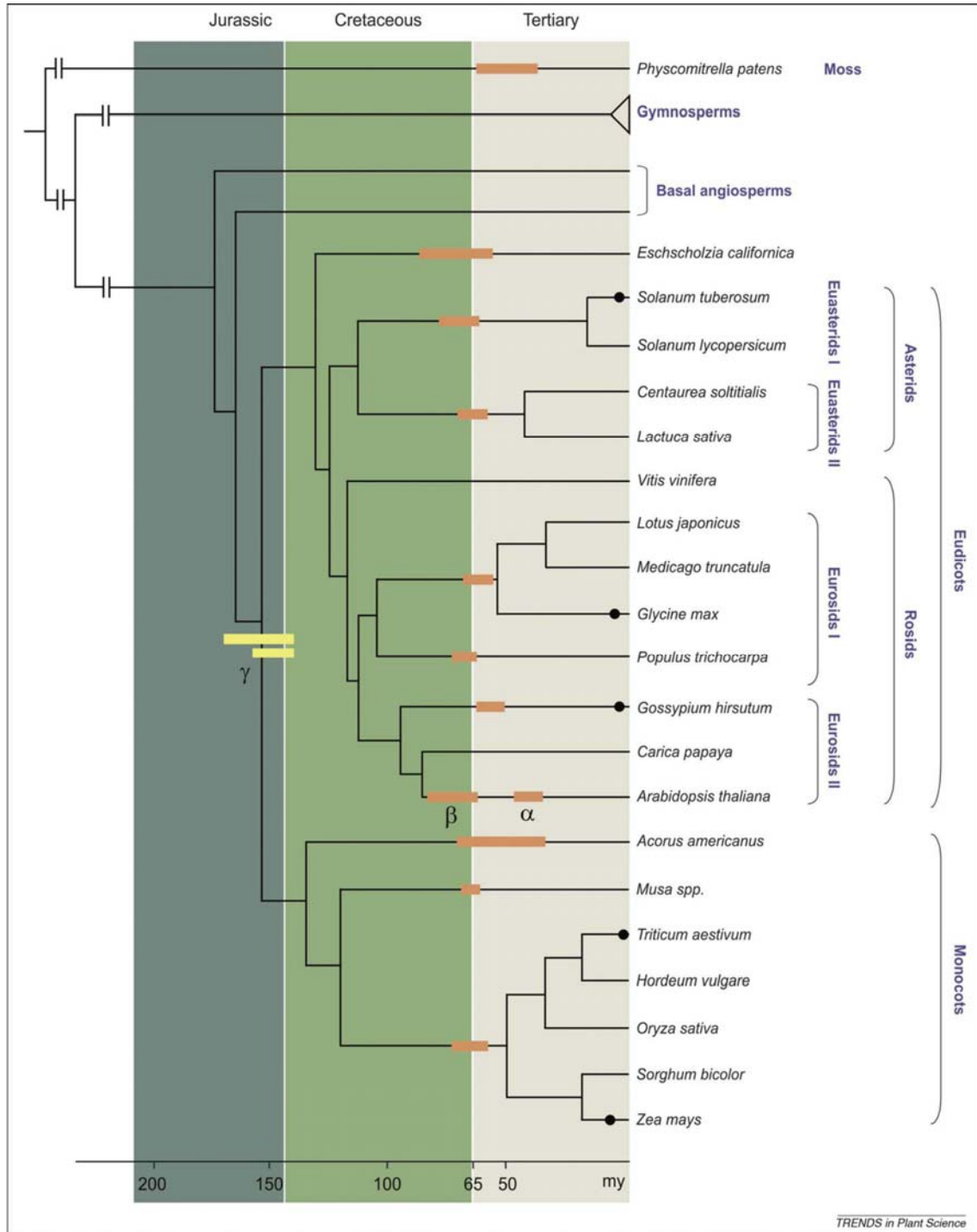
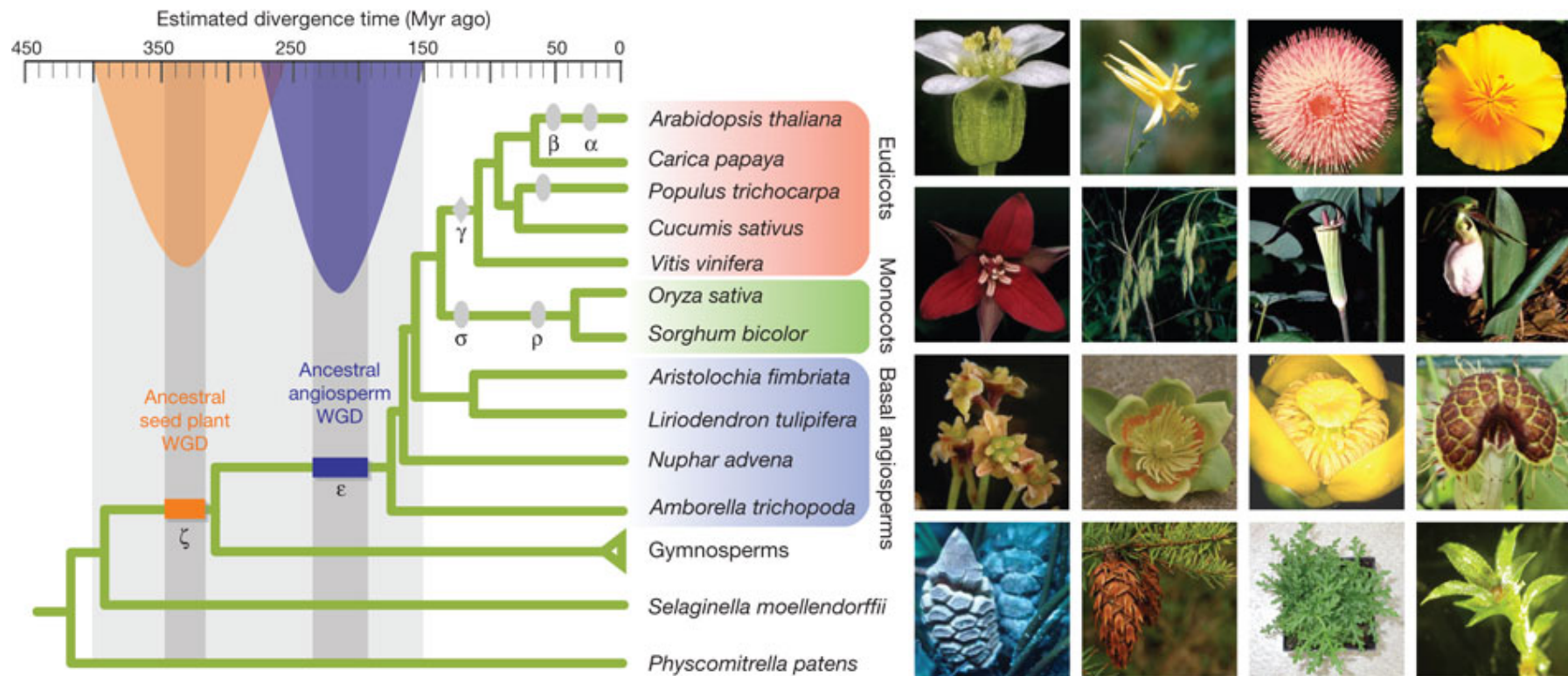


Figure 2 . Phylogenetic tree of flowering plants (eudicots and monocots). WGDs, inferred from recent studies [28–30], are indicated by horizontal bars. Yellow bars denote the hexaploidy event. More recent WGDs appear to be clustered around the KT boundary [29]. The black dots indicate recent polyploidy events [~1–2 mya in cotton (*Gossypium hirsutum*), <10 mya in potato (*Solanum tuberosum*), ~10–15 mya in soybean (*Glycine max*), ~10 mya in maize (*Zea mays*), and <1 mya in wheat (*Triticum aestivum*)]. Alpha, beta and gamma denote the generally accepted duplication events in *Arabidopsis* [5–7,36] (see main text for details). Modified with permission from [29].

Figure 8: Ancestral polyploidy events in seed plants and angiosperms. [Jiao et al (2011) Nature 473:97]



Original figure legend from manuscript. Two ancestral duplications identified by integration of phylogenomic evidence and molecular time clock for land plant evolution. Ovals indicate the generally accepted genome duplications identified in sequenced genomes (see text). The diamond refers to the triplication event probably shared by all core eudicots. Horizontal bars denote confidence regions for ancestral seed plant WGD and ancestral angiosperm WGD, and are drawn to reflect upper and lower bounds of mean estimates from [Fig. 2](#) (more orthogroups) and [Supplementary Fig. 5](#) (more taxa). The photographs provide examples of the reproductive diversity of eudicots (top row, left to right: *Arabidopsis thaliana*, *Aquilegia chrysantha*, *Cirsium pumilum*, *Eschscholzia californica*), monocots (second row, left to right: *Trillium erectum*, *Bromus kalmii*, *Arisaema triphyllum*, *Cypripedium acaule*), basal angiosperms (third row, left to right: *Amborella trichopoda*, *Liriodendron tulipifera*, *Nuphar advena*, *Aristolochia fimbriata*), gymnosperms (fourth row, first and second from left: *Zamia vazequezii*, *Pseudotsuga menziesii*) and the outgroups *Selaginella moellendorffii* (vegetative; fourth row, third from left) and *Physcomitrella patens* (fourth row, right). See [Supplementary Table 4](#) for photo credits.

Figure 9. Macrosyteny between tomato and eggplant, including a QTL for a shared domestication trait. (from: Doganlar et al. 2002. Genetics 161:1713.)

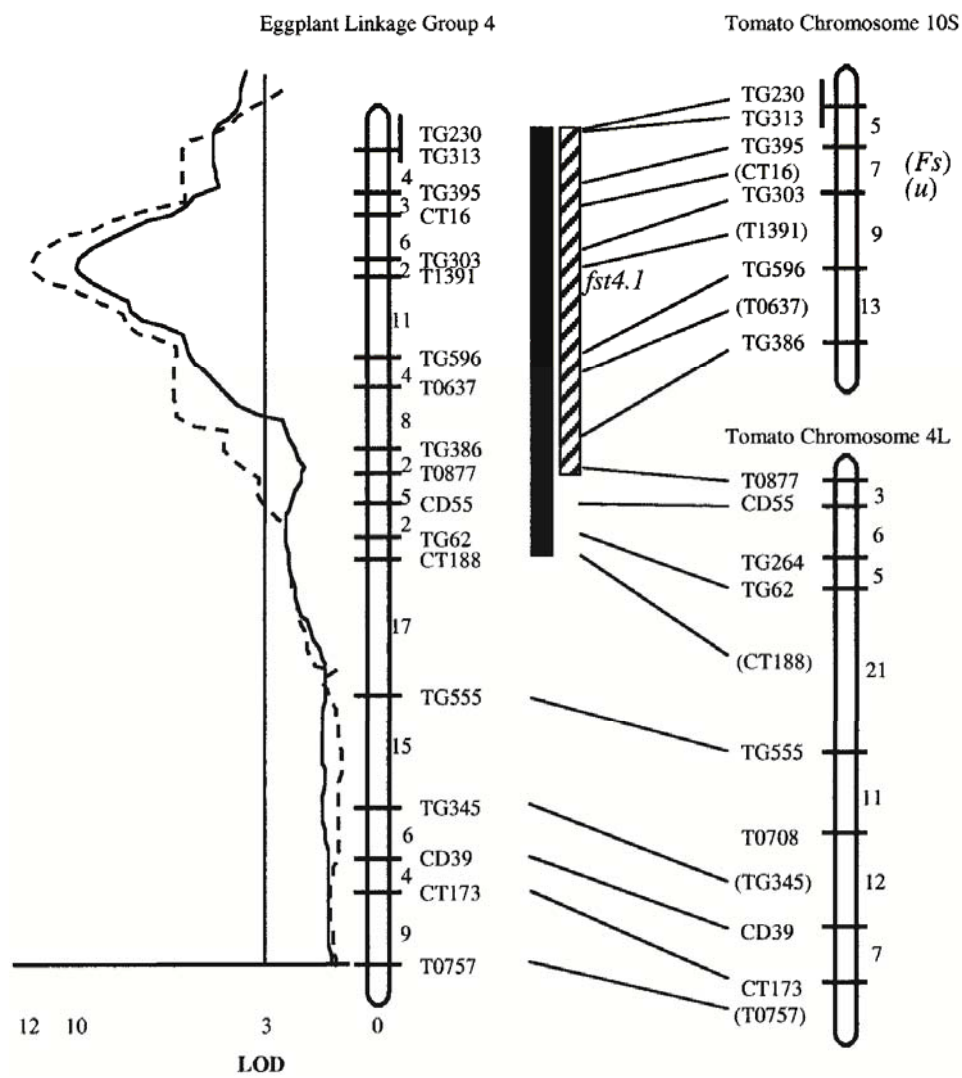


Figure 3.—Comparative mapping of fruit stripe locus on eggplant linkage group 4. Simple interval analysis for *fst4.1* is shown to the left of the molecular map of eggplant linkage group 4 (solid line for NY data, dashed line for FR data). Bars to the right of the linkage group represent the position of the QTL as determined by single-point regression analysis ($P \leq 0.05$; see Table 1 for details; solid bar for NY data, hatched bar for FR data). Molecular maps for tomato chromosome arms are from Tanksley et al (1992).