

Bioinformatics tools to analyze complex genomes



Yves Van de Peer
Ghent University/VIB

TREE OF LIFE

Metazoan
Vertebrata

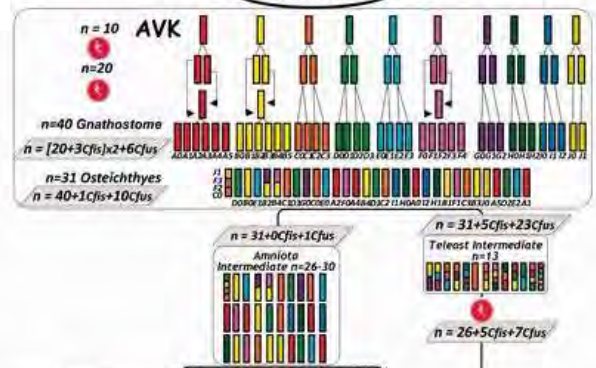
Choanozoan

Fungi

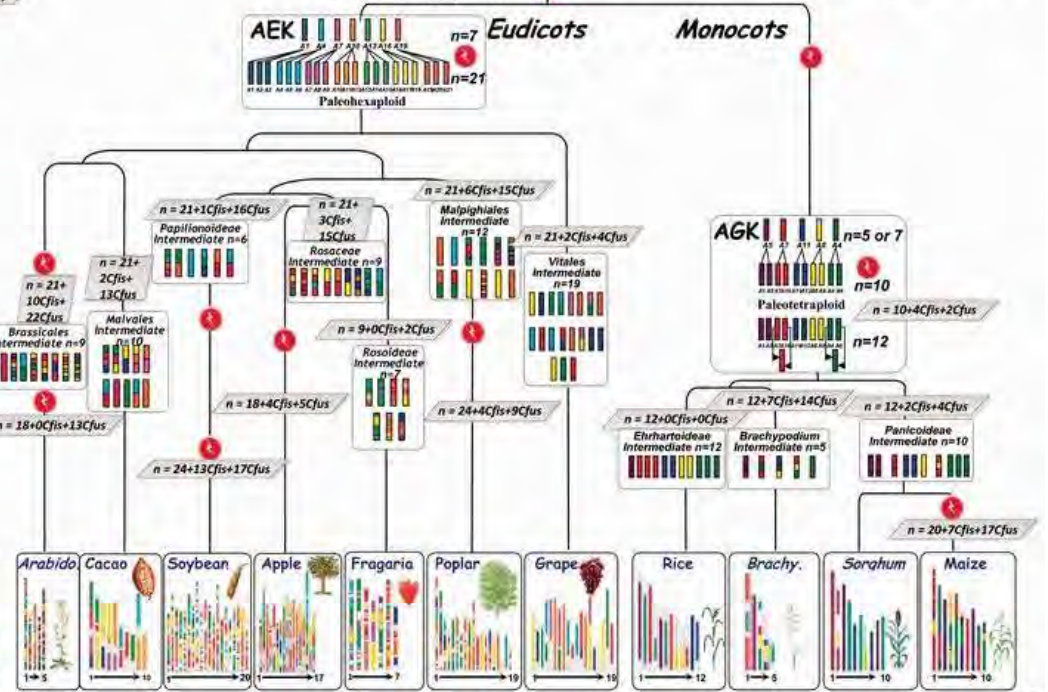
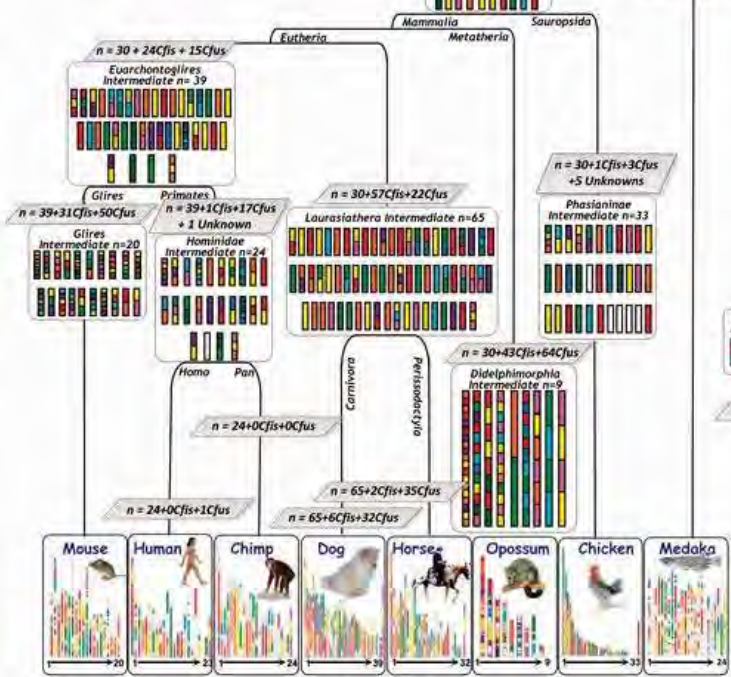
Plantae

Archaea

Bacteria



- Aquatic organisms
- Nonvascular plants
- Vascular Plants
- Flowering plants
- Chlorophytes (*Chlamydomonas...*)
- Bryophytes (mosses, *Physcomitrella...*)
- Lycophytes (*Selaginella...*)
- Ferns
- Gymnosperms (Pine, Cycad...)
- Basal Angiosperms (*Amborella, Nuphar...*)



ANIMALS

PLANTS

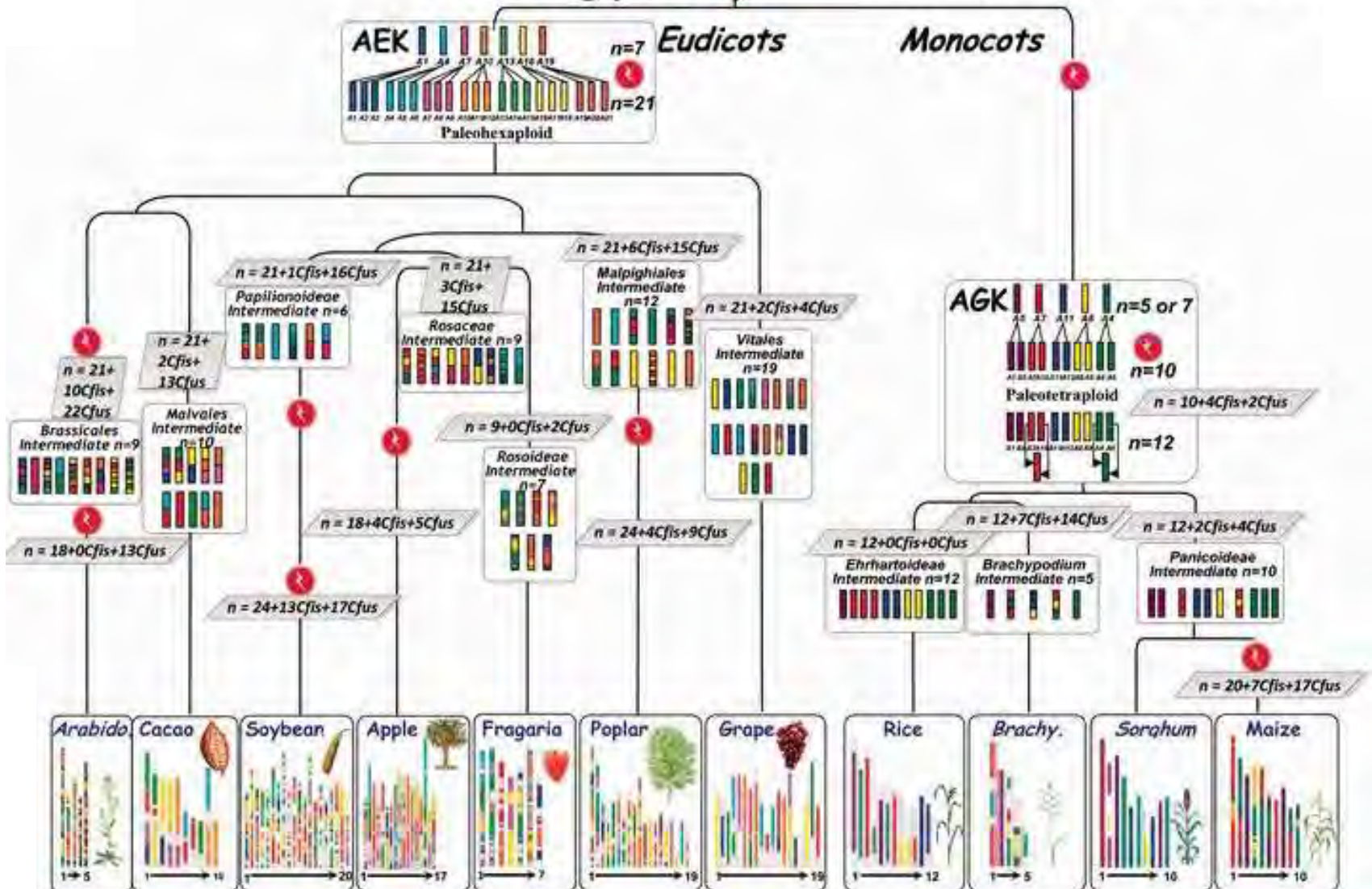
Vascular Plants

Ferns

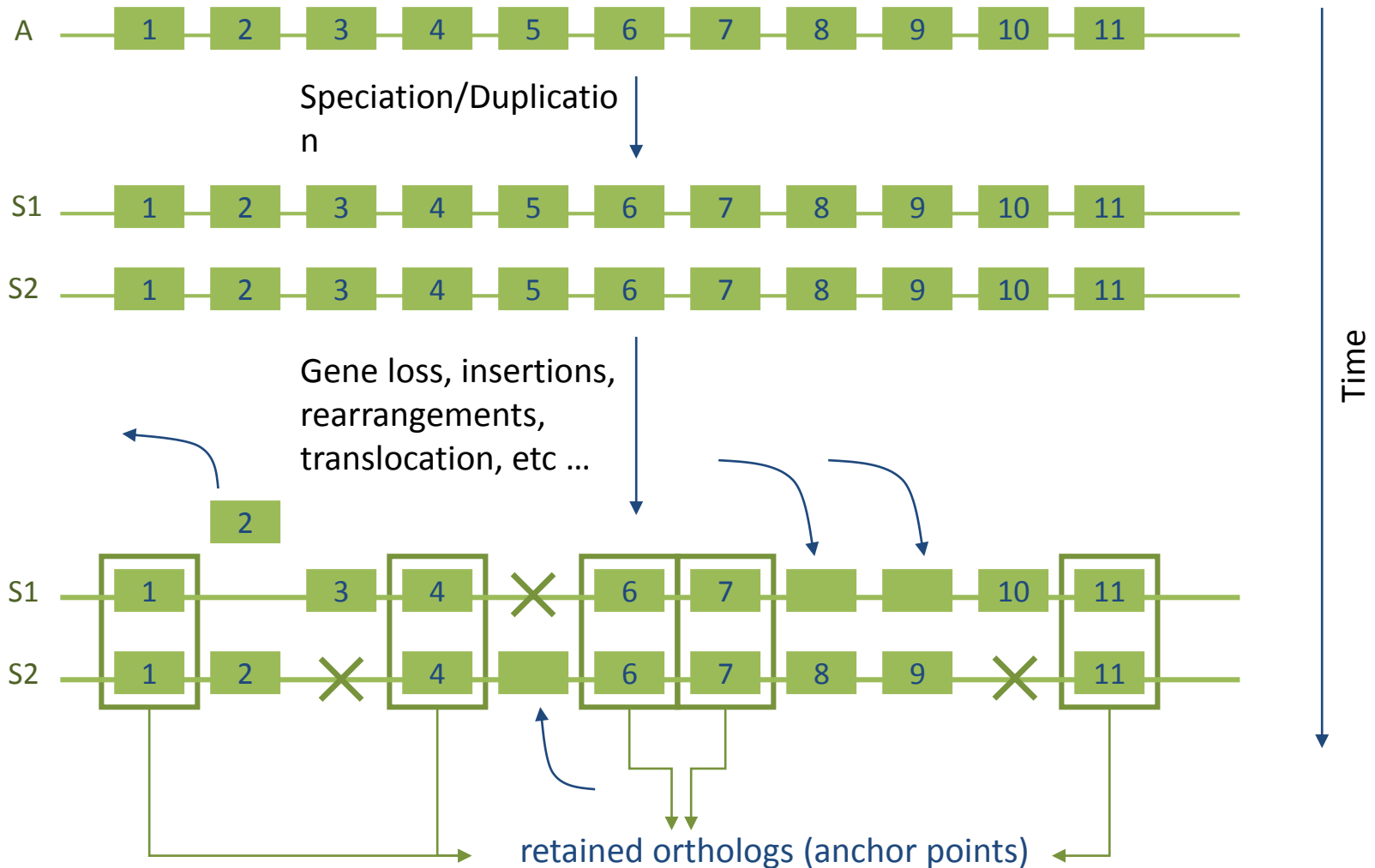
Gymnosperms (Pine, Cycad...)

Flowering plants

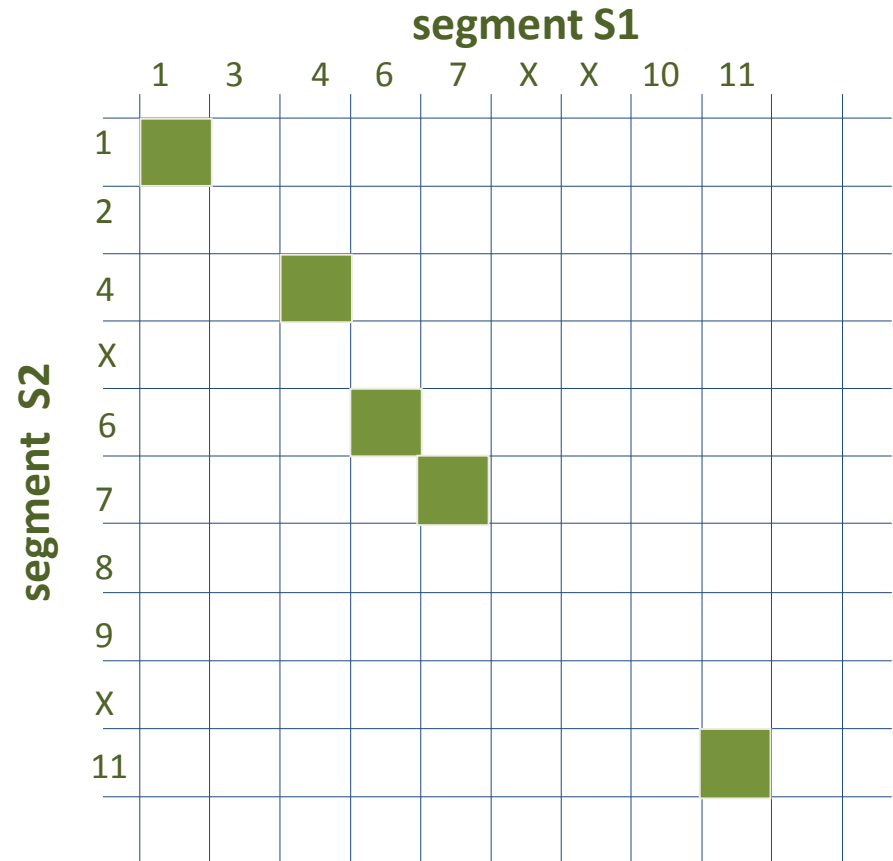
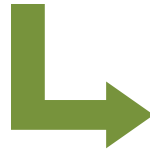
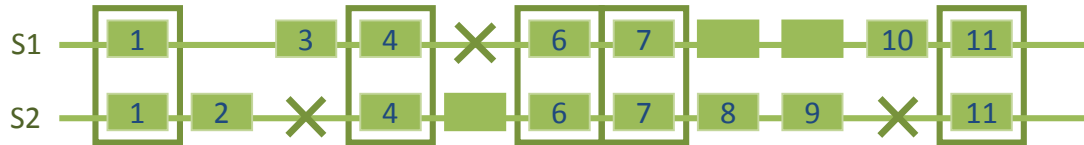
Basal Angiosperms (Amborella, Nuphar...)



Detecting colinearity and large-scale gene duplications



Detecting colinearity and large-scale gene duplications

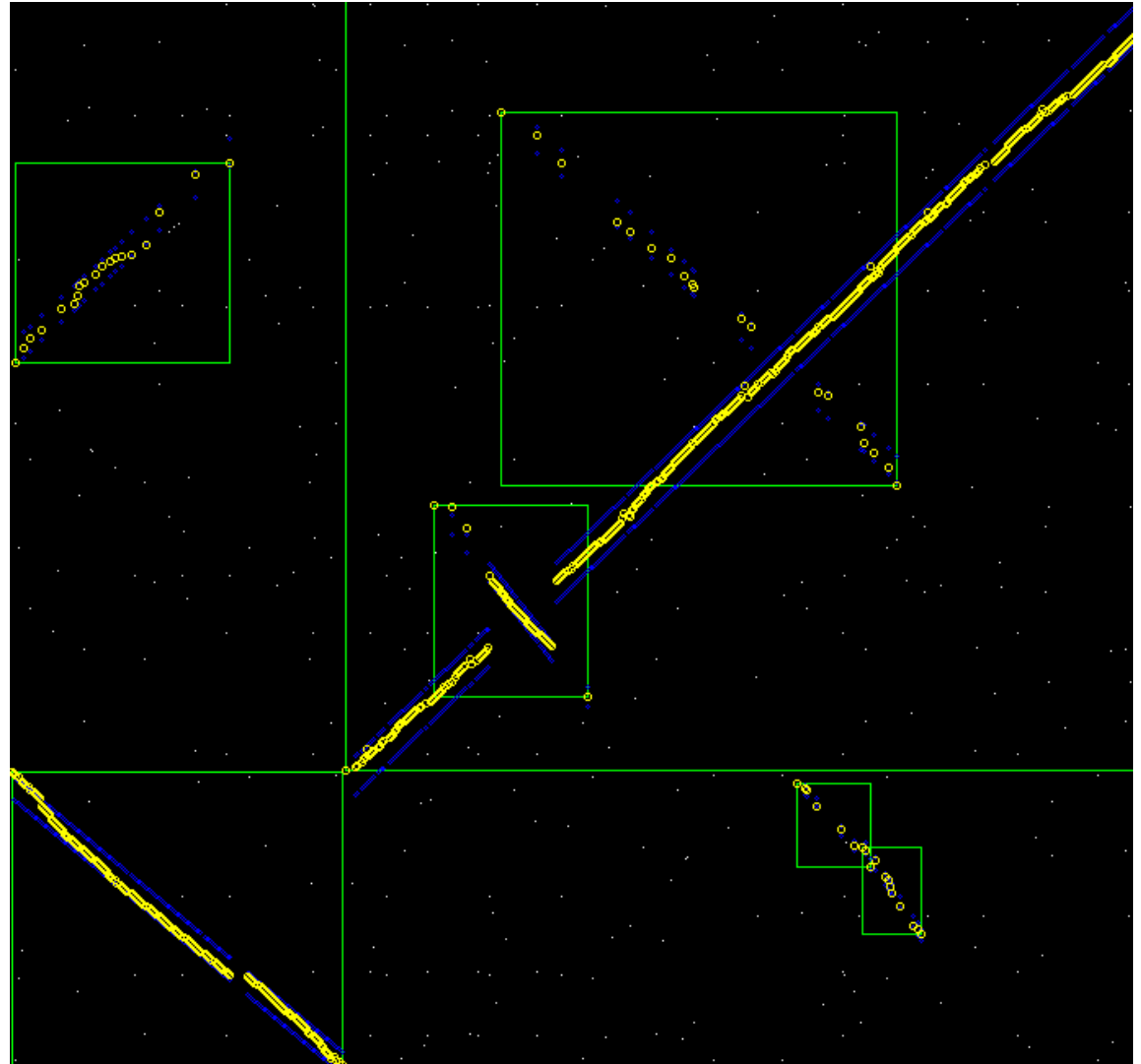
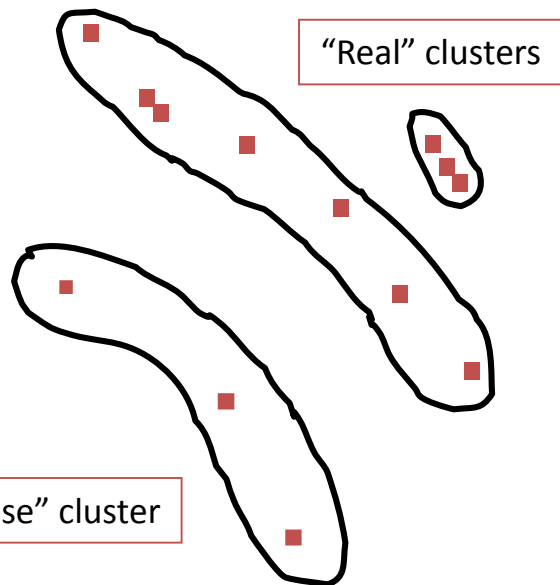


Gene Homology Matrix:
 Conserved gene content and order
 Collinear regions appear as diagonals

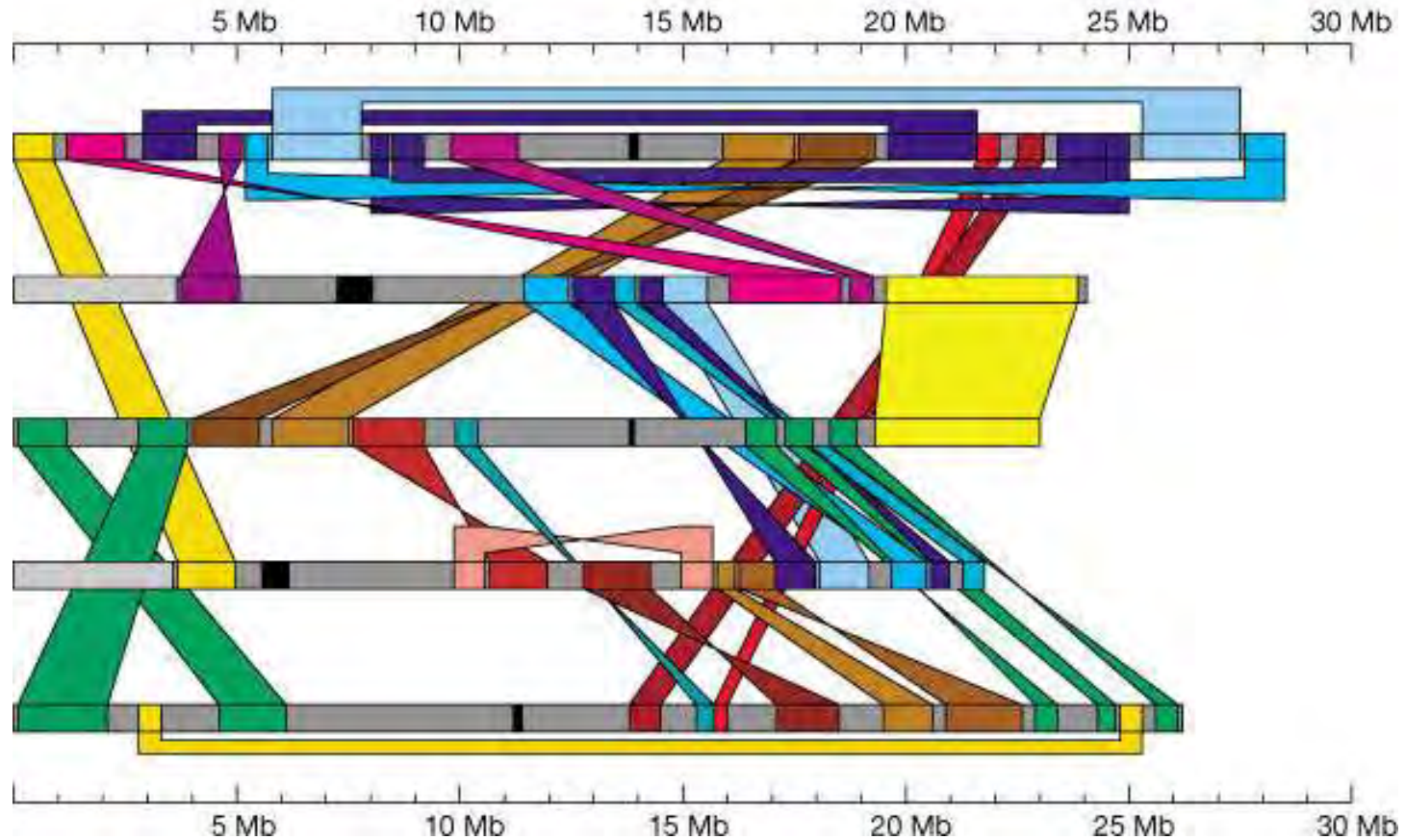
Detecting colinearity and large-scale gene duplications

In an actual genome
this becomes complex

Good statistical model
to find biologically
relevant regions

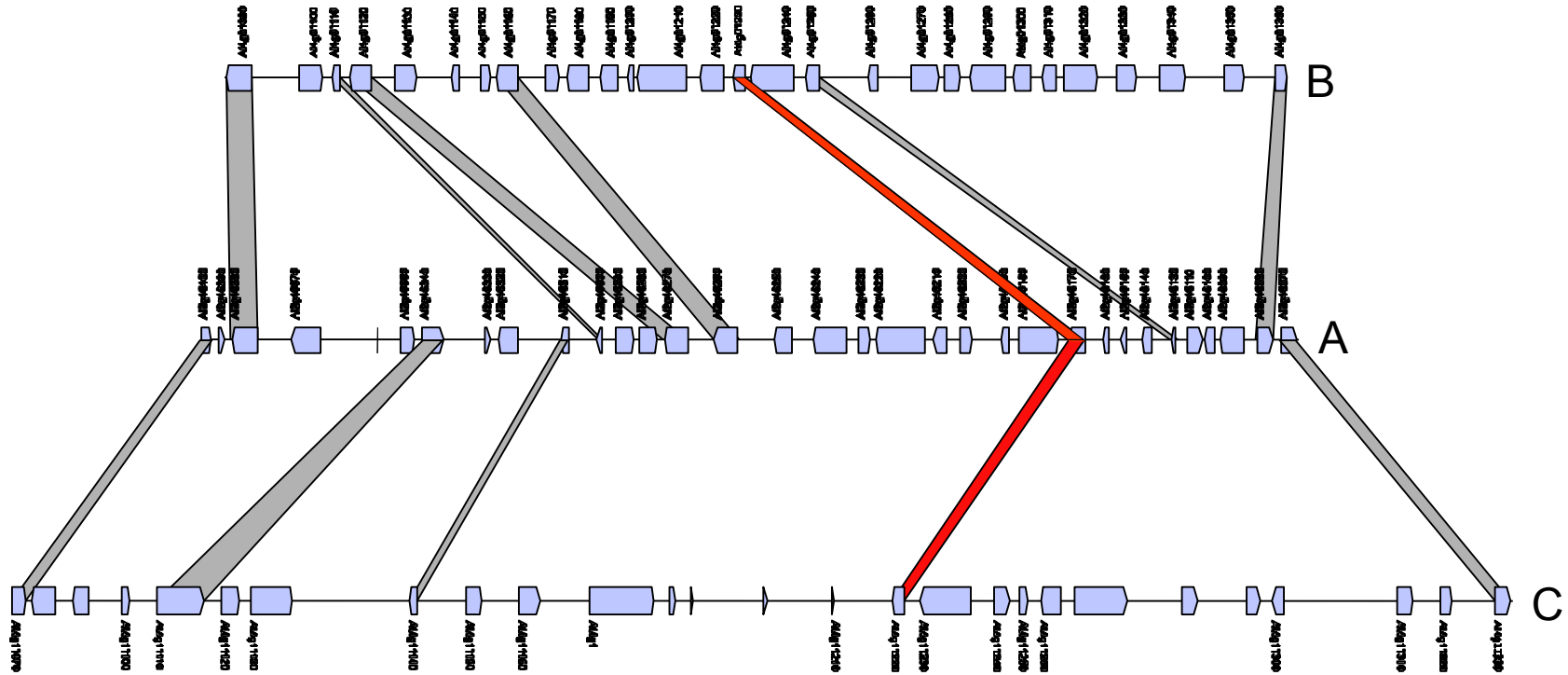


The Arabidopsis genome duplication



The Arabidopsis Genome Initiative (2000)

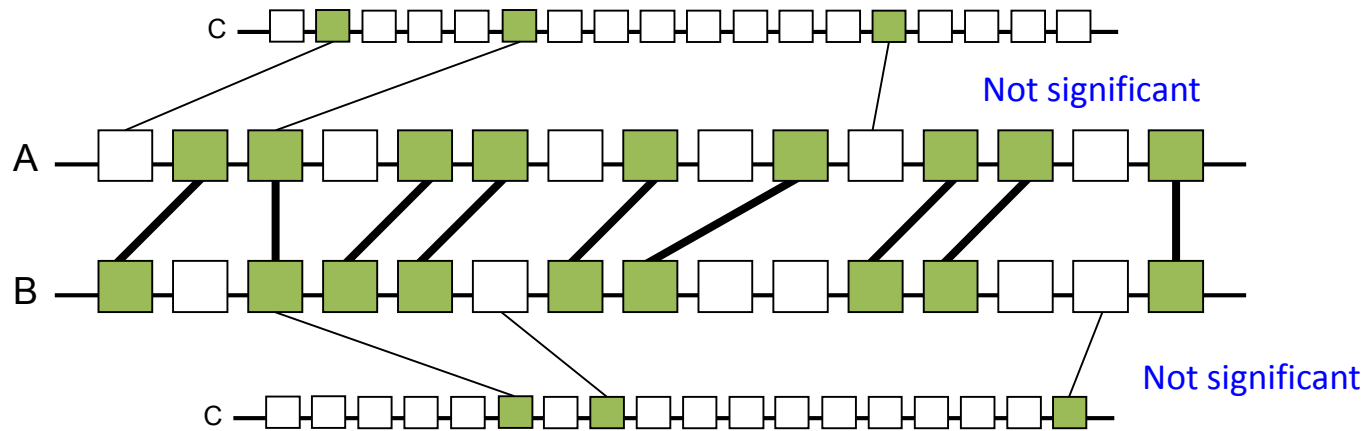
Transitive homology



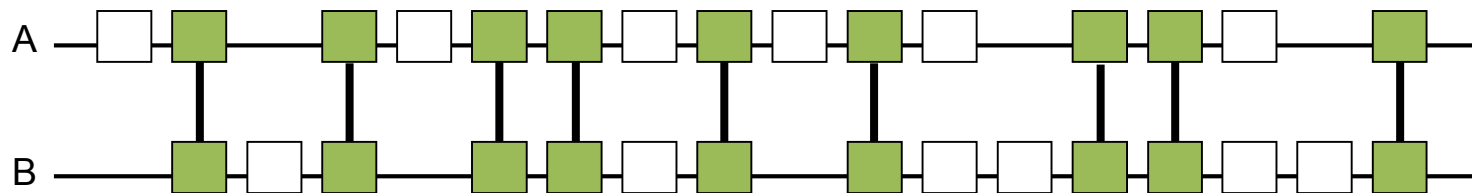
Segments B and C are homologous because both are homologous to the segment A

However, segments B and C share only 1 gene because of “differential gene loss”

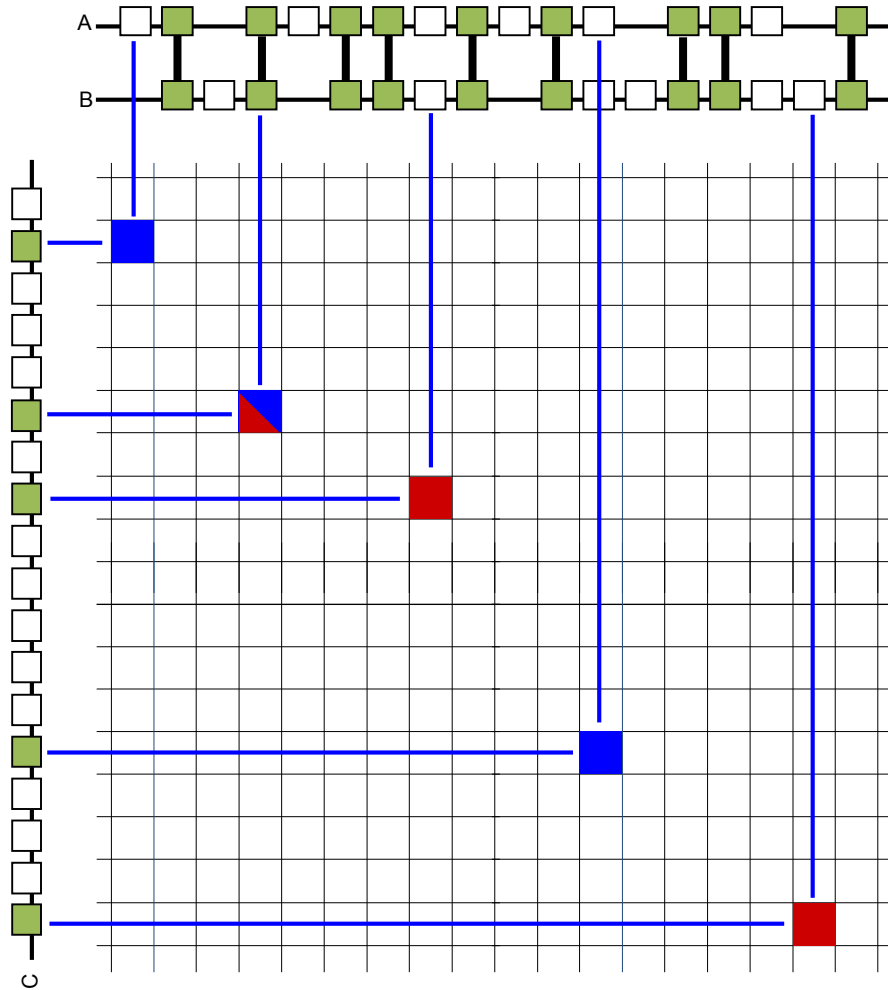
Detecting homology in the 'twilight zone': genomic profiles



Building genomic profiles

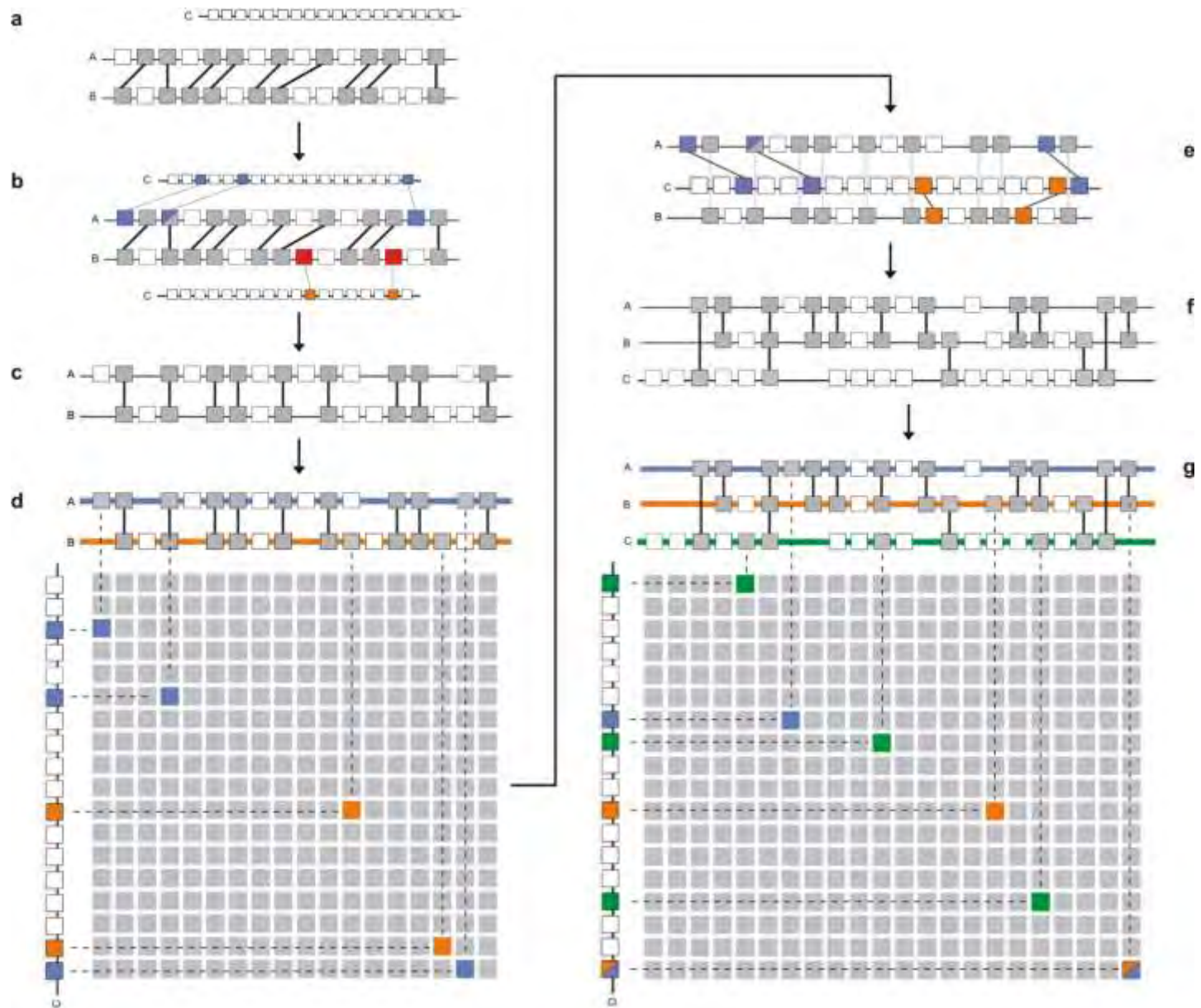


Searching for diagonals ... revisited

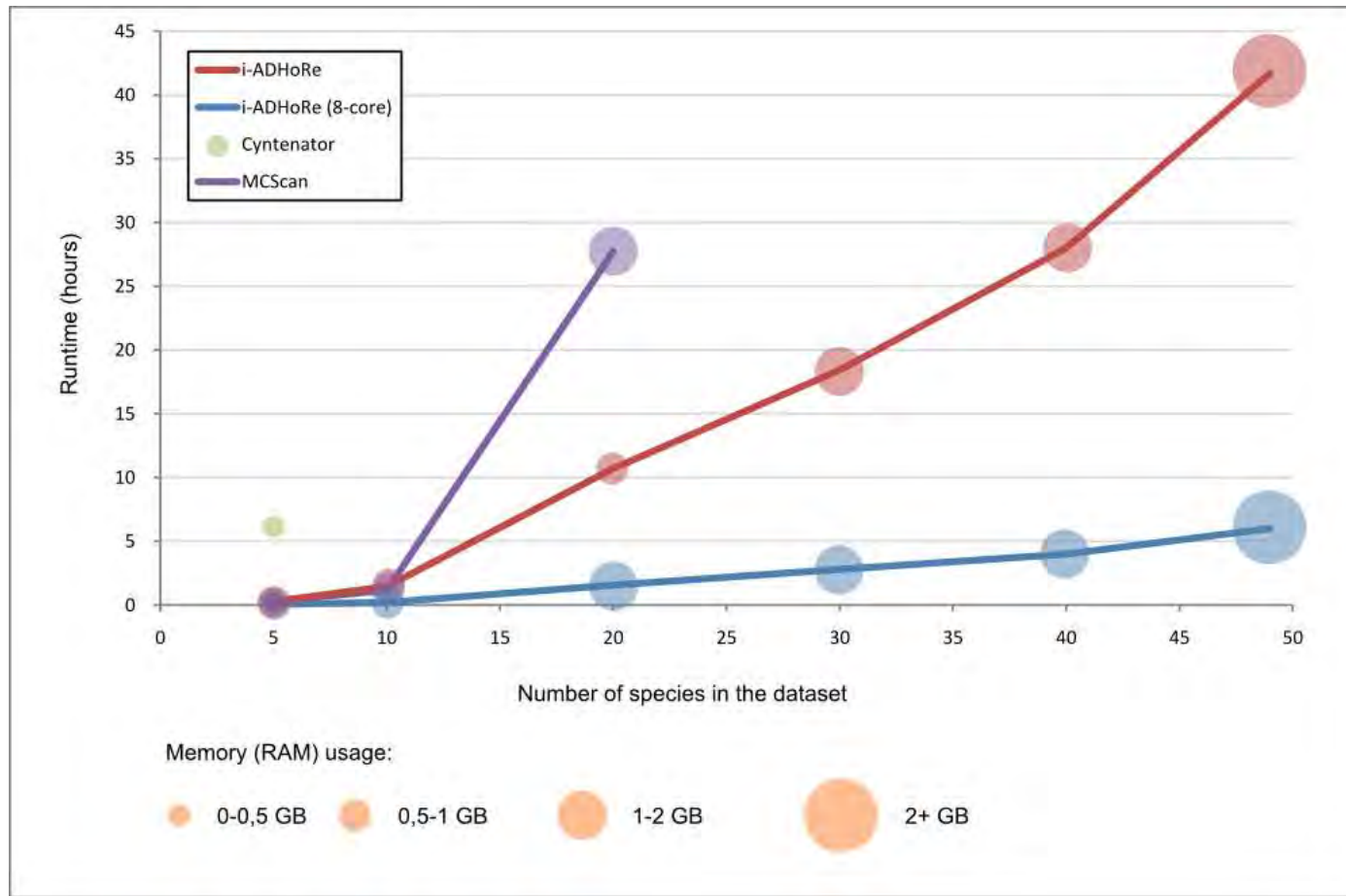


significant homology!

Genomic profiles: increasing the multiplication level



i-ADHoRe 3.0

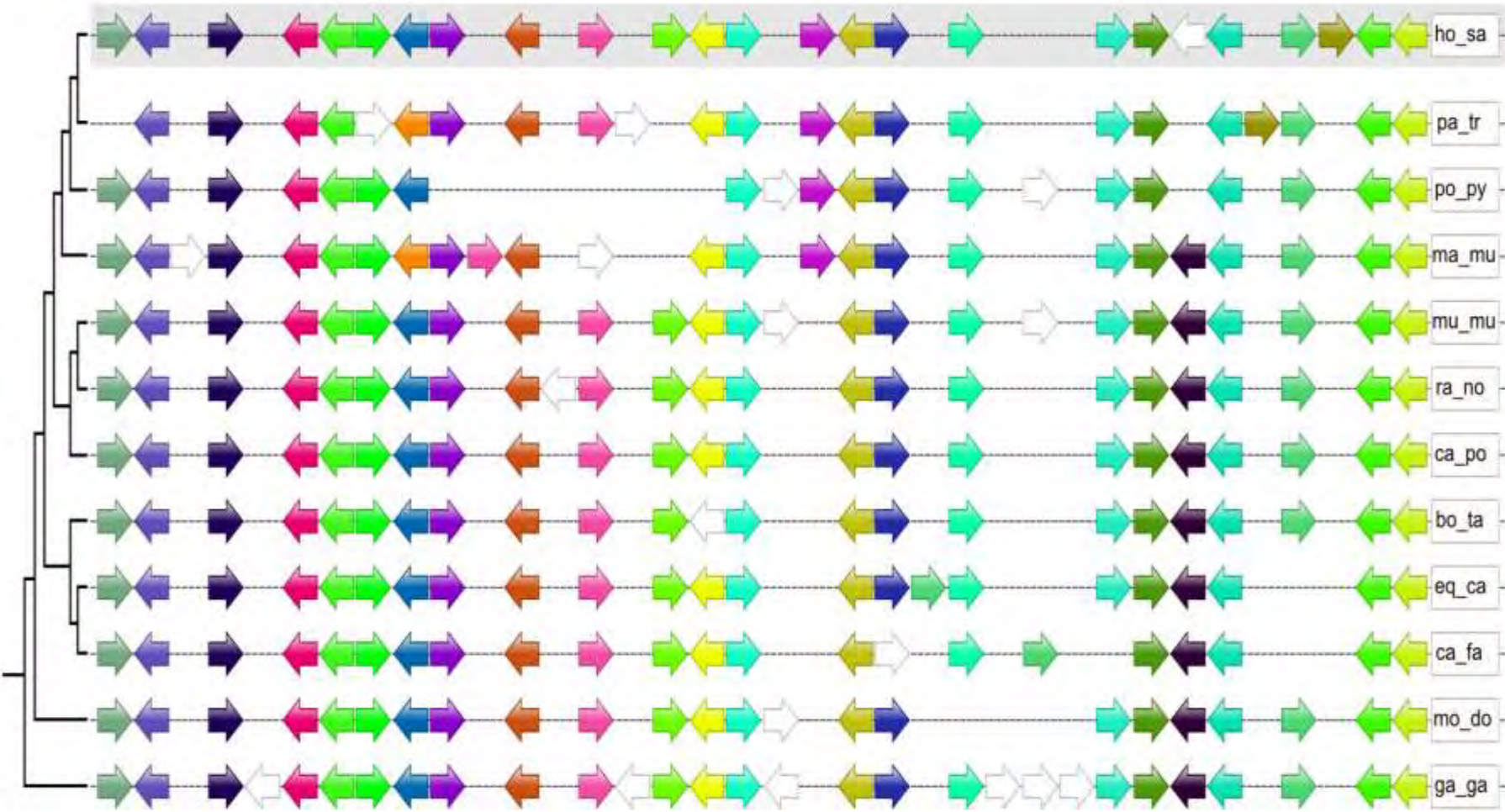


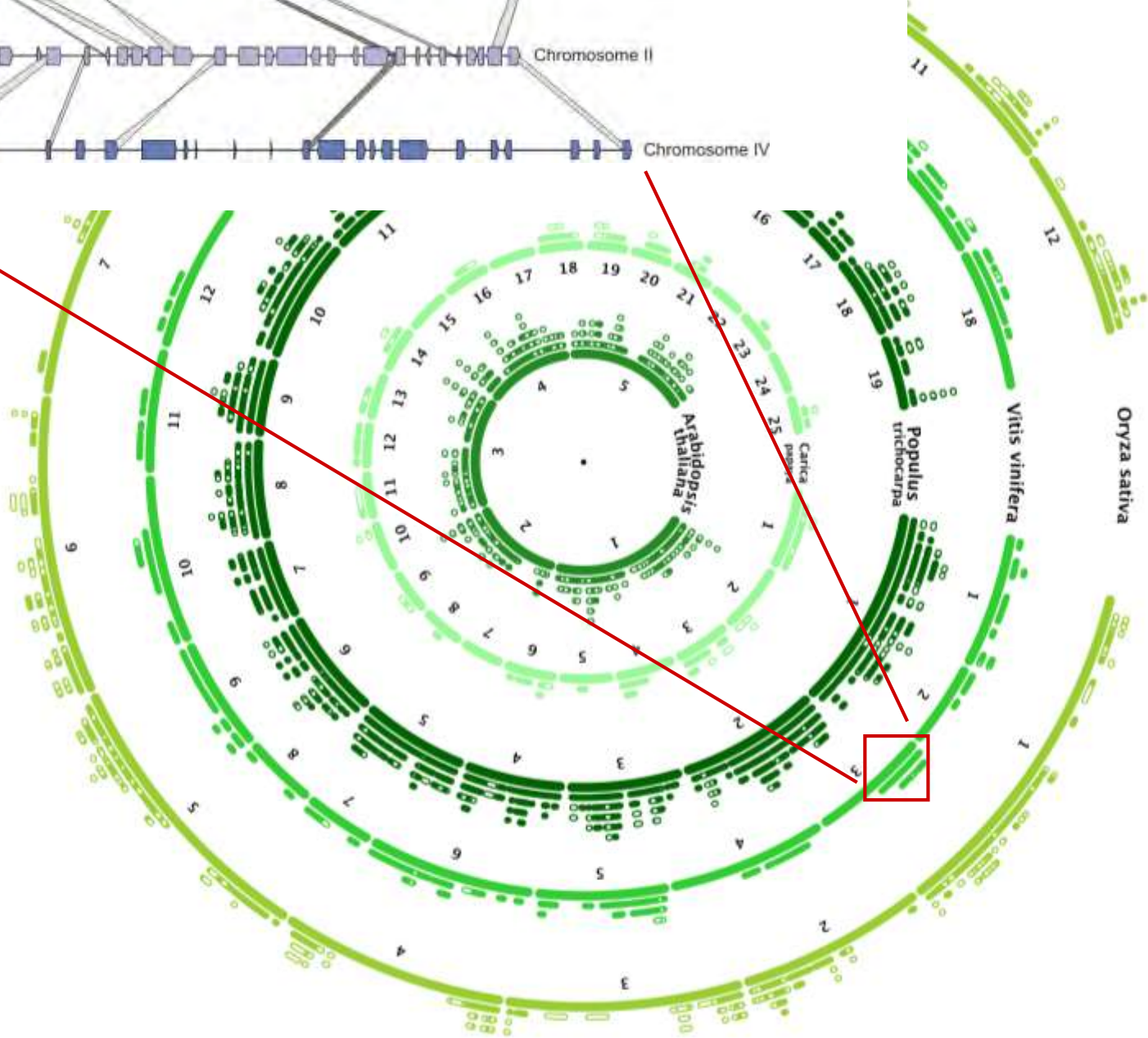
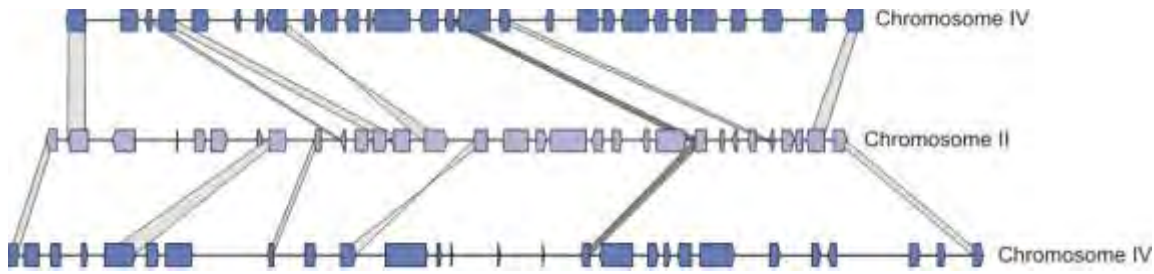
MCSCan: Tang et al., 2008

Cyntenator: Rödelsperger et al., 2010

Fostier et al., 2011

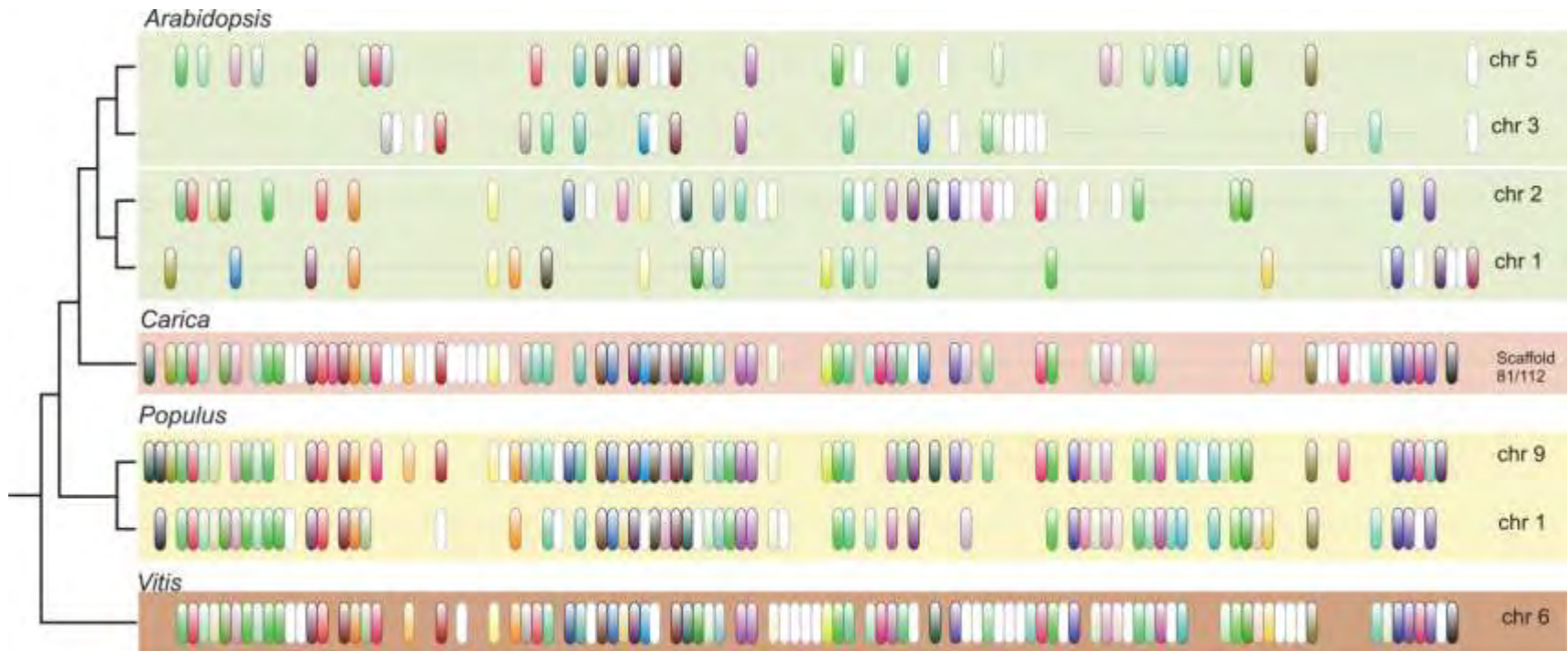
i-ADHoRe 3.0 is extremely efficient in analyzing large datasets



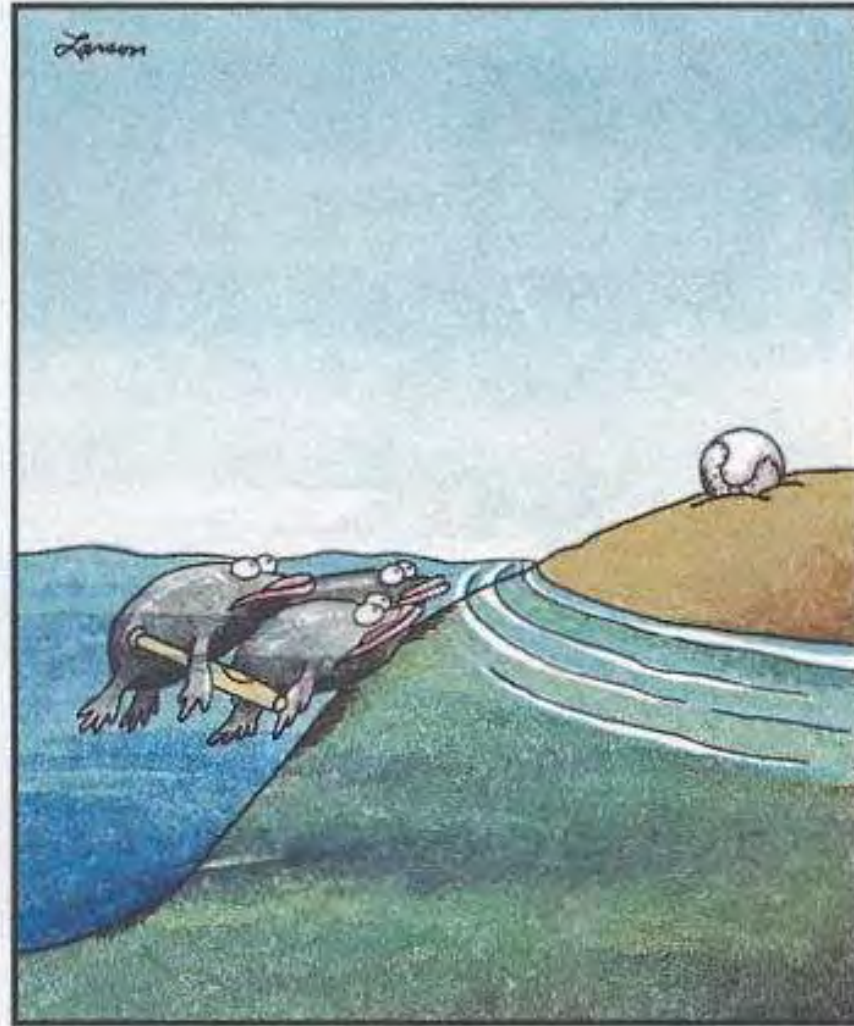


Gene loss and colinearity/synteny

WGDs add to the complexity of plant genomes (both in numbers of genes and genome structure) ...



Linking WGDs with great or decisive moments in evolution ...

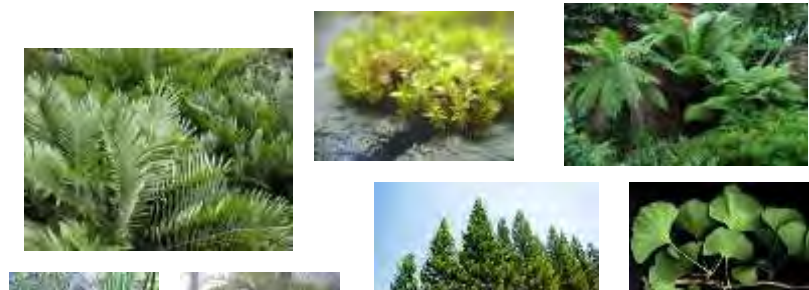


Great moments in evolution

Darwin's abominable mystery ...



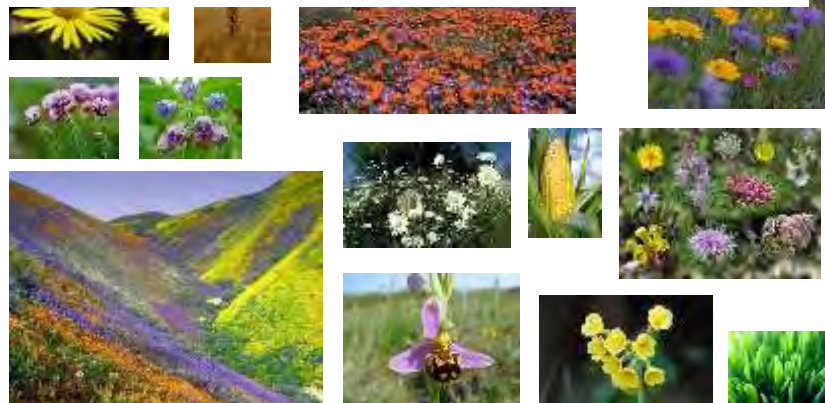
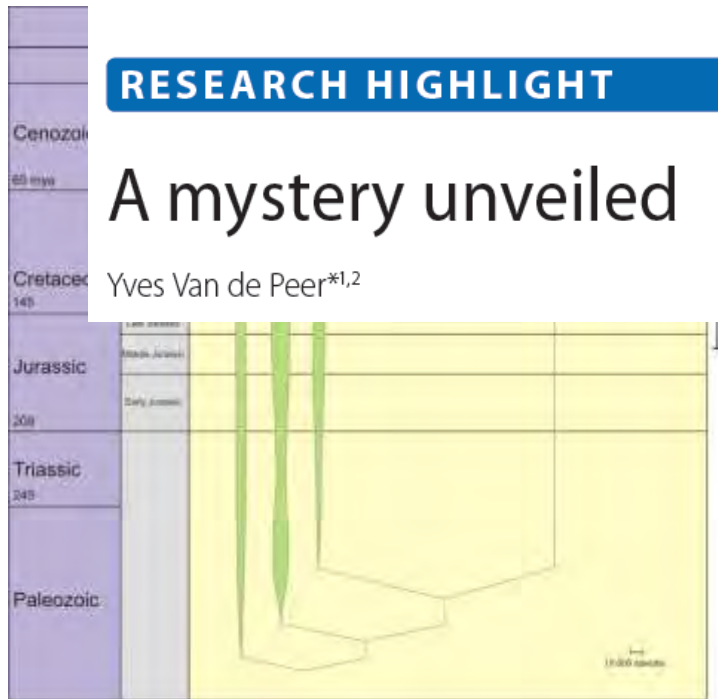
Van de Peer *Genome Biology* 2011, 12:113
<http://genomebiology.com/2011/12/5/113>



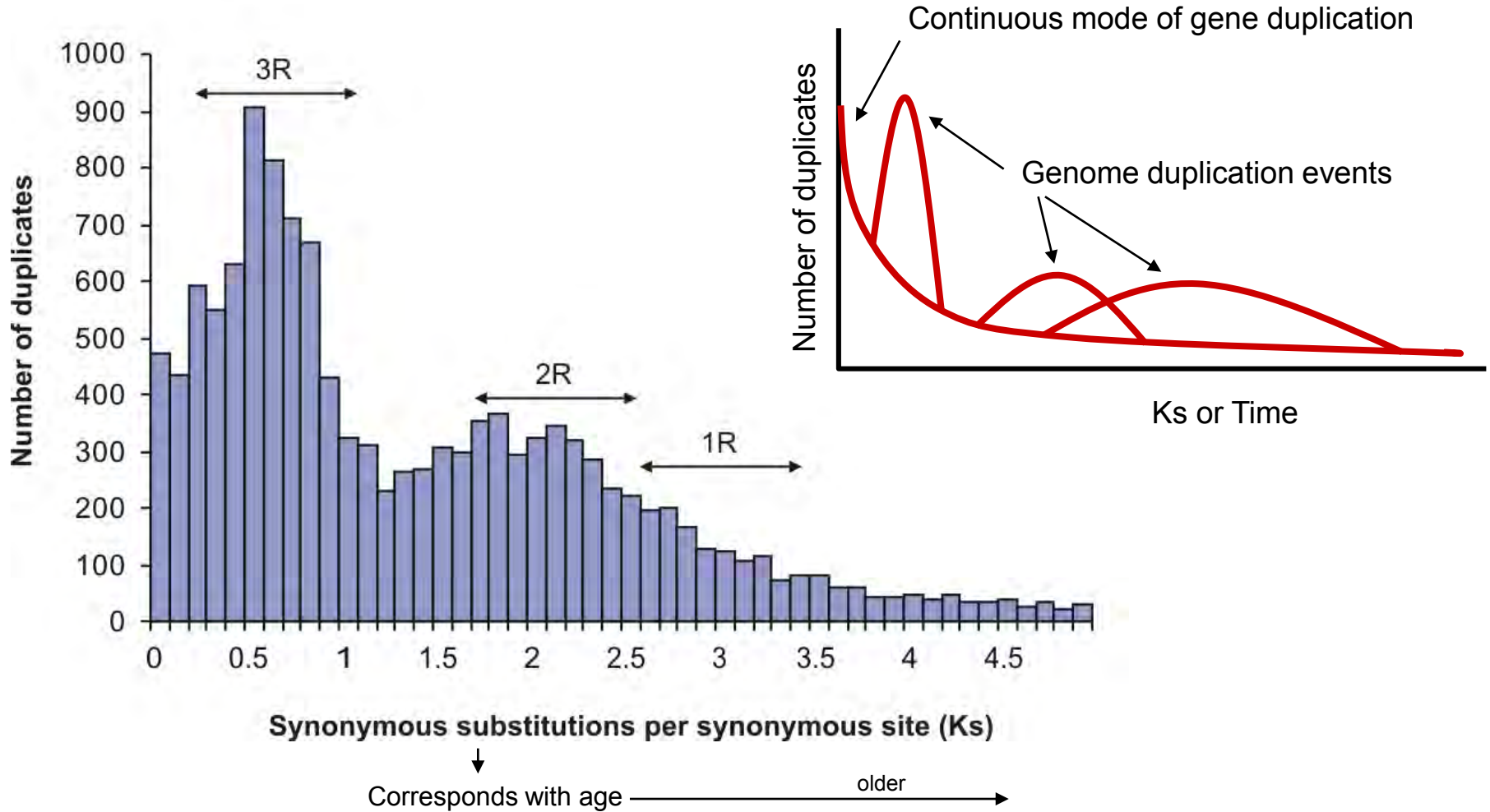
RESEARCH HIGHLIGHT

A mystery unveiled

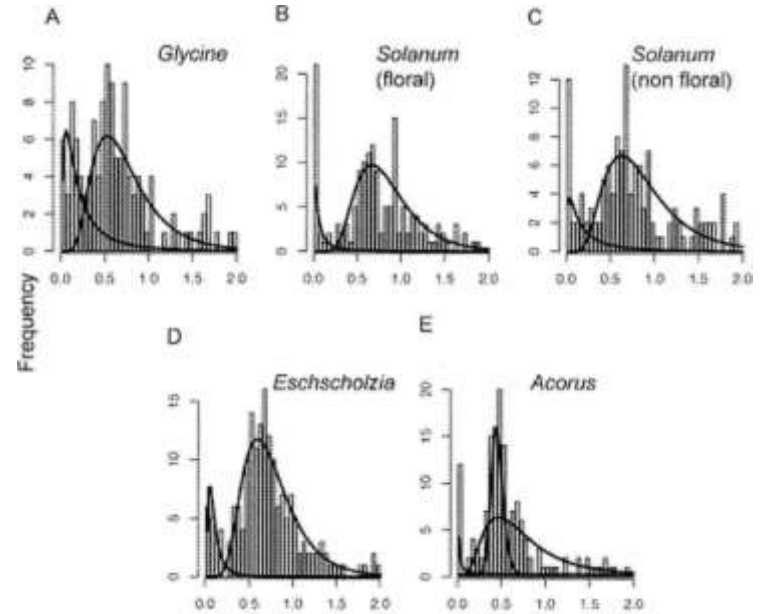
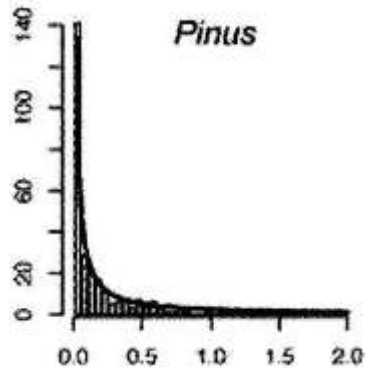
Yves Van de Peer*^{1,2}



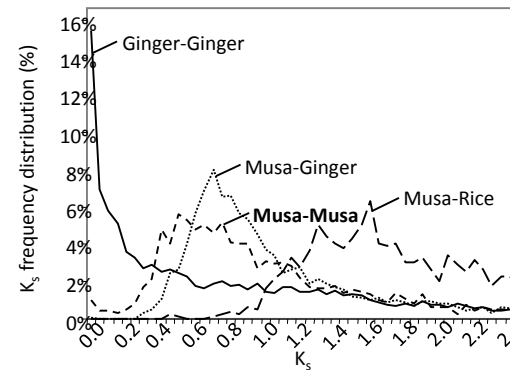
Another way of unveiling WGDs: K_S age distributions



Many plants have undergone a 'recent' genome duplication ...



Cui et al., Genome Research (2006)



Lescot et al. (2008)
BMC Genomics (2008)

The tomato genome sequence provides insights into fleshy fruit evolution

The Tomato Genome Consortium*

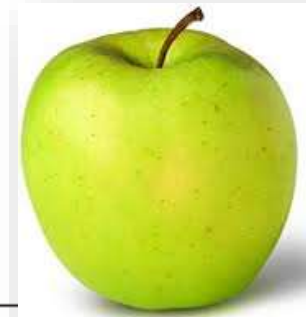


The *Solanum* lineage has experienced two consecutive genome duplications: one that is ancient and shared with rosids, and a more recent one. These duplications set the stage for the neofunctionalization of genes controlling fruit characteristics, such as colour and fleshiness.

The *Medicago* genome provides insight into the evolution of rhizobial symbioses



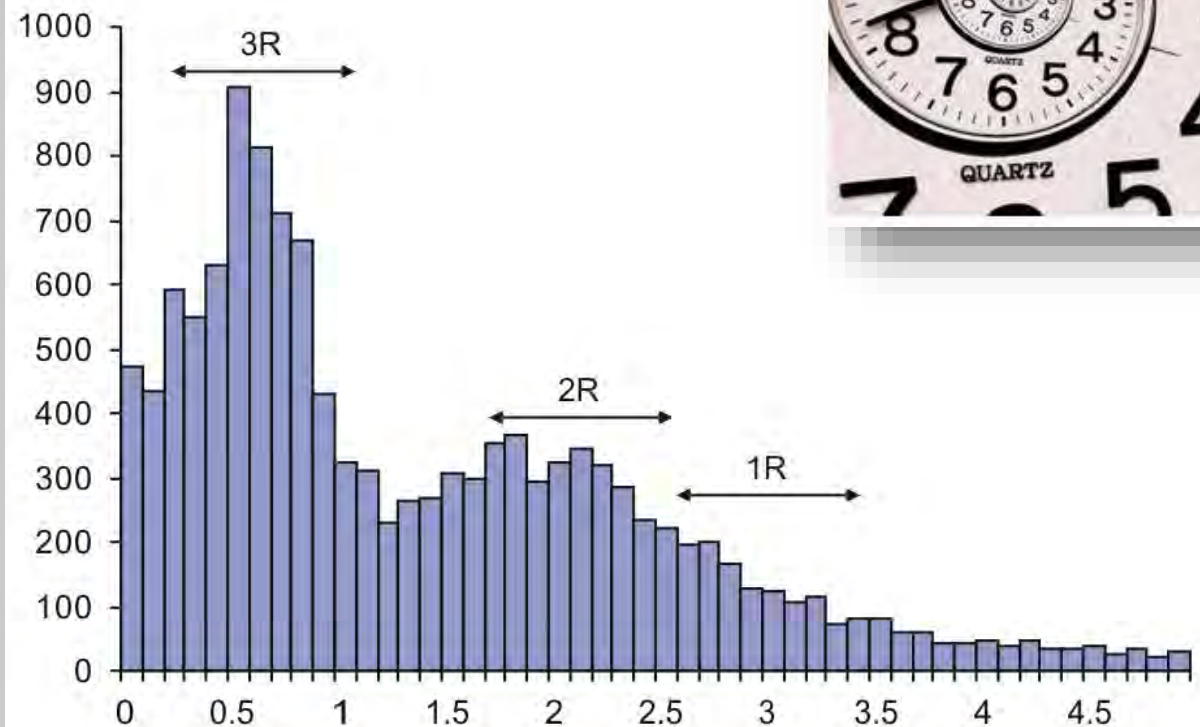
A whole-genome duplication (WGD) approximately 58 Myr ago had a major role in shaping the *M. truncatula* genome and thereby contributed to the evolution of endosymbiotic nitrogen fixation.



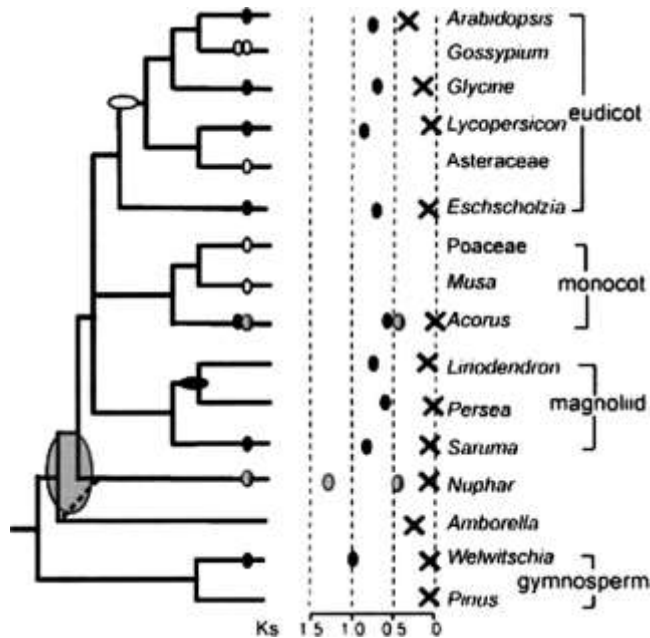
The genome of the domesticated apple (*Malus × domestica* Borkh.)

An intriguing aspect of the apple's biology concerns its characteristic fruit, the pome, which is found only in the Pyreae tribe. This indicates that the pome probably evolved after a relatively recent Pyreae-specific GWD, a polyploidization step that we hypothesize has contributed to the apple's developmental and metabolic specificity.

Calibrating the molecular clock



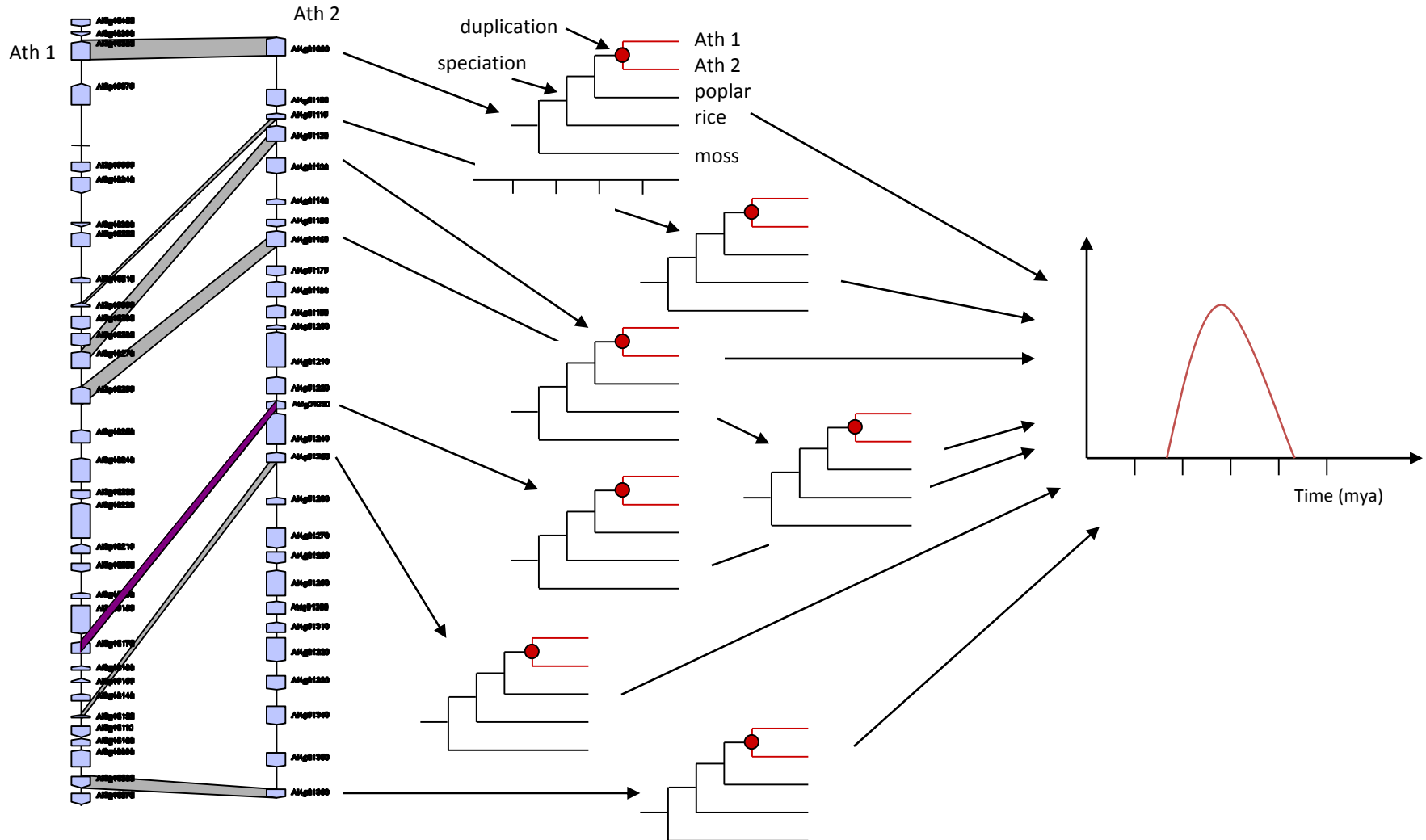
Many plants have undergone a 'recent' genome duplication ...



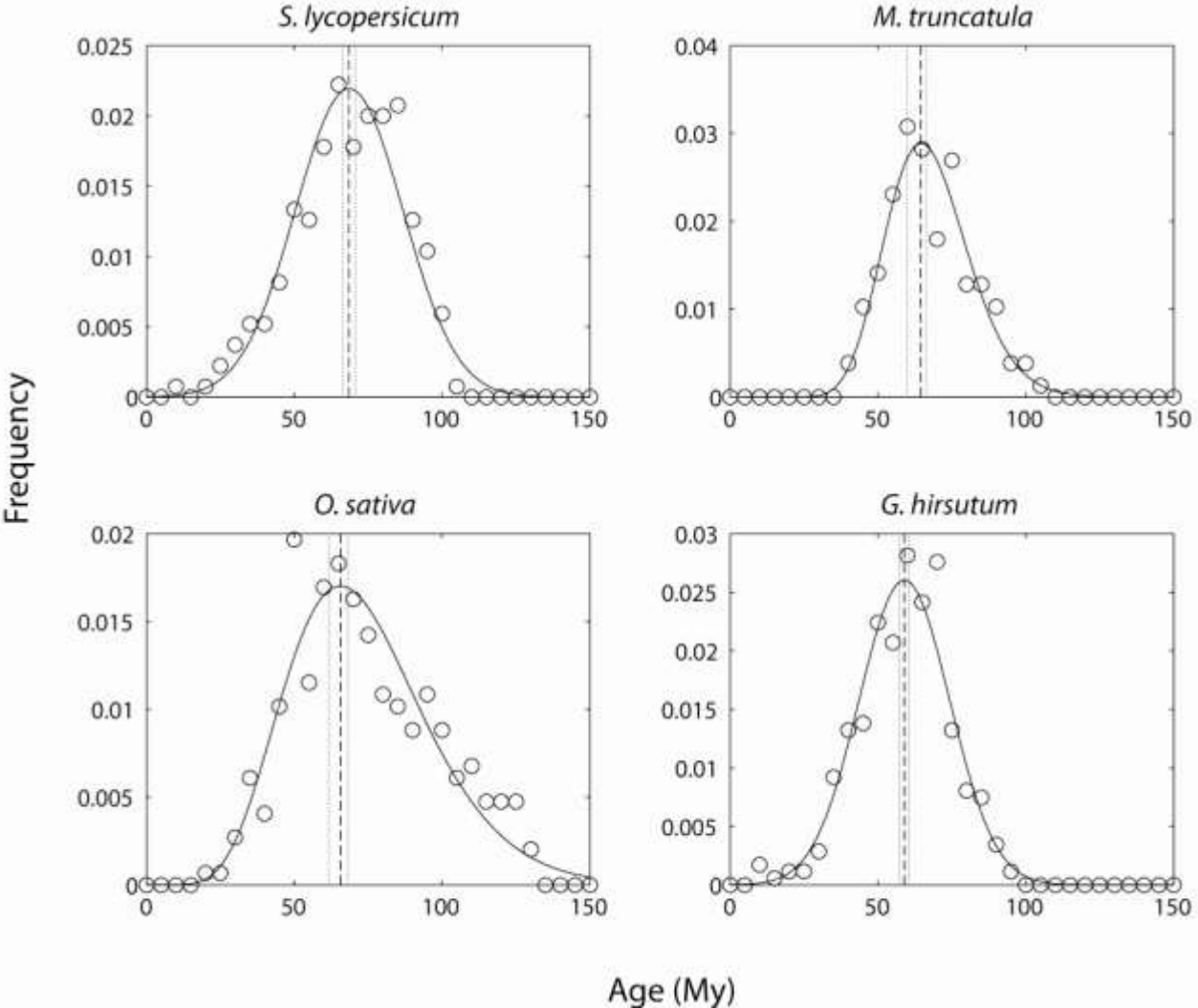
Organism	Molecular clock rates (K_s)
<i>Arabidopsis thaliana</i>	25 – 30 (8), 65 mya (11)
<i>Populus trichocarpa</i>	13 mya (3)
<i>Vitis vinifera</i>	No recent duplication
<i>Medicago truncatula</i>	> 50 mya (29); 58 mya (10)
<i>Lotus japonicus^b</i>	> 50 mya (29)
<i>Glycine max</i>	44 mya (10)
<i>Gossypium hirsutum</i>	13 – 15 mya (8)
<i>Carica papaya</i>	No recent duplication
<i>Solanum tuberosum^b</i>	50 – 52 mya (10)
<i>Solanum lycopersicum</i>	50 - 52 mya (10)
<i>Eschscholzia californica</i>	Unknown
<i>Musa spp,</i>	61 mya (13)
<i>Oryza sativa</i>	50 – 60 mya (10); 70 mya (6)
<i>Sorghum bicolor^b</i>	50 – 60 mya (10); 70 mya (6)
<i>Acorus americanus</i>	Unknown
<i>Physcomitrella patens</i>	Unknown

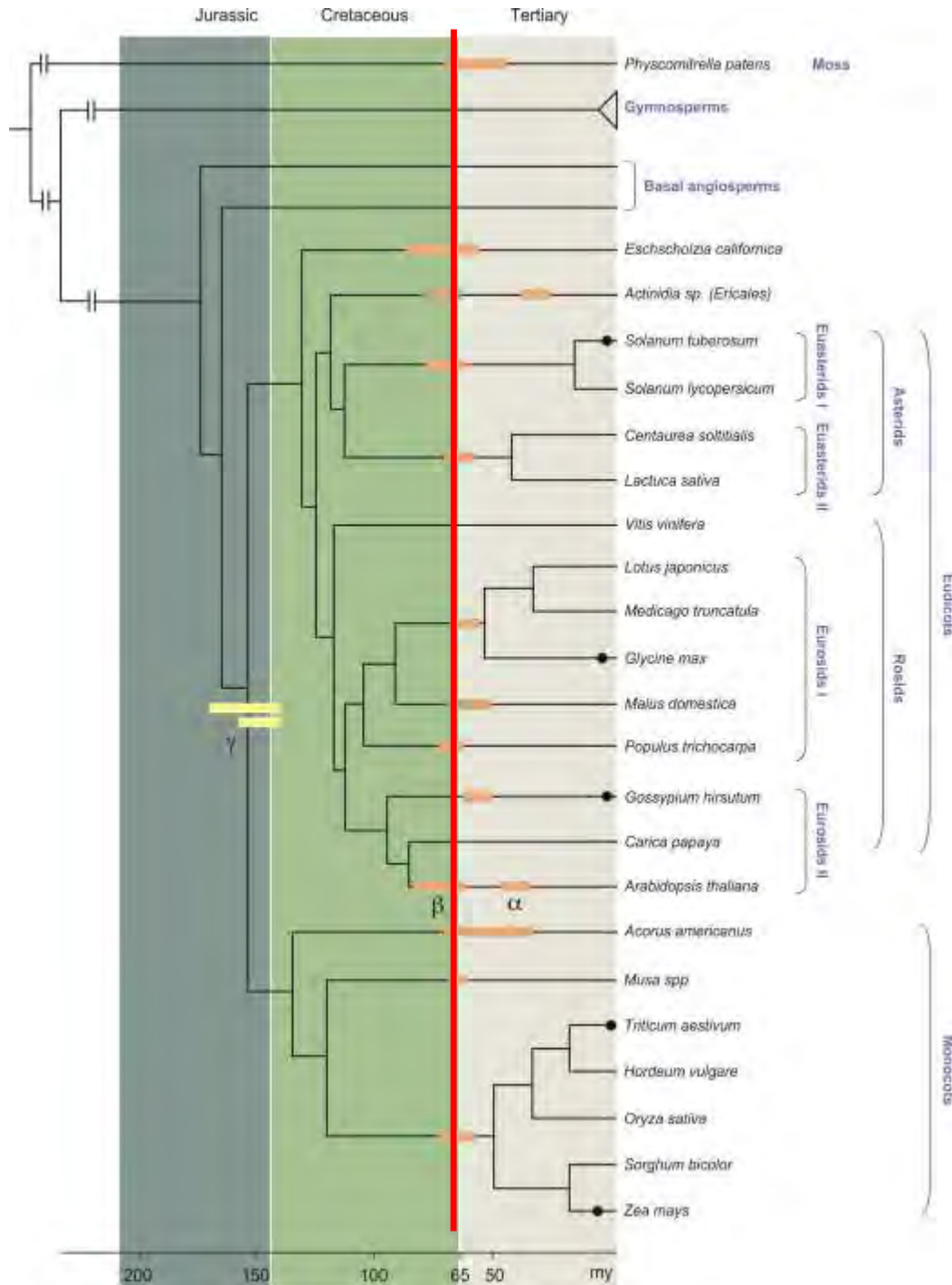
Cui et al., Genome Research (2006)

Absolute dating through the construction of phylogenetic trees

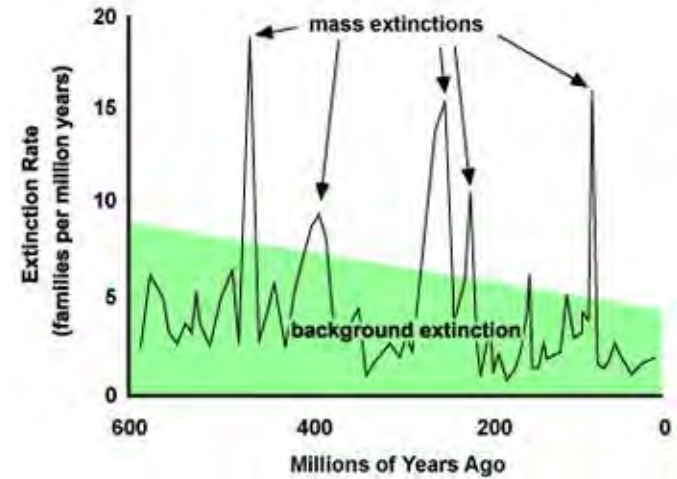


Absolute dating through the construction of phylogenetic trees





Duplications are clustered in time



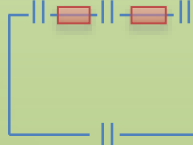
The **Cretaceous–Tertiary (KT) extinction event** is known as the most recent large-scale mass extinction of animal and plant species in a geologically short period of time, **approximately 65 mya**.

Fawcett et al. (2009) *PNAS*
 Vanneste et al. (2014) *Genome Research*
 (accepted, pending revisions)

Jurassic

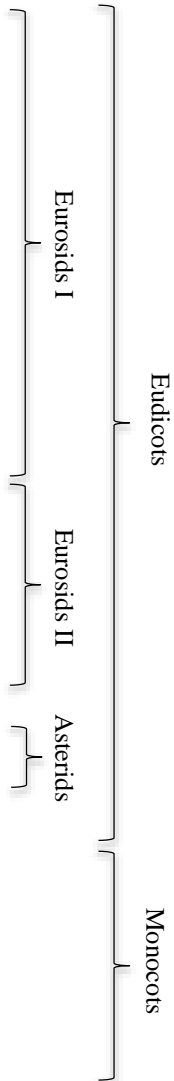
Cretaceous

Tertiary



175 150 125 100 75 50 25 mya

- Cucumis melo*
- Cucumis sativus*
- Citrullus lanatus*
- Malus domestica*
- Pyrus bretschneideri*
- Prunus persica*
- Prunus mume*
- Fragaria vesca*
- Glycine max*
- Cajanus cajan*
- Medicago truncatula*
- Cicer arietinum*
- Lotus japonicus*
- Ricinus communis*
- Manihot esculenta*
- Jatropha curcas*
- Linum usitatissimum*
- Populus trichocarpa*
- Brassica rapa*
- Thelungiella parvula*
- Arabidopsis thaliana*
- Arabidopsis lyrata*
- Carica papaya*
- Theobroma cacao*
- Gossypium raimondii*
- Eucalyptus grandis*
- Vitis vinifera*
- Solanum lycopersicum*
- Solanum tuberosum*
- Lactuca sativa*
- Nelumbo nucifera*
- Aquilegia formosa x pubescens*
- Brachypodium distachyon*
- Hordeum vulgare*
- Oryza sativa*
- Zea mays*
- Sorghum bicolor*
- Setaria italica*
- Musa acuminata*
- Phoenix dactylifera*
- Phaeodactylum equestris*
- Nuphar advena*
- Physcomitrella patens*



65 mya: the Cretaceous–Tertiary (KT) extinction event ...



Hiroshima: 15 kiloton
Tsar Bomba: 57 megaton
K-T impactor: 2 million times Tsar Bomba



65 mya: the Cretaceous–Tertiary (KT) extinction event ...

These catastrophic events have:

- caused huge wildfires
- reduced sunlight and prolonged darkness
- caused temperatures to drop
- caused acid rain

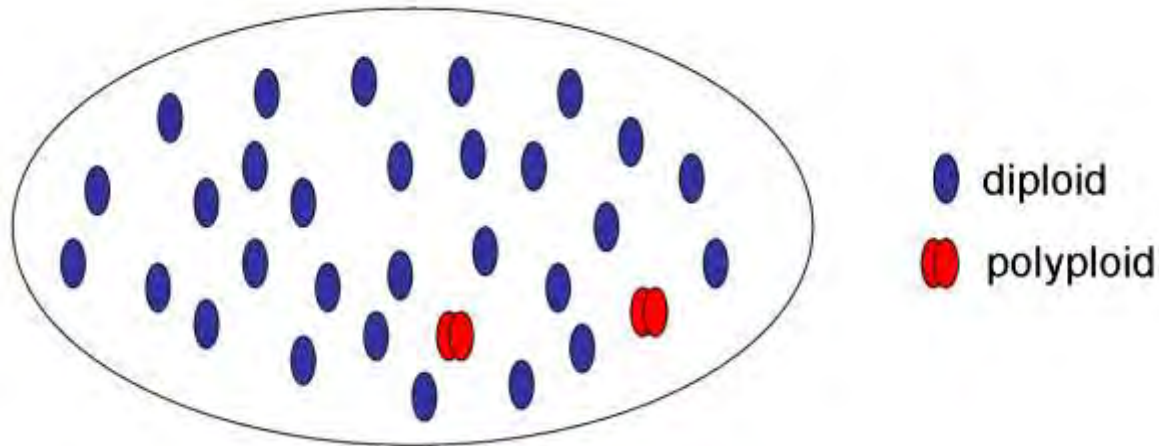
As a result, there was:

- hindered photosynthesis
- reduced germination of seeds
- abrupt extinction or local disappearance of terrestrial vegetation

Paleobotanical studies of fossil pollen, spores and leaves from North American localities have shown that 18-30% of plant genera and families and **up to 60% of plant species disappeared** at the KT boundary.

Polyploids: evolutionary dead ends or hopeful monsters?

Although polyploids occur frequently in the wild under natural conditions

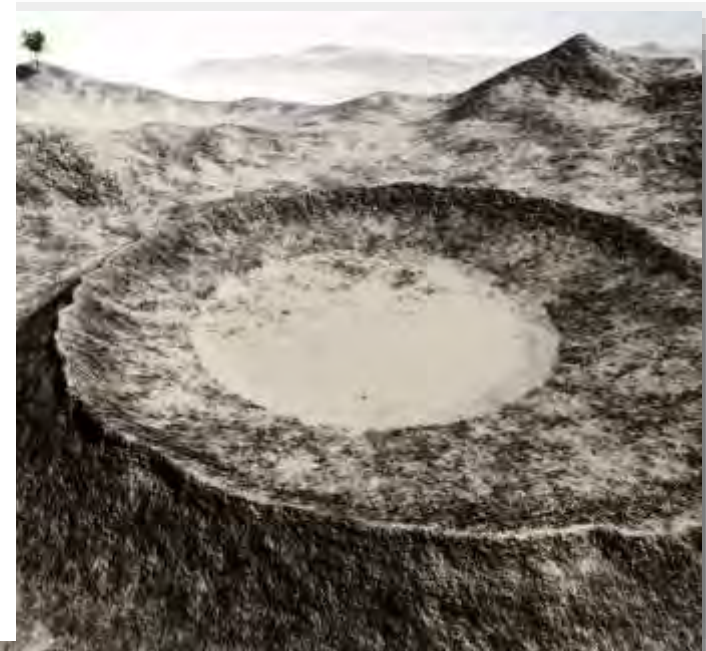
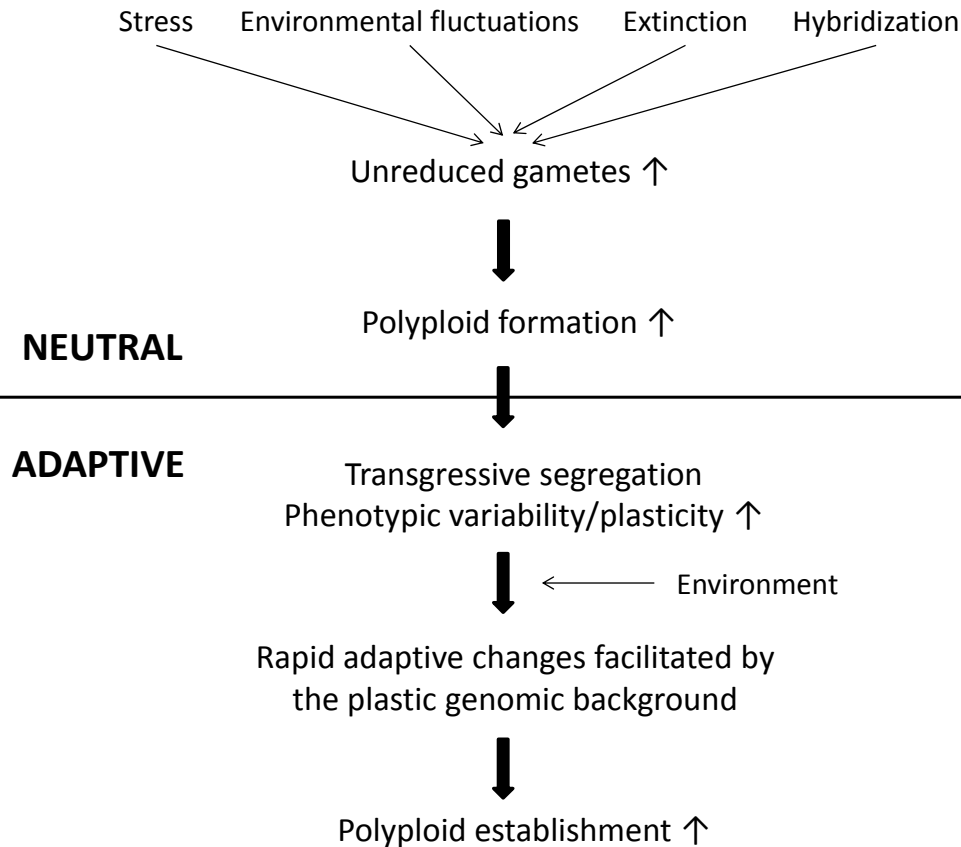


Under 'normal' circumstances

Not many survive, not many get established, not many proliferate

Many polyploids have reduced fertility, reduced fitness, ...

Polyploids: evolutionary dead ends or hopeful monsters?



Acknowledgements

