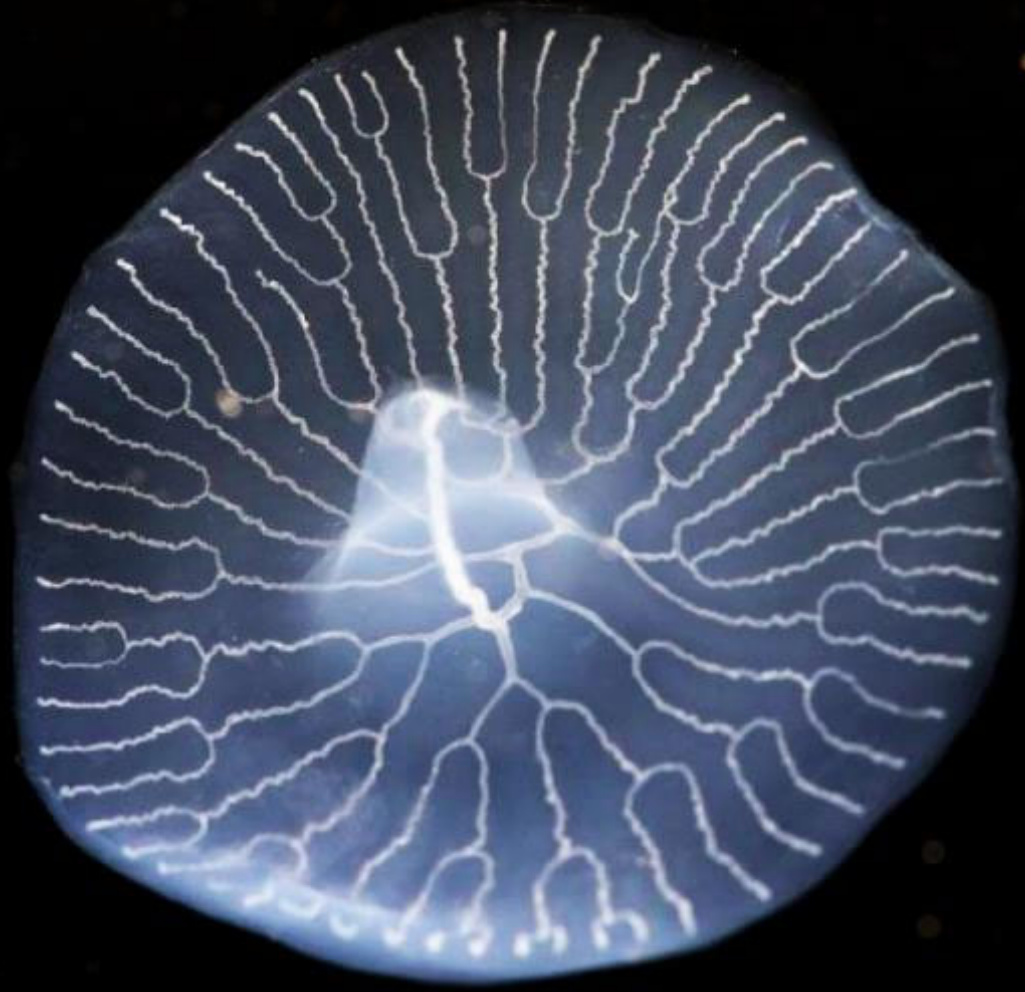
# *Evolutionary Genomics*



**Antonis Rokas**
***Department of Biological Sciences, Vanderbilt University***
**http://www.rokaslab.org**
**@RokasLab**

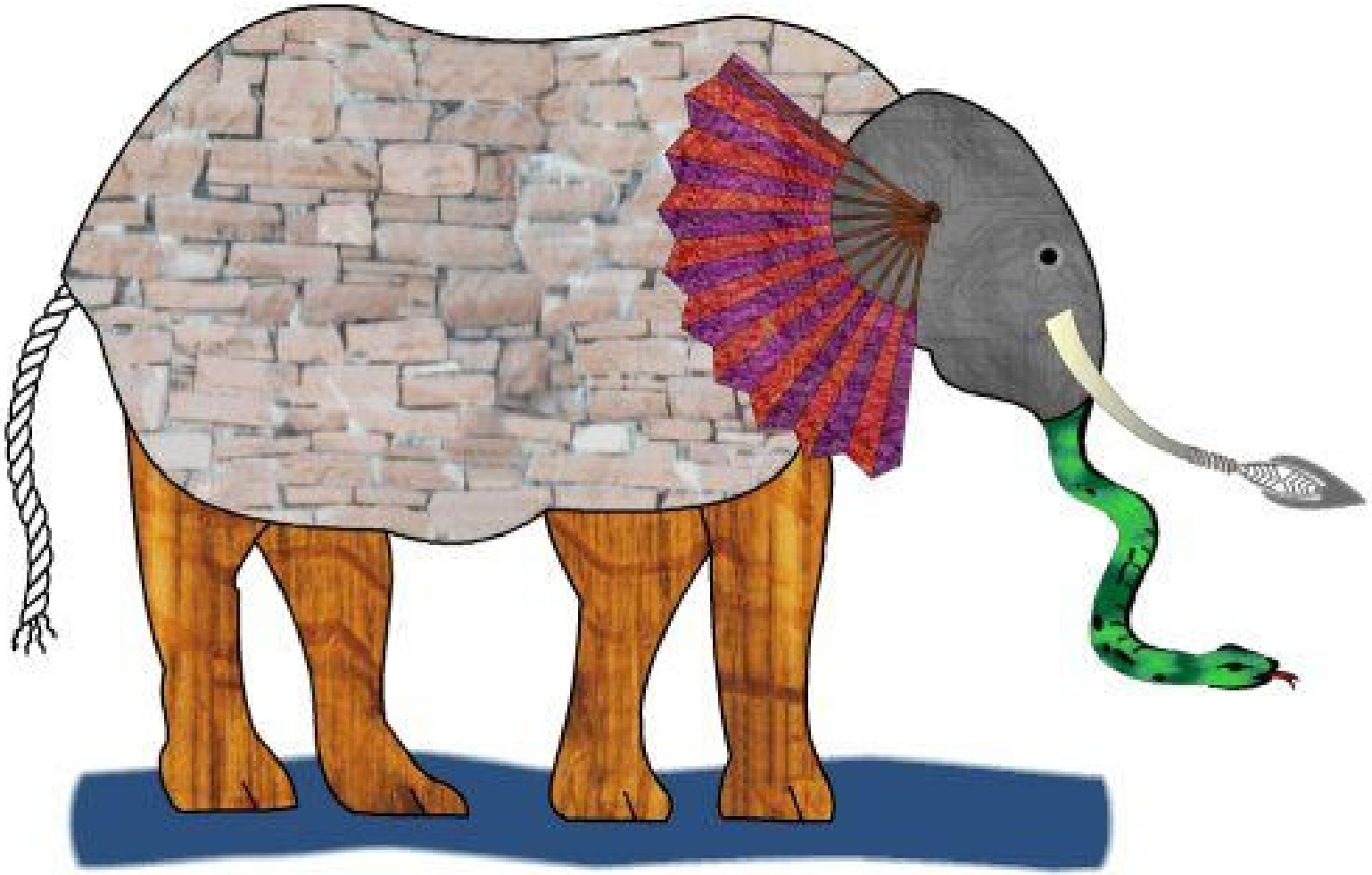
# *Lecture Outline*

- **Introduction to evolutionary genomics**
- **Phylogenomics**

**---------- Coffee Break ----------**

- **Using genomes to understand ecology and evolution**

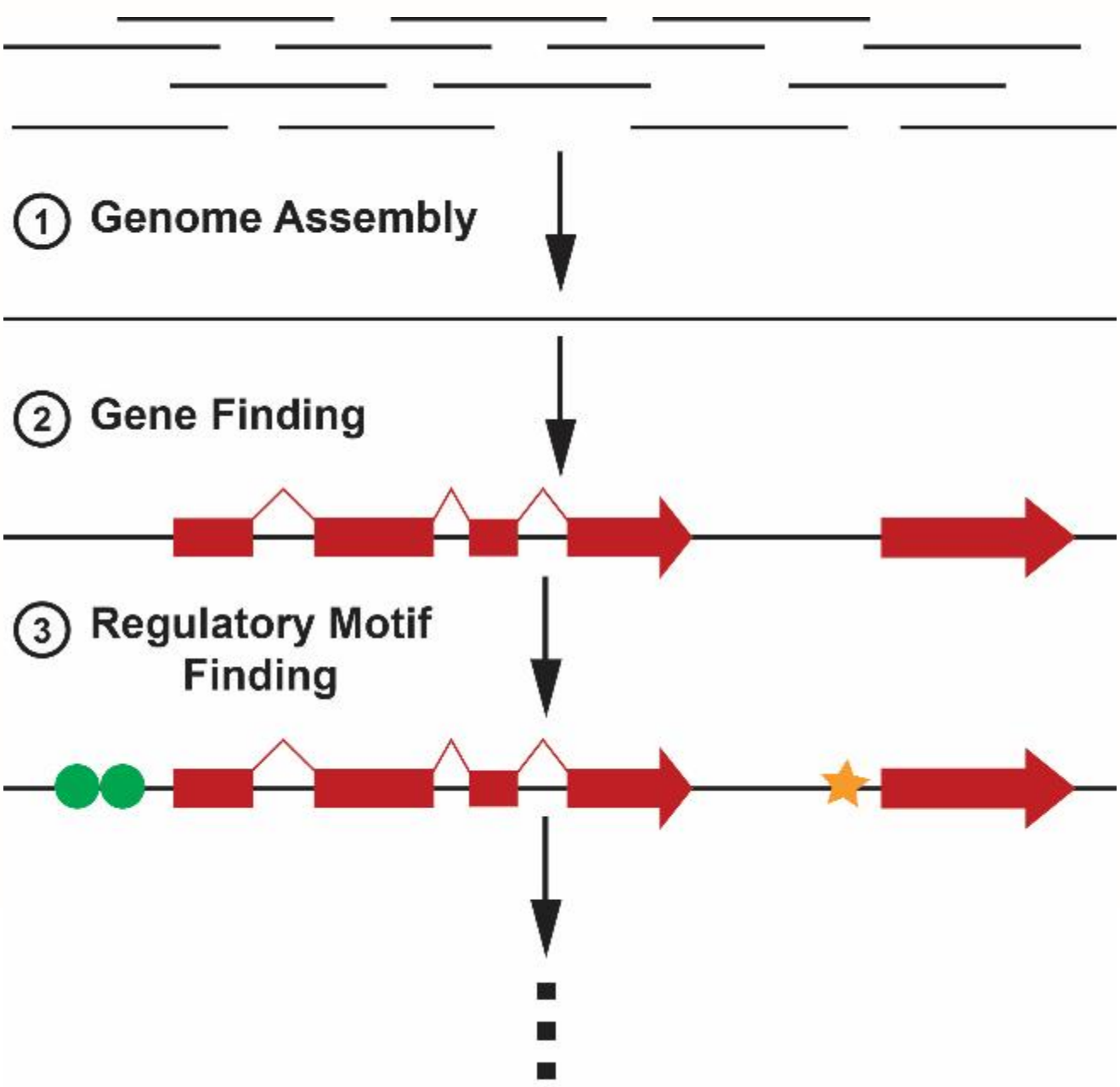
# *What is an Elephant Like?*



# *What is a Genome Like?*

ACAACCCCTCCACCTCATGTACCTGCGGACTCTCCTCCAGTCACAGCTCAGGCAGTCCACTTTGCAACCCCTAAACCTCAAAACCGGTTT
GACGTTCTGTTAGACGAACAACTATGATATATCGACCCCGCCTAAGAACGGAGCCTCTGTCAGTGCTCCAGCTGAACGTAGGCCGCGGG
CCAGCCACTCATGAAATCGCCCTTTCATTAGCATACTCTAGCGGCATTGATATCATCCTTATACAGGAGCCATACATATATACTGACCTCA
GCCGGCAAATCACAAAAAGGCACCCATCATACGAGTGCTTCTCCCCAACAGACAGCTGGCTTGTAAGCGGTCGACCCCGGGTCCTCACC
TATGTCCGGAAAAAGATGGGCATTCGGGCCTCTCAGCTCCGCCCTCAGCCAATAGATCAAGATGTTCTCTCAGACCTTCTTCTACTACAG
ATCCTCTCCCGCTCTGGACAATCTGCATTGATAATCAACATCTATAATGCTCCAATCGGCTCAATCAGGTCAGGTGAGGCTGCAAAAGCG
CTTACACTCCTGCCTGACTCCTACTTTTCCCAGCCTACCGTGCTTGCCGGCGACTTCAACCTACTACATAGCAGGTGGCAGCCATCACTG
CATTGCAGCCCTACCACCTTTGCTGAGCCATTTGTTGACTGGCTTGATCGCCTAGGGCTGGTTCTTATCTCCGAGATAGACCAGCCTACAC
ACGATAGAGGCAACGTTCTTGACCTCACTTTCGCCTCCAGCTCCCTAGCACTGGCAGGGTCGAGTACCAGGATAGCAAGTCATTTAGAGT
CAACATCAGATCATCGGCCACTCCTCACCACCATGCCATGGAGCCAGAGATTCACAGAGGCAGCTCAGAAACTGAGATTTGATACATTA
GACCACCCTCGCTTCCTCTCACTACTCAGTTCCCACCTTGCTGTCATTGAATGCTCAGCTACAACAGAAGAGGGCCTGGACAGTCTAGCT
CATGGGTTAACCTTAGCAACTGCTAGTGCGTATAAAGGCTCTGCTAGGAGCTCCTTGGCGCAGGGAATAGGTCAGCCATGGTGGAATATT
GACTGCAGAAAAGCGTTGCAAGACTTCCGCTTAGGTCTCTGTTCAAGAAACGACTTCCGTCGGATAACTAGACGGTCTAAATAGCAGTTC
TGGCGAGATAAACTTACCGCAGTGACACAGATCAAAGATGTCTTTGACATAAGCAAGTGACATAAGTTTACAGGATCTTATCGAAACCCT
CCACTAAACGACCCTTTAAGGCCAAACAGCCCTCCAGCAGGGGCTCTGAATGAGAAACAAGACGTATTAGTCCGTAATCTTCTTCAGAAT
ACTGCTGAAGCGGGTGATATTGTCATAGGCTATGGCCTGGGCTGTGGTTGTCAGCCATGCCCTCAACCATAGAACATTCTAGAAGAACCA
TCGGGAAGAGGTTGGAACCCAGTGGAAGTTTGGGAACATGTATATAAGAAGGAGAGGGAGATGTATCTGCCTATTTCTCTCTCCAAGTCT
GCGATATTCGTTAATACATTATACAGGATTGCCAGTTGAAAACAATACTGCCTACGCCCGTCACAGGTACTGCAGTTTCCAACAAGAATC
AACGCTCGACCCGGCAATTATGGCTCAAGGTTAGACTACGTCCTGTGTAGCCTTGATATGCAAGATTAGTTCTGCGATTTGAATATCTAAG
AGGATCTAATGGTAAGCCCCAAGGCTGCCATGGCTTTATTGTAGATTGATTTTCTAGCTGACAATATGCAATTTGGGACAGGGATCTGATG
ATTGTCCGGTTTATGCTGTCTTCAAAAATGTTATACGCCTCGGCGAAGAAGAGGTCAACATTAAATGAGCCCTCCTGGGATGTTTAAAGAT
GGCGAGCGTCAGCAGGAATACTCTACTAAATATCTTCTGCCTACATCAGGGCGCTTAATACCAGAATTTAACAAGCGGAGGAGGATCAA
GGACATGTTCTTGCGTAAACCATCAGCCAACGTATAGAGACCGACGACGAACATCCTGACATTGAGATATTTTACCTCTAGTCAGGAAAA
GGGAACAGCACCCGCTATTTTGGAGAGTGCTGCCAGCGTCATAGCTACCTGCCAGCCTGTAGTAGCTGCTGACAGCACTCAAATGAAAG
AAGTTATTCGTAAGAGCTCTCAGAAATATGAGACAGGTTCCCCTGTCTCAGTCCAGTATTTGACATCGGGTTCAGCCCAATCATCAACAC
CCCCCACTGCTGGACAGAGGACTCTAAAGGGGTTCTTCAAACTTAAAAGTGGTCTAGCCAGCCAAATGGCCATAGCCCAGGATCCTGCA
ACAGTGTCTACTATGCCAACGAAACAACCAGCCGCATCCCCTACAAAATCTACCCCAGTTACAGAACCTCCTGCACTGGAAGCATTACTG
ACAGCTCCCGCTGGTGAAGCTTCTCCAGGAGAACAGCCAAATTCCGCGACTCCTACAGCTCCCGCTTCACCCCAAAGCAATGATACTATT
ATCGATCCCATTGTCAGCAAGGAAGATTGGTCAAAGCTCTTCACTAAAAAGCCCATTCCCAAGTGCGAGGGCCACCAGGAACCATGTTT
CAGTCTGACAACTAAGAAGCCTGGCATCAACTGCGGAAGATCGTTCTGGATCTGTTTGAGACCCCTTGGGCCCAGCGGAAACAAGGAAA
AGGGGATACAGTGGCGATTTCCTACATTCATATGGGCCAGCGATTGGAACCCTTCCGCTCCGTAGATTTTCTGTCTGGGGCAACTTCTTTT
TGCGATAGTGTAACGATACCCGGTTTTATACTTAGAAGGCTACGAATGGTATGATGTATCATGGTTTCAATGATAAGACATTTCGTCAAGT

# *Understanding the Genome Requires Tools*

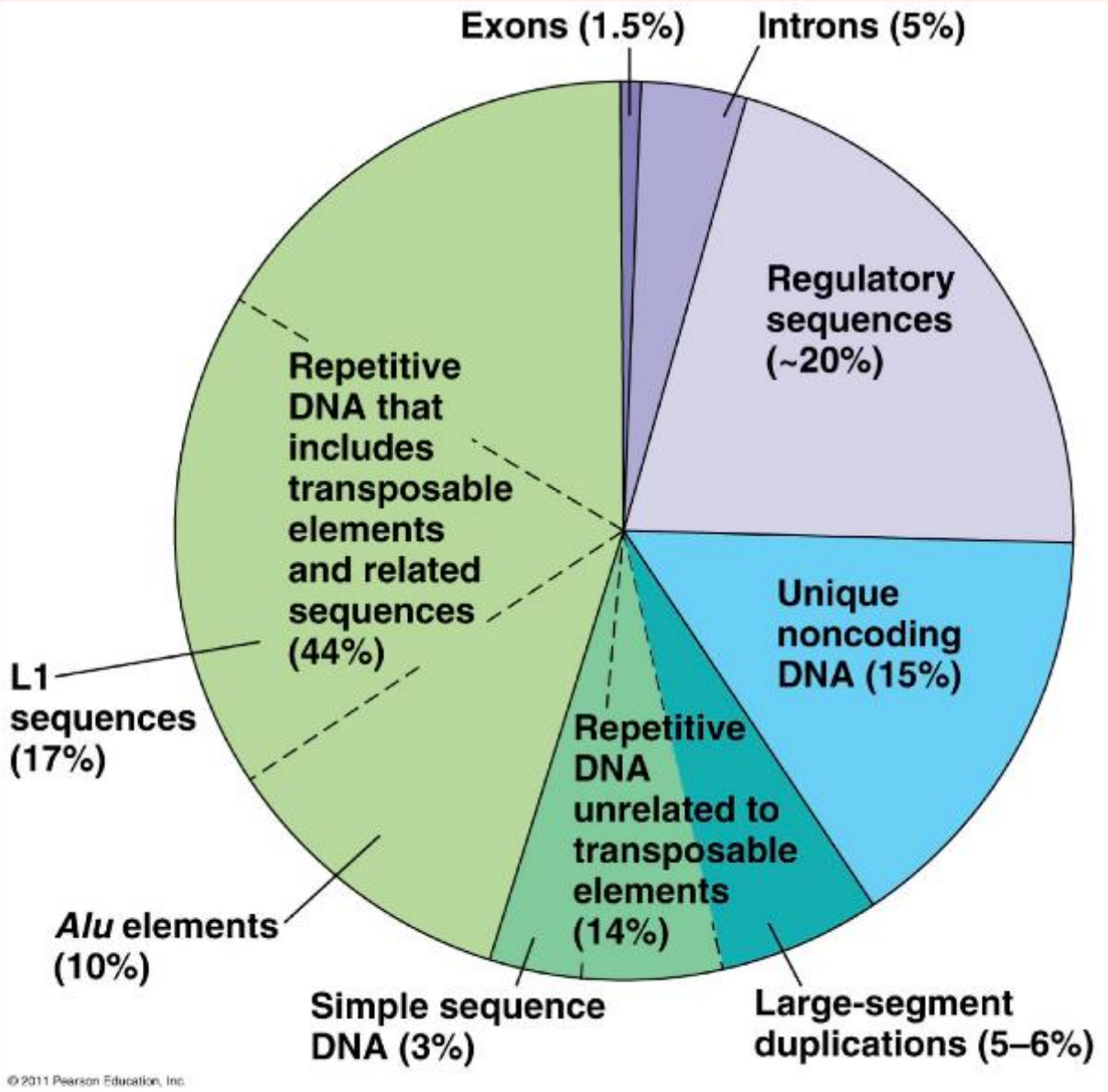


# *What is a Genome Like?*

ACAACCCCTCCACCTCATGTACCTGCGGACTCTCCTCCAGTCACAGCTCAGGCAGTCCACTTTGCAACCCCTAAACCTCAAAACCGGTTT
GACGTTCTGTTAGACGAACAACTATGATATATCGACCCCGCCTAAGAACGGAGCCTCTGTCAGTGCTCCAGCTGAACGTAGGCCGCGGG
CCAGCCACTCATGAAATCGCCCTTTCATTAGCATACTCTAGCGGCATTGATATCATCCTTATACAGGAGCCATACATATATACTGACCTCA
GCCGGCAAATCACAAAAAGGCACCCATCATACGAGTGCTTCTCCCCAACAGACAGCTGGCTTGTAAGCGGTCGACCCCGGGTCCTCACC
TATGTCCGGAAAAAGATGGGCATTCGGGCCTCTCAGCTCCGCCCTCAGCCAATAGATCAAGATGTTCTCTCAGACCTTCTTCTACTACAG
ATCCTCTCCCGCTCTGGACAATCTGCATTGATAATCAACATCTATAATGCTCCAATCGGCTCAATCAGGTCAGGTGAGGCTGCAAAAGCG
CTTACACTCCTGCCTGACTCCTACTTTTCCCAGCCTACCGTGCTTGCCGGCGACTTCAACCTACTACATAGCAGGTGGCAGCCATCACTG
CATTGCAGCCCTACCACCTTTGCTGAGCCATTTGTTGACTGGCTTGATCGCCTAGGGCTGGTTCTTATCTCCGAGATAGACCAGCCTACAC
ACGATAGAGGCAACGTTCTTGACCTCACTTTCGCCTCCAGCTCCCTAGCACTGGCAGGGTCGAGTACCAGGATAGCAAGTCATTTAGAGT
CAACATCAGATCATCGGCCACTCCTCACCACCATGCCATGGAGCCAGAGATTCACAGAGGCAGCTCAGAAACTGAGATTTGATACATTA
GACCACCCTCGCTTCCTCTCACTACTCAGTTCCCACCTTGCTGTCATTGAATGCTCAGCTACAACAGAAGAGGGCCTGGACAGTCTAGCT
CATGGGTTAACCTTAGCAACTGCTAGTGCGTATAAAGGCTCTGCTAGGAGCTCCTTGGCGCAGGGAATAGGTCAGCCATGGTGGAATATT
GACTGCAGAAAAGCGTTGCAAGACTTCCGCTTAGGTCTCTGTTCAAGAAACGACTTCCGTCGGATAACTAGACGGTCTAAATAGCAGTTC
TGGCGAGATAAACTTACCGCAGTGACACAGATCAAAGATGTCTTTGACATAAGCAAGTGACATAAGTTTACAGGATCTTATCGAAACCCT
CCACTAAACGACCCTTTAAGGCCAAACAGCCCTCCAGCAGGGGCTCTGAATGAGAAACAAGACGTATTAGTCCGTAATCTTCTTCAGAAT
ACTGCTGAAGCGGGTGATATTGTCATAGGCTATGGCCTGGGCTGTGGTTGTCAGCCATGCCCTCAACCATAGAACATTCTAGAAGAACCA
TCGGGAAGAGGTTGGAACCCAGTGGAAGTTTGGGAACATGTATATAAGAAGGAGAGGGAGATGTATCTGCCTATTTCTCTCTCCAAGTCT
GCGATATTCGTTAATACATTATACAGGATTGCCAGTTGAAAACAATACTGCCTACGCCCGTCACAGGTACTGCAGTTTCCAACAAGAATC
AACGCTCGACCCGGCAATTATGGCTCAAGGTTAGACTACGTCCTGTGTAGCCTTGATATGCAAGATTAGTTCTGCGATTTGAATATCTAAG
AGGATCTAATGGTAAGCCCCAAGGCTGCCATGGCTTTATTGTAGATTGATTTTCTAGCTGACAATATGCAATTTGGGACAGGGATCTGATG
ATTGTCCGGTTTATGCTGTCTTCAAAAATGTTATACGCCTCGGCGAAGAAGAGGTCAACATTAAATGAGCCCTCCTGGGATGTTTAAAGAT
GGCGAGCGTCAGCAGGAATACTCTACTAAATATCTTCTGCCTACATCAGGGCGCTTAATACCAGAATTTAACAAGCGGAGGAGGATCAA
GGACATGTTCTTGCGTAAACCATCAGCCAACGTATAGAGACCGACGACGAACATCCTGACATTGAGATATTTTACCTCTAGTCAGGAAAA
GGGAACAGCACCCGCTATTTTGGAGAGTGCTGCCAGCGTCATAGCTACCTGCCAGCCTGTAGTAGCTGCTGACAGCACTCAAATGAAAG
AAGTTATTCGTAAGAGCTCTCAGAAATATGAGACAGGTTCCCCTGTCTCAGTCCAGTATTTGACATCGGGTTCAGCCCAATCATCAACAC
CCCCCACTGCTGGACAGAGGACTCTAAAGGGGTTCTTCAAACTTAAAAGTGGTCTAGCCAGCCAAATGGCCATAGCCCAGGATCCTGCA
ACAGTGTCTACTATGCCAACGAAACAACCAGCCGCATCCCCTACAAAATCTACCCCAGTTACAGAACCTCCTGCACTGGAAGCATTACTG
ACAGCTCCCGCTGGTGAAGCTTCTCCAGGAGAACAGCCAAATTCCGCGACTCCTACAGCTCCCGCTTCACCCCAAAGCAATGATACTATT
ATCGATCCCATTGTCAGCAAGGAAGATTGGTCAAAGCTCTTCACTAAAAAGCCCATTCCCAAGTGCGAGGGCCACCAGGAACCATGTTT
CAGTCTGACAACTAAGAAGCCTGGCATCAACTGCGGAAGATCGTTCTGGATCTGTTTGAGACCCCTTGGGCCCAGCGGAAACAAGGAAA
AGGGGATACAGTGGCGATTTCCTACATTCATATGGGCCAGCGATTGGAACCCTTCCGCTCCGTAGATTTTCTGTCTGGGGCAACTTCTTTT
TGCGATAGTGTAACGATACCCGGTTTTATACTTAGAAGGCTACGAATGGTATGATGTATCATGGTTTCAATGATAAGACATTTCGTCAAGT

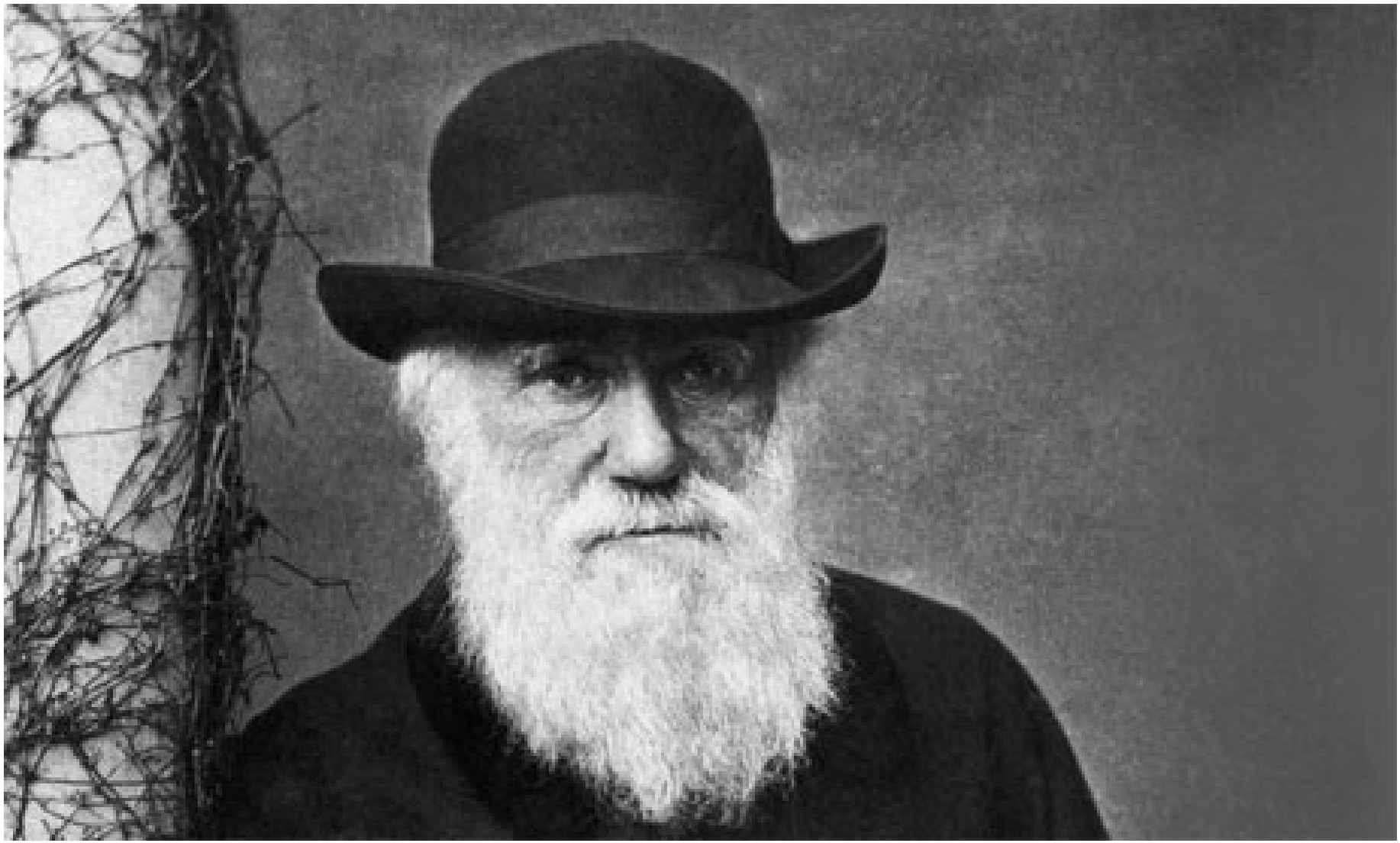
**Transposon** **Protein Binding Site** **Exon** **Intron**

# *Organization of the Human Genome*

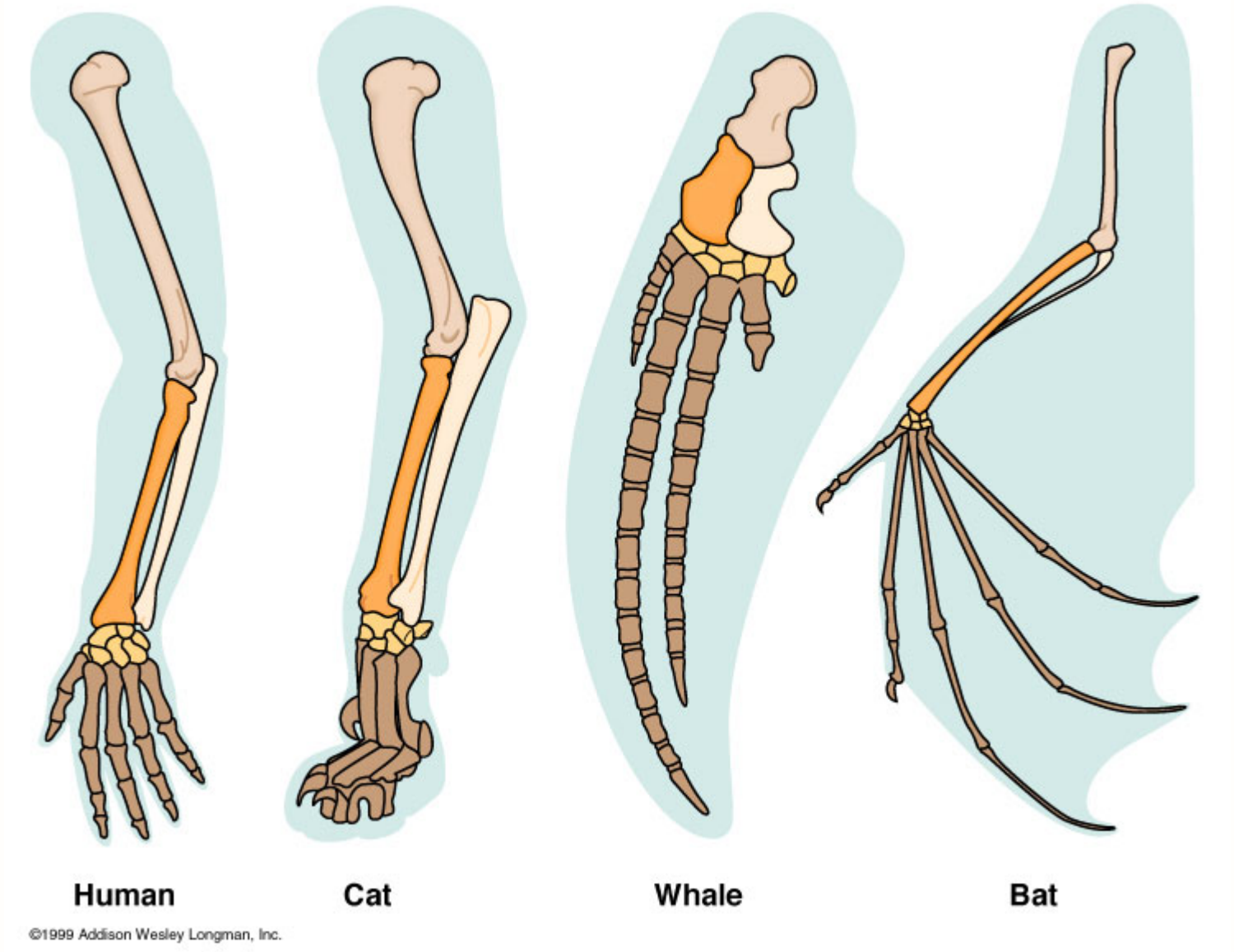



*http://todd.jackman.villanova.edu/HumanGenome.jpg*

# *Understanding the Genome Requires a Theory*



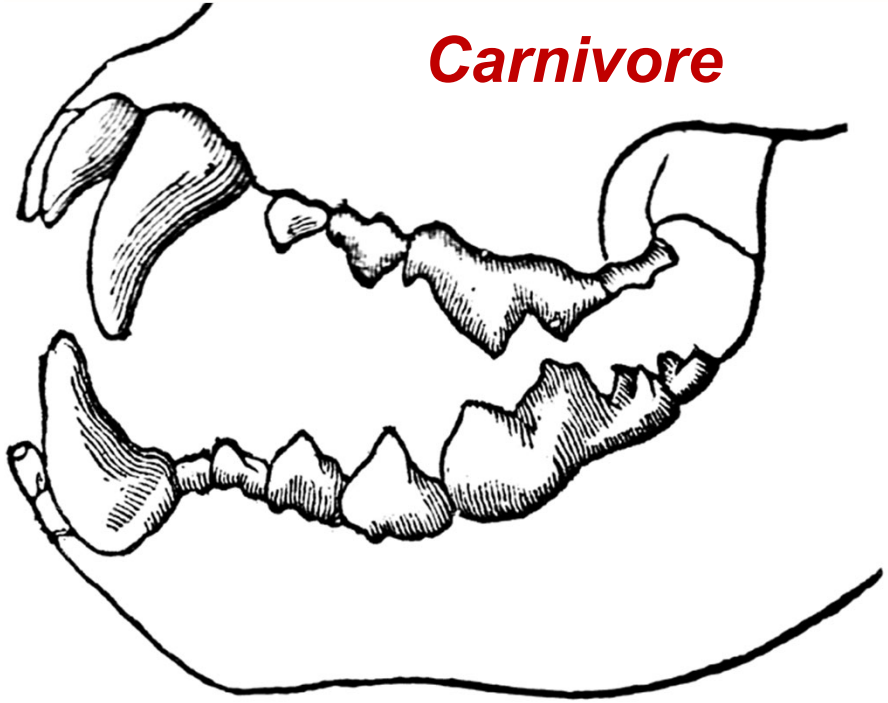
# *Similarity in Anatomy Suggests Common Origins*



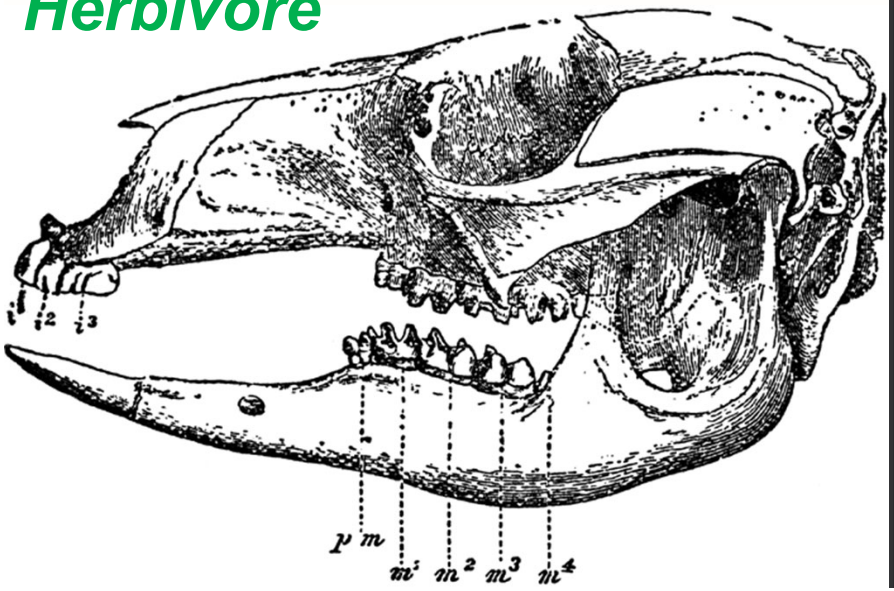

*http://www.mun.ca/biology/scarr/139393_forelimb_homology.jpg*

# *Differences in Anatomy Suggest Adaptations*

**Carnivore**



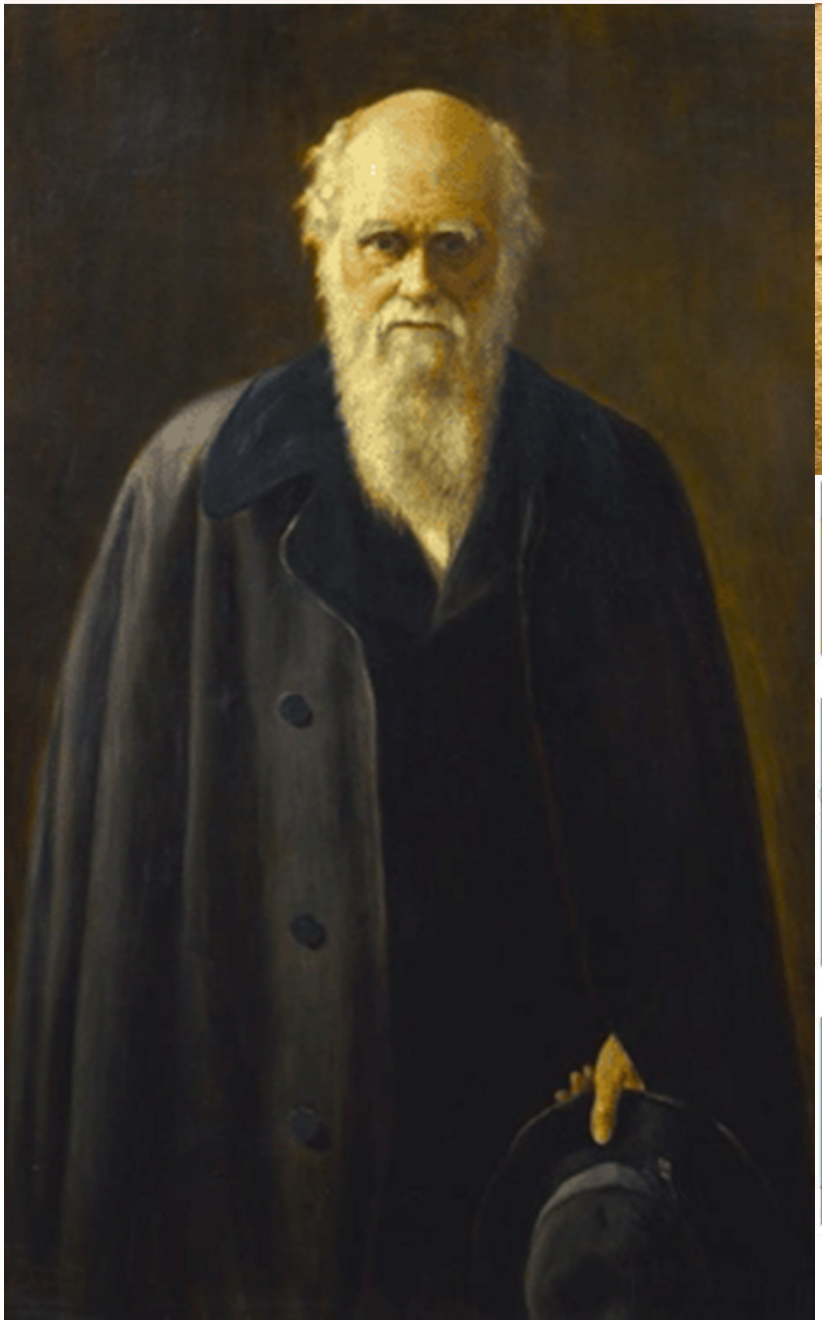
**Herbivore**

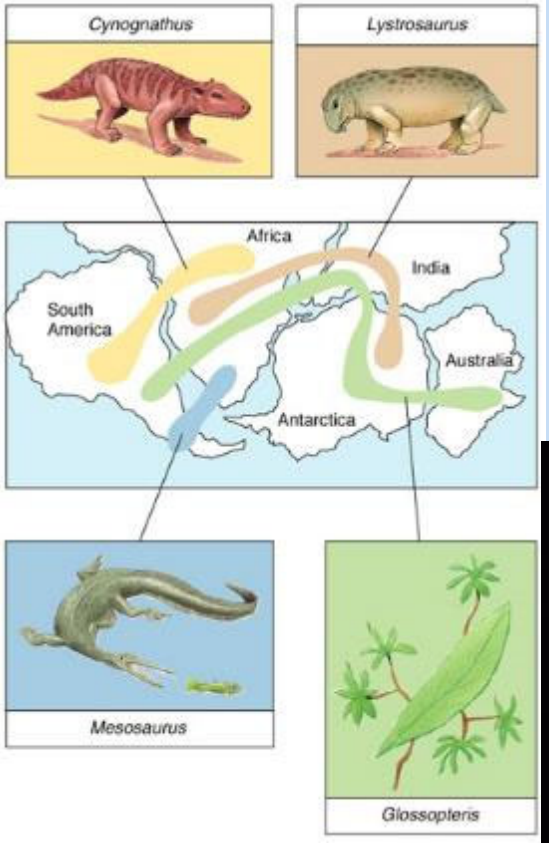


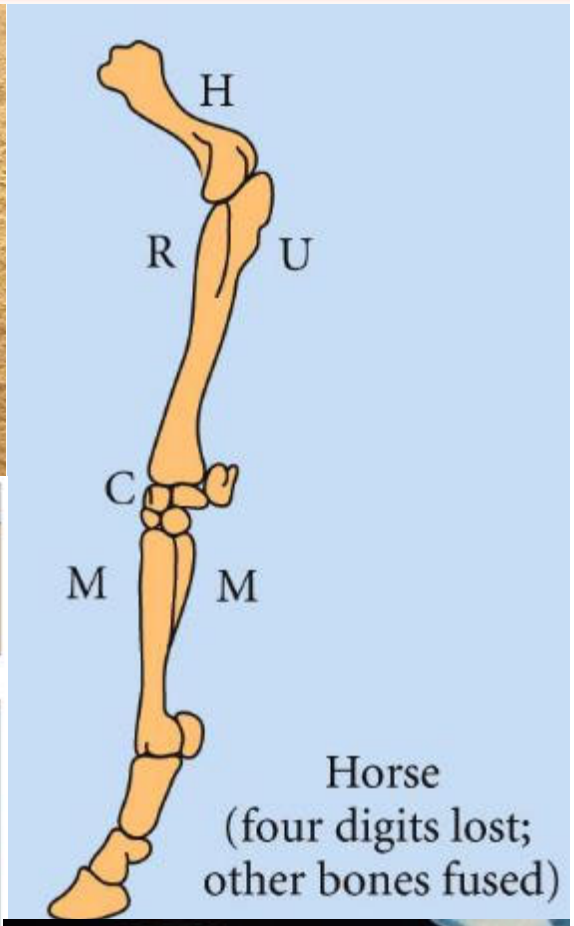

***Incisivosaurus***

**Carnivore or Herbivore?**

# *Darwin's Data*

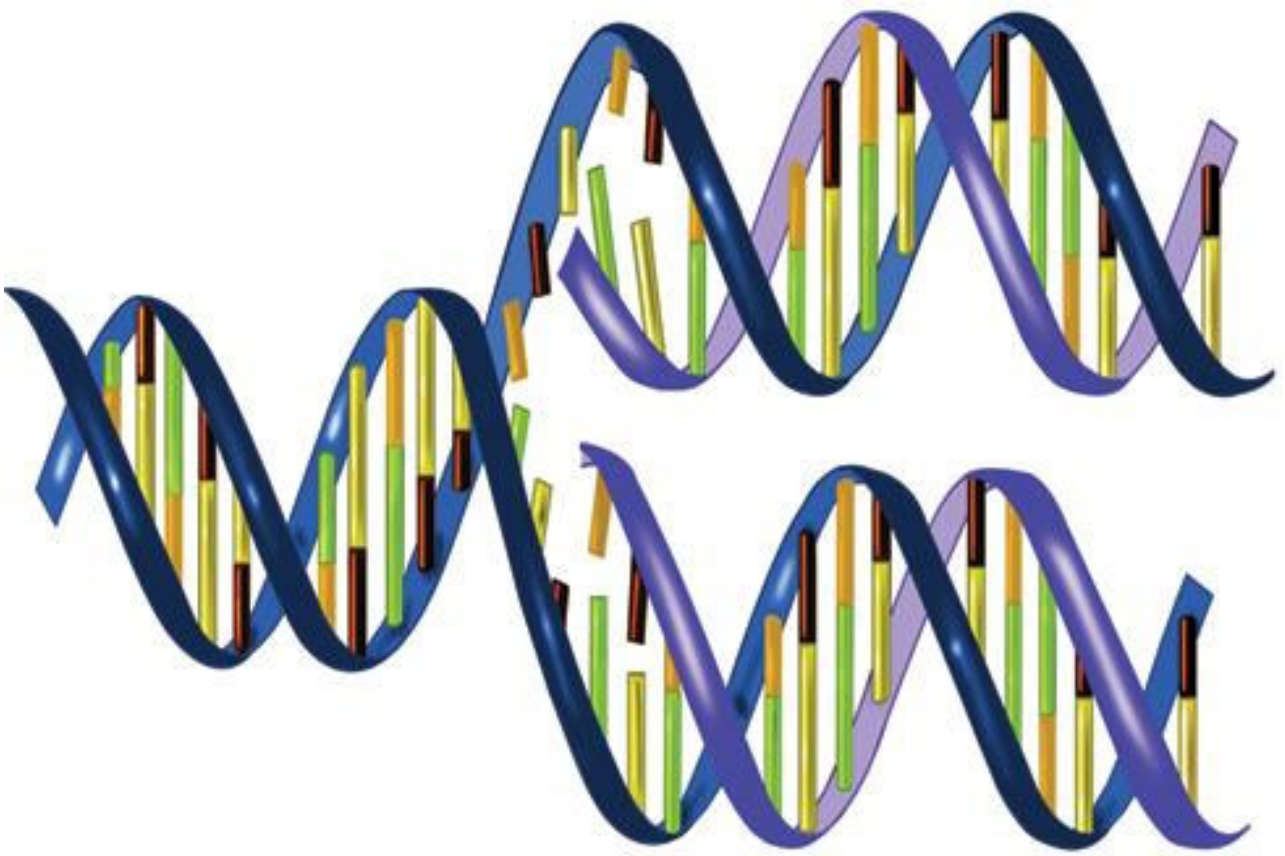





# *The DNA Record*

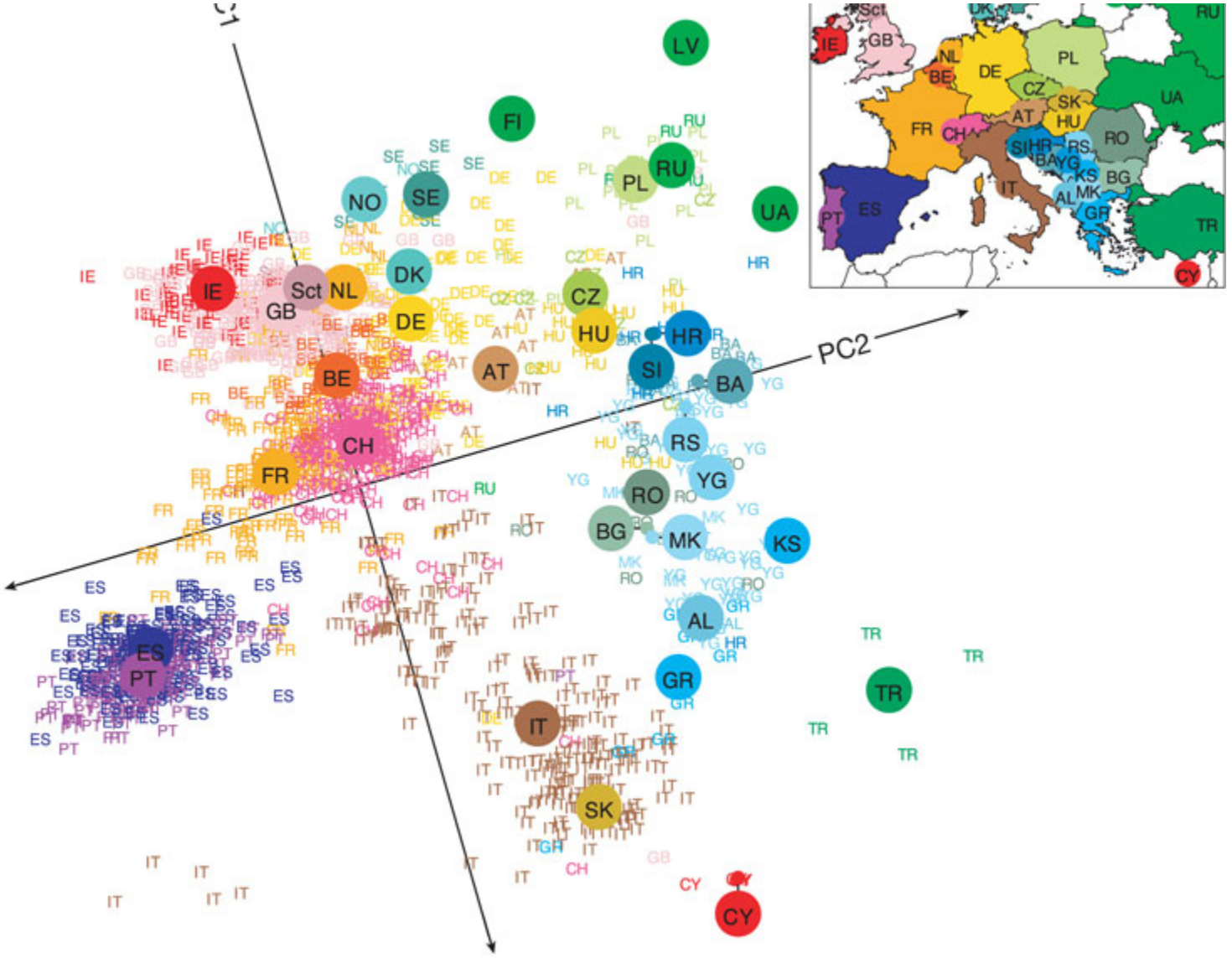
**The DNA record contains important clues about organisms' biological past, and their history of change and adaptation**



**Similarities in the DNA record suggest common origin**

**Differences in the DNA record (might) suggest adaptations**

# *Human Genes Mirror Geography*




*Novembre et al. (2008) Nature*

# *Recent Positive Selection in Human Populations*

in the Asian Population, involved in hair follicle development

**The twenty-two strongest candidates for natural selection**

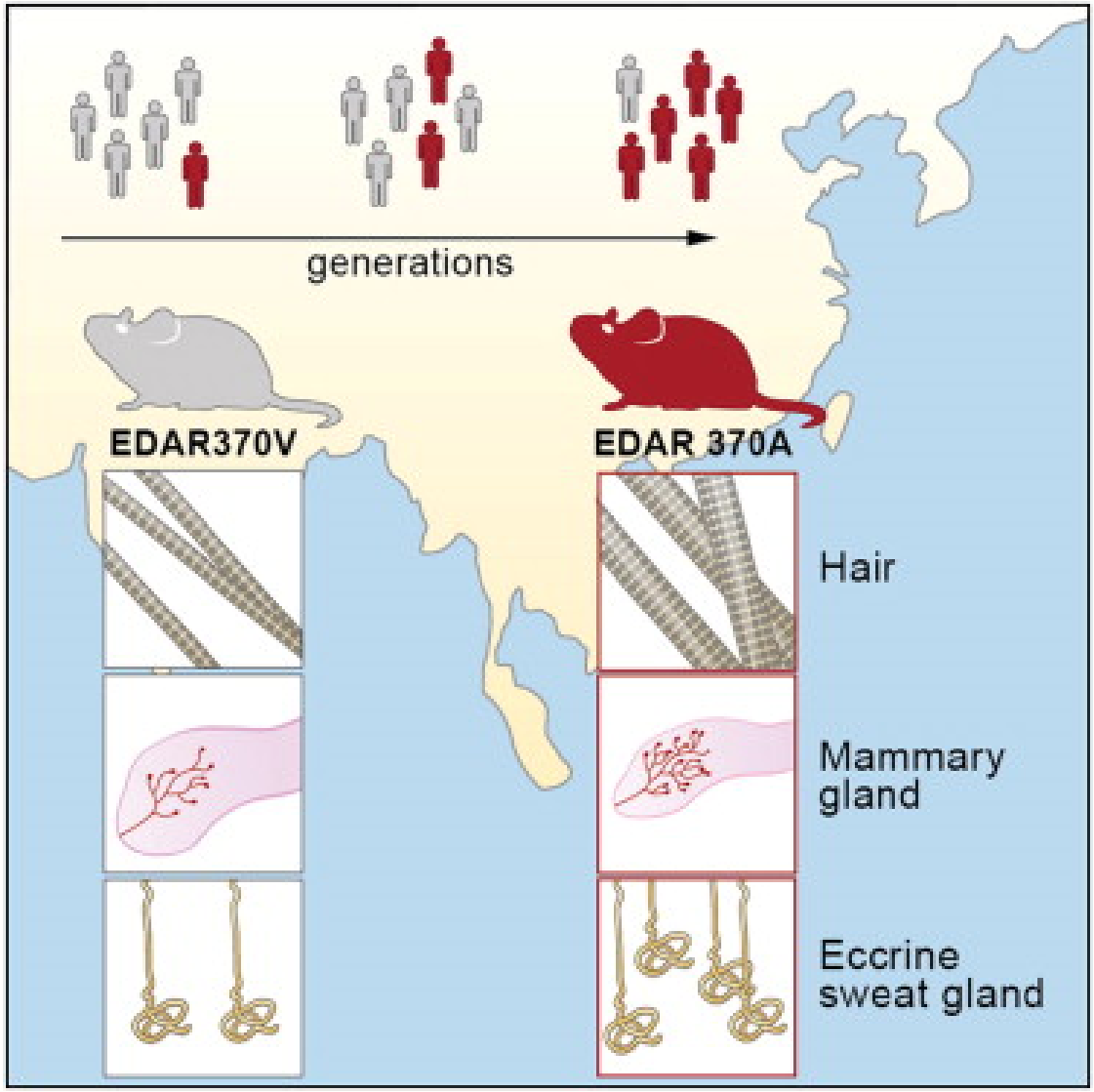
| Chr:position (MB, HG17) | Selected population | Long Haplotype Test | Size (Mb) | Total SNPs with Long Haplotype Signal | Subset of SNPs that fulfil criteria 1 | Subset of SNPs that fulfil criteria 1 and 2 | Subset of SNPs that fulfil criteria 1, 2 and 3 | Genes at or near SNPs that fulfil all three criteria |
|---|---|---|---|---|---|---|---|---|
| chr1:166 | CHB + JPT | LRH, iHS | 0.4 | 92 | 39 | 30 | 2 | *BLZF1, SLC19A2* |
| chr2:72.6 | CHB + JPT | XP-EHH | 0.8 | 732 | 250 | 0 | 0 | |
| chr2:108.7 | CHB + JPT | LRH, iHS, XP-EHH | 1.0 | 972 | 265 | 7 | 1 | *EDAR* |
| chr2:136.1 | CEU | LRH, iHS, XP-EHH | 2.4 | 1,213 | 282 | 24 | 3 | *RAB3GAP1, R3HDM1, LCT* |
| chr2:177.9 | CEU,CHB + JPT | LRH, iHS, XP-EHH | 1.2 | 1,388 | 399 | 79 | 9 | *PDE11A* |
| chr4:33.9 | CEU,YRI, CHB + JPT | LRH, iHS | 1.7 | 413 | 161 | 33 | 0 | |
| chr4:42 | CHB + JPT | LRH, iHS, XP-EHH | 0.3 | 249 | 94 | 65 | 6 | *SLC30A9* |
| chr4:159 | CHB + JPT | LRH, iHS, XP-EHH | 0.3 | 233 | 67 | 34 | 1 | |
| chr10:3 | CEU | LRH, iHS, XP-EHH | 0.3 | 179 | 63 | 16 | 1 | |
| chr10:22.7 | CEU, CHB + JPT | XP-EHH | 0.3 | 254 | 93 | 0 | 0 | |
| chr10:55.7 | CHB+JPT | LRH, iHS, XP-EHH | 0.4 | 735 | 221 | 5 | 2 | *PCDH15* |
| chr12:78.3 | YRI | LRH, iHS | 0.8 | 151 | 91 | 25 | 0 | |
| chr15:46.4 | CEU | XP-EHH | 0.6 | 867 | 233 | 5 | 1 | *SLC24A5* |
| chr15:61.8 | CHB + JPT | XP-EHH | 0.2 | 252 | 73 | 40 | 6 | *HERC1* |
| chr16:64.3 | CHB + JPT | XP-EHH | 0.4 | 484 | 137 | 2 | 0 | |
| chr16:74.3 | CHB + JPT, YRI | LRH, iHS | 0.6 | 55 | 35 | 28 | 3 | *CHST5, ADAT1, KARS* |
| chr17:53.3 | CHB + JPT | XP-EHH | 0.2 | 143 | 41 | 0 | 0 | |
| chr17:56.4 | CEU | XP-EHH | 0.4 | 290 | 98 | 26 | 3 | *BCAS3* |
| chr19:43.5 | YRI | LRH, iHS, XP-EHH | 0.3 | 83 | 30 | 0 | 0 | |
| chr22:32.5 | YRI | LRH | 0.4 | 318 | 188 | 35 | 3 | *LARGE* |
| chr23:35.1 | YRI | LRH, iHS | 0.6 | 50 | 35 | 25 | 0 | |
| chr23:63.5 | YRI | LRH, iHS | 3.5 | 13 | 3 | 1 | 0 | |
| | Total SNPs | | 16.74 | 9,166 | 2,898 | 480 | 41 | |

In the European population, involved in skin pigmentation

In the West African population, related to Lassa virus infection

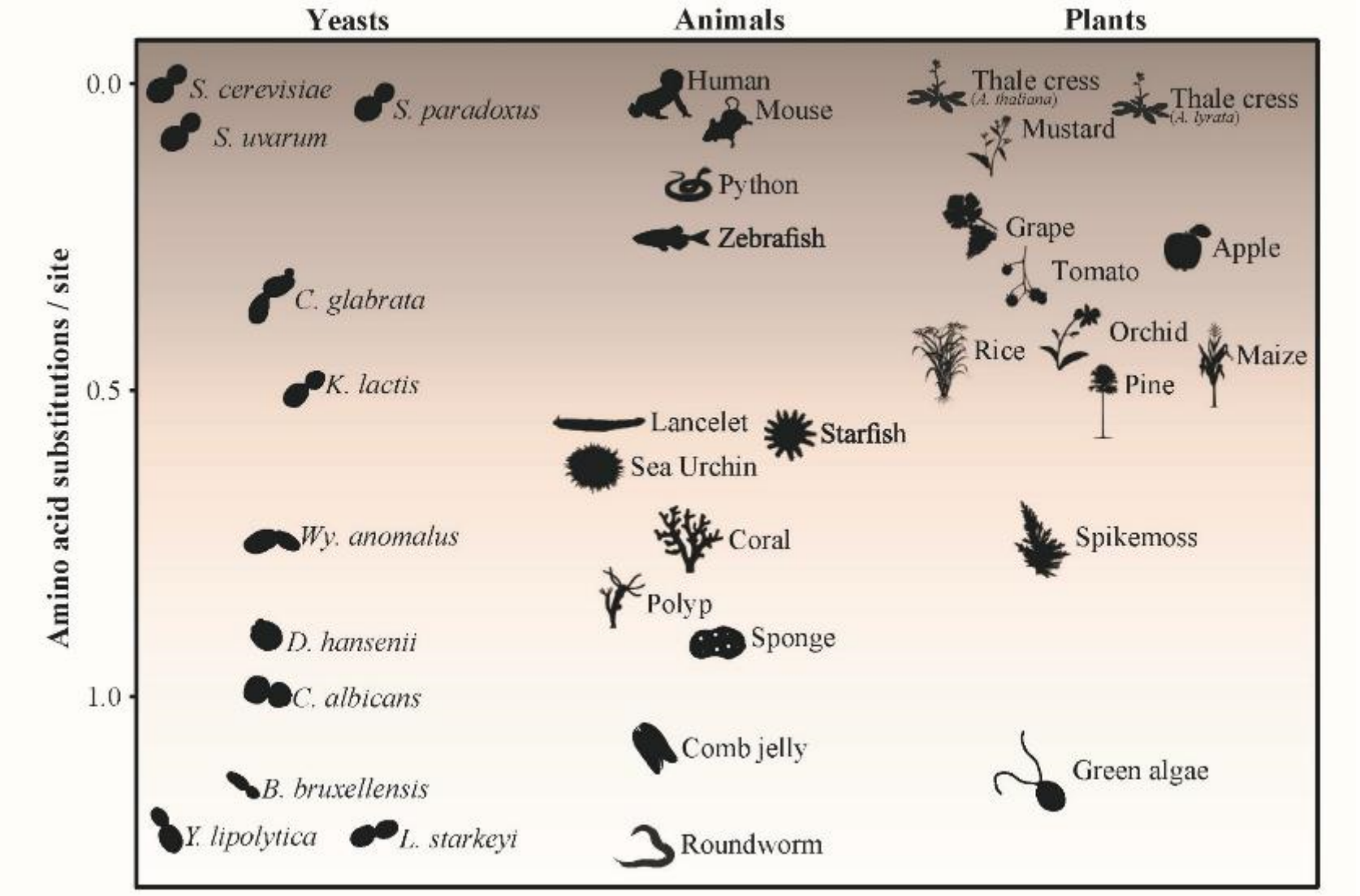
*Sabeti et al. (2007) Nature*

# *Phenotypic Effects of Recent Positive Selection*

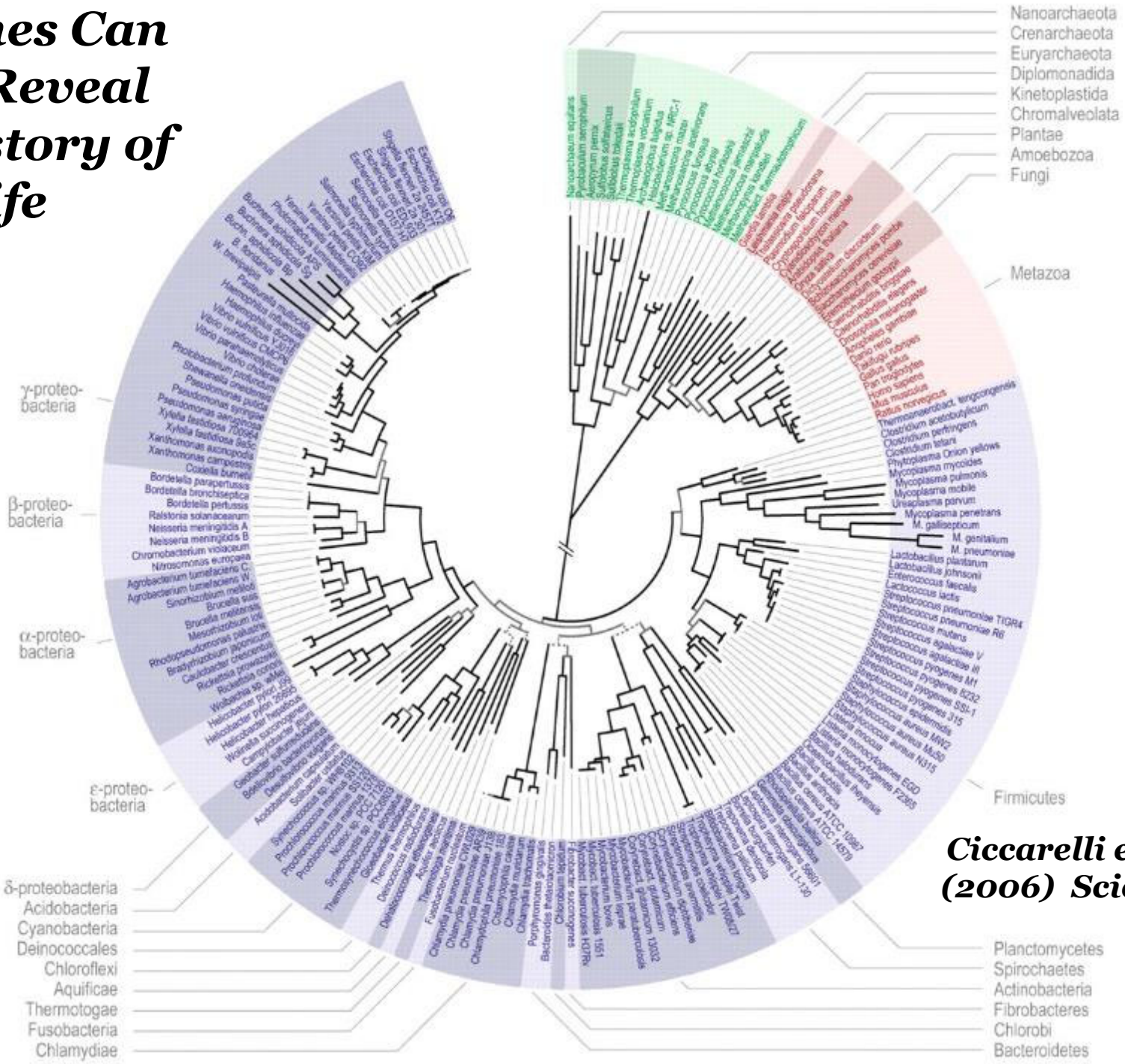



*Kamberov et al. (2013) Cell*

# *Genomes Provide a Common Yardstick for Comparison*



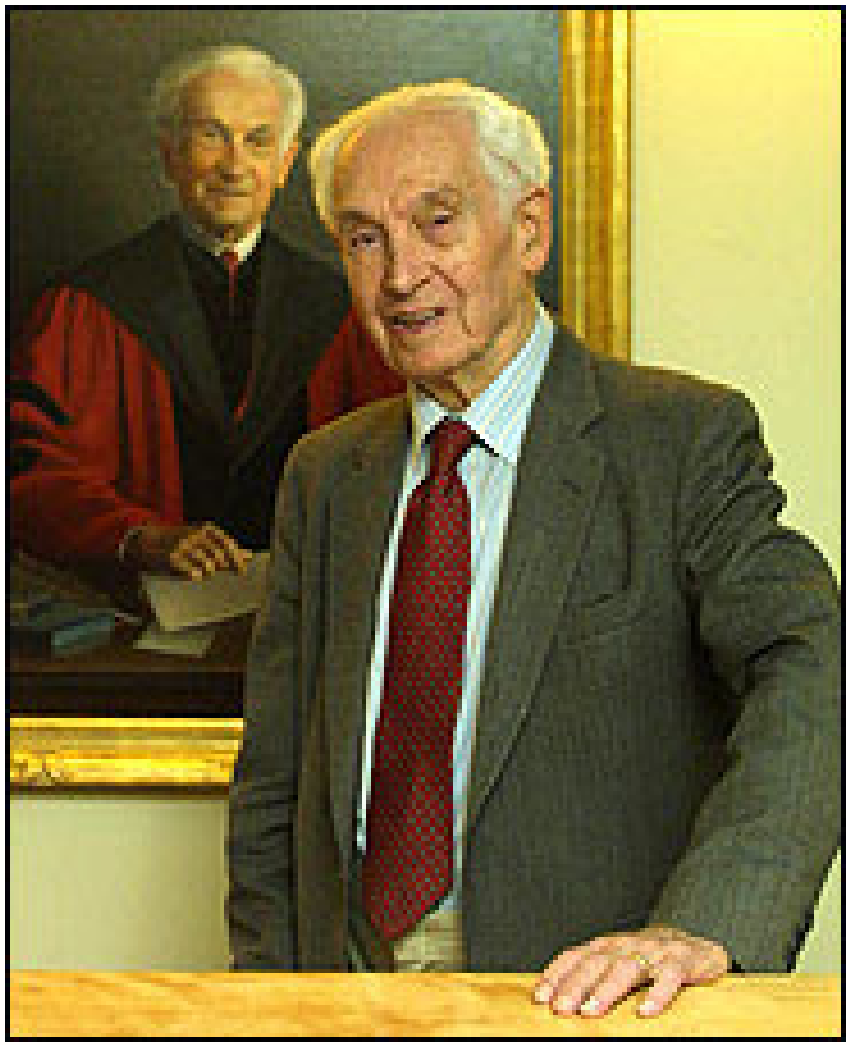
# *Genomes Can Help Reveal the History of Life*



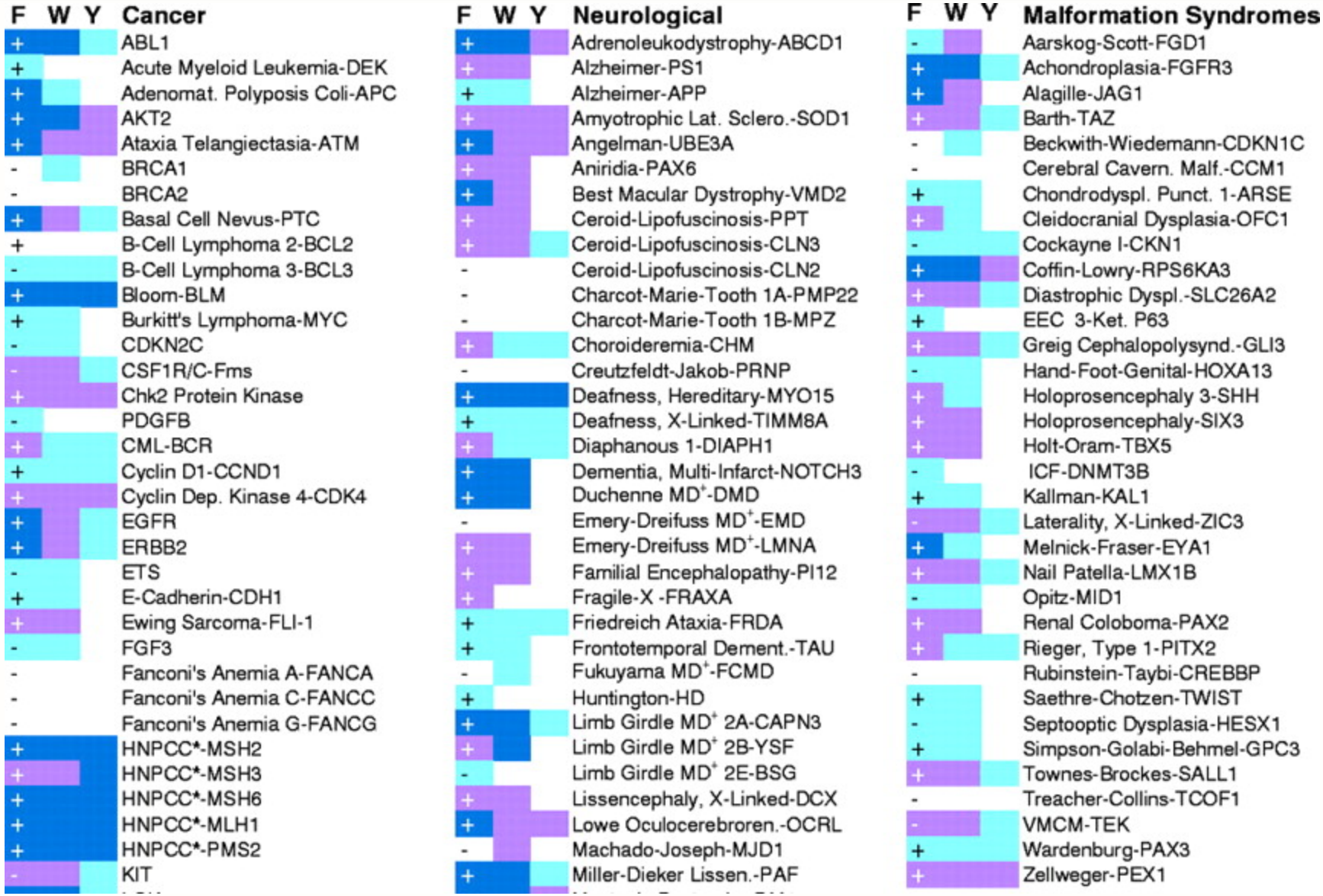

***Ciccarelli et al. (2006) Science***

**“…the search for homologous genes is quite futile except in very close relatives”**

**Ernst Mayr, 1963**



# *What Makes Us Sick Is the Stuff of Life*



| F | W | Y | Cancer |
|---|---|---|---|
| + | + | | ABL1 |
| + | | | Acute Myeloid Leukemia-DEK |
| + | | | Adenomat. Polyposis Coli-APC |
| + | + | | AKT2 |
| + | | | Ataxia Telangiectasia-ATM |
| - | | | BRCA1 |
| - | | | BRCA2 |
| + | | | Basal Cell Nevus-PTC |
| + | | | B-Cell Lymphoma 2-BCL2 |
| - | | | B-Cell Lymphoma 3-BCL3 |
| + | + | | Bloom-BLM |
| + | | | Burkitt's Lymphoma-MYC |
| - | | | CDKN2C |
| - | | | CSF1R/C-Fms |
| + | | | Chk2 Protein Kinase |
| - | | | PDGFB |
| + | | | CML-BCR |
| + | | | Cyclin D1-CCND1 |
| + | | | Cyclin Dep. Kinase 4-CDK4 |
| + | | | EGFR |
| + | | | ERBB2 |
| - | | | ETS |
| + | | | E-Cadherin-CDH1 |
| + | | | Ewing Sarcoma-FLI-1 |
| - | | | FGF3 |
| - | | | Fanconi's Anemia A-FANCA |
| - | | | Fanconi's Anemia C-FANCC |
| - | | | Fanconi's Anemia G-FANCG |
| + | + | | HNPCC*-MSH2 |
| + | | | HNPCC*-MSH3 |
| + | + | | HNPCC*-MSH6 |
| + | + | | HNPCC*-MLH1 |
| + | + | | HNPCC*-PMS2 |
| - | | | KIT |

| F | W | Y | Neurological |
|---|---|---|---|
| + | | | Adrenoleukodystrophy-ABCD1 |
| + | | | Alzheimer-PS1 |
| + | | | Alzheimer-APP |
| + | | | Amyotrophic Lat. Sclero.-SOD1 |
| + | | | Angelman-UBE3A |
| + | | | Aniridia-PAX6 |
| + | | | Best Macular Dystrophy-VMD2 |
| + | | | Ceroid-Lipofuscinosis-PPT |
| + | | | Ceroid-Lipofuscinosis-CLN3 |
| - | | | Ceroid-Lipofuscinosis-CLN2 |
| - | | | Charcot-Marie-Tooth 1A-PMP22 |
| - | | | Charcot-Marie-Tooth 1B-MPZ |
| + | | | Choroideremia-CHM |
| - | | | Creutzfeldt-Jakob-PRNP |
| + | + | | Deafness, Hereditary-MYO15 |
| + | | | Deafness, X-Linked-TIMM8A |
| + | | | Diaphanous 1-DIAPH1 |
| + | | | Dementia, Multi-Infarct-NOTCH3 |
| + | | | Duchenne MD$^+$-DMD |
| - | | | Emery-Dreifuss MD$^+$-EMD |
| + | | | Emery-Dreifuss MD$^+$-LMNA |
| + | | | Familial Encephalopathy-PI12 |
| + | | | Fragile-X -FRAXA |
| + | | | Friedreich Ataxia-FRDA |
| + | | | Frontotemporal Dement.-TAU |
| - | | | Fukuyama MD$^+$-FCMD |
| + | | | Huntington-HD |
| + | | | Limb Girdle MD$^+$ 2A-CAPN3 |
| + | | | Limb Girdle MD$^+$ 2B-YSF |
| - | | | Limb Girdle MD$^+$ 2E-BSG |
| + | | | Lissencephaly, X-Linked-DCX |
| + | | | Lowe Oculocerebroren.-OCRL |
| - | | | Machado-Joseph-MJD1 |
| + | | | Miller-Dieker Lissen.-PAF |

| F | W | Y | Malformation Syndromes |
|---|---|---|---|
| - | | | Aarskog-Scott-FGD1 |
| + | + | | Achondroplasia-FGFR3 |
| + | | | Alagille-JAG1 |
| + | | | Barth-TAZ |
| - | | | Beckwith-Wiedemann-CDKN1C |
| - | | | Cerebral Cavern. Malf.-CCM1 |
| + | | | Chondrodyspl. Punct. 1-ARSE |
| + | | | Cleidocranial Dysplasia-OFC1 |
| - | | | Cockayne I-CKN1 |
| + | + | | Coffin-Lowry-RPS6KA3 |
| + | | | Diastrophic Dyspl.-SLC26A2 |
| + | | | EEC 3-Ket. P63 |
| + | | | Greig Cephalopolysynd.-GLI3 |
| - | | | Hand-Foot-Genital-HOXA13 |
| + | | | Holoprosencephaly 3-SHH |
| + | | | Holoprosencephaly-SIX3 |
| + | | | Holt-Oram-TBX5 |
| - | | | ICF-DNMT3B |
| + | | | Kallman-KAL1 |
| - | | | Laterality, X-Linked-ZIC3 |
| + | | | Melnick-Fraser-EYA1 |
| + | | | Nail Patella-LMX1B |
| - | | | Opitz-MID1 |
| + | | | Renal Coloboma-PAX2 |
| + | | | Rieger, Type 1-PITX2 |
| - | | | Rubinstein-Taybi-CREBBP |
| + | | | Saethre-Chotzen-TWIST |
| - | | | Septooptic Dysplasia-HESX1 |
| + | | | Simpson-Golabi-Behmel-GPC3 |
| + | | | Townes-Brockes-SALL1 |
| - | | | Treacher-Collins-TCOF1 |
| - | | | VMCM-TEK |
| + | | | Wardenburg-PAX3 |
| + | | | Zellweger-PEX1 |

***Human disease-associated genes shared with flies (F), worms (W), and Yeast (Y); from Rubin et al. (2000) Science***

# *Genomics Used to Be "Big Science"…*

wellcome trust sanger institute

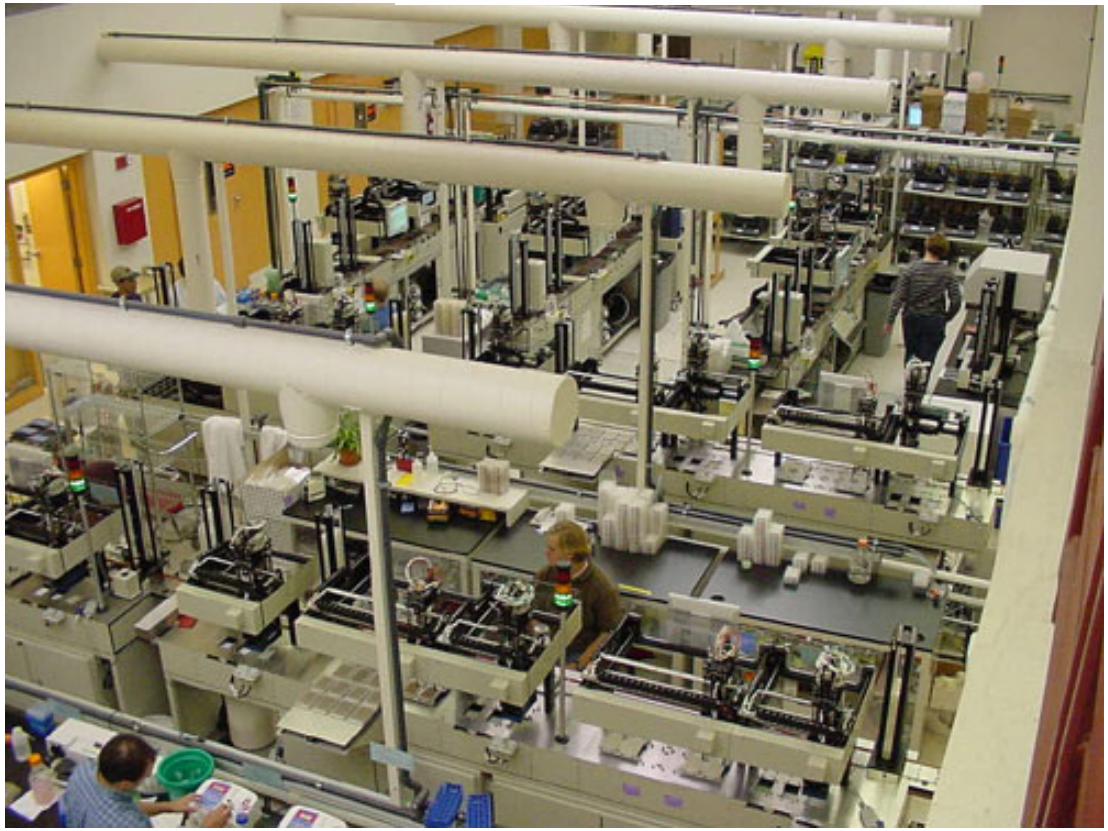
BCM
Baylor College of Medicine

BROAD INSTITUTE

JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

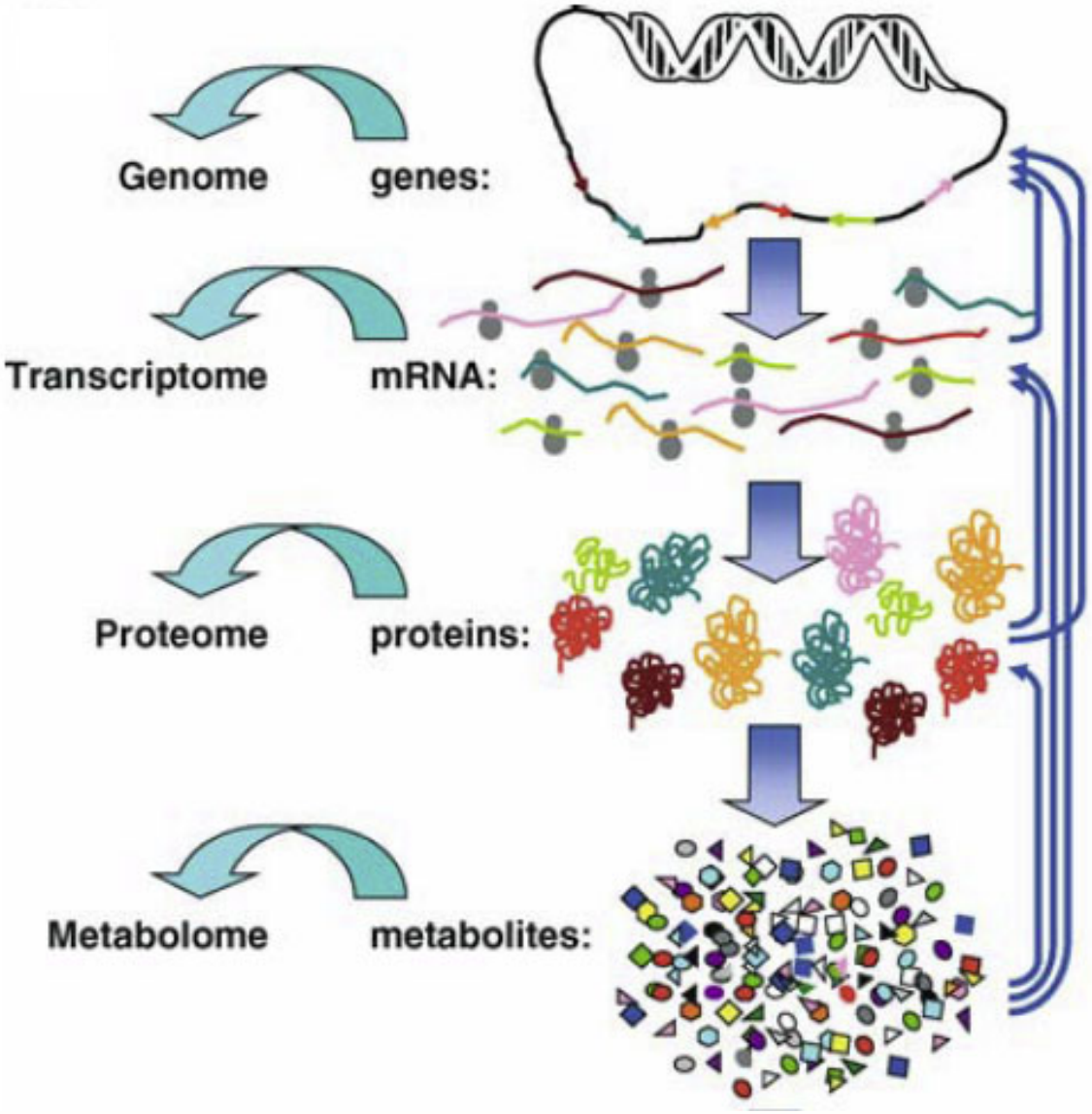
J. Craig Venter INSTITUTE

THE Genome CENTER
AT WASHINGTON UNIVERSITY



# *… But is now Accessible to Every Lab*

# *And So Are Many High Throughput Technologies*


*Goodacre (2005) Metabolomics*

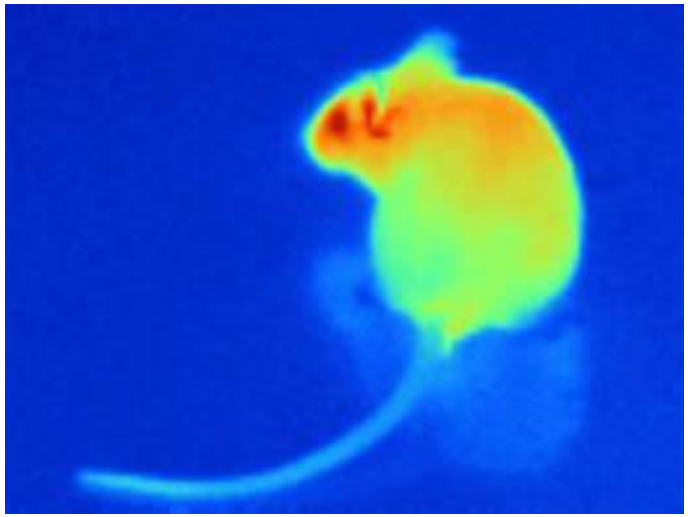
# *Novel Ways to Probe Gene Function in Any Organism*

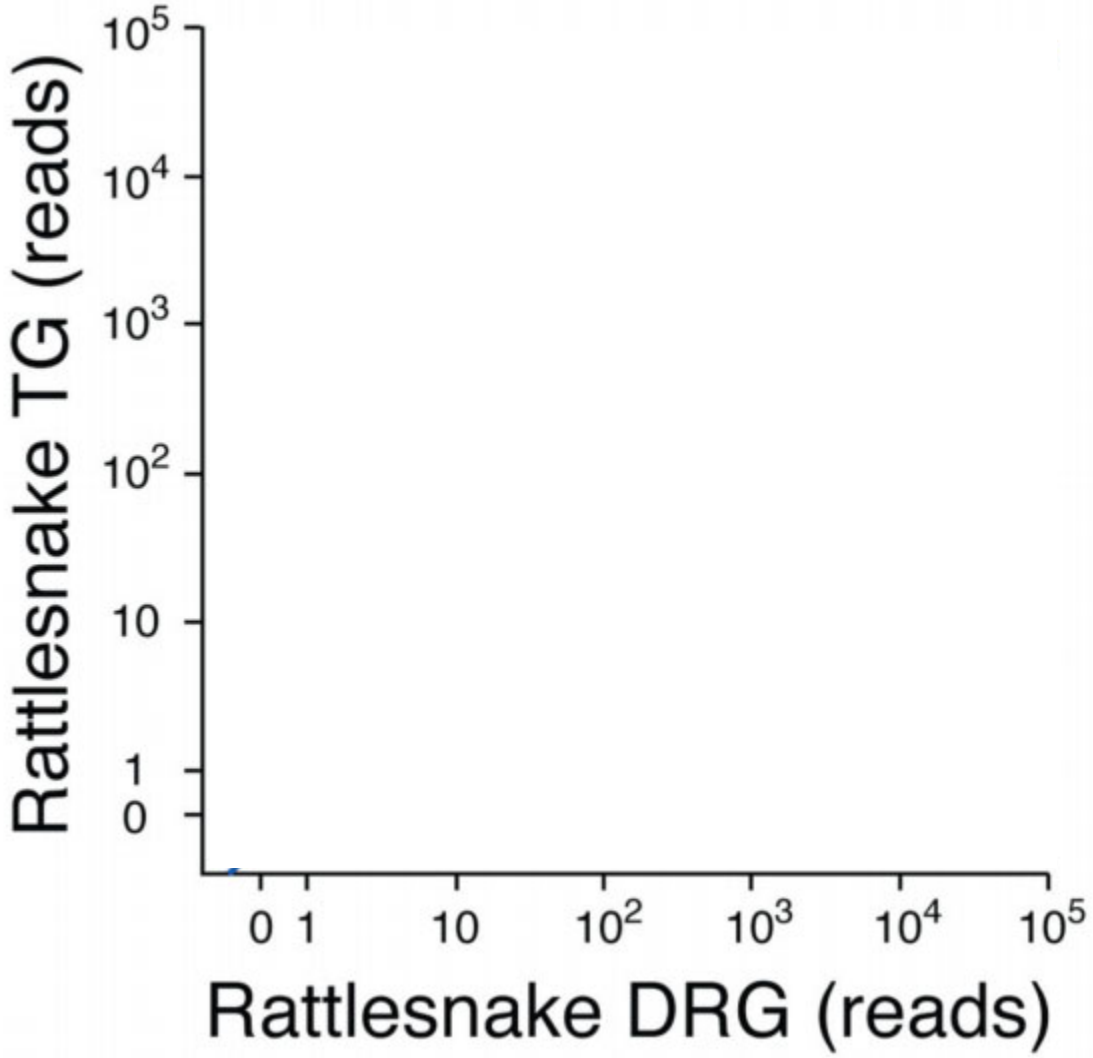
**RNAi** **TALENs / ZFNs and other nucleases / CRISPRs**

# *The Genomes of Non-Model Organisms are the New Frontiers*

*Rokas & Abbot (2009) Trends Ecol. Evol.*

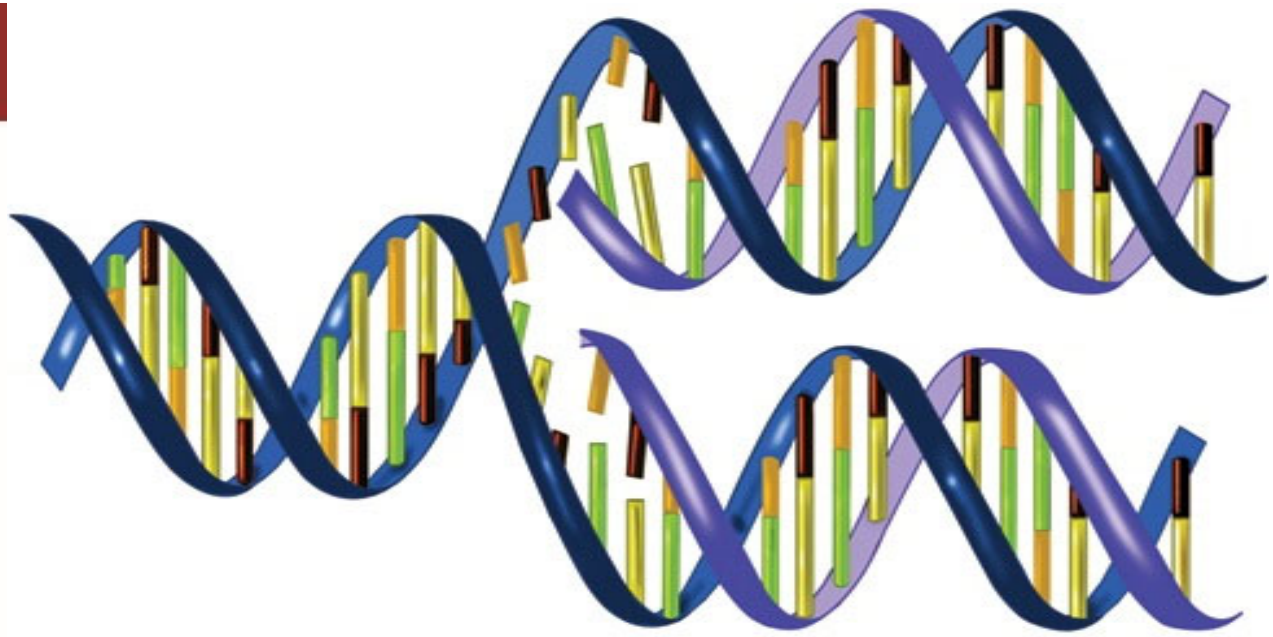
# *Snake Infrared Vision*






*Gracheva et al. (2010) Nature*

# *The DNA Record*

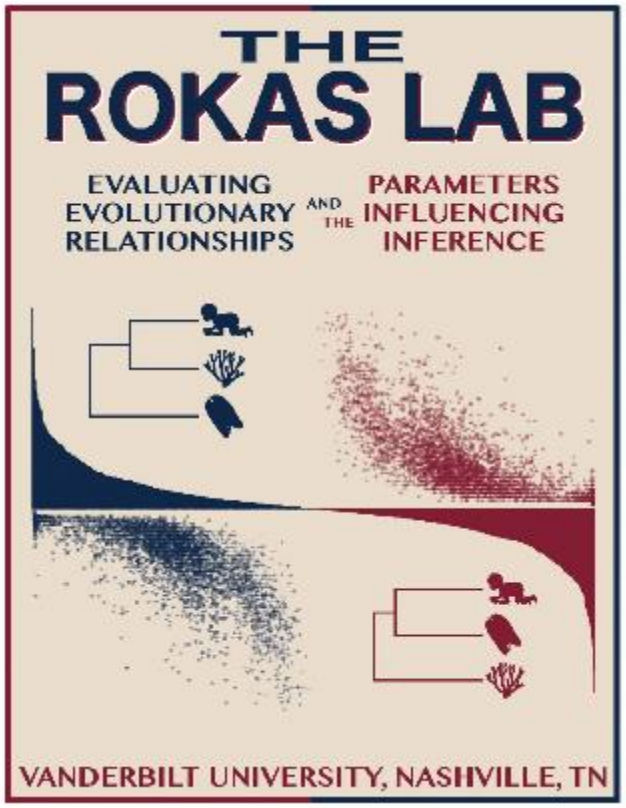


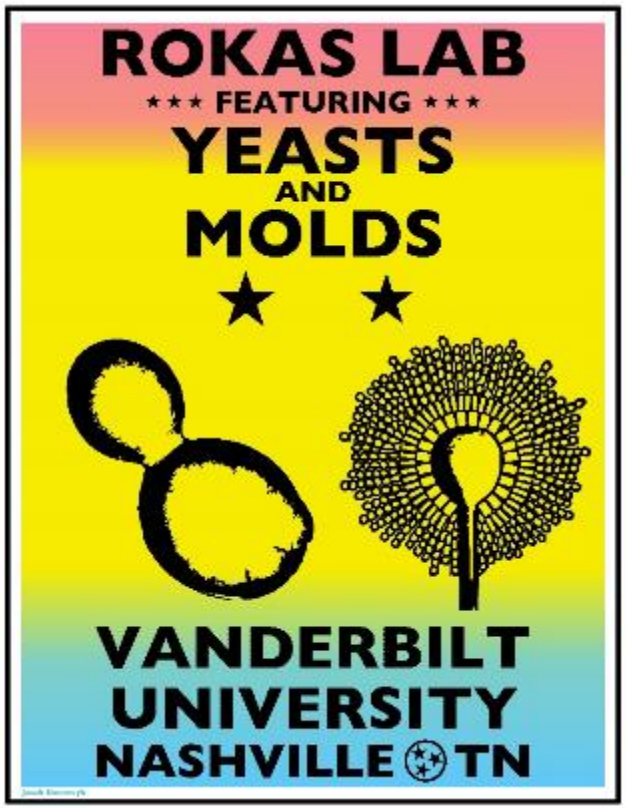
**“The genome is, it's a fossil record; the genome is a landscape; the genome is a whole geography of distributions. [...] The genome is a storybook that's been edited for a couple of billion years, and you could take it to bed, like *A Thousand and One Arabian Nights*, and read a different story, in the genome, every night.”**

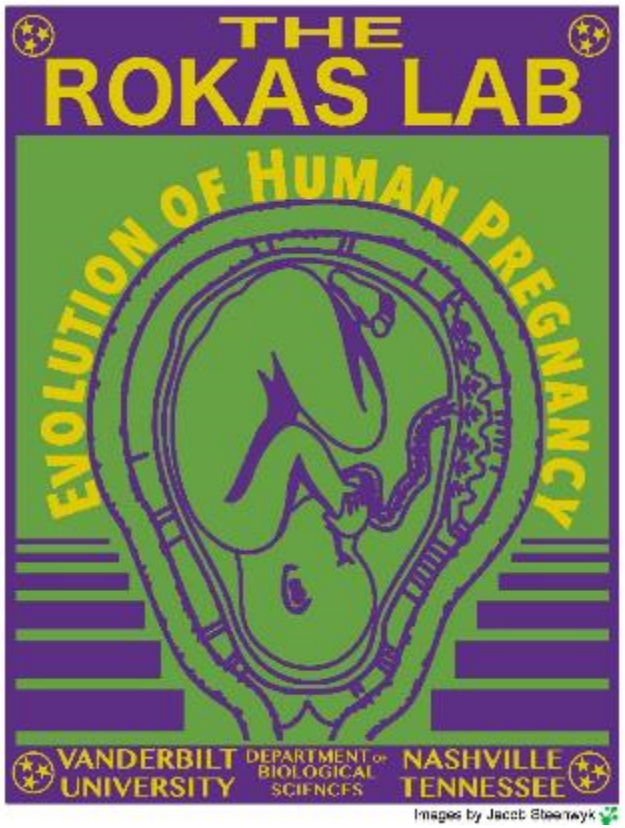
**Eric Lander**

# *The Rokas Lab*

**We study the DNA record to gain insight into evolutionary patterns and processes using computational and experimental approaches**



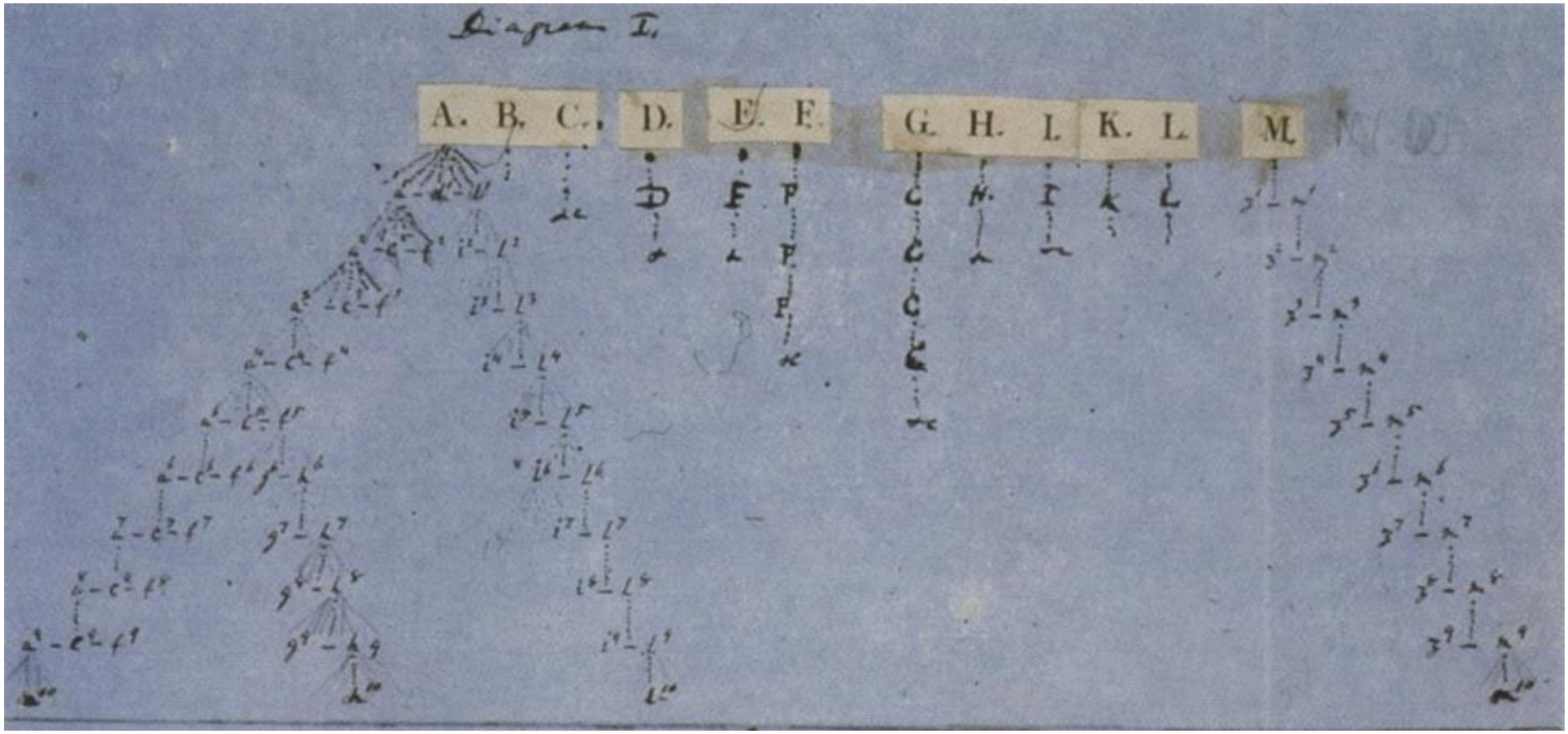




**Phylogenomics**

**The evolution of primary & secondary metabolism in fungi**

**The evolution of mammalian pregnancy**

# *Darwin's Tree*



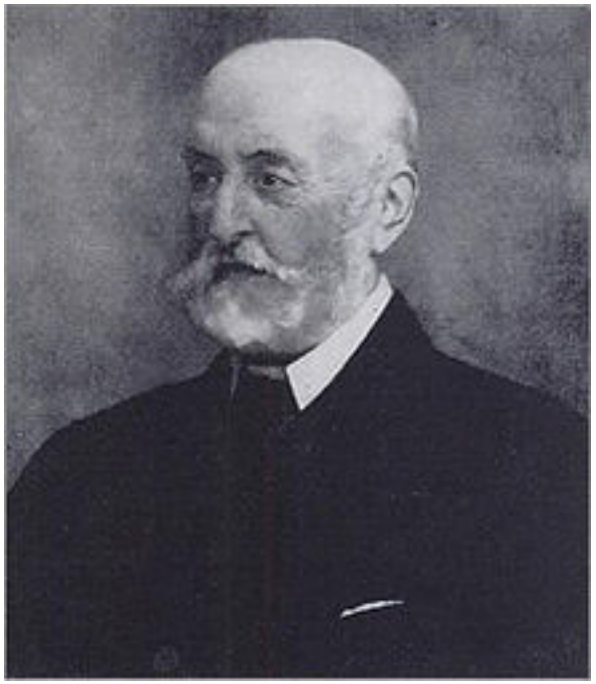

**Darwin's hand-made proof of the famous diagram from his *Origin of Species***

and instinct as the summing up of many contrivances, each useful to the possessor, nearly in the same way as when we look at any great mechanical invention as the summing up of the labour, the experience, the reason, and even the blunders of numerous workmen; when we thus view each organic being, how far more interesting, I speak from experience, will the study of natural history become!

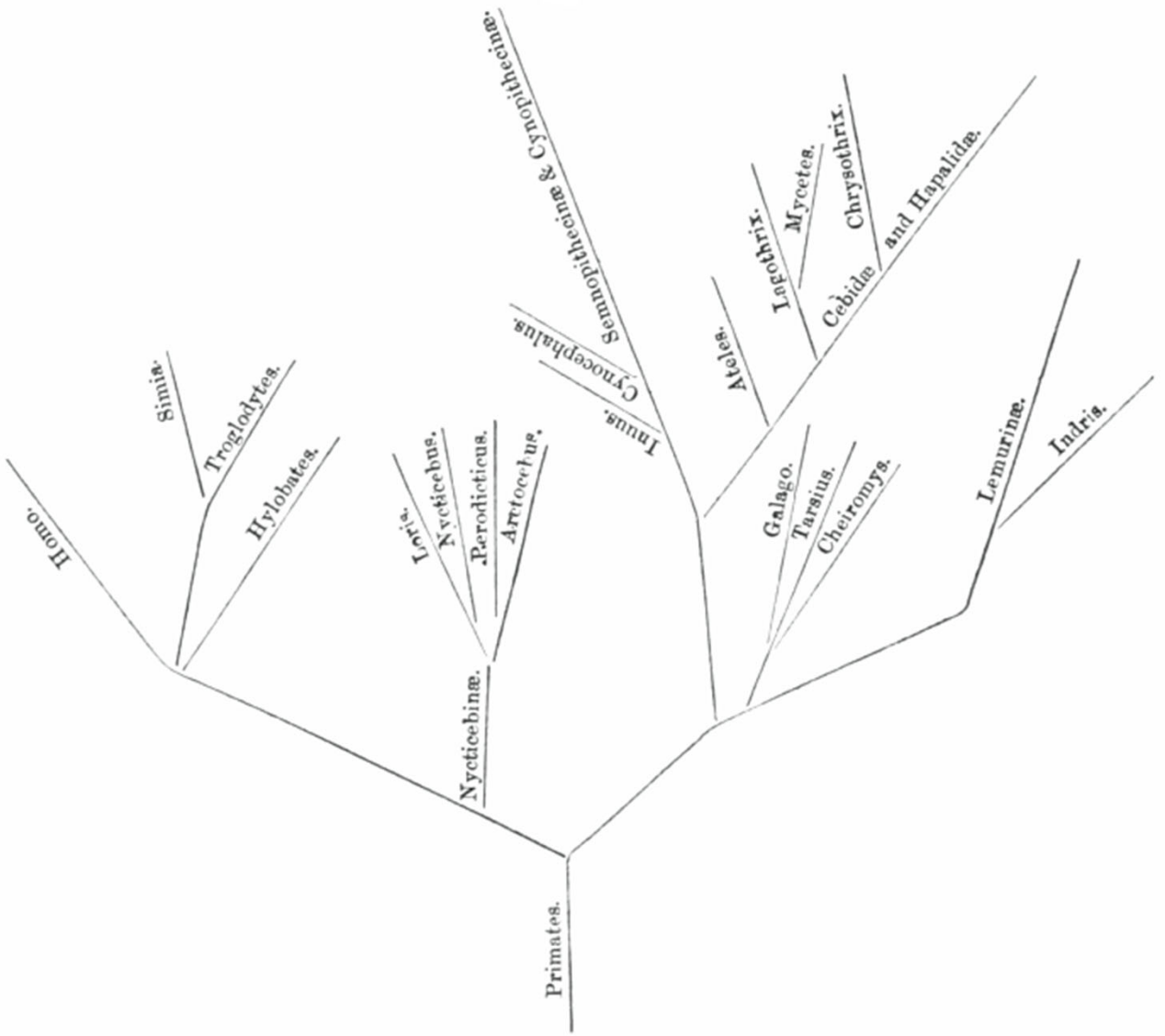
A grand and almost untrodden field of inquiry will be opened, on the causes and laws of variation, on correlation of growth, on the effects of use and disuse, on the direct action of external conditions, and so forth. The study of domestic productions will rise immensely in value. A new variety raised by man will be a far more important and interesting subject for study than one more species added to the infinitude of already recorded species. Our classifications will come to be, as far as they can be so made, genealogies; and will then truly give what may be called the plan of creation. The rules for classifying will no doubt become simpler when we have a definite object in view. We possess no pedigrees or armorial bearings; and we have to discover and trace the many diverging lines of descent in our natural genealogies, by characters of any kind which have long been inherited. Rudimentary organs will speak infallibly with respect to the nature of long-lost structures. Species and groups of species, which are called aberrant, and which may fancifully be called living fossils, will aid us in forming a picture of the ancient forms of life. Embryology will reveal to us the structure, in some degree obscured, of the prototypes of each great class.

When we can feel assured that all the individuals of the same species, and all the closely allied species of most genera, have within a not very remote period de-

# *The First Published Phylogeny*

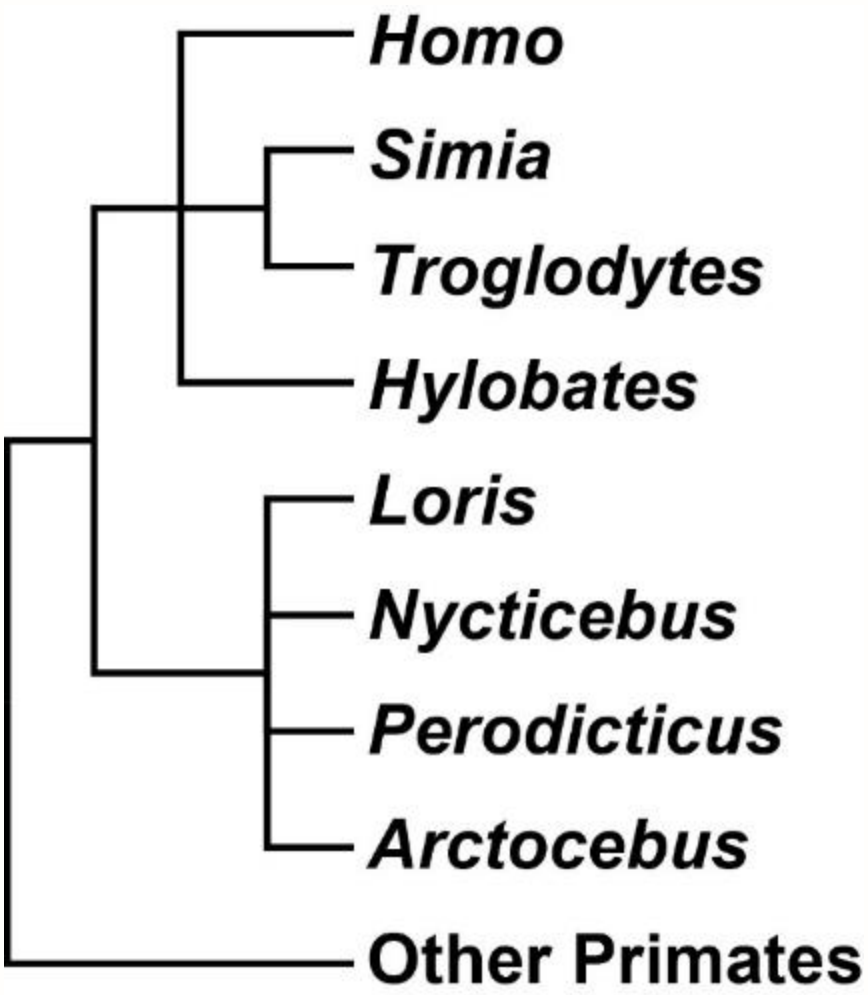


**St. George Jackson Mivart**

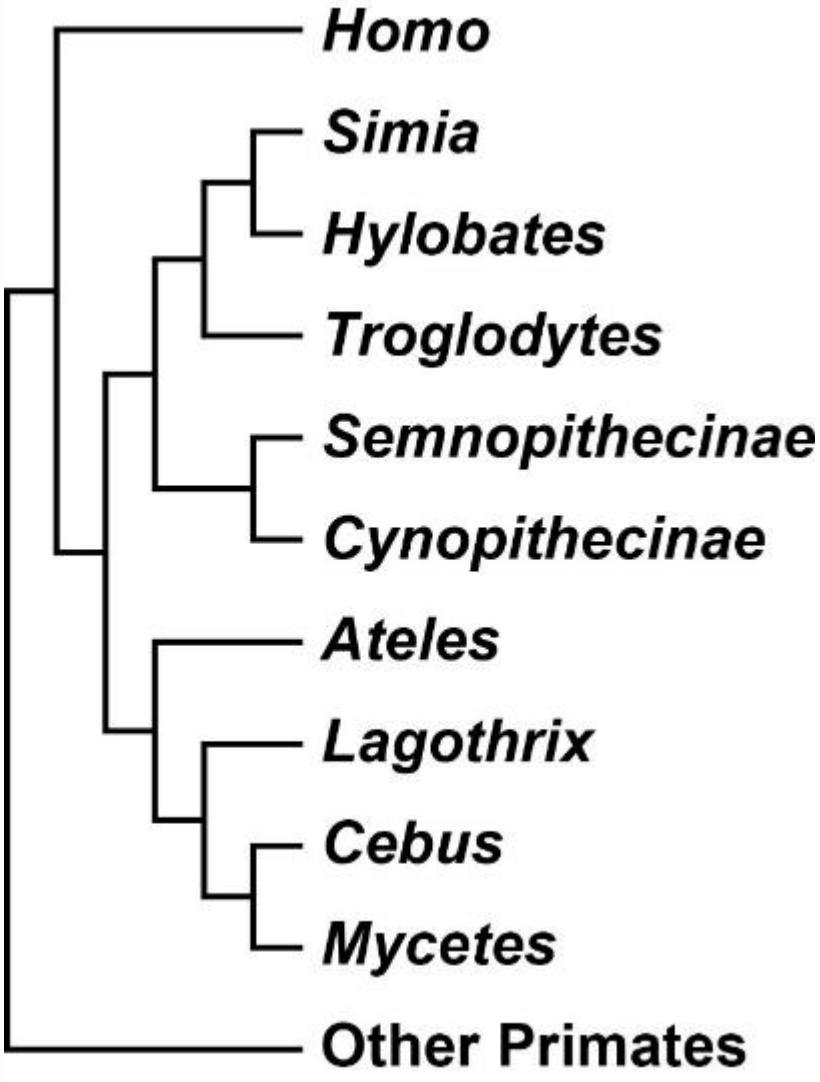


# *Discordance Between Trees There from the Beginning*

## 1865: SPINAL COLUMN




## 1867: LIMBS

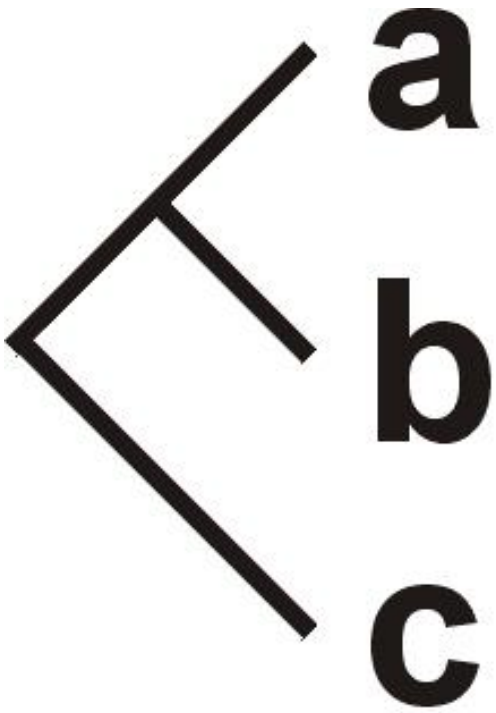



*http://phylonetworks.blogspot.se/2012/09/the-first-network-from-conflicting.html*

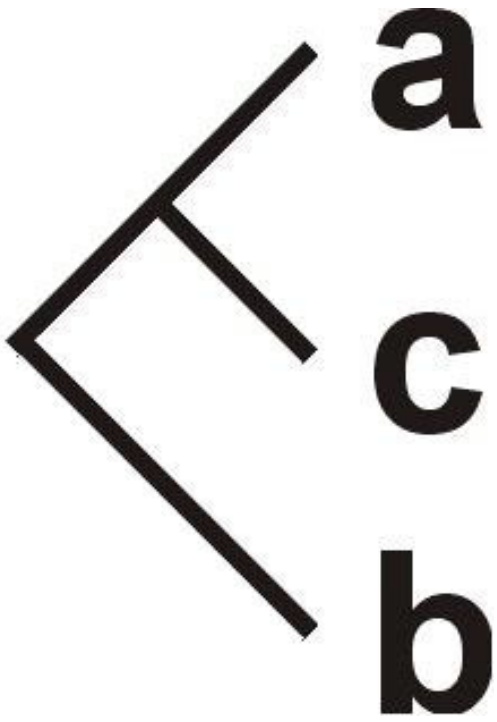
In some M.S. [… I say] that on genealogical principles alone, & considering whole organisation man probably diverged from the Catarhine stem a little below the branch of the anthropo:apes […]. I have then added in my M.S. that this is your opinion […]. Is this your opinion?

I have really expressed no opinion as to Man's origin nor am I prepared to do so at this moment. The [1865] diagram […] expresses what I believe to be the degree of resemblance as regards the spinal column *only*. The [1867] diagram expresses what I believe to be the degree of resemblance as regards the appendicular skeleton *only*

# *The Problem of Incongruence*



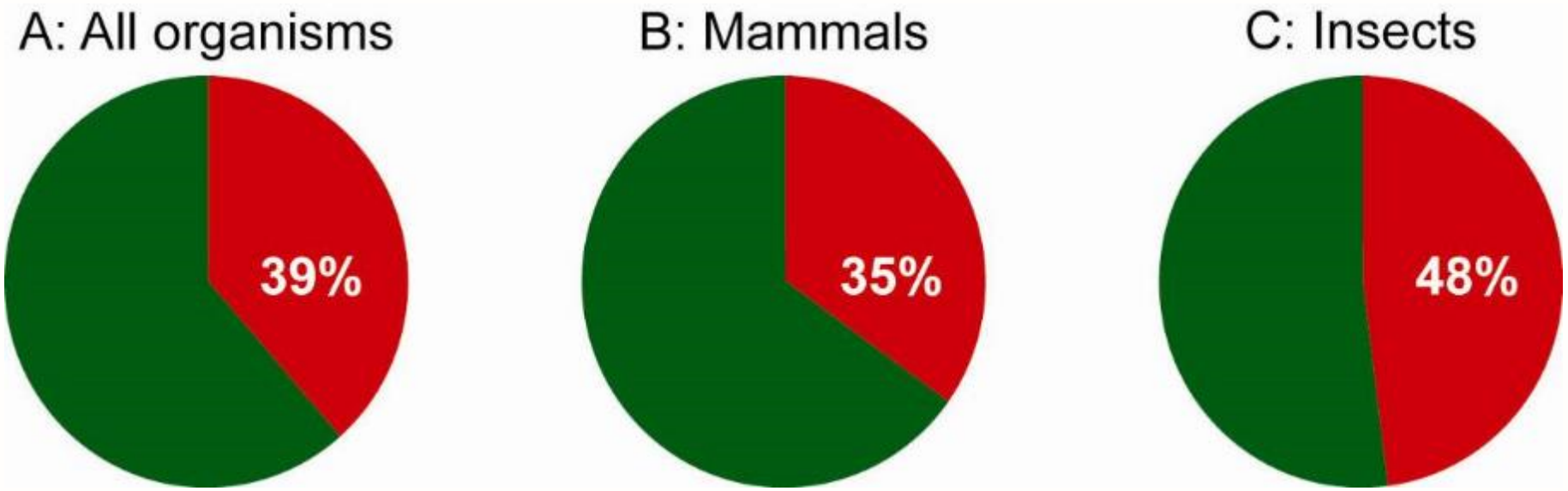

Gene X




Gene Y

**Species phylogeny?**

# *Incongruence is Pervasive in the Phylogenetics Literature*




*Rokas & Chatzimanolis (2008) in Phylogenomics (W. J. Murphy, Ed.)*

# *A Systematic Evaluation of Single Gene Phylogenies*

***S. cerevisiae***

***S. paradoxus***

***S. mikatae***

***S. kudriavzevii***

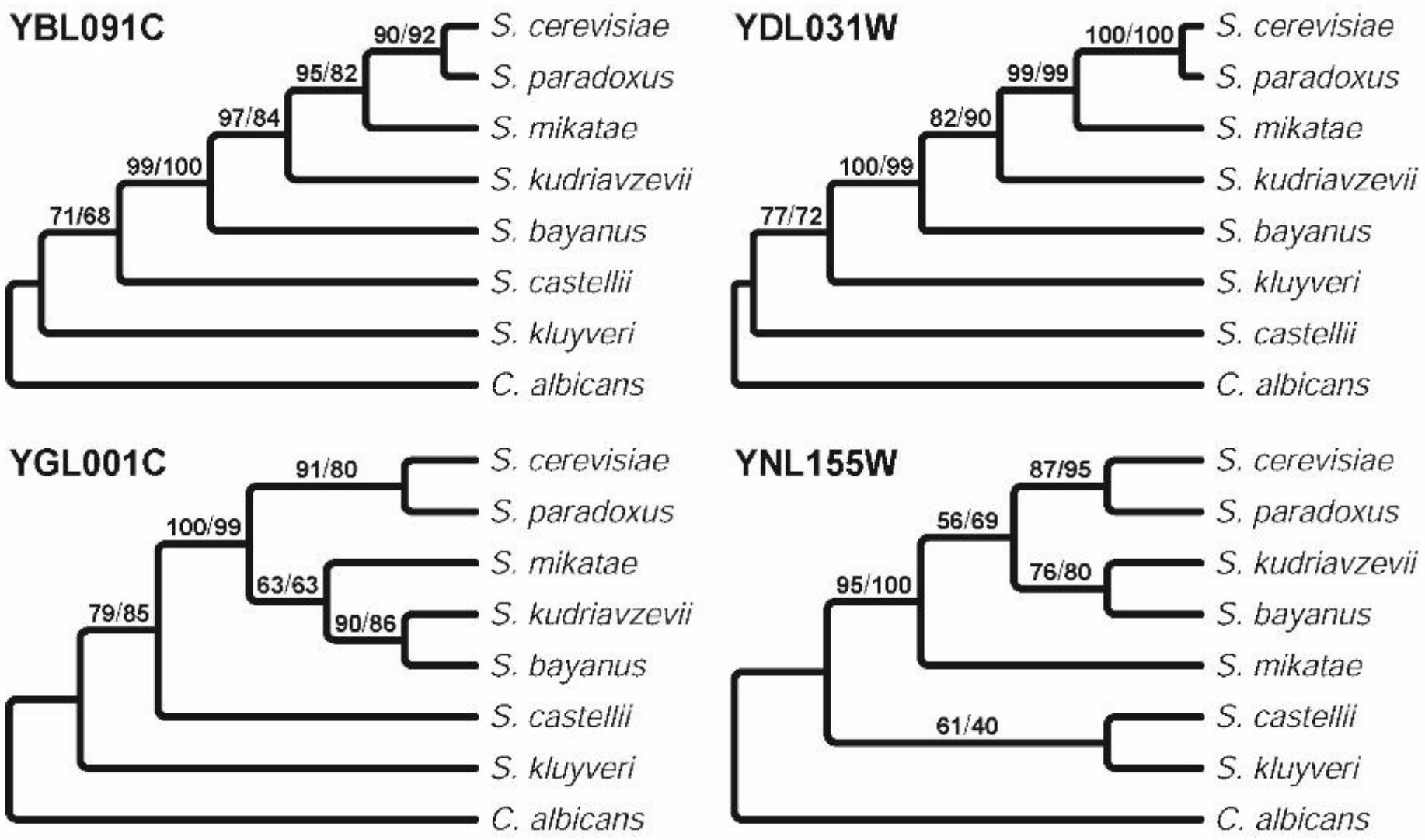
***S. bayanus***

***S. castellii***

***S. kluyveri***

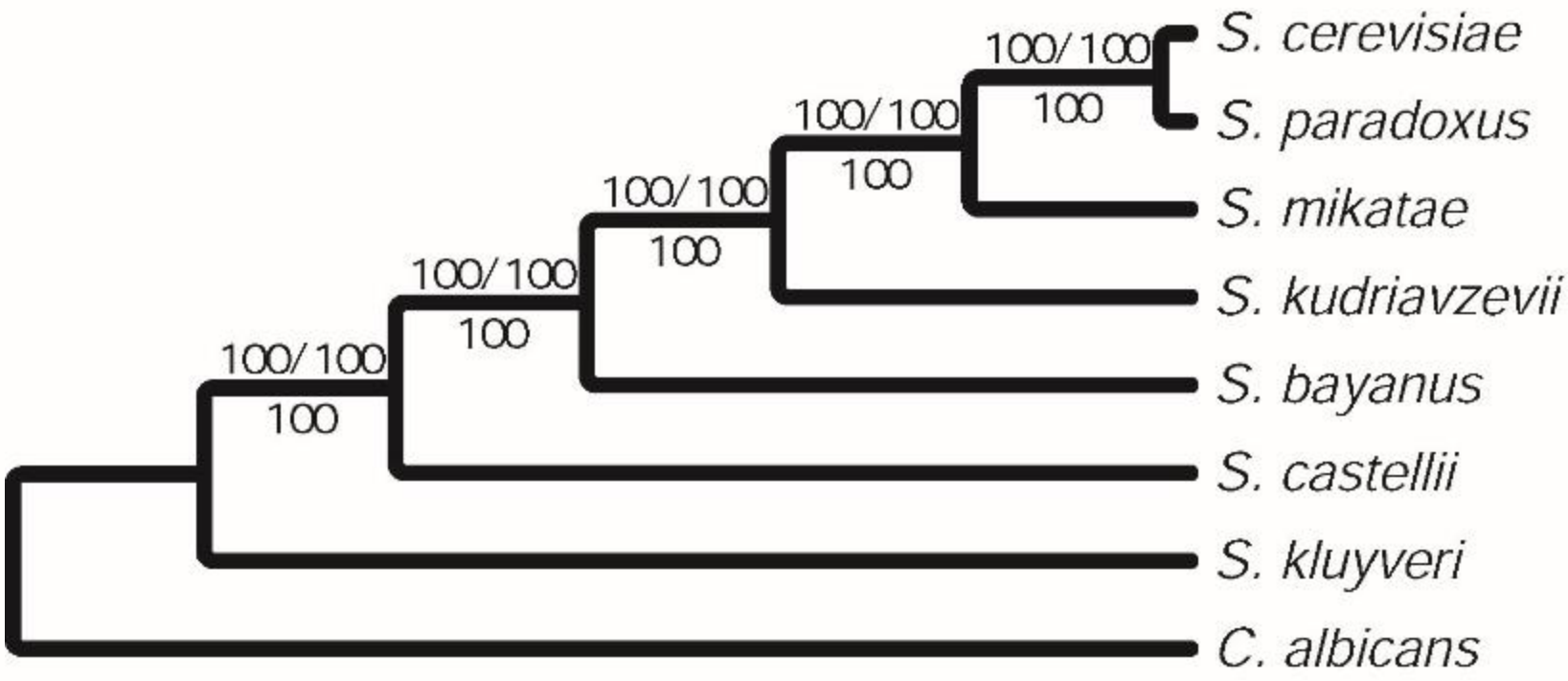
***Candida glabrata***

# *Incongruence at the Single Gene Level*




**ML / MP** *Anonymous Reviewer for Nature (2003)* *Rokas et al. (2003) Nature*

# *Concatenation of 106 Genes Yields a Single Yeast Phylogeny*




**ML / MP on nt**
**MP on aa**

*Rokas et al. (2003) Nature*

# *The Phylogenomics Era – "Resolving" the Tree of Life*

*Syst. Biol.* 61(1):150–164, 2012

DOI:10.1093/sysbio/syr089
Advance Access publication on September 7, 2011

LETT
LETT

## Phylogenomic Analysis Resolves the Interordinal Relationships and Rapid Diversification of the Laurasiatherian Mammals

XUMING ZHOU, SHIXIA XU, JUNXIAO XU, BINGYAO CHEN, KAIYA ZHOU, AND GUANG YANG*

*Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China; *Correspondence to be sent to: Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China; E-mail: gyang@njnu.edu.cn.*

## Resolving the evolutionary relationships of molluscs with phylogenomic tools

Stephen A. Smith[1,2], Nerida G. Wilson[3,4], Freya
Gonzalo Giribet[5] & Casey W. Dunn[1]

nature

LETTERS

*Syst. Biol.* 57(6):920–938, 2008

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150802570791

## Toward Resolving t
## Tree: The Phylogen
## of Jakobids and Cercozoans

## Resolving Arthropod Phylogeny: Exploring Phylogenetic Signal within 41 kb of Protein-Coding Nuclear Gene Sequence

JEROME C. REGIER,[1] JEFFREY W. SHULTZ,[2] AUSTEN R. D. GANLEY,[3,6] APRIL HUSSEY,[1] DIANE SHI,[1] BERNARD BALL,[3] ANDREAS ZWICK,[1] JASON E. STAJICH,[3,7] MICHAEL P. CUMMINGS,[4] JOEL W. MARTIN,[5] AND CLIFFORD W. CUNNINGHAM[3]

Yeast

An

OPEN ACCESS Free

## Toward Resolving Priors

## Prion-Like Proteins in the Fungal Kingdom

Edgar M. Medina · Gary W. Jones · David A. Fitzpatrick

## Towards

*Renae C. Pratt,* Gillian C. Gibb,* Mary Morgan-Richards,* Matthew J. Phillips,† Michael D. Hendy,* and David Penny**

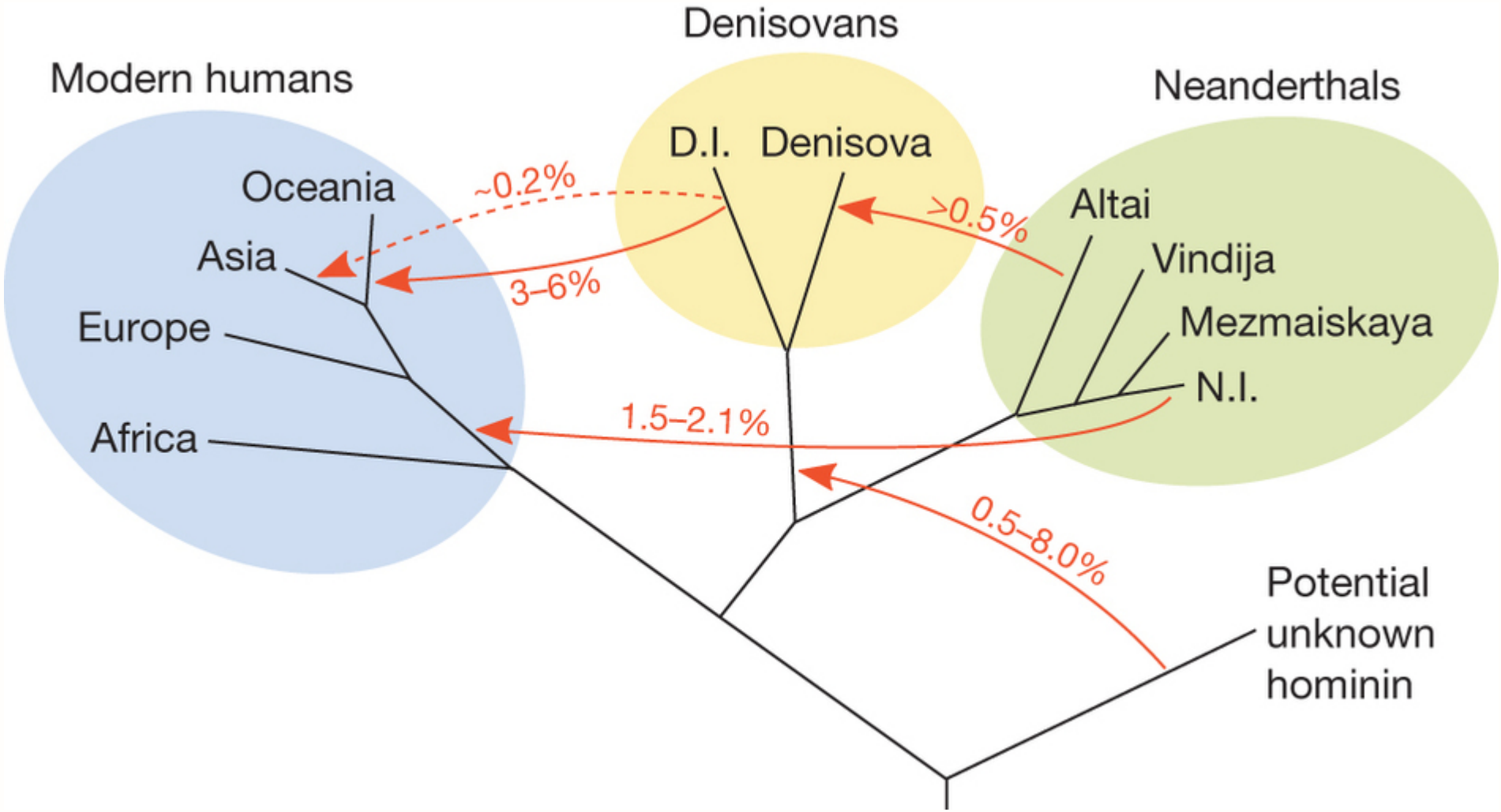
Samuli Lehtonen

Department of Biology, U

*Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand; and †Centre for Macroevolution and Macroecology, School of Botany and Zoology, Australian National University, Canberra ACT, Australia

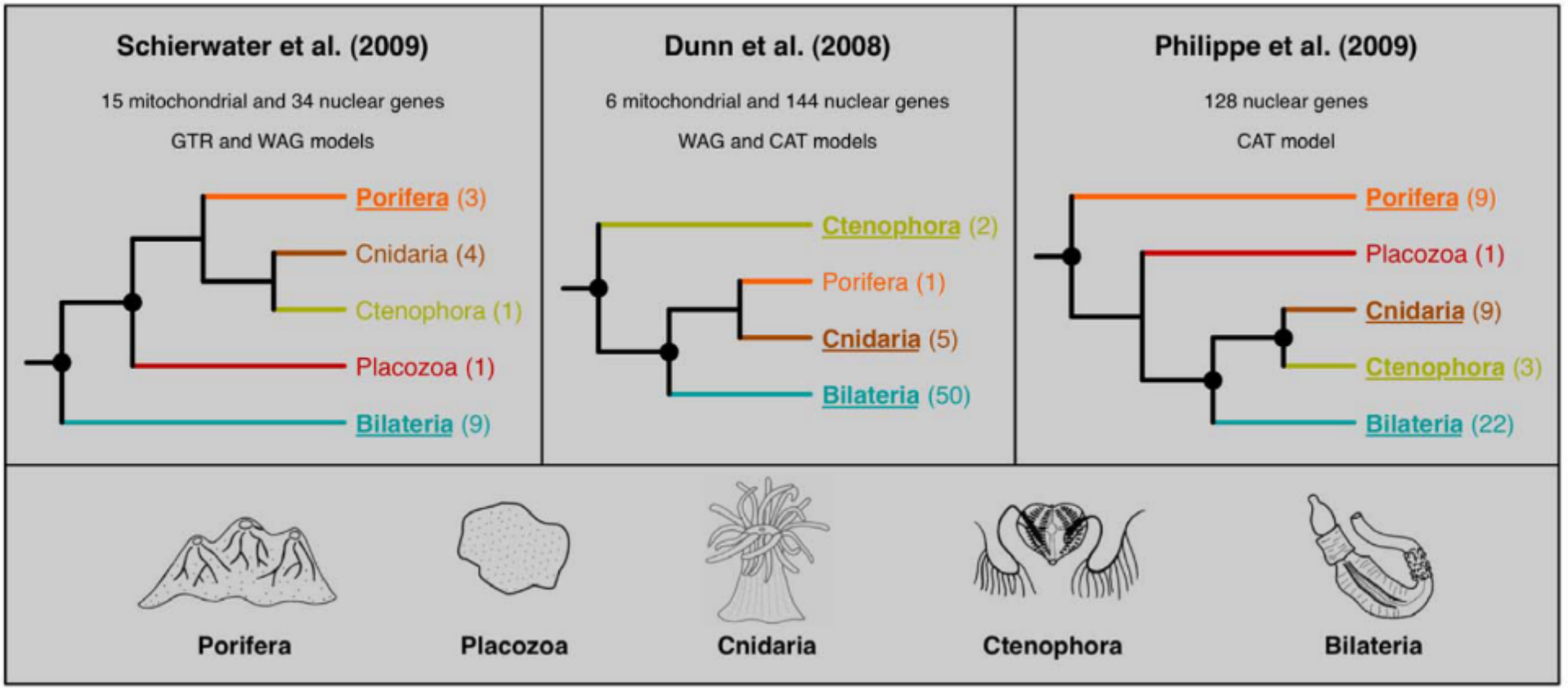
# Have we eliminated incongruence?

# *At "Shallow" Depths, True History is Easier Seen & Quantified*




*Prufer et al. (2014) Nature*

# *Incongruence in Deep Time is More Challenging*



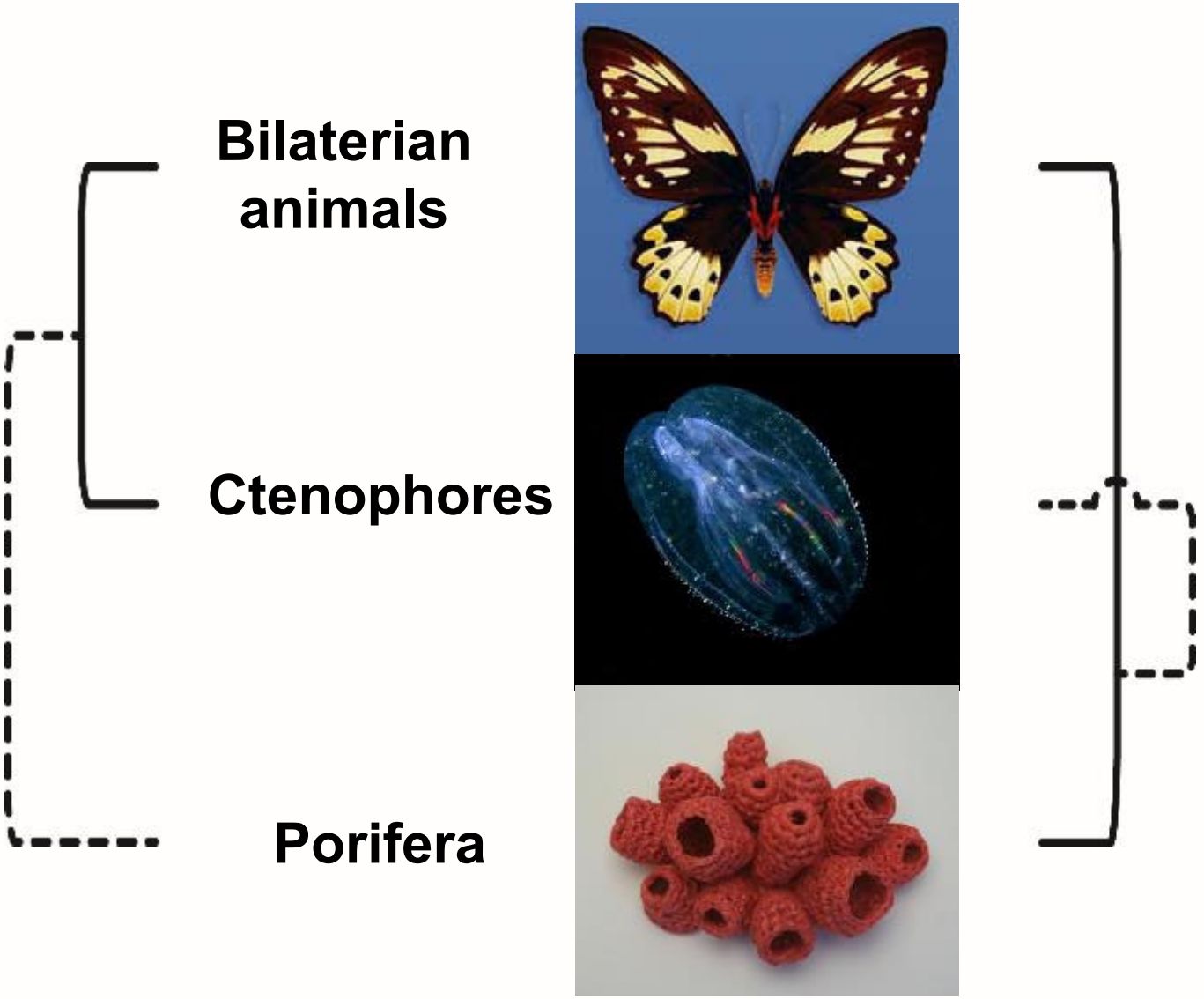

*Philippe et al. (2011) PLoS Biol.*

# *Incongruence in Deep Time is More Challenging*


*Kocot et al. (2011) Nature*

*Smith et al. (2011) Nature*

# *Incongruence in Deep Time is More Challenging*




*Pisani et al. (2015) PNAS*

*Chang et al. (2015) PNAS*

# Why the disconnect?

# *An Expanded Yeast Data Matrix*

## Yeast Gene Order Browser (YGOB)

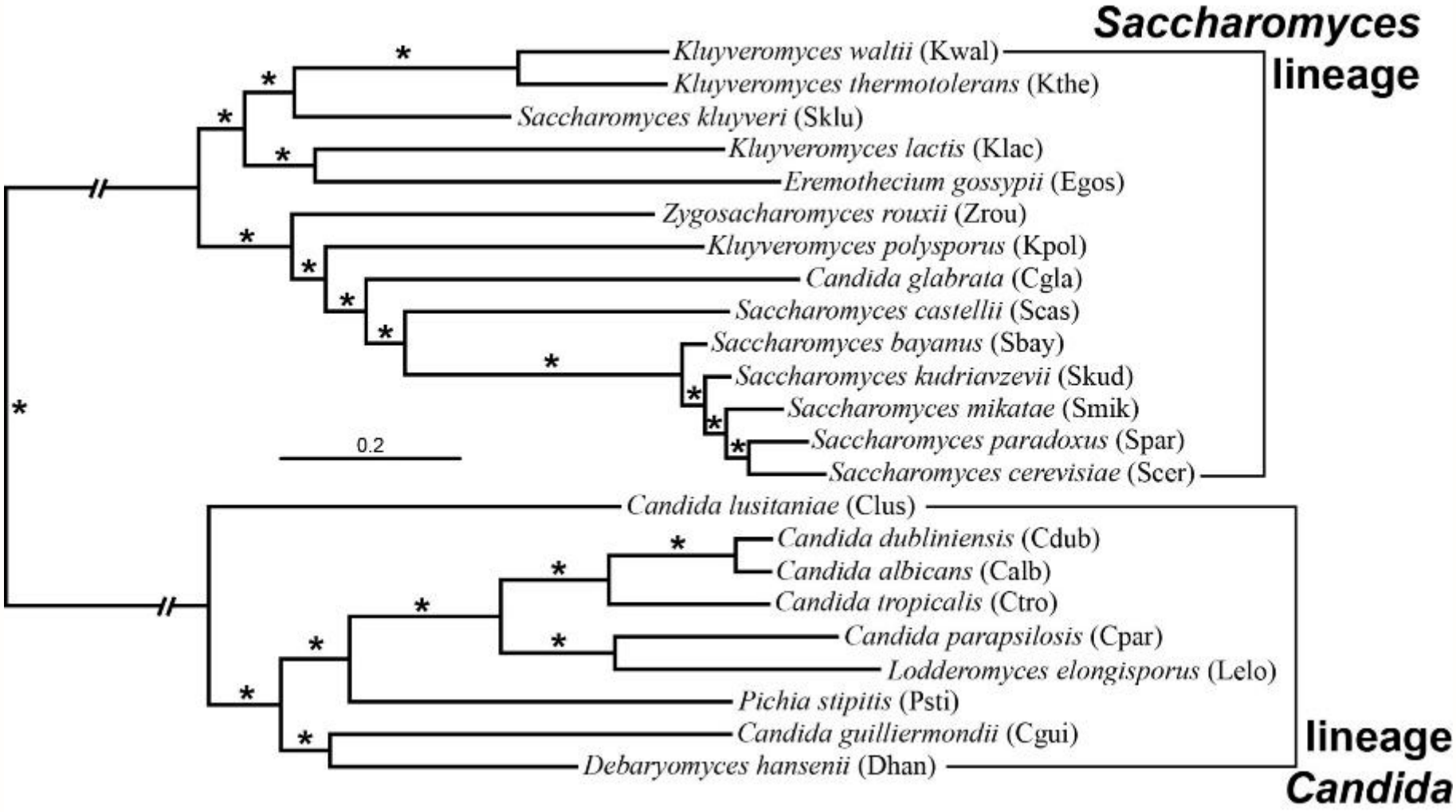
## *Candida* Gene Order Browser (CGOB)

***Saccharomyces* lineage**

***Candida* lineage**

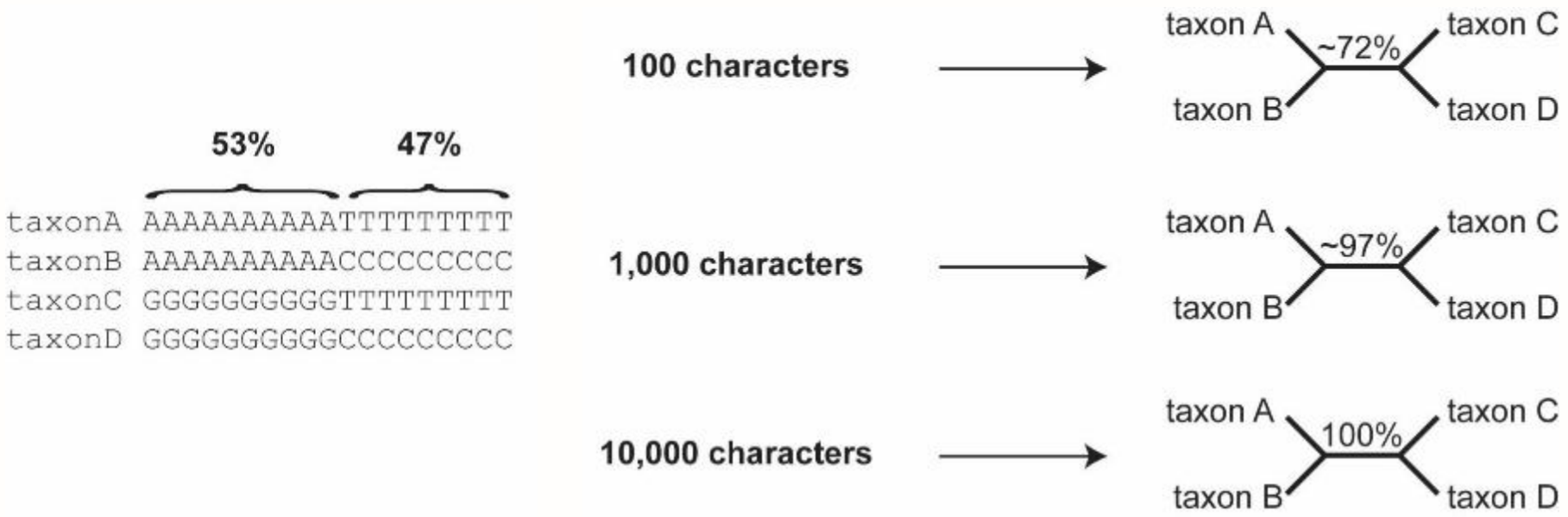
**1,070 genes**
**23 taxa**
**no missing data**

# *Concatenation Yields an Absolutely Supported Phylogeny*



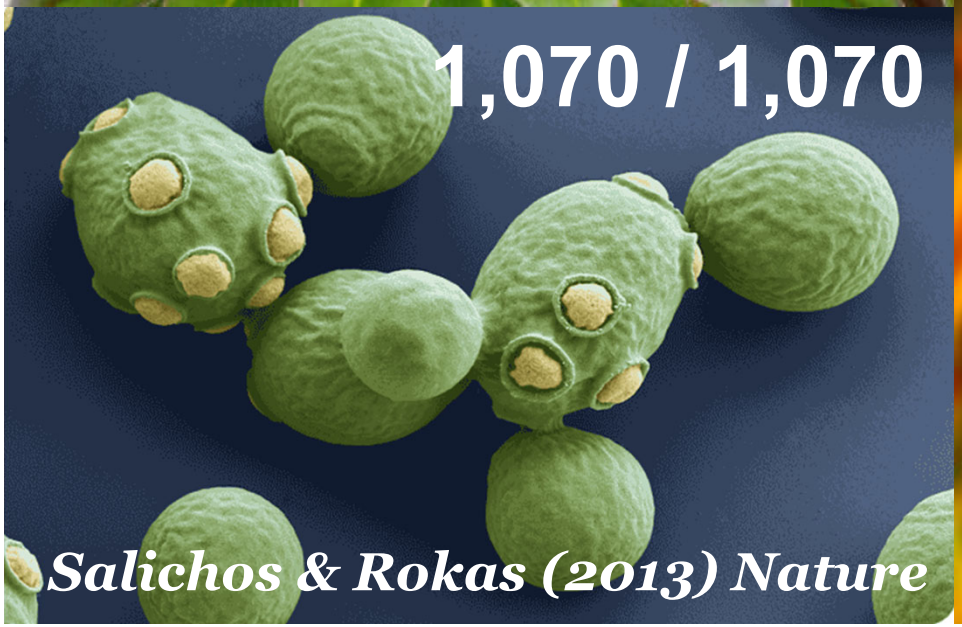

*Salichos & Rokas (2013) Nature*

# *Bootstrap Support is Misleading When Used in Large Datasets*

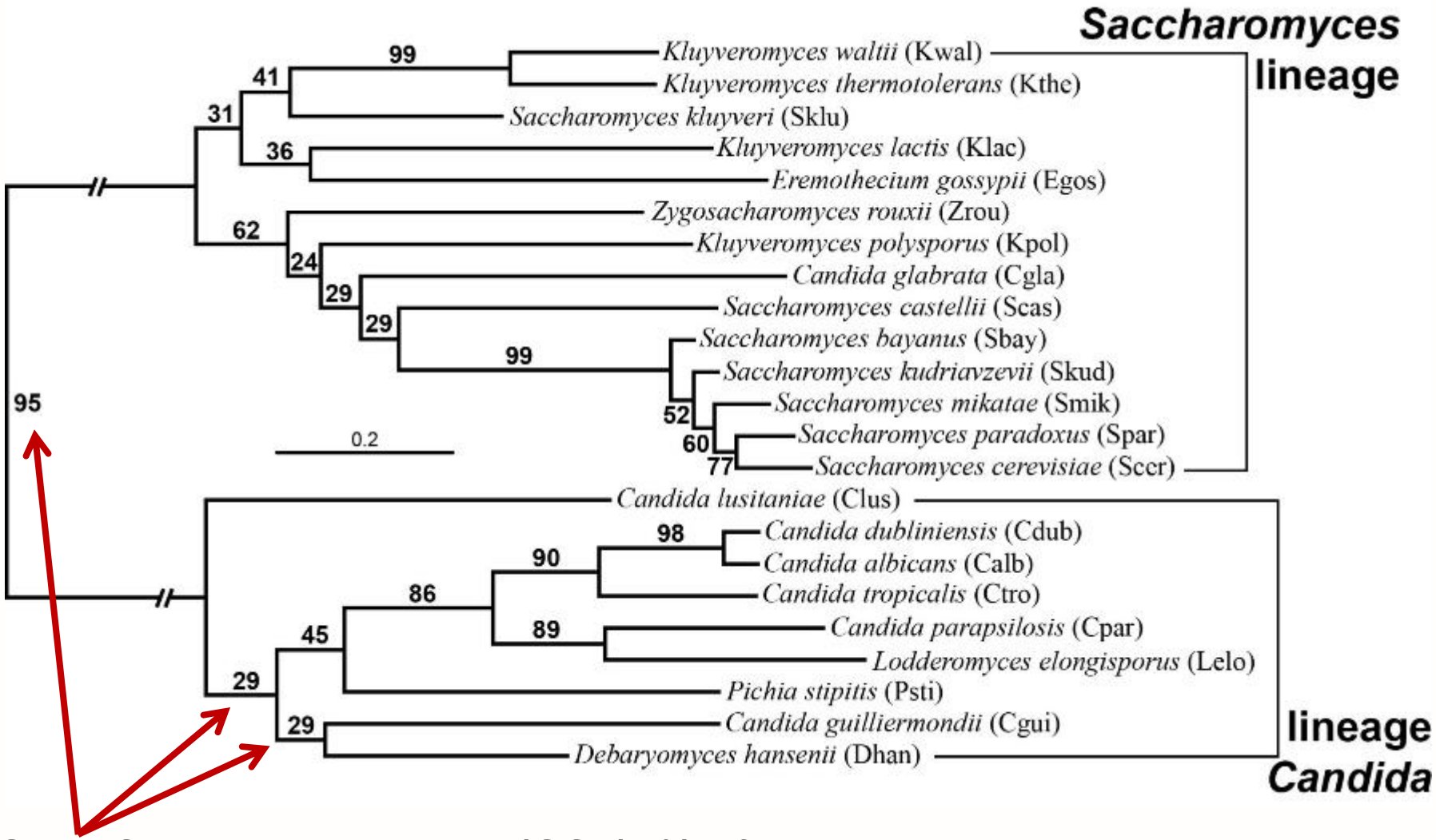


# *Gene Trees are Incongruent in Most Datasets*






# *The Yeast Phylogeny Inferred by Majority-Rule Consensus*



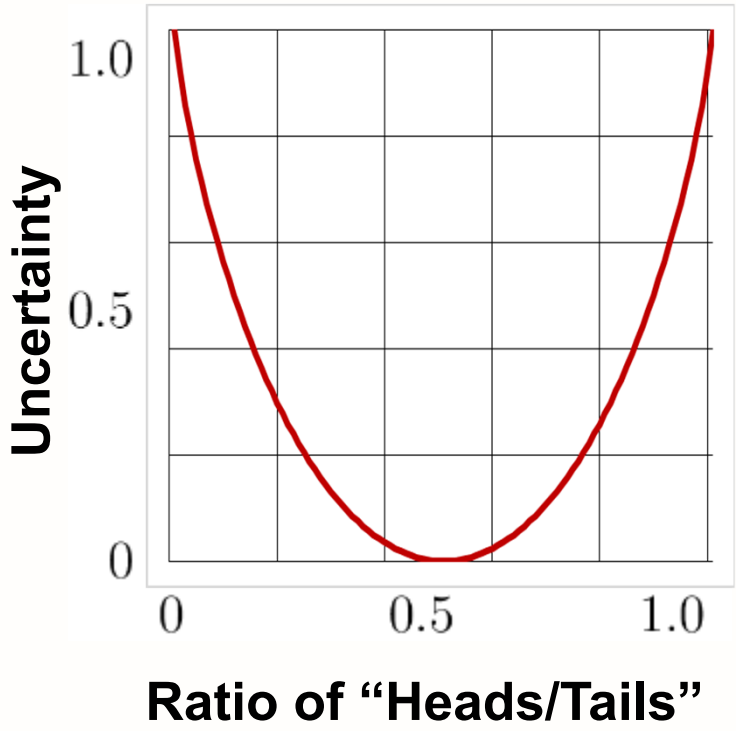

**Gene Support Frequency (GSF): % of single gene trees supporting a given internode**

# *Quantifying Incongruence*

**Internode Certainty (IC)**: a measure of the support for a given internode by considering its frequency in a given set of trees jointly with that of the most prevalent conflicting internode in the same set of trees

**Tree Certainty (TC)**: the sum of IC across all internodes

**IC and TC are implemented in the latest versions of RAxML**

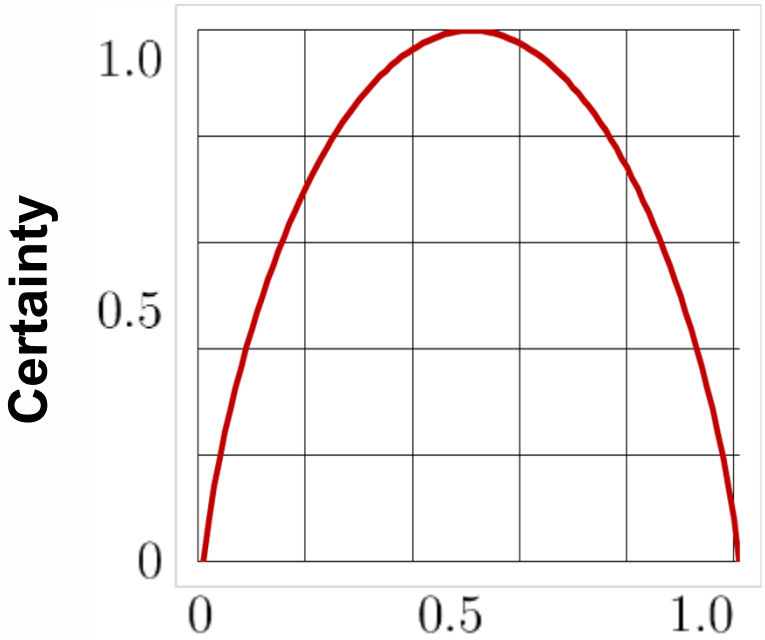


# *Quantifying Incongruence*

**Internode Certainty (IC)**: a measure of the support for a given internode by considering its frequency in a given set of trees jointly with that of the most prevalent conflicting internode in the same set of trees

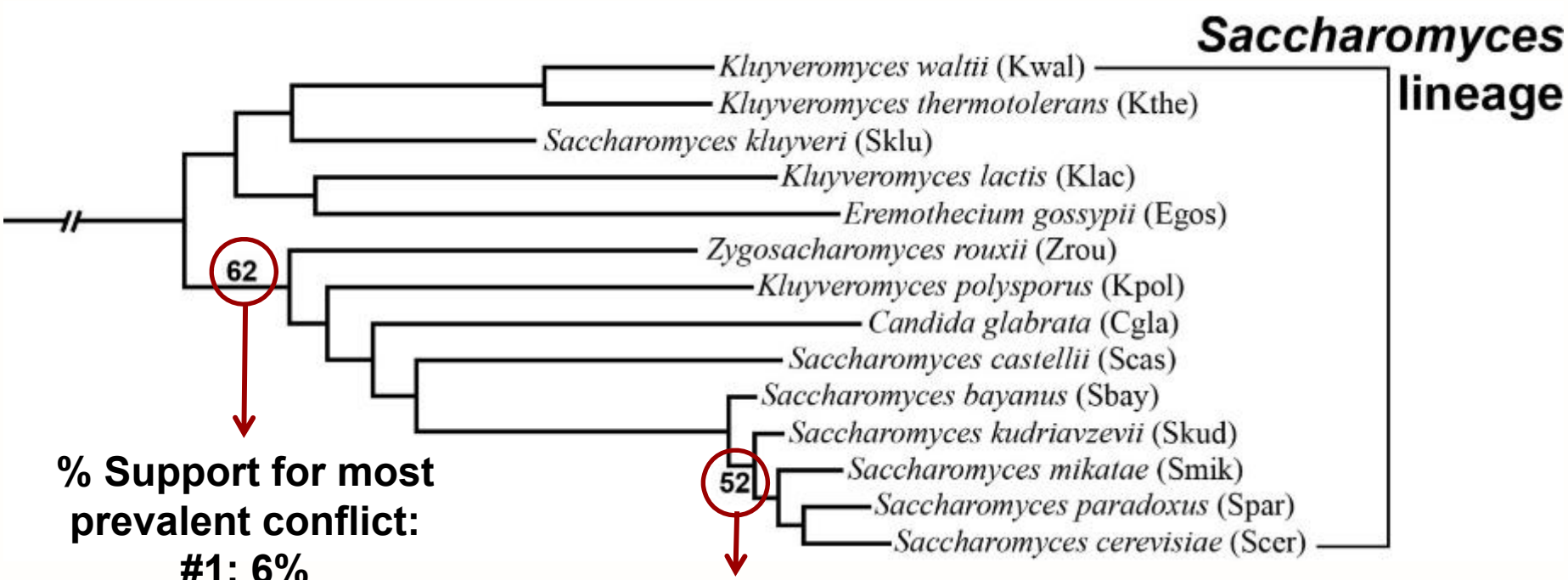
**Tree Certainty (TC)**: the sum of IC across all internodes

**IC and TC are implemented in the latest versions of RAxML**




**Ratio of Support for Two Conflicting Internodes**

# *A More Informative Measure of Branch Support*




**% Support for most prevalent conflict: #1: 6%**

**IC value: 0.59**

**% Support for most prevalent conflict: #1: 29%**

**IC value: 0.06**

# *Similar Results in Other Lineages*

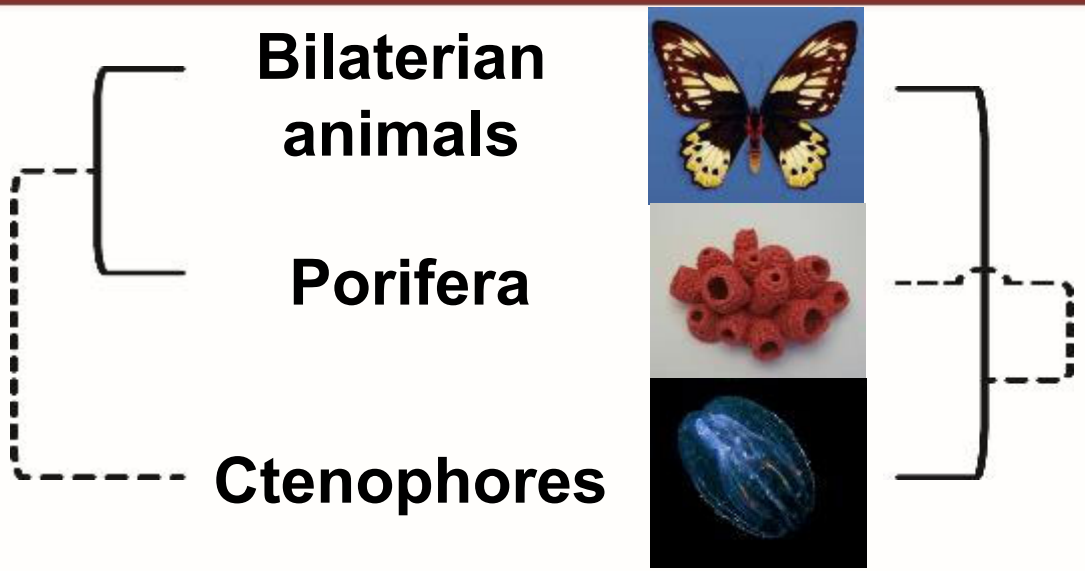
**Vertebrates**
**(1,086 genes, 18 taxa)**

**Animals**
**(225 genes, 21 taxa)**

**Mosquitoes**
**(2,007 genes, 20 taxa)**

*Salichos & Rokas (2013) Nature; Wang et al. (2015) Genome Biol. Evol.*

# *Incongruence in Phylogenomic Datasets*




**These debates often concern relationships poorly supported by individual gene trees**

# What is the phylogenetic signal in branches of the tree of life that are challenging to resolve?

# *Definitions of Phylogenetic Signal*

**A measure of the statistical dependence among species' trait values due to their phylogenetic relationships / the tendency of related species to resemble each other more than species drawn at random from the same tree**

**Revell et al. (2008) Syst. Biol.**
**Münkemüller et al. (2012) Methods Ecol. Evol.**

**The amount of support for a particular topology, e.g., the relative number of resolved internodes in a consensus tree**

**Sanderson (2008) Science**

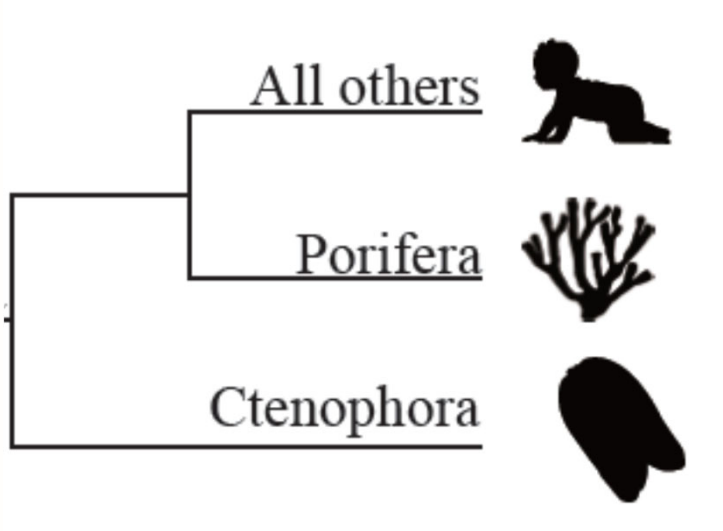
**A measure of the substitutions occurring along a given branch of the evolutionary tree. In parsimony methods, the signal is encoded in shared derived characters. In probabilistic methods, the amount of phylogenetic signal actually extracted from a given dataset depends on the model and is expected to increase with the fit of the model to the data**

**Philippe et al. (2011) PLoS Biol.**
**Townsend et al. (2012) Syst. Biol.**

# *Our Definition*

**Maximum Likelihood tree**

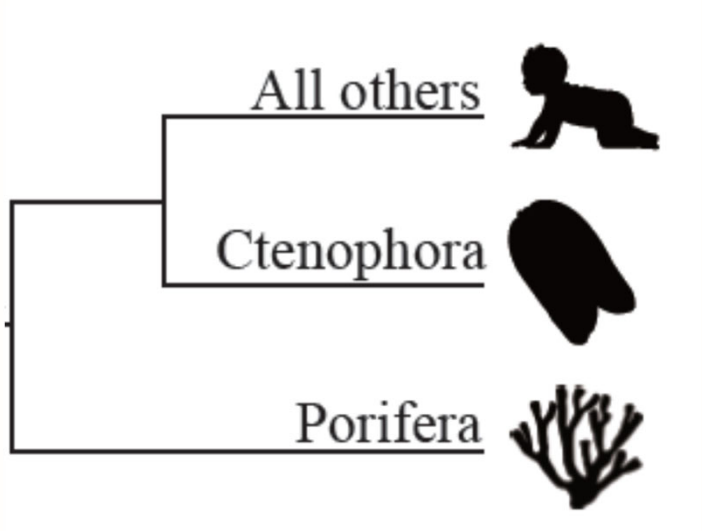
**(T1)**




$$\ln(T_1|X_i) = -100$$

**Conflicting tree**

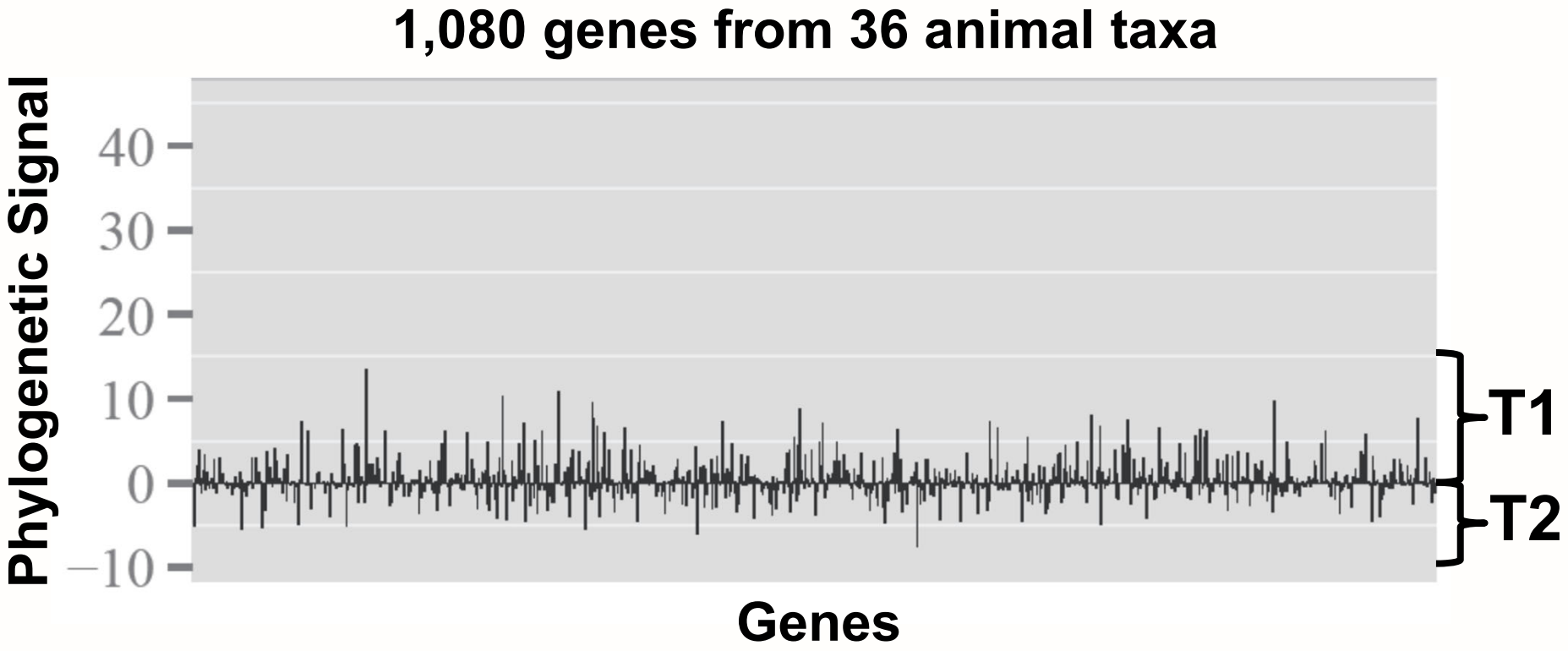
**(T2)**




$$\ln(T_2|X_i) = -150$$

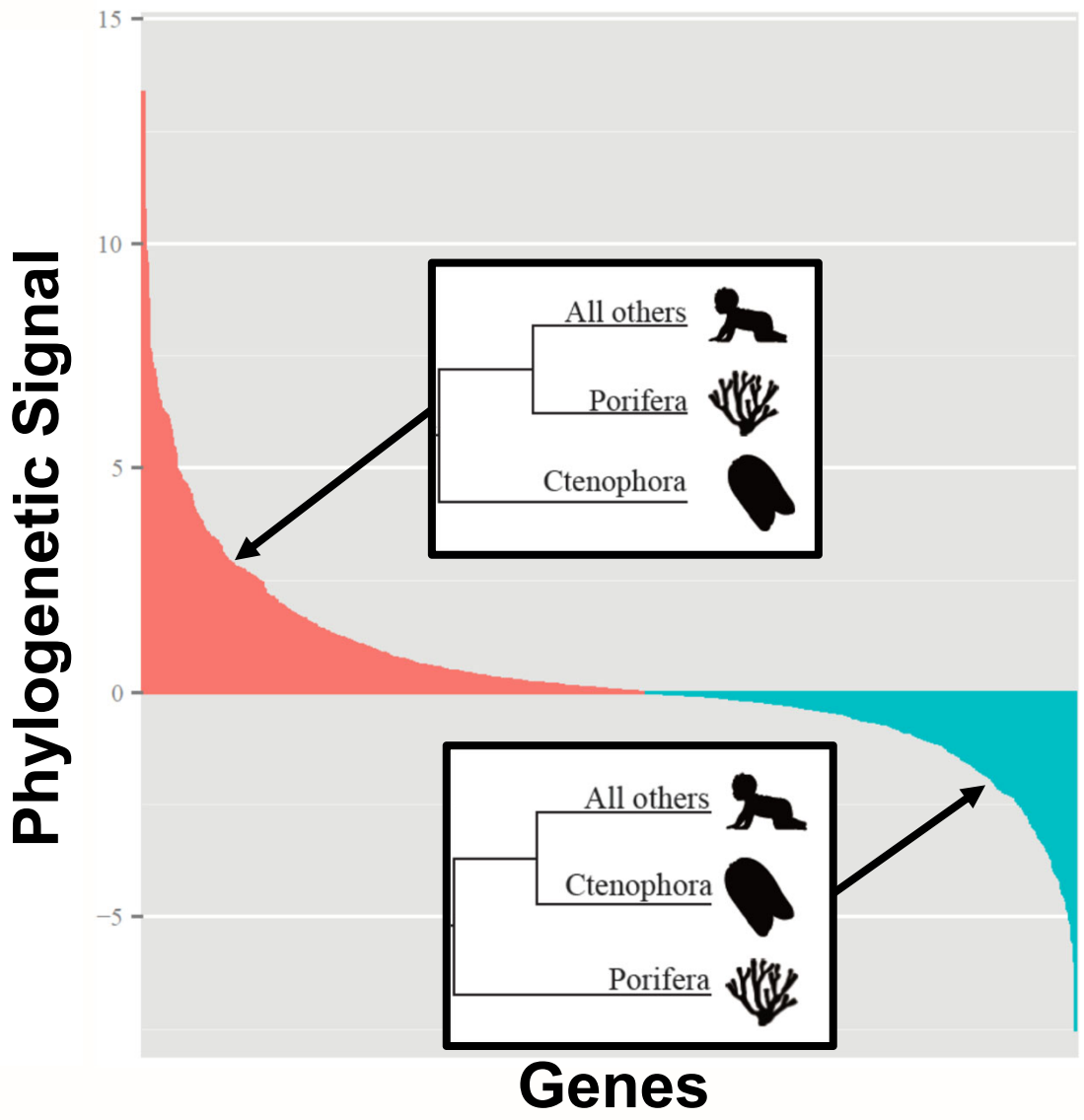
$$\textit{Phylogenetic Signal} = -(\ln(T_1|X_i) - \ln(T_2|X_i))$$

# *Signal of the Genes in a Phylogenomic Data Matrix*



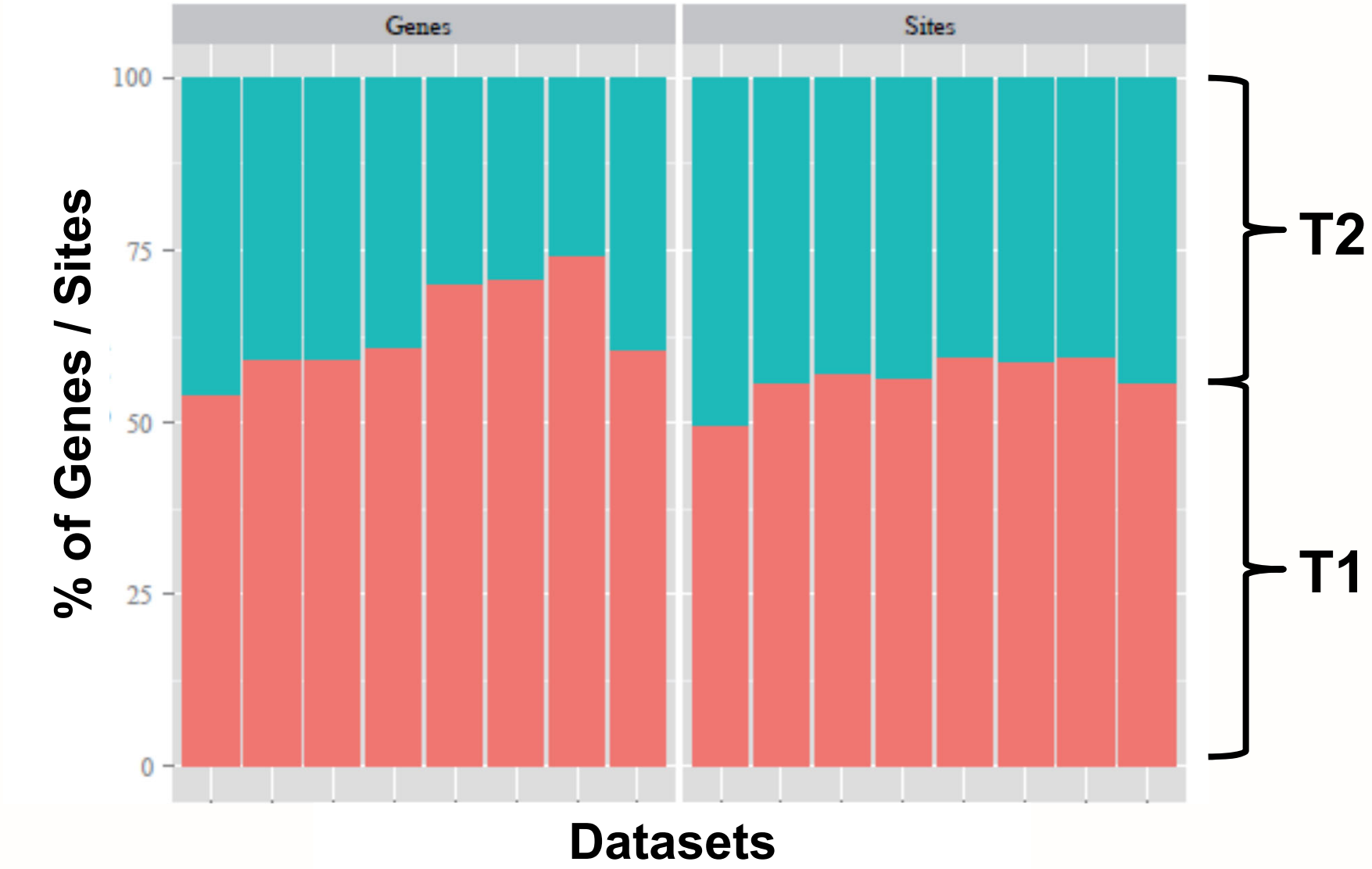

*Shen et al. (2017) Nature Ecol. Evol.; data from Borowiec et al. (2015) BMC Genomics*

# *Signal of the Genes in a Phylogenomic Data Matrix*



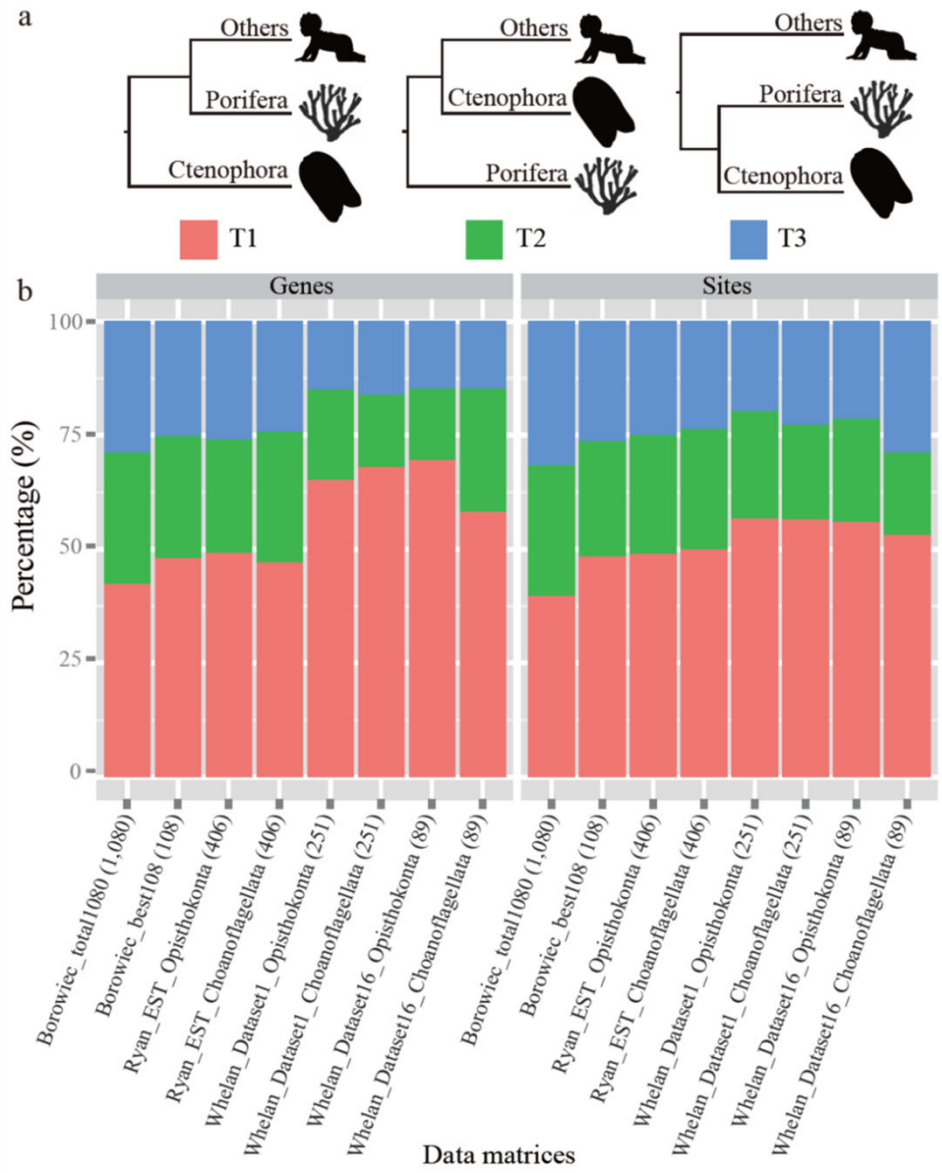

*Shen et al. (2017) Nature Ecol. Evol.*

# *Summarizing Phylogenetic Signal Across Genes and Sites*



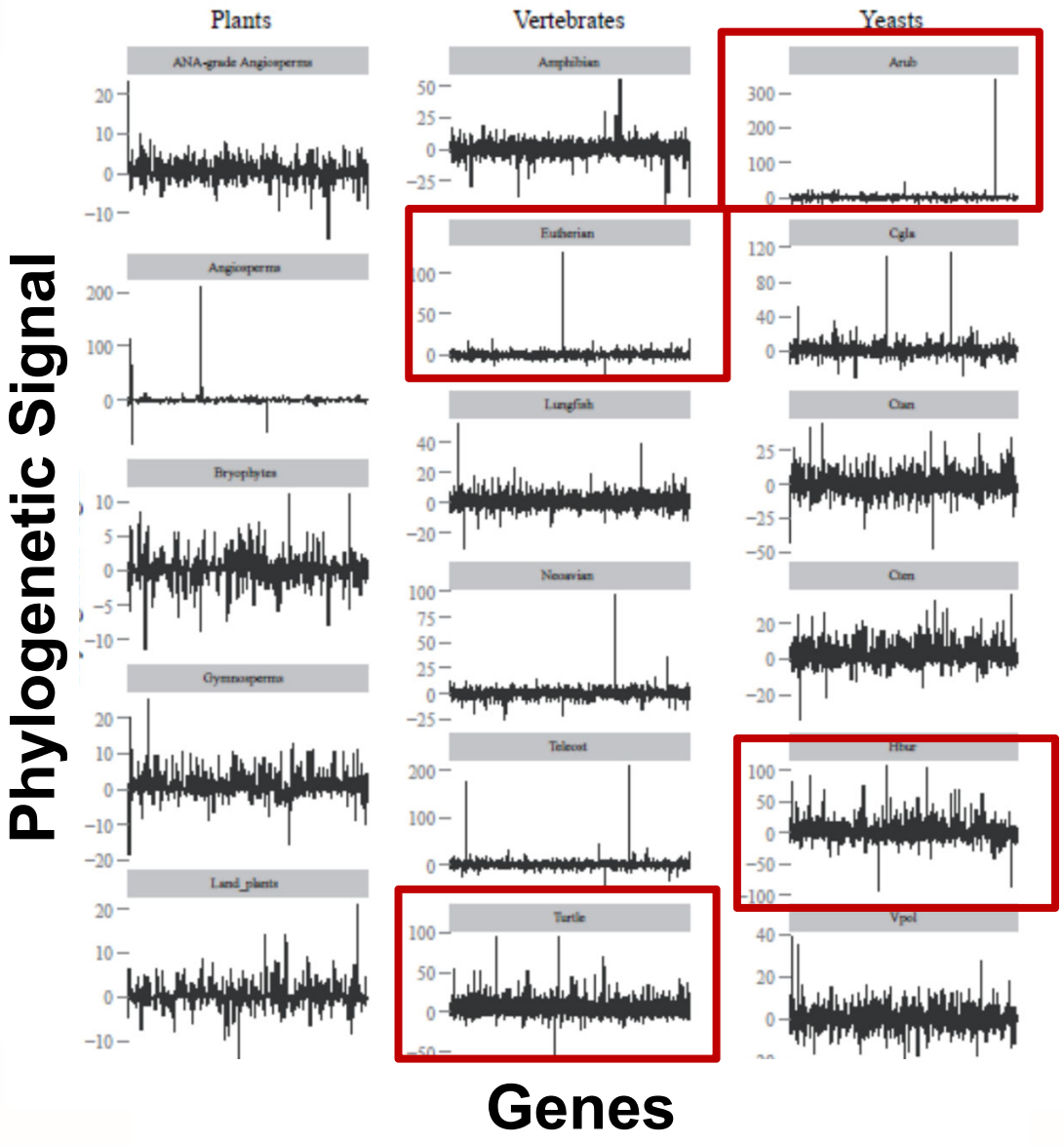

*Shen et al. (2017) Nature Ecol. Evol.*

# *Summarizing the Signal Across All 3 Possible Topologies*

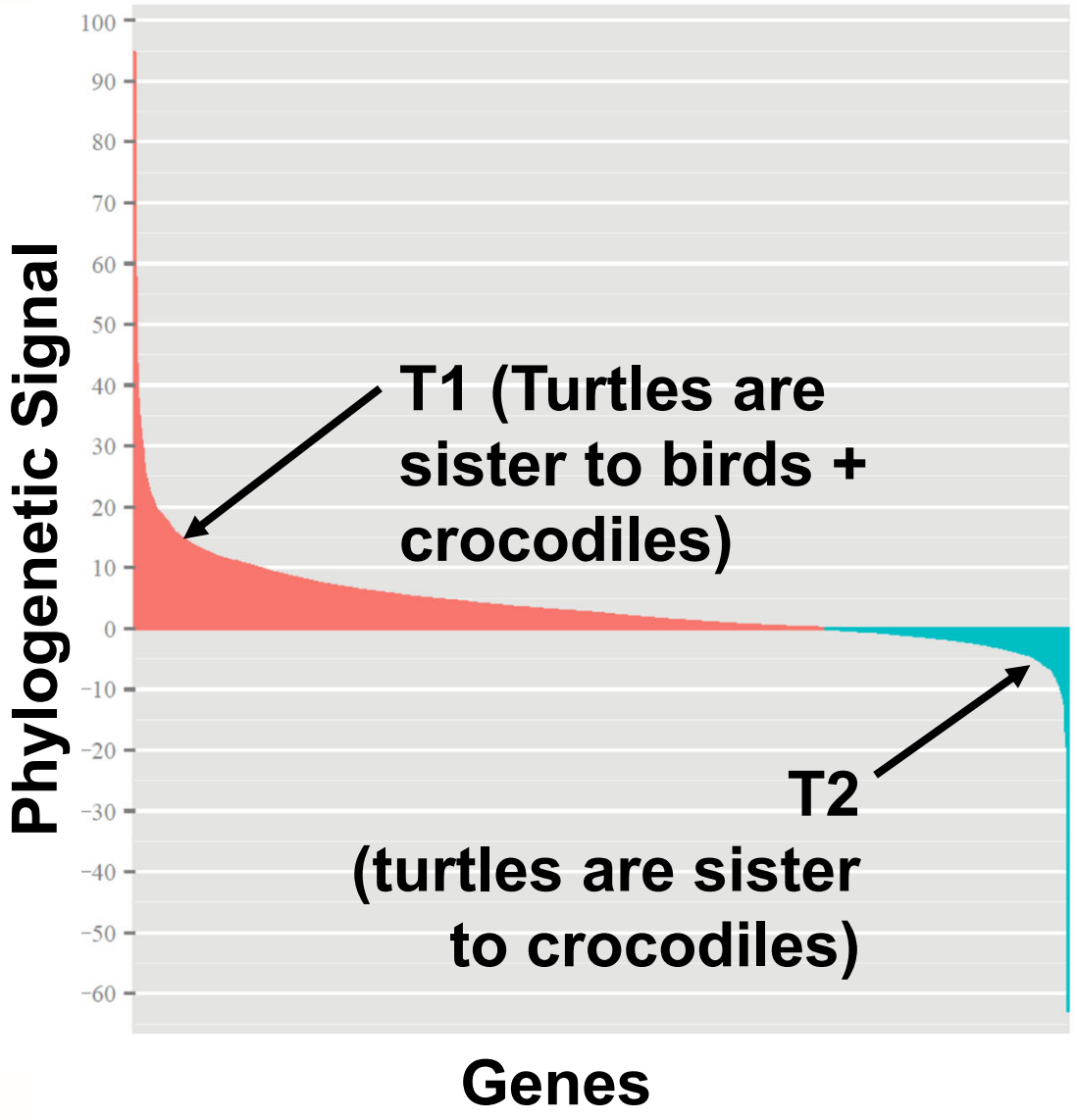



*Shen et al. (2017) Nature Ecol. Evol.*

# *Phylogenetic Signal in Contentious Branches of the ToL*

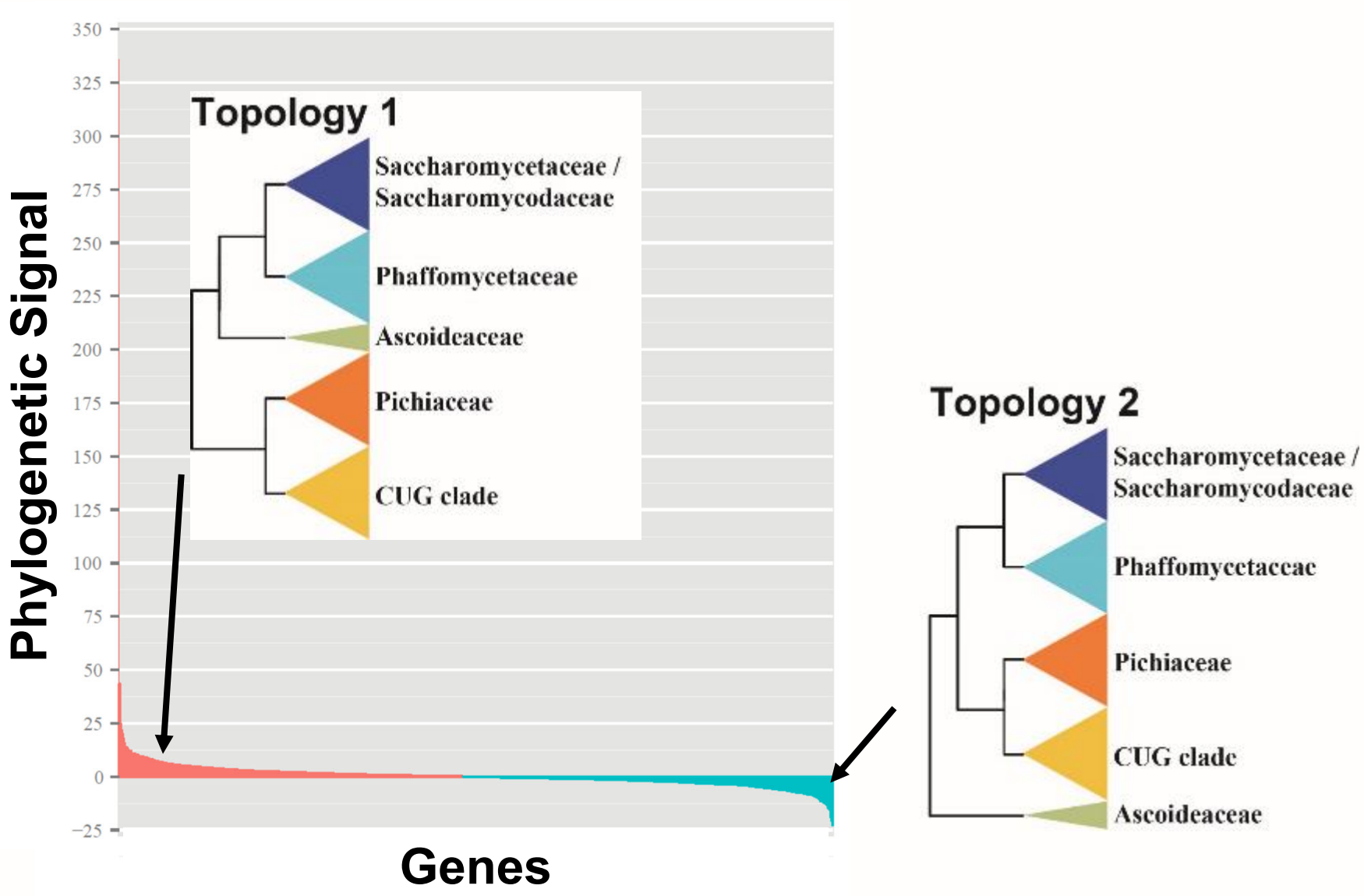


# *The Signal in Some Branches is Very Strong…*



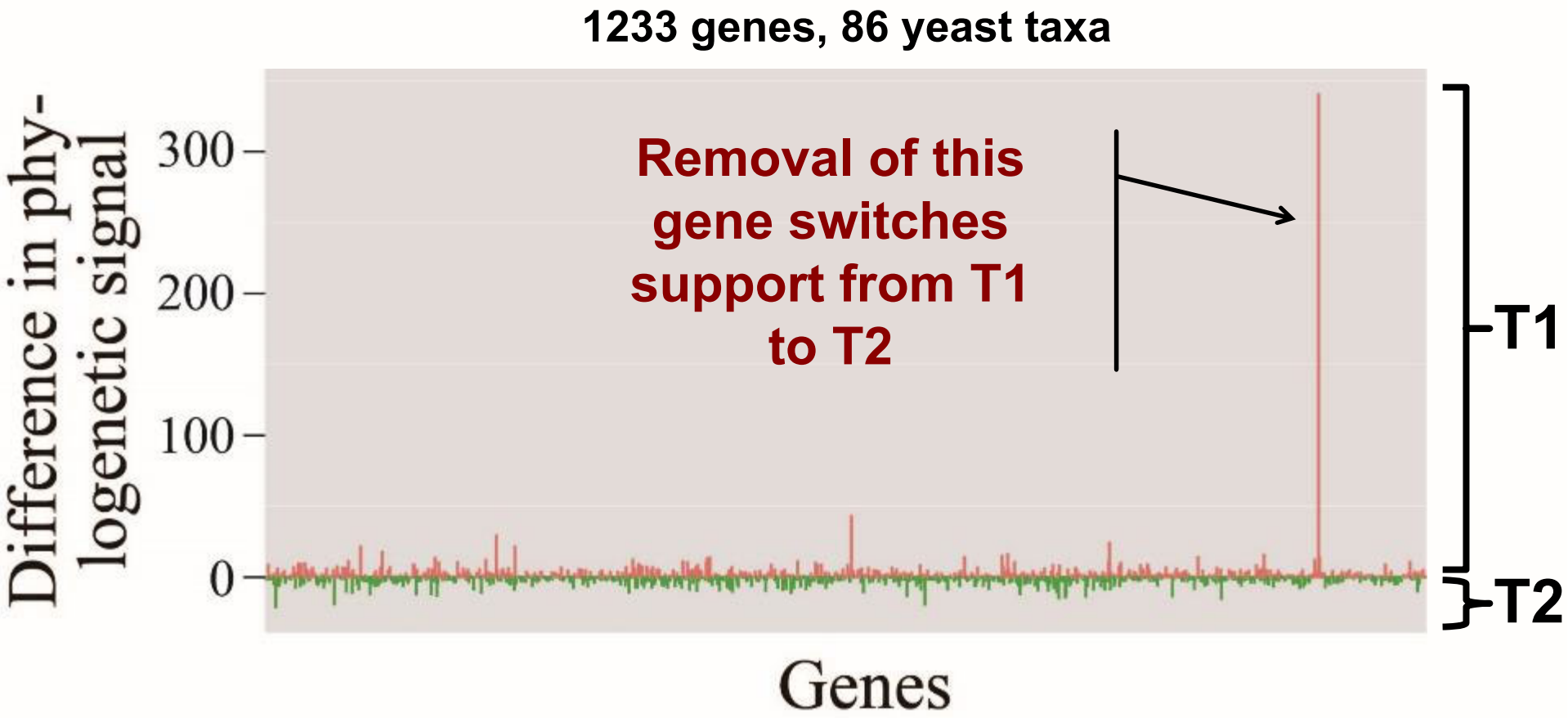

*Shen et al. (2017) Nature Ecol. Evol.*

# ...But in Others It Stems from One or Two Genes

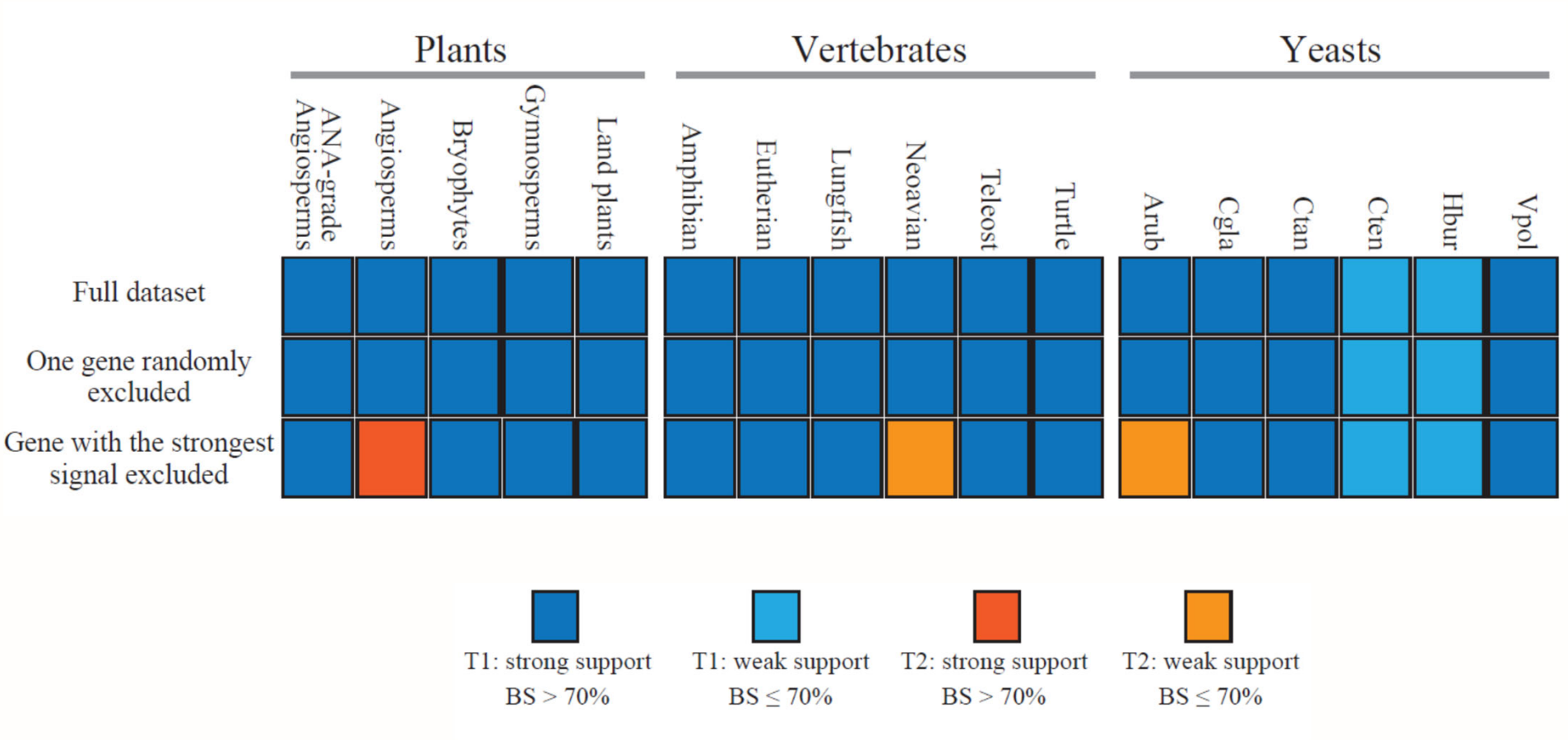



*Shen et al. (2017) Nature Ecol. Evol.*

# *Phylogenetic Signal per Gene for the Two Hypotheses*

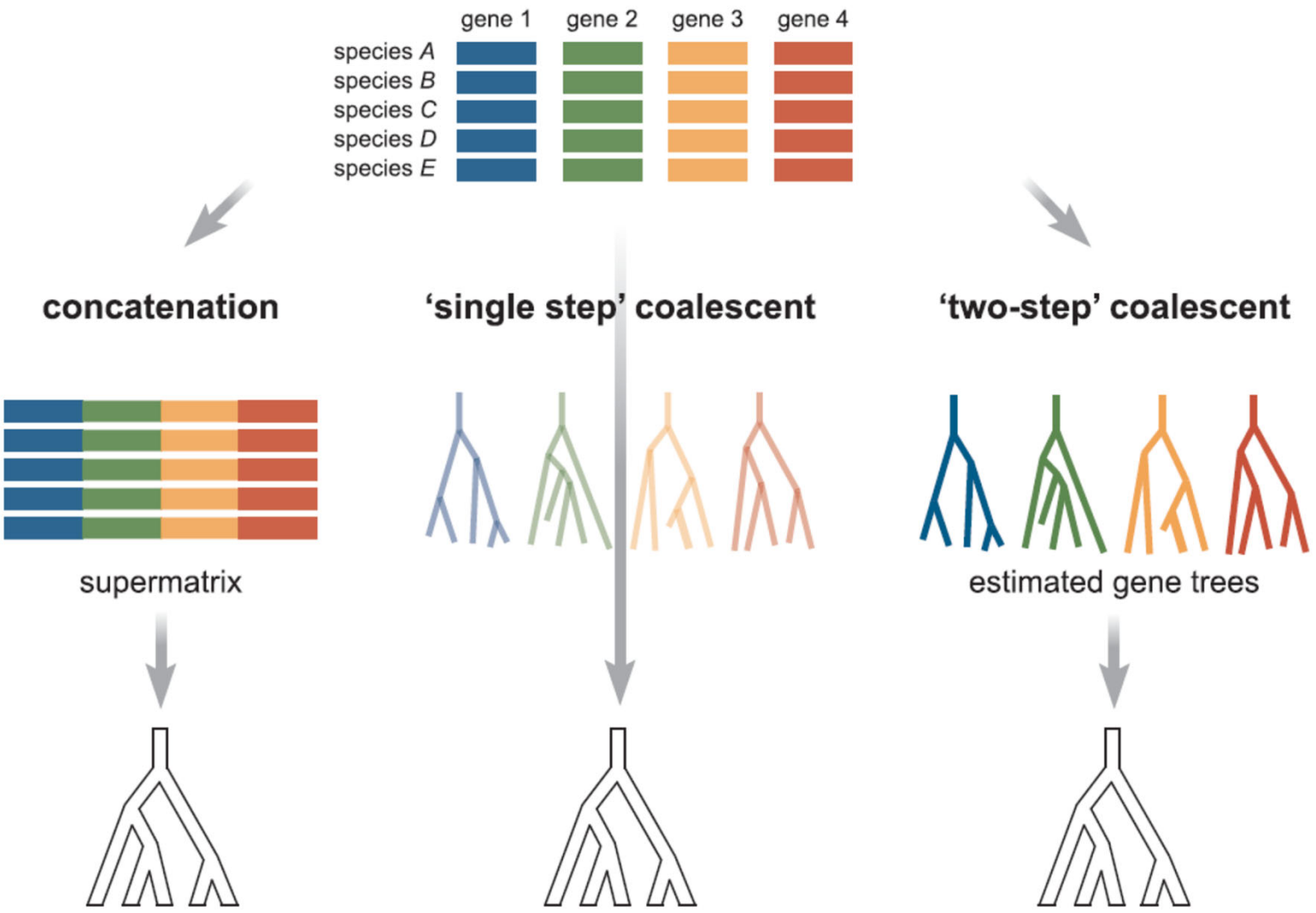


# *Removing One Gene Alters the Topology*



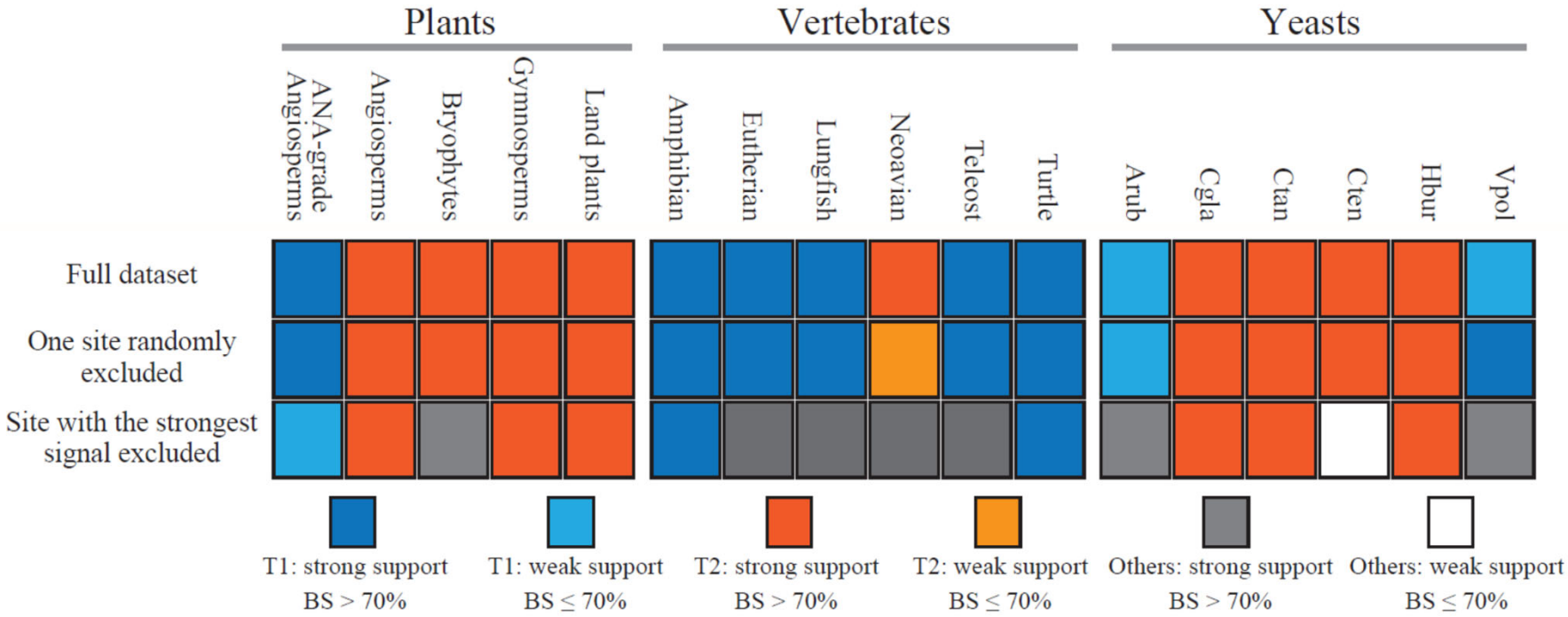

*Shen et al. (2017) Nature Ecol. Evol.*

# *Methods for Phylogenomic Inference*

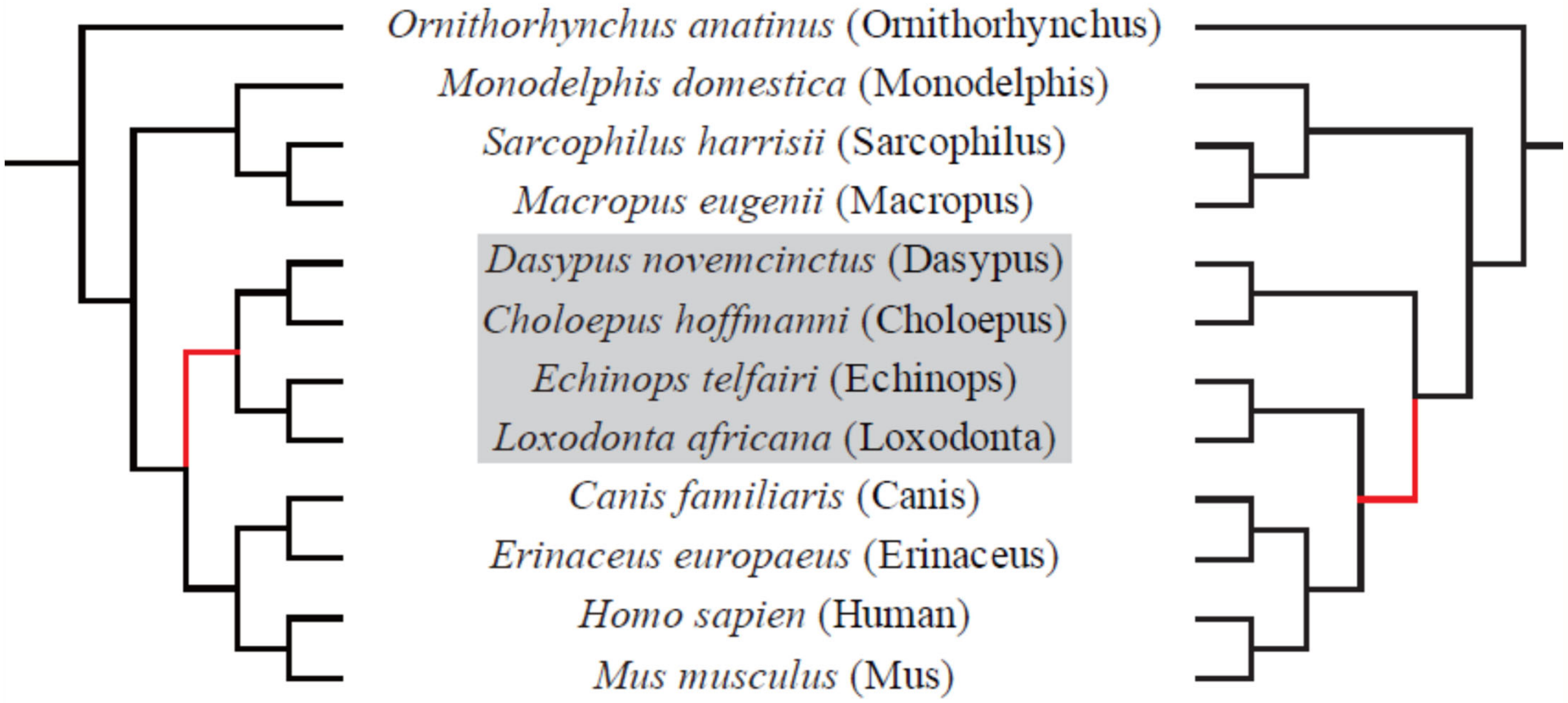



*Liu et al. (2015) Ann. N. Y. Acad. Sci.*

# *What Happens if we Remove One Site from Every Gene?*



# *Removing 1 Site Alters the Topology*



# *What's Going On?*

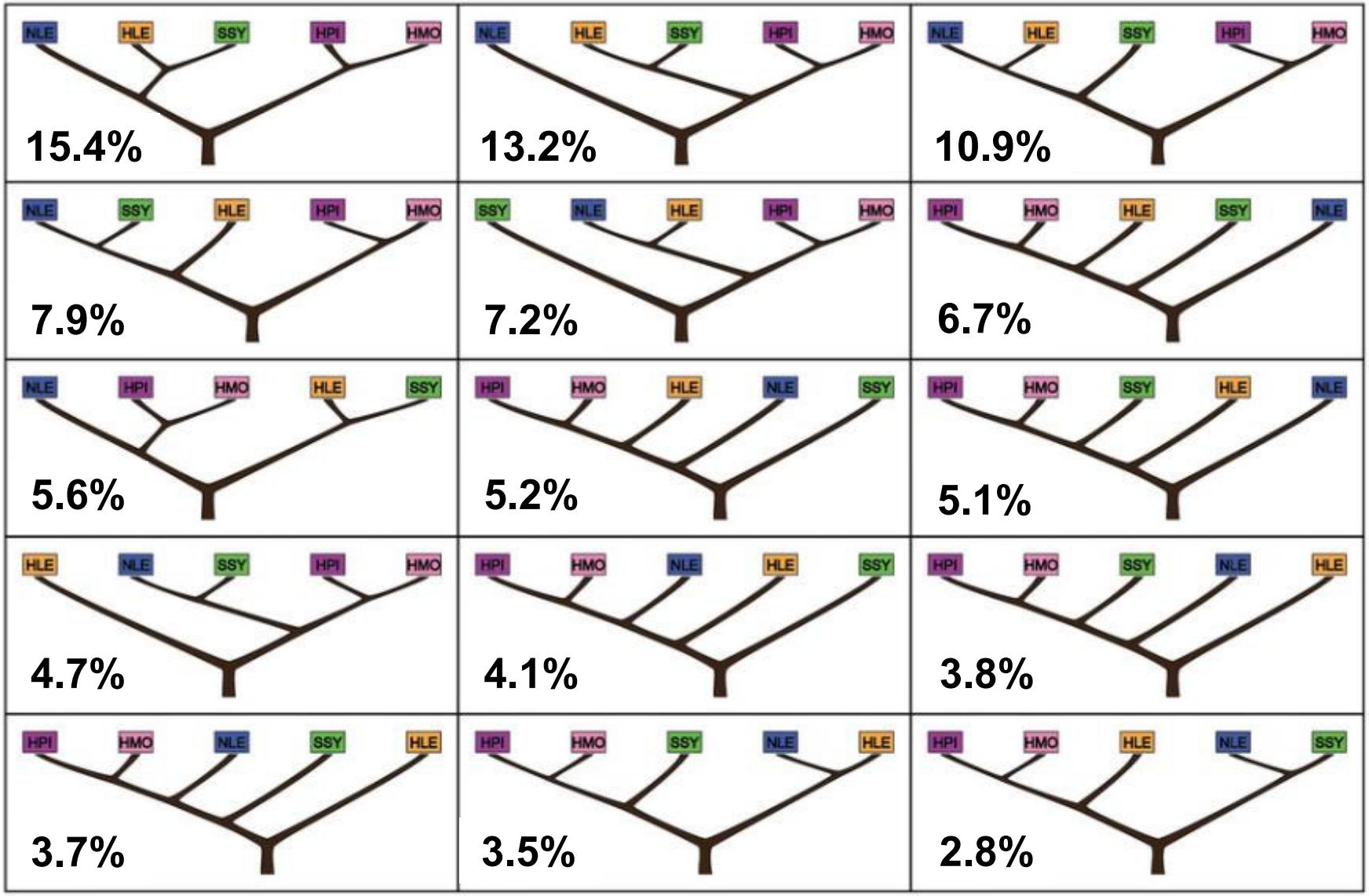
**Explanation #1: Parts of the tree of life are bush-like / network-like (rather than tree-like)**

**Explanation #2: Our models don't fit well our data**

# *The Phylogeny of Primate Genera*

| *Nomascus leucogenys* | *Hoolock leuconedys* | *Symphalangus syndactylus* | *Hylobates pileatus* | *Hylobates moloch* |
|---|---|---|---|---|
|  |  |  |  |  |
| NLE | HLE | SSY | HPI | HMO |

# *Which is the Phylogeny of These 4 Genera?*




*Carbone et al. (2014) Nature*

# *Genomfart?*

- **Parts of the tree of life are more likely to resemble a bush rather than a tree – do we expect that we can confidently infer every branch and twig?**
- **Bootstrap-based measures not useful in large data sets**
- **Methods evaluating conflict among data subsets (e.g., internode certainty among genes or sites) are preferable**
- **Explicitly identify internodes that, despite the use of genome-scale data sets, robust study designs and powerful algorithms, are poorly supported**

# Coffee Break

# *Lecture Outline*

- Introduction to evolutionary genomics
- Phylogenomics

---------- Coffee Break ----------

- Using genomes to understand ecology and evolution

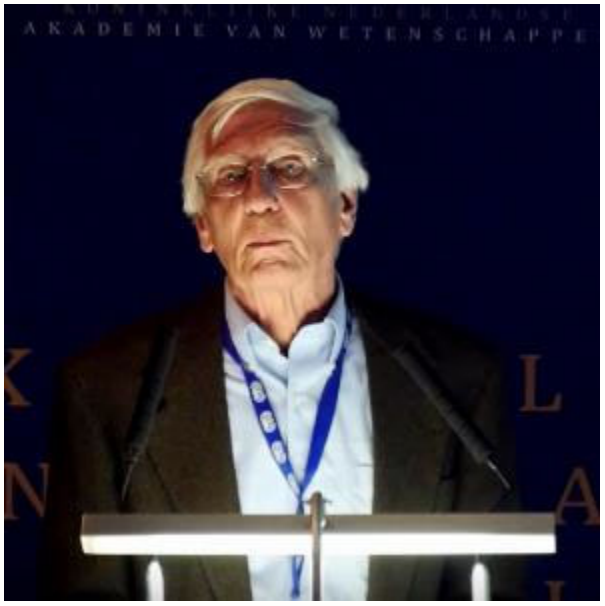
# *The Making of Biodiversity across the Yeast Subphylum*

THE UNIVERSITY of WISCONSIN MADISON

USDA

VANDERBILT UNIVERSITY

**Hittinger lab**




**Kurtzman lab**

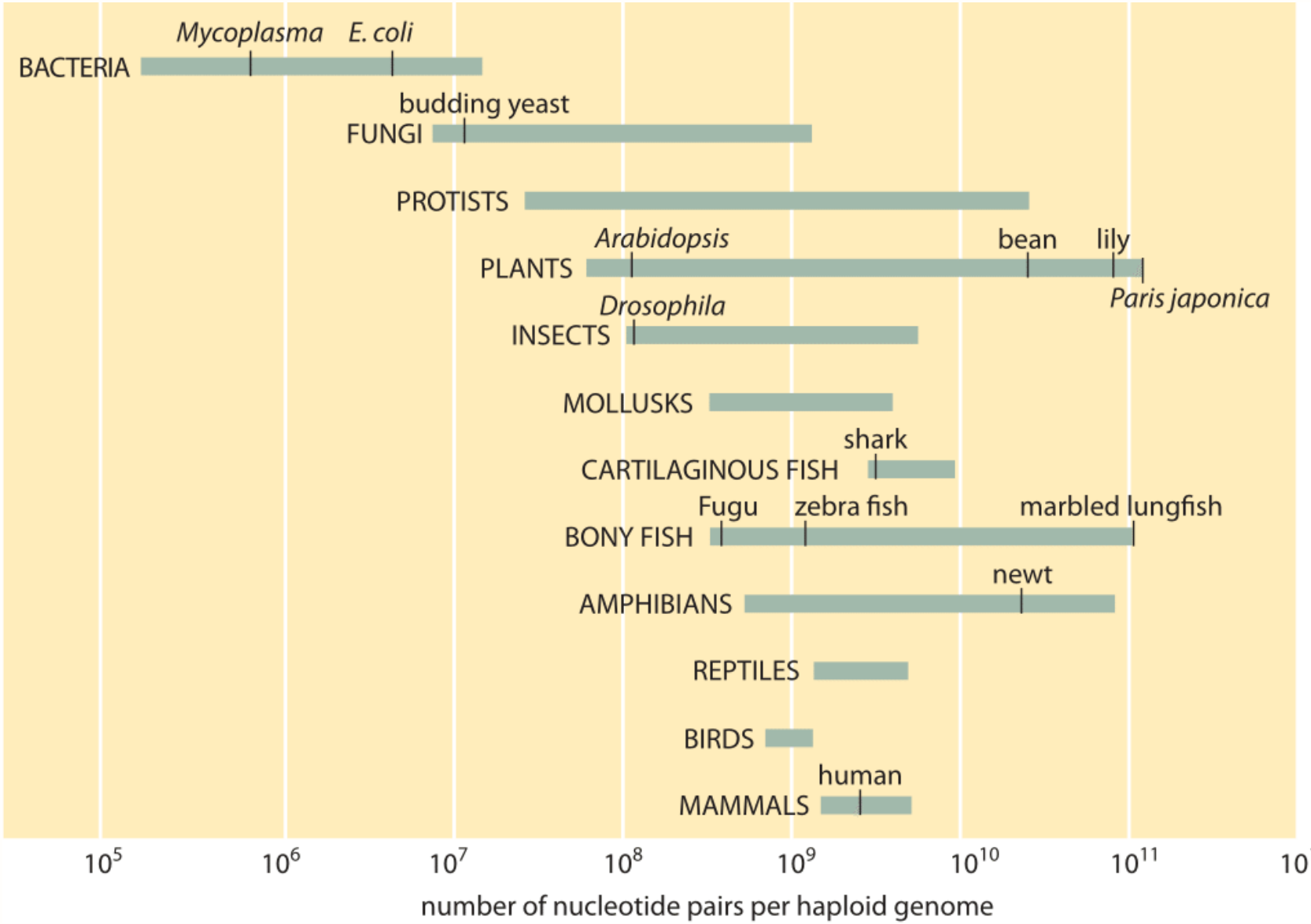
**Rokas lab**

# *The Making of Biodiversity across the Yeast Subphylum*

- Sequence the genomes of all ~1,000+ known budding yeast species
- Construct their definitive phylogeny and timetree
- Examine the impact of metabolism on yeast diversification
- Revise their taxonomy

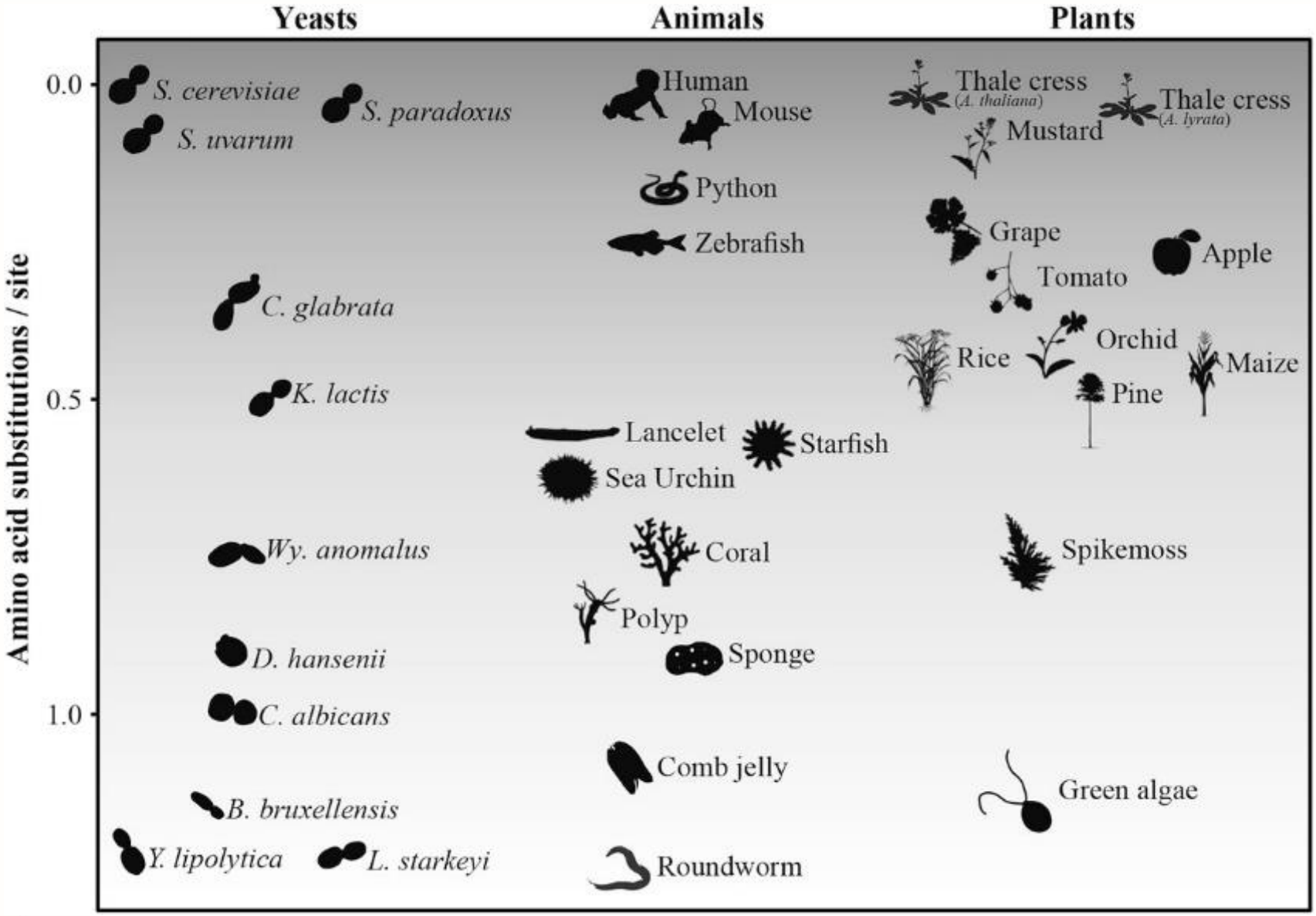

# why budding yeasts?

# *Budding Yeasts Have Very Small Genomes*

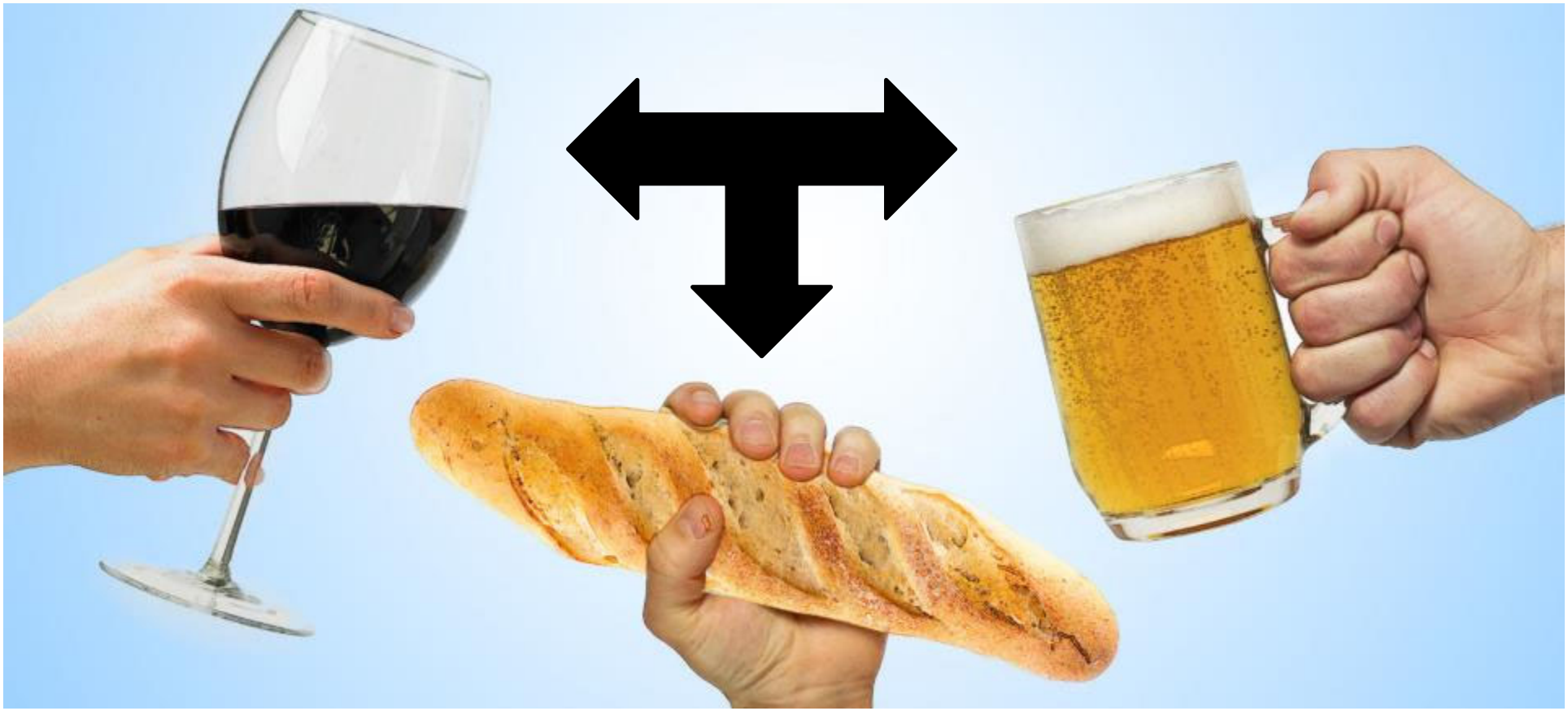



***http://book.bionumbers.org/how-big-are-genomes/***

# *Budding Yeasts Exhibit Striking Genomic Diversity*



# S. cerevisiae, *Cornerstone of Wine, Baking, and Brewing Industries*



# *Several Other Genera Critical to the Food Industry*

**Kefir**

**Soy sauce**

**Sourdough**

**Lambic beers**

**Kimchi**

**Dietary supplements**

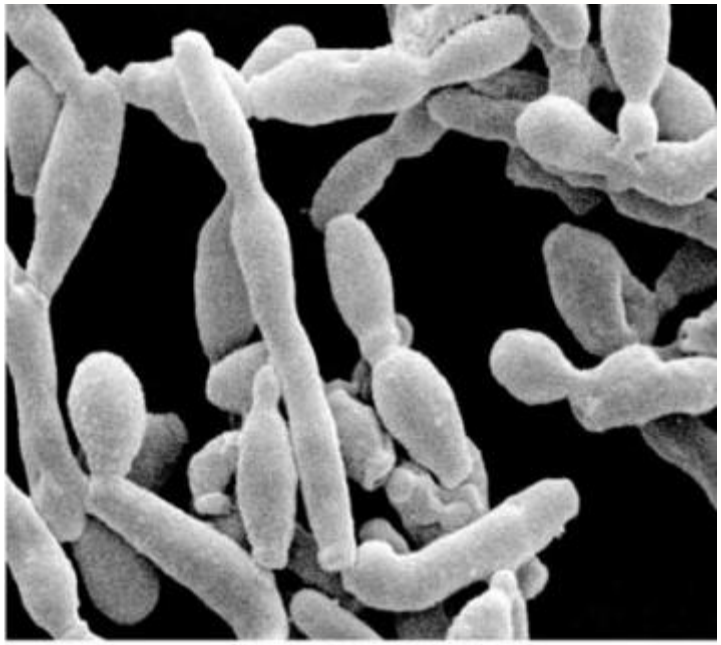
**Probiotics**

**Cheeses**

**Genera involved: *Saccharomyces*, *Kluyveromyces*, *Zygosaccharomyces*, *Candida*, *Kazachstania*, *Pichia*, and *Dekkera* (*Brettanomyces*)**

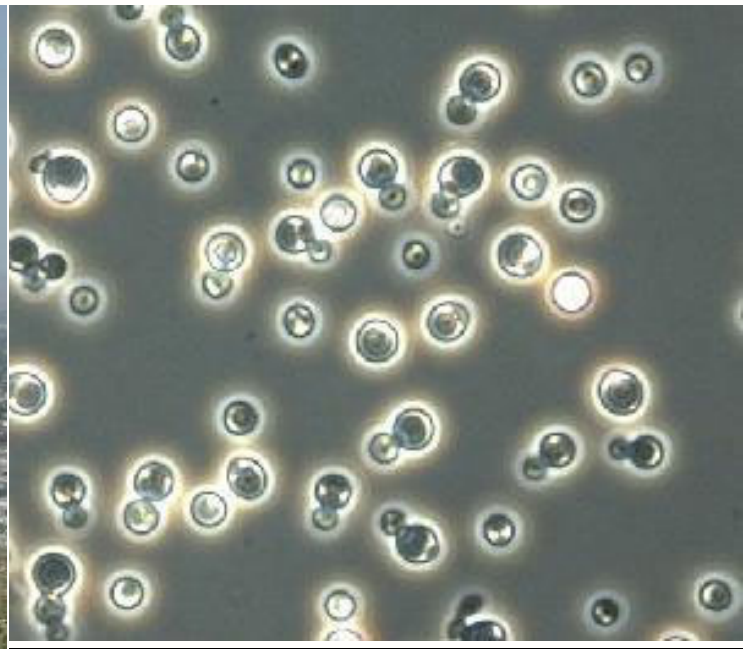
# *The Metabolisms of Budding Yeasts Vary Widely*

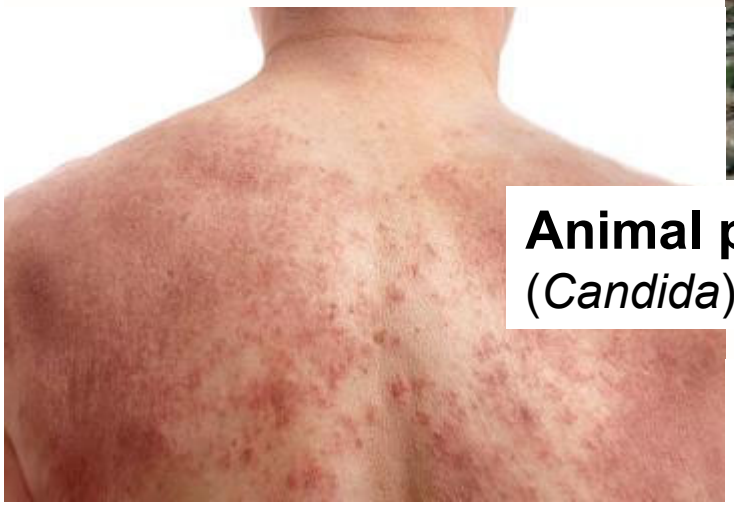
**Xylose fermenters**
(*Scheffersomyces (Pichia) stipitis*)



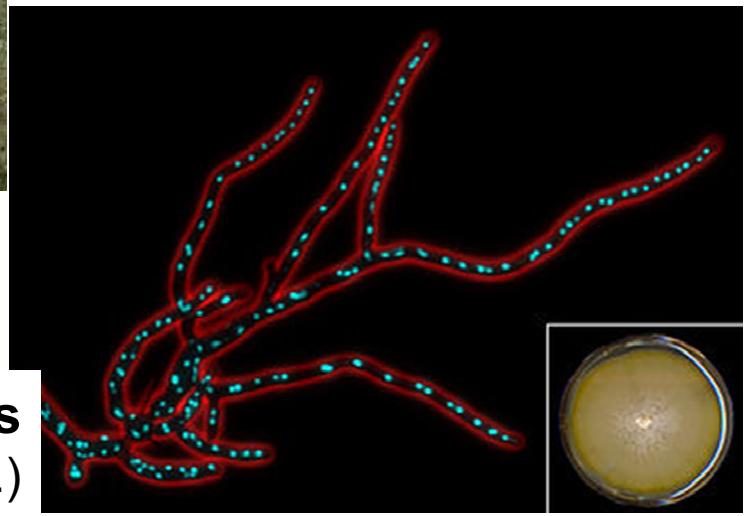
**Cactophilic yeasts**

**Oil producers**
(*Lipomyces, Yarrowia*)





**Animal pathogens**
(*Candida*)



**Plant pathogens**
(*Eremothecium* sp.)

i. **Developing pipelines for genomic and phylogenomic data analyses**

ii. **Inferring a comprehensive & robust genome-scale phylogeny of budding yeasts**

iii. **Using the phylogeny to gain insights into budding yeast evolution**

# *Developing Pipelines for Genome Assembly*

**iWGS: *in silico* Whole Genome Sequencer & Analyzer**

**INPUT: Experimental Design** + Reference Genome (Optional)

↓

(Optional) Step 1: Simulation of Illumina / PacBio Data

↓

**Step 2: Quality Control (Quality / Adaptor Trimming; Error Correction)**

↓

**Step 3: Assembly (Illumina-only (10), PacBio-only (3), Hybrid (4))**

↓

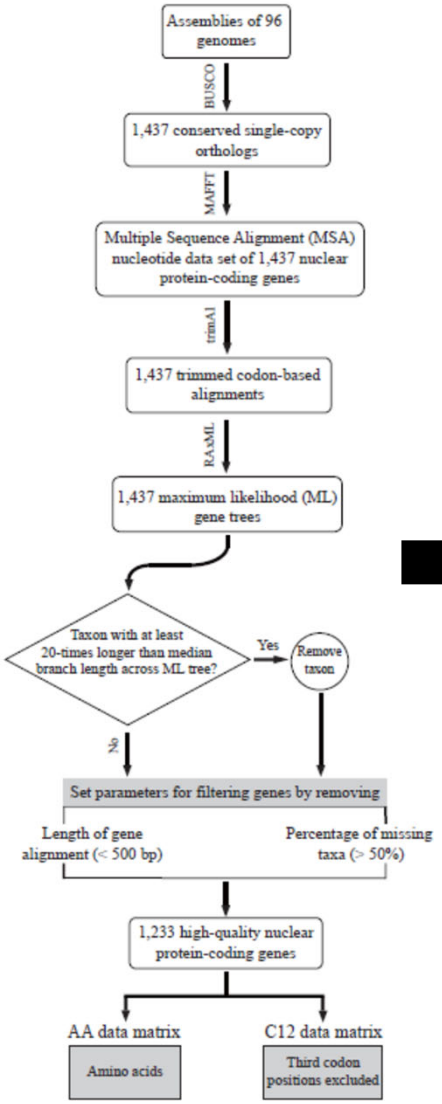
**Step 4: QUAST Evaluation (Standalone** or vs the Reference Genome**)**

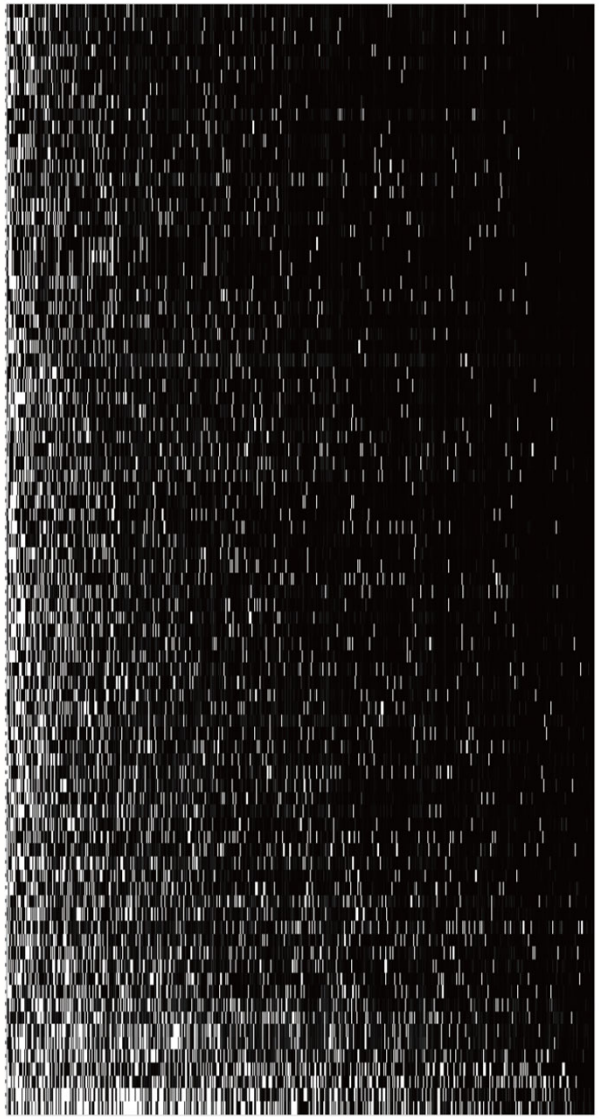
↓

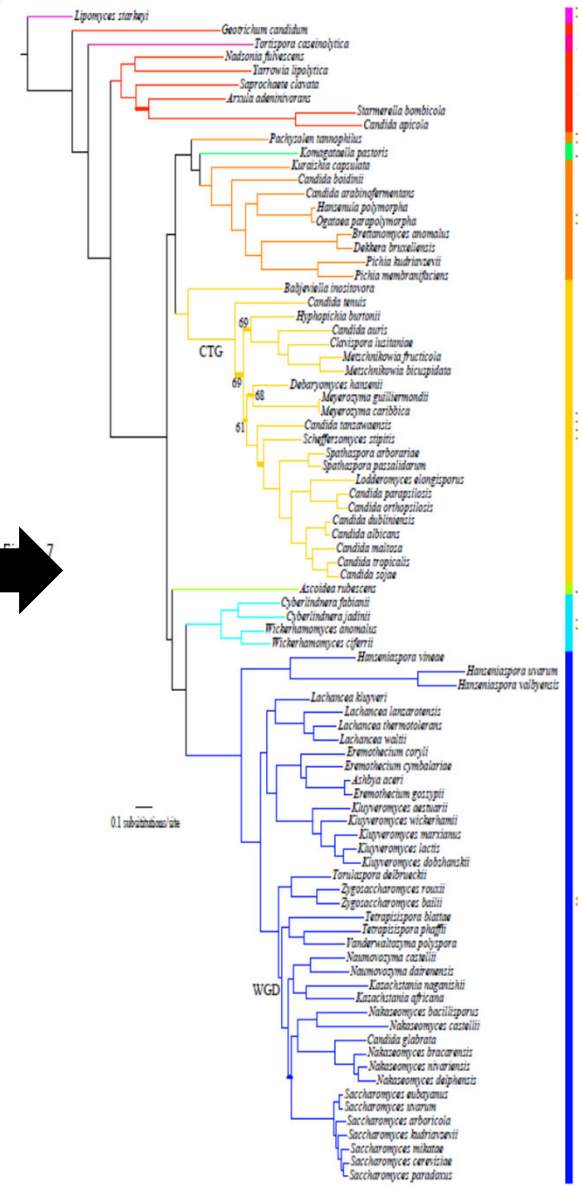
**OUTPUT: Evaluation Report, Ranking of Experimental Designs Tested**

*Zhou et al. (2016) G3*

# *Developing Pipelines for Phylogenomic Inference*

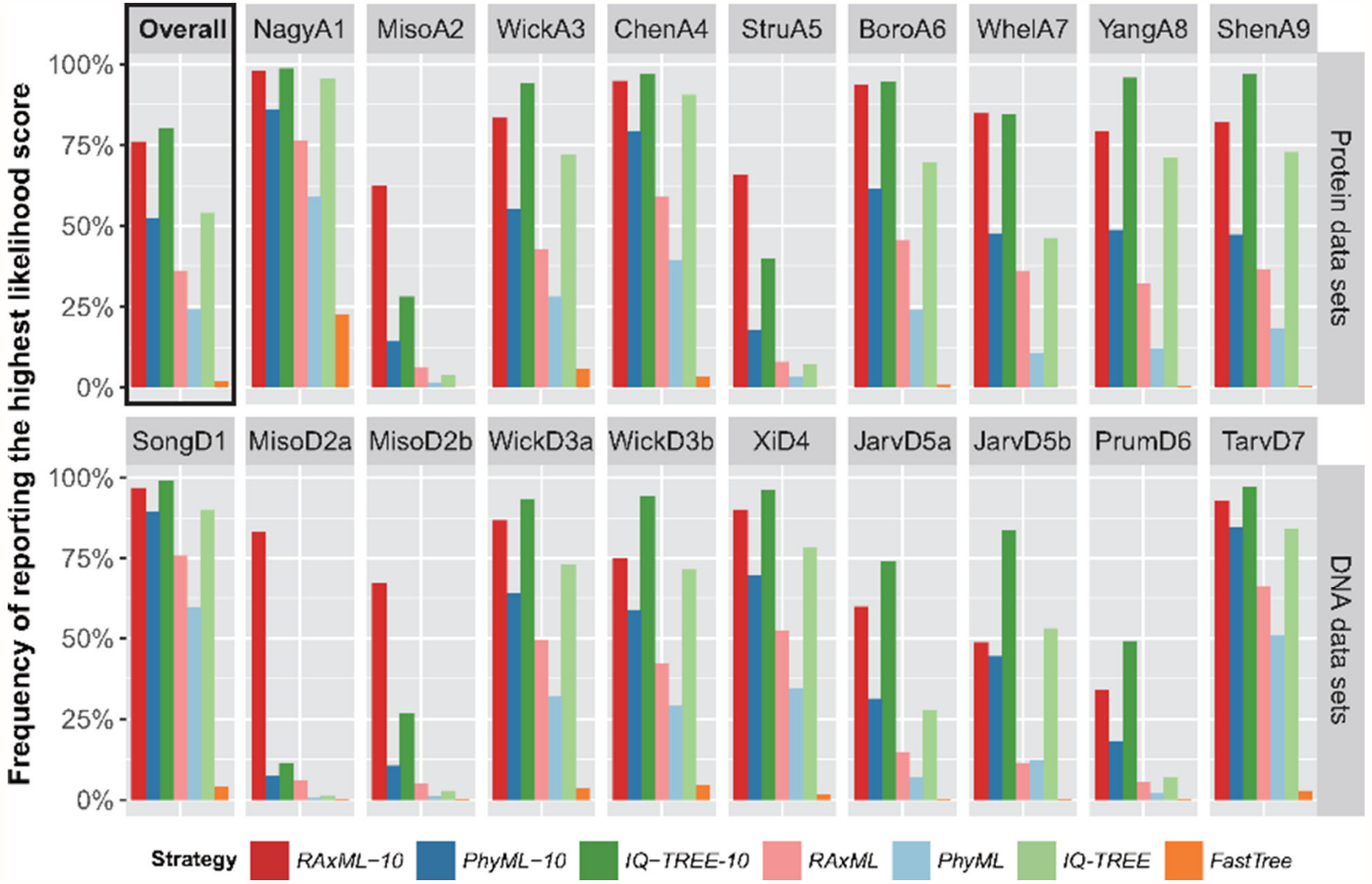



*Shen et al. (2016) G3*

# *Assessing Speed and Accuracy of Phylogenomic Software*




*Zhou et al. (2018) Mol. Biol. Evol.*

i. Developing pipelines for genomic and phylogenomic data analyses

**ii. Inferring a comprehensive & robust genome-scale phylogeny of budding yeasts**

iii. Using the phylogeny to gain insights into budding yeast evolution

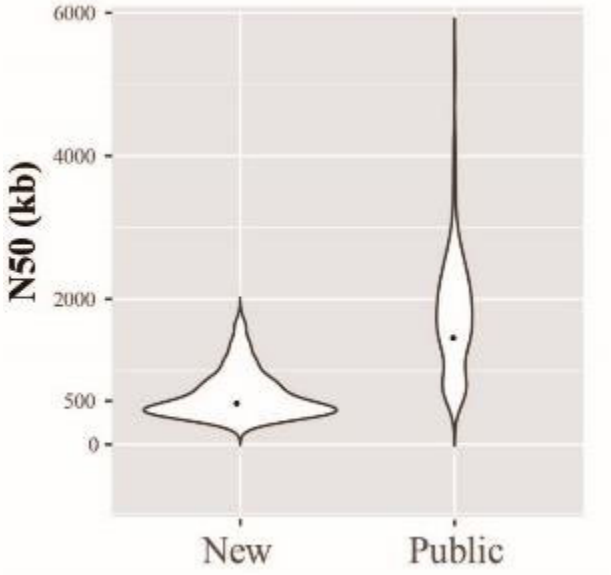
❖ **Sequenced the genomes of 220 species (196 Y1000+ species + 24 RIKEN genomes); most of them are from type strains**

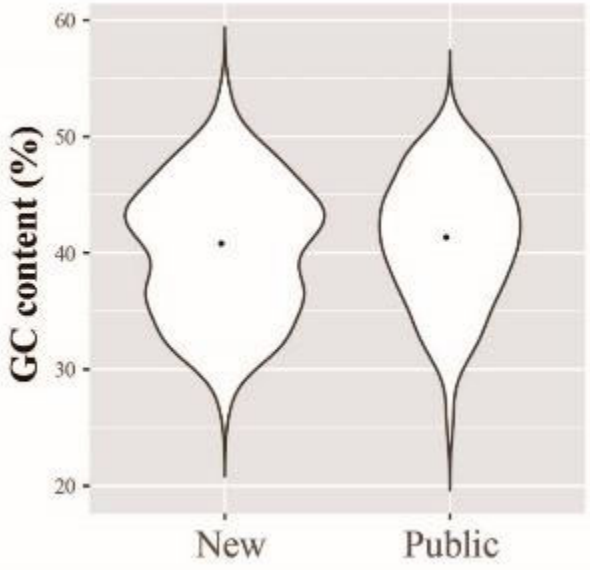
❖ **+ 112 publicly available genomes -> 332 genomes**

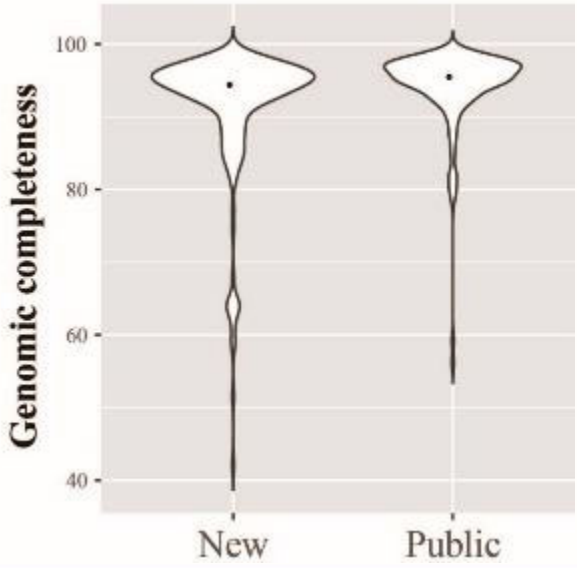
❖ **Sampled taxa from 79 / 92 genera (~85%)**

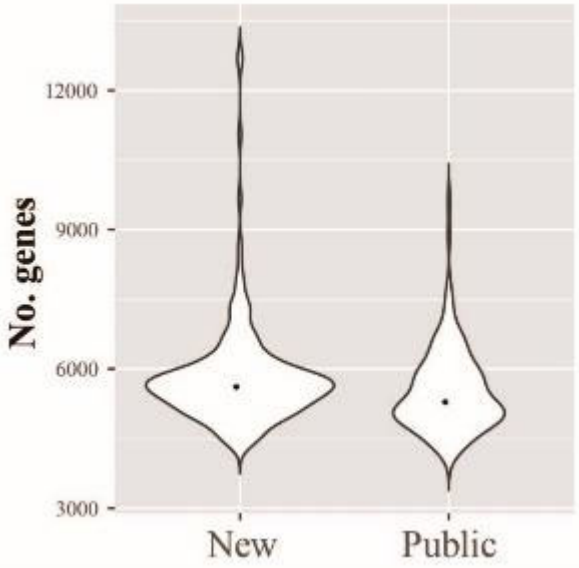
***Shen, Opulente, Kominek, Zhou et al. (2018) Cell***

# *The Quality of the New Genomes is High*



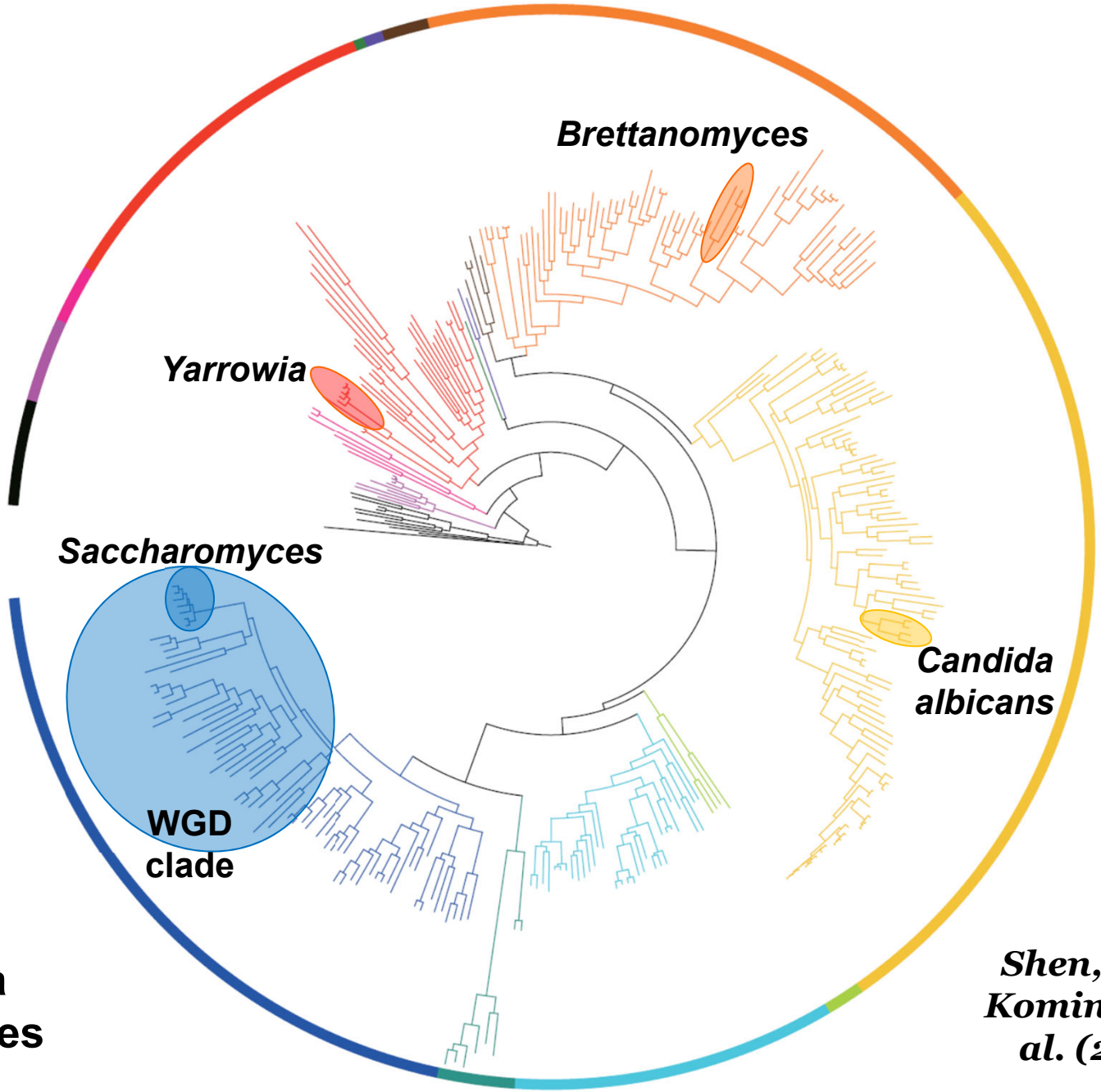




*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

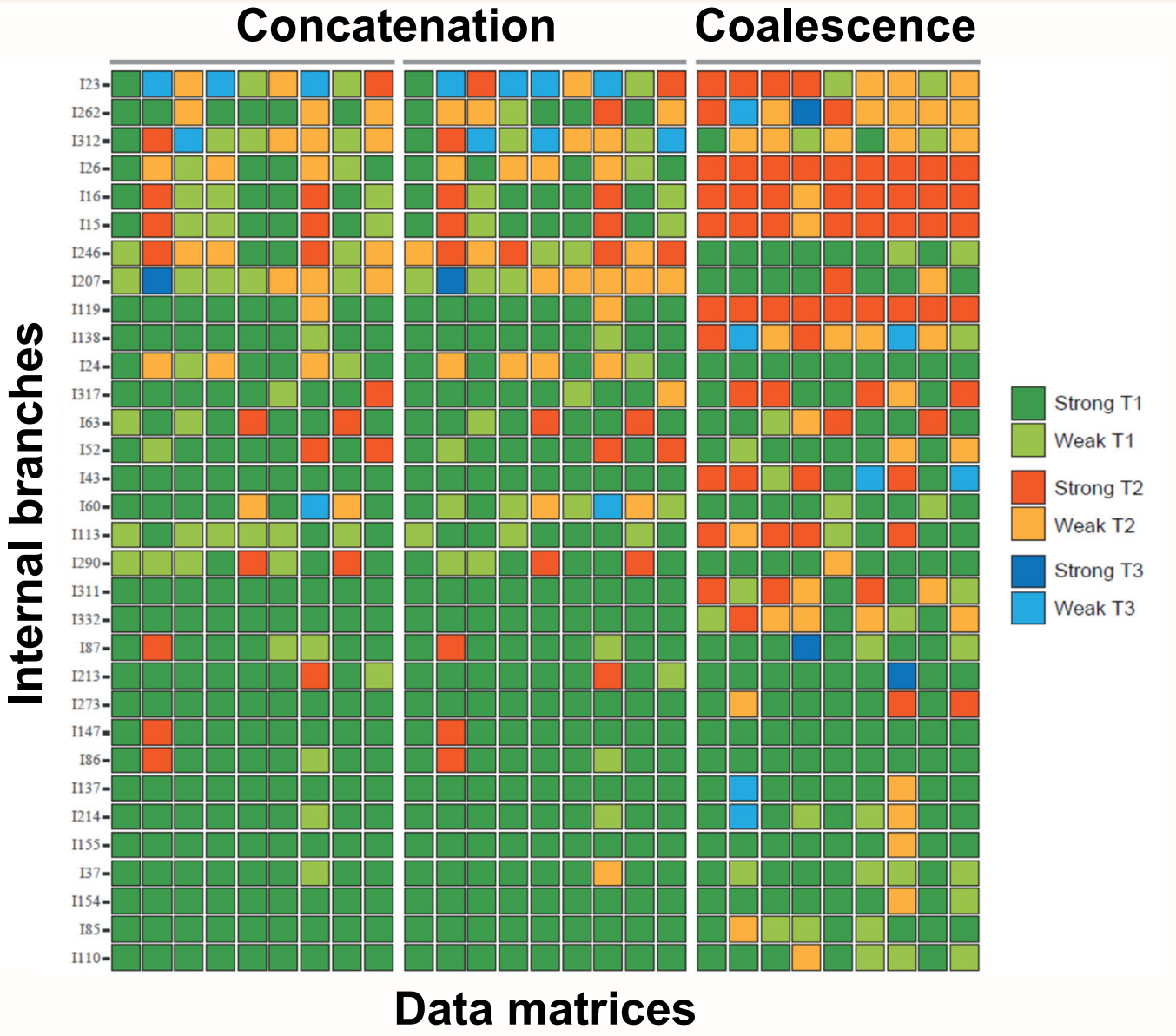
# *Genome-Scale Phylogeny of Budding Yeasts*




**332 taxa**
**2,408 genes**

***Shen, Opulente, Kominek, Zhou et al. (2018) Cell***

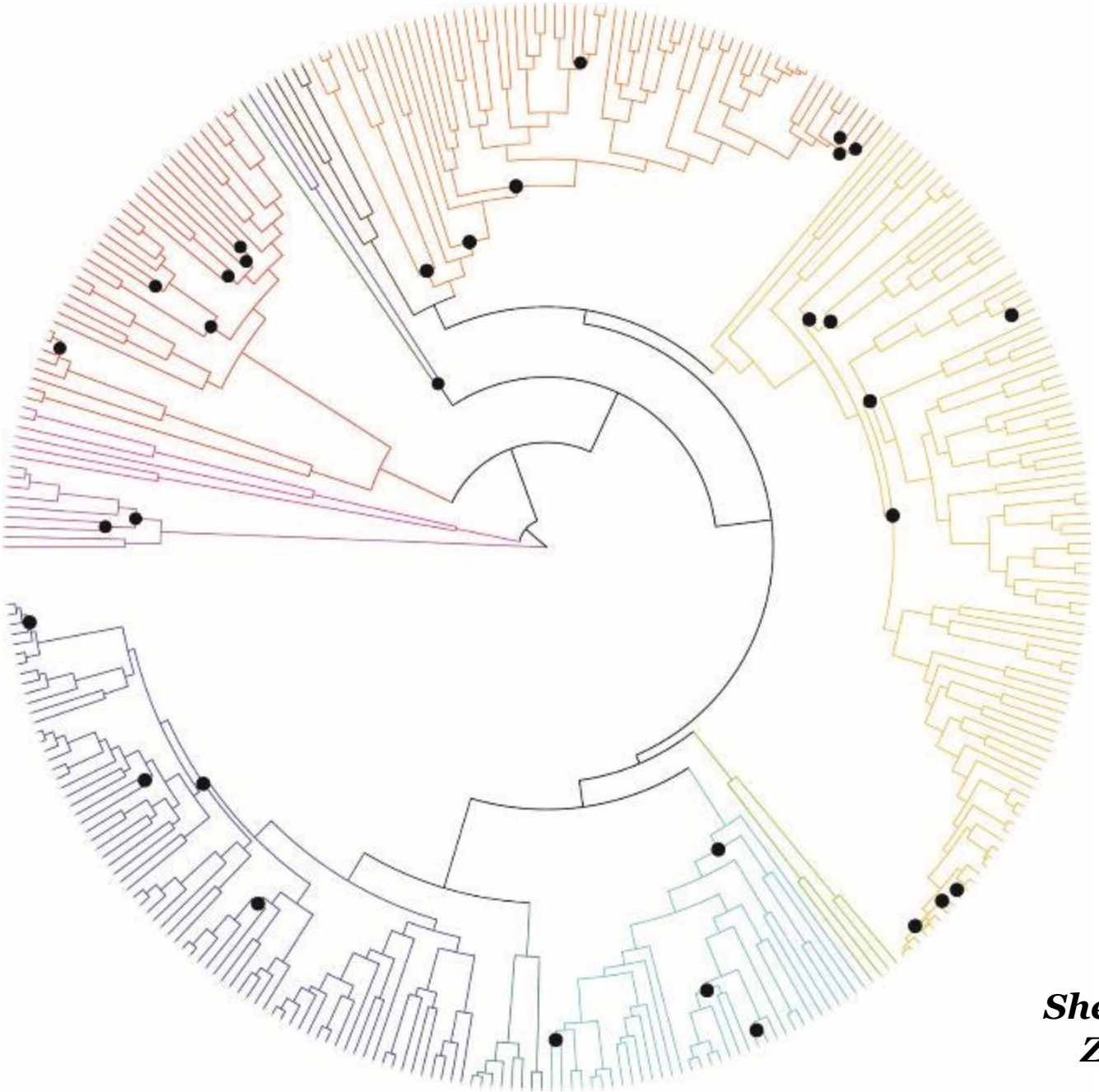
# *The 32 Conflicting Branches in the Yeast Phylogeny*




**~10% (32 / 331) of internal branches show conflict between analyses**

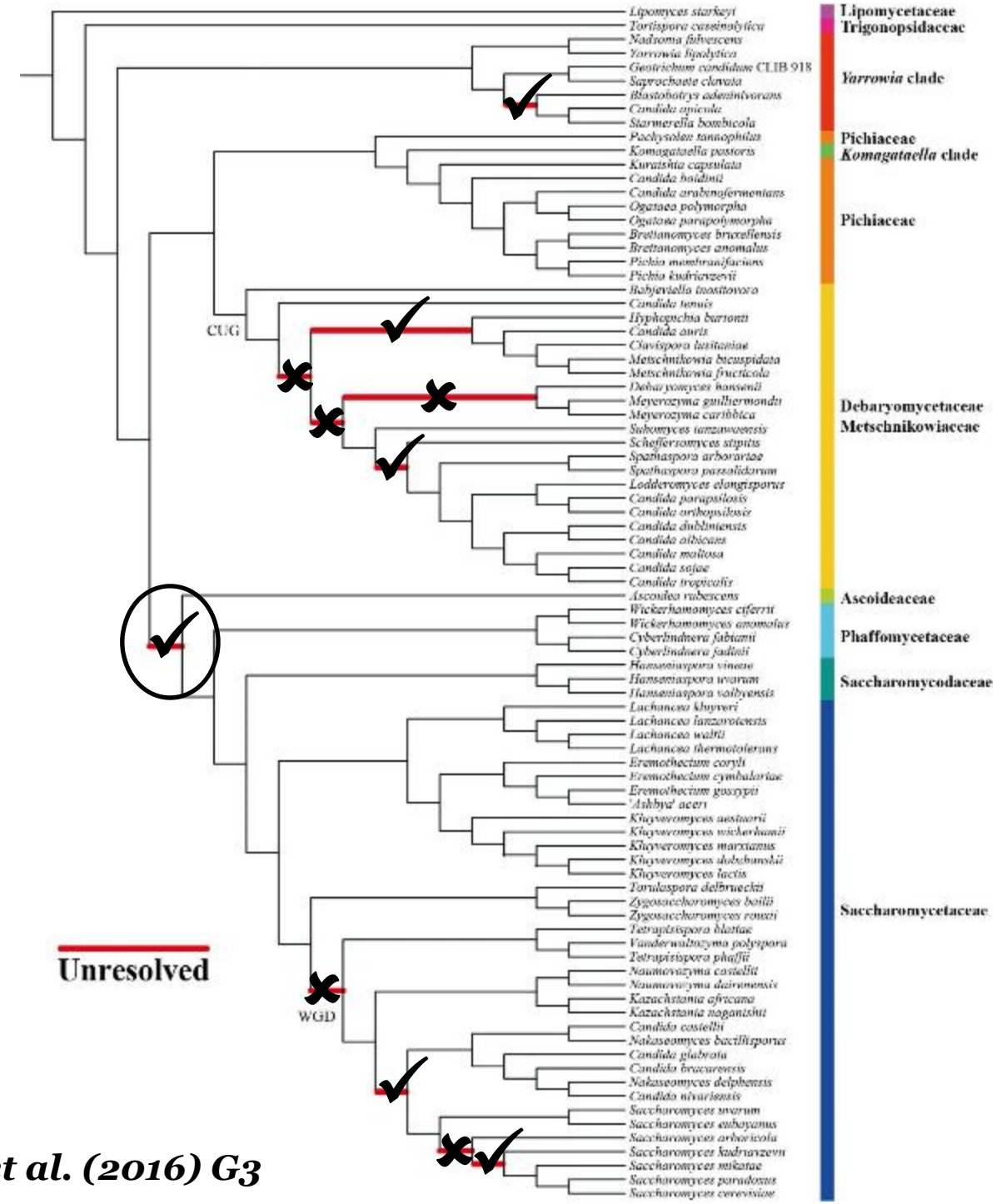
*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

# *Distribution of Conflict on the Yeast Phylogeny*



- Lipomycetaceae
- Trigonopsidaceae
- Dipodascaceae / Trichomonascaceae
- Alloascoideaceae
- *Sporopachydermia* clade
- CUG-Ala clade
- Pichiaceae
- CUG-Ser1 clade
- CUG-Ser2 clade
- Phaffomycetaceae
- Saccharomycodaceae
- Saccharomycetaceae

***Shen, Opulente, Kominek, Zhou et al. (2018) Cell***



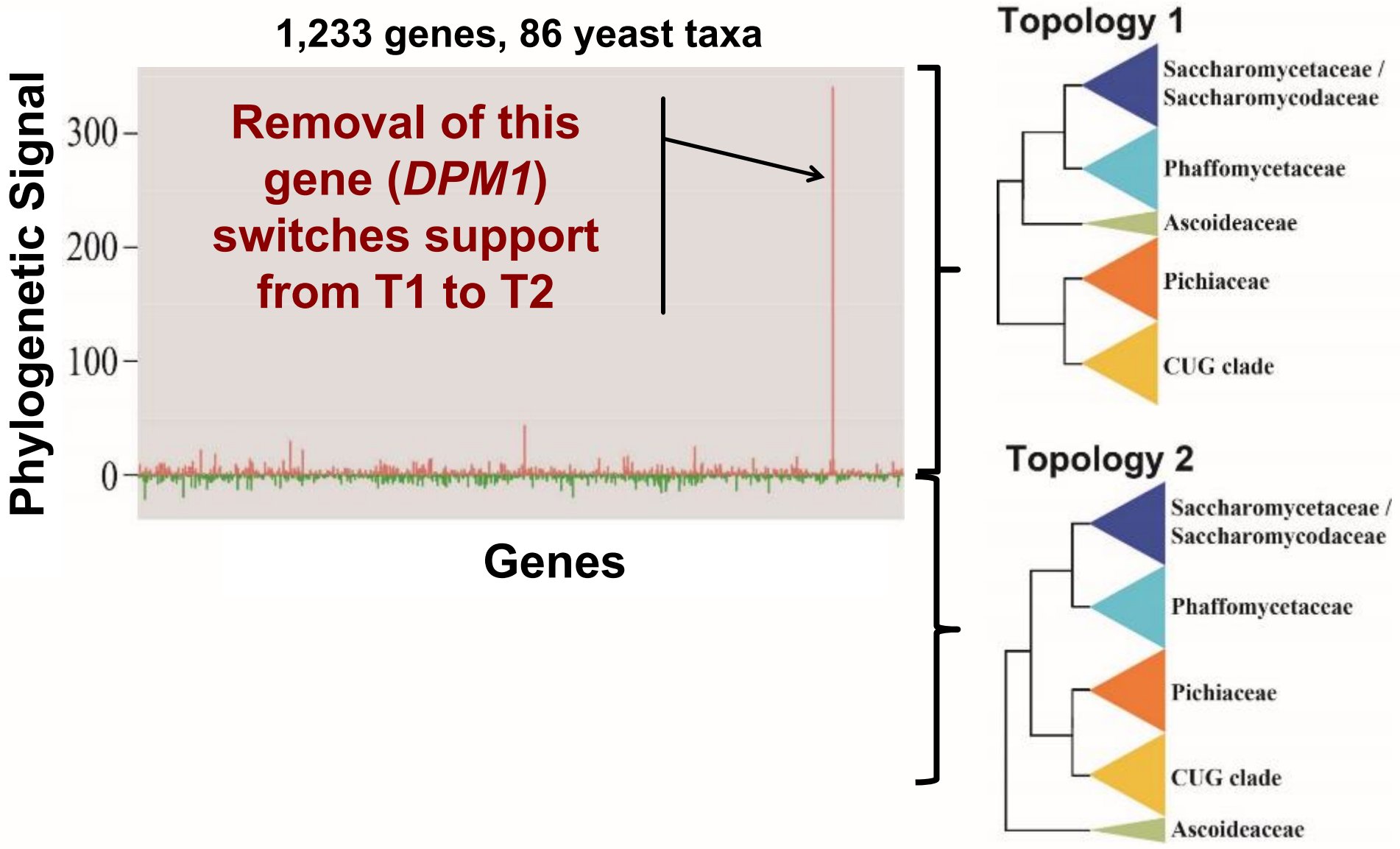

*Shen et al. (2016) G3*

**1,233-gene, 86-taxon data matrix**

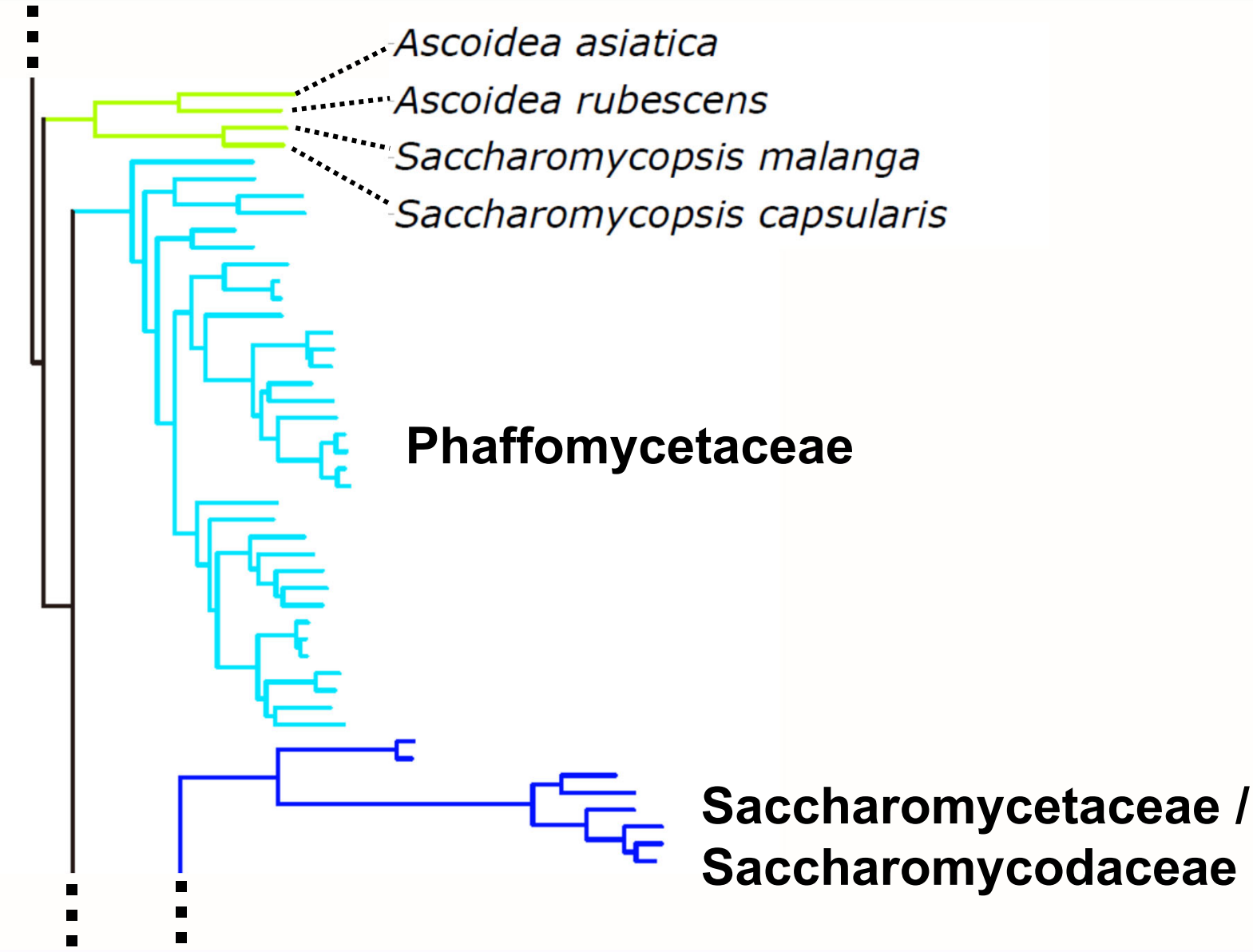
**~13% (11 / 85) of internal branches conflict between analyses**

**Despite increasing # internal branches ~4X, (85 -> 331), conflict decreased**

# *A Single Gene Governs the Placement of* Ascoideaceae



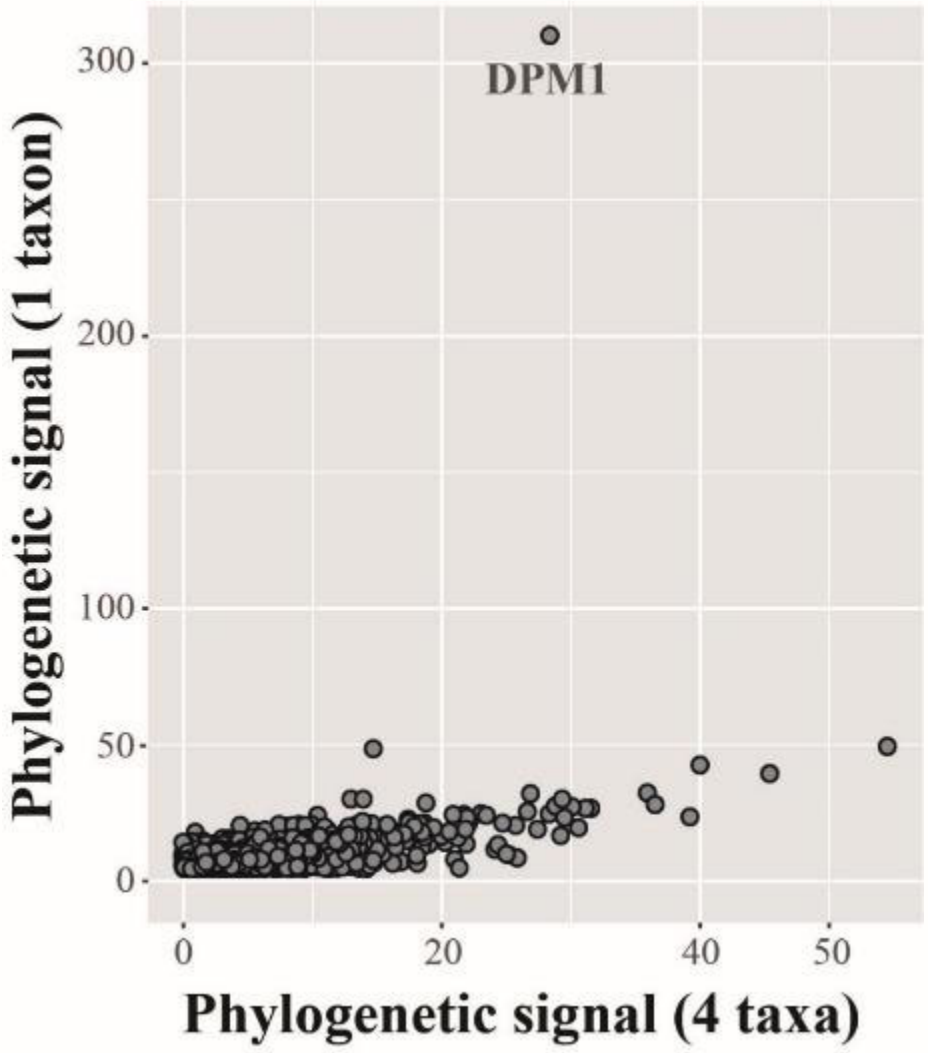
# *Sampling of 3 Additional Taxa "Breaks" the Long Branch*




*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

# *Sampling of 3 Additional Taxa Decreases Gene's Signal*

**2,408 genes, 329 – 332 yeast taxa**

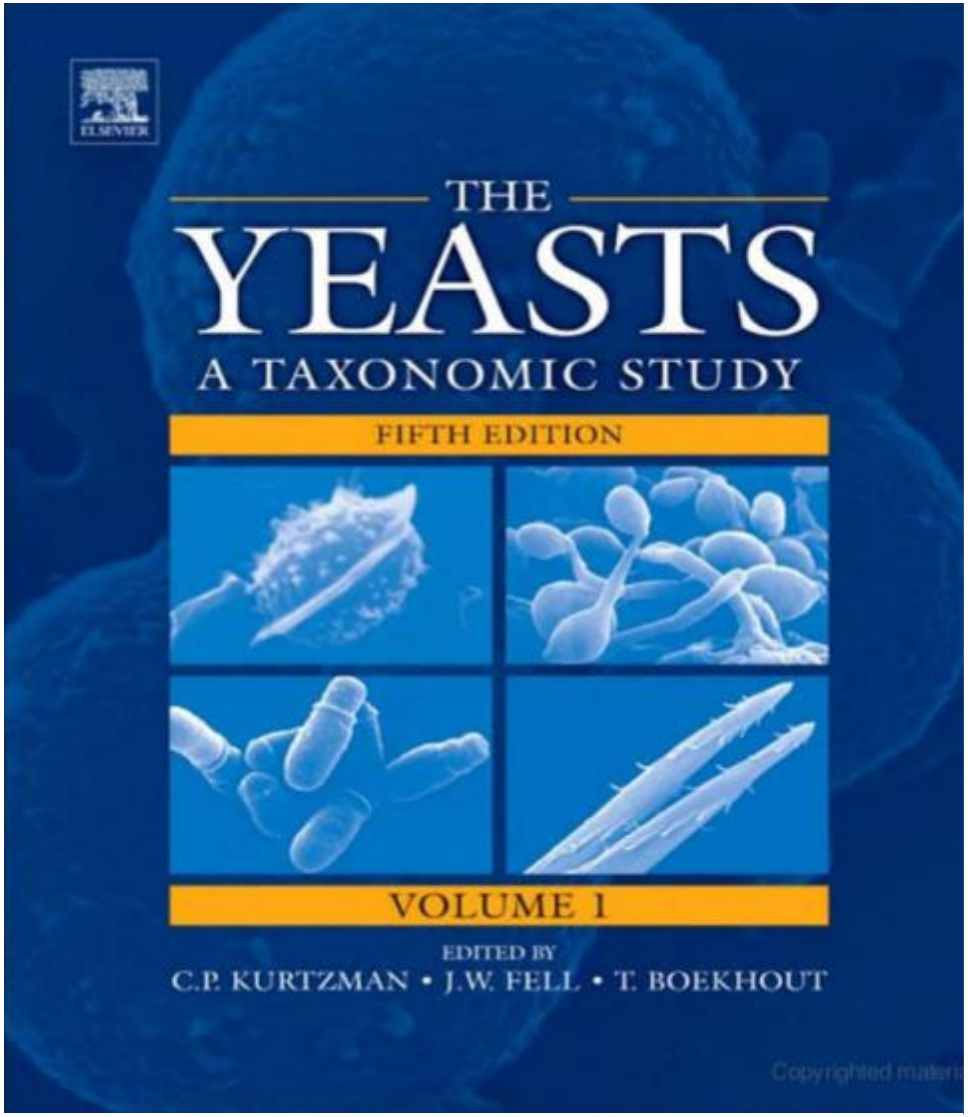



*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

i. Developing pipelines for genomic and phylogenomic data analyses

ii. Inferring a comprehensive & robust genome-scale phylogeny of budding yeasts

**iii. Using the phylogeny to gain insights into budding yeast evolution**




## Fermentation

| | | | |
|---|---|---|---|
| Glucose | + | Lactose | − |
| Galactose | − | Raffinose | − |
| Sucrose | − | Trehalose | − |
| Maltose | − | | |

## Growth (in Liquid Media)

| | | | |
|---|---|---|---|
| Glucose | + | D-Ribose | − |
| Inulin | − | Methanol | − |
| Sucrose | − | Ethanol | − |
| Raffinose | − | Glycerol | − |
| Melibiose | − | Erythritol | − |
| Galactose | − | Ribitol | − |
| Lactose | − | Galactitol | − |
| Trehalose | − | D-Mannitol | − |
| Maltose | − | D-Glucitol | − |
| Melezitose | − | *myo*-Inositol | − |
| Methyl-α-D-glucoside | − | DL-Lactate | − |
| Soluble starch | − | Succinate | − |
| Cellobiose | + | Citrate | − |
| Salicin | + | D-Gluconate | + |
| L-Sorbose | − | D-Glucosamine | − |
| L-Rhamnose | − | *N*-Acetyl-D-glucosamine | n |
| D-Xylose | − | Hexadecane | n |
| L-Arabinose | − | Nitrate | − |
| D-Arabinose | − | Vitamin-free | − |

## Additional Growth Tests and Other Characteristics

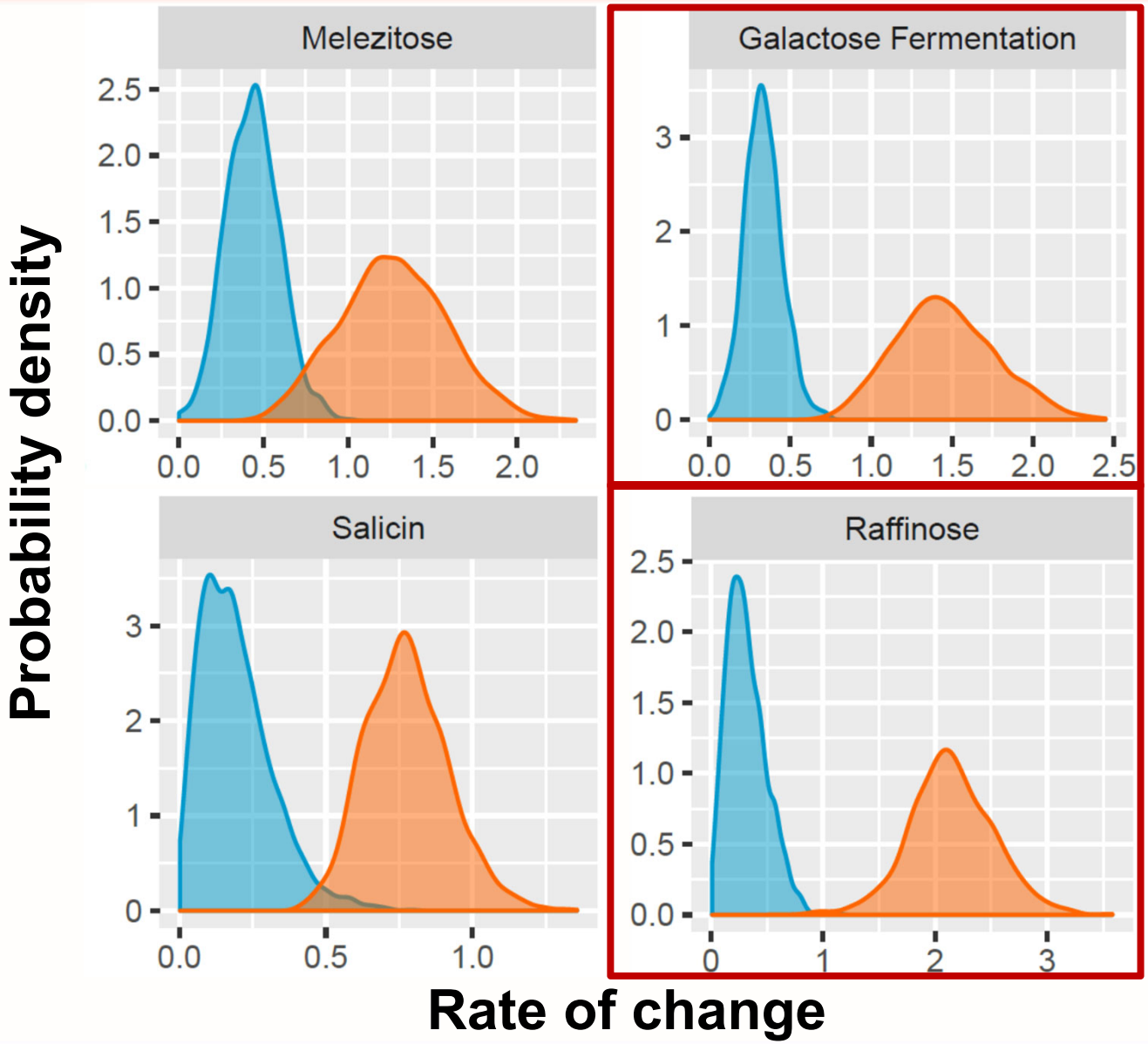
| | | | |
|---|---|---|---|
| 2-Keto-D-gluconate | + | Growth at 25°C | + |
| Cycloheximide 0.01% | + | Growth at 30°C | − |
| Starch formation | − | | |

# *Loss Exceeds Gain for 38 /45 Metabolic Traits*

**Rate of trait gain**

**Rate of trait loss**

**For 17 / 38, the difference is statistically significant**

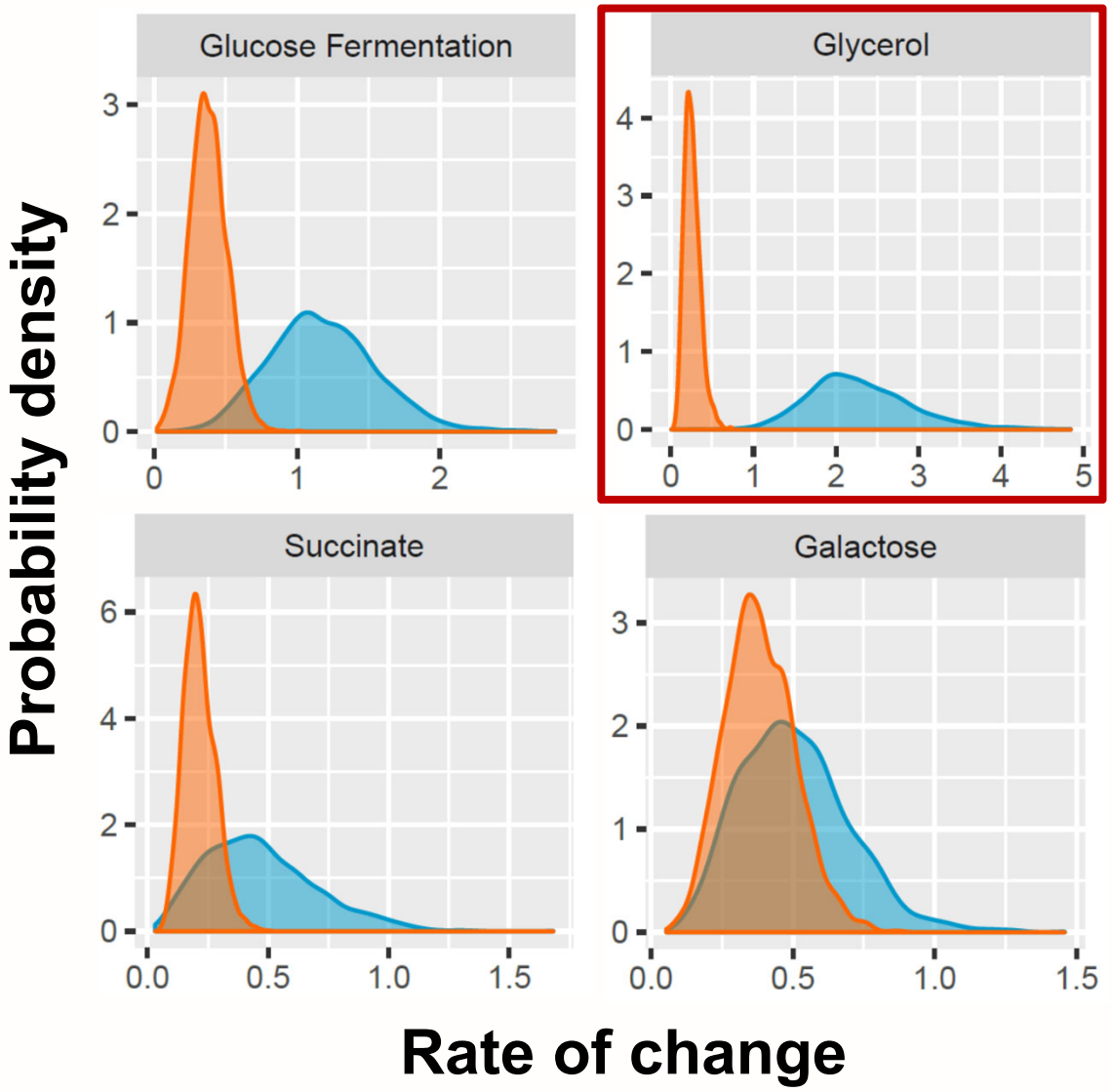


# *Gain Exceeds Loss for the Remaining 7 Traits*

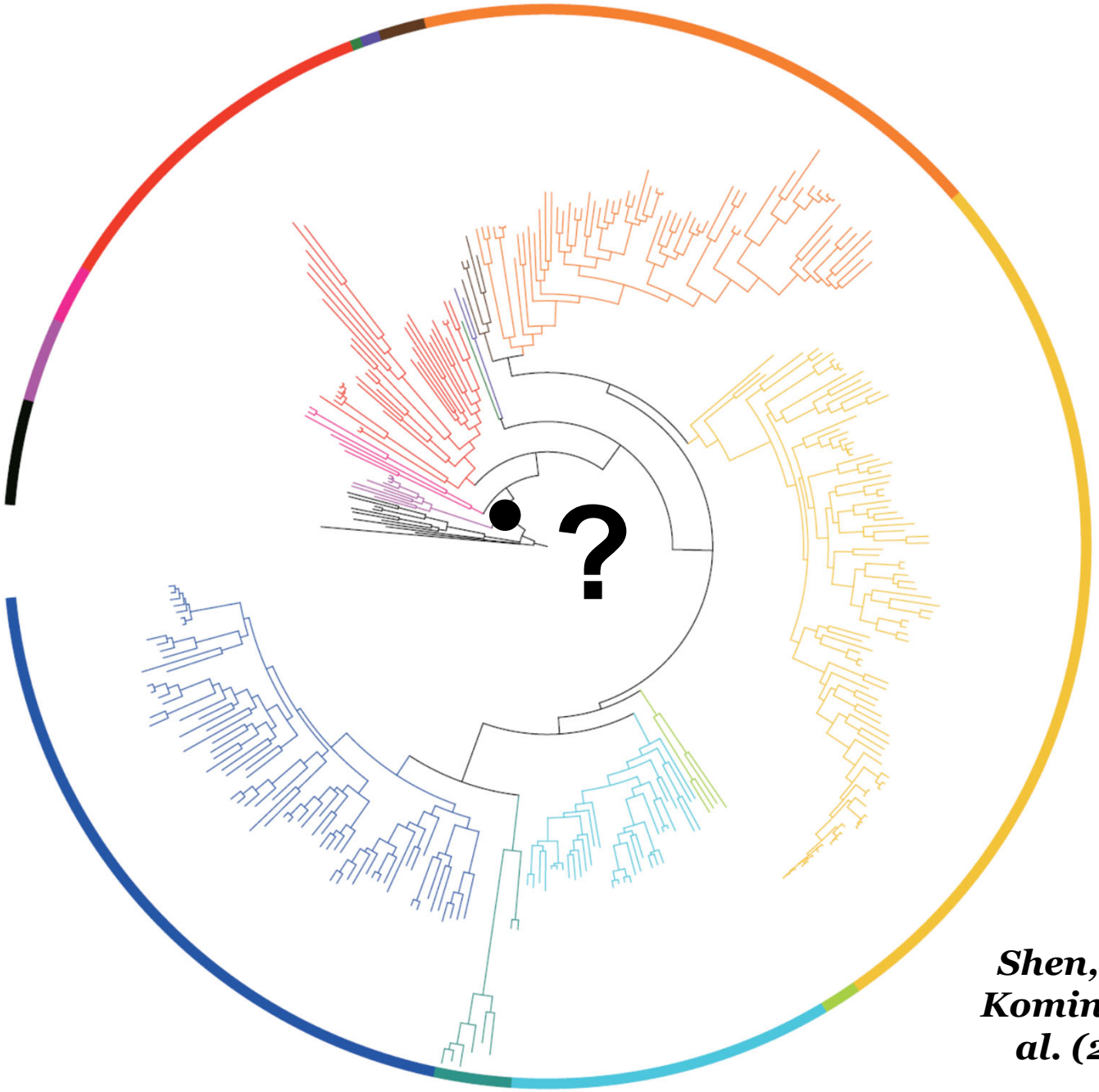
**Rate of trait gain**

**Rate of trait loss**

**For 1 / 7, the difference is statistically significant**



# *Inferring the Metabolic Capabilities of Yeast Ancestors*




***Shen, Opulente, Kominek, Zhou et al. (2018) Cell***

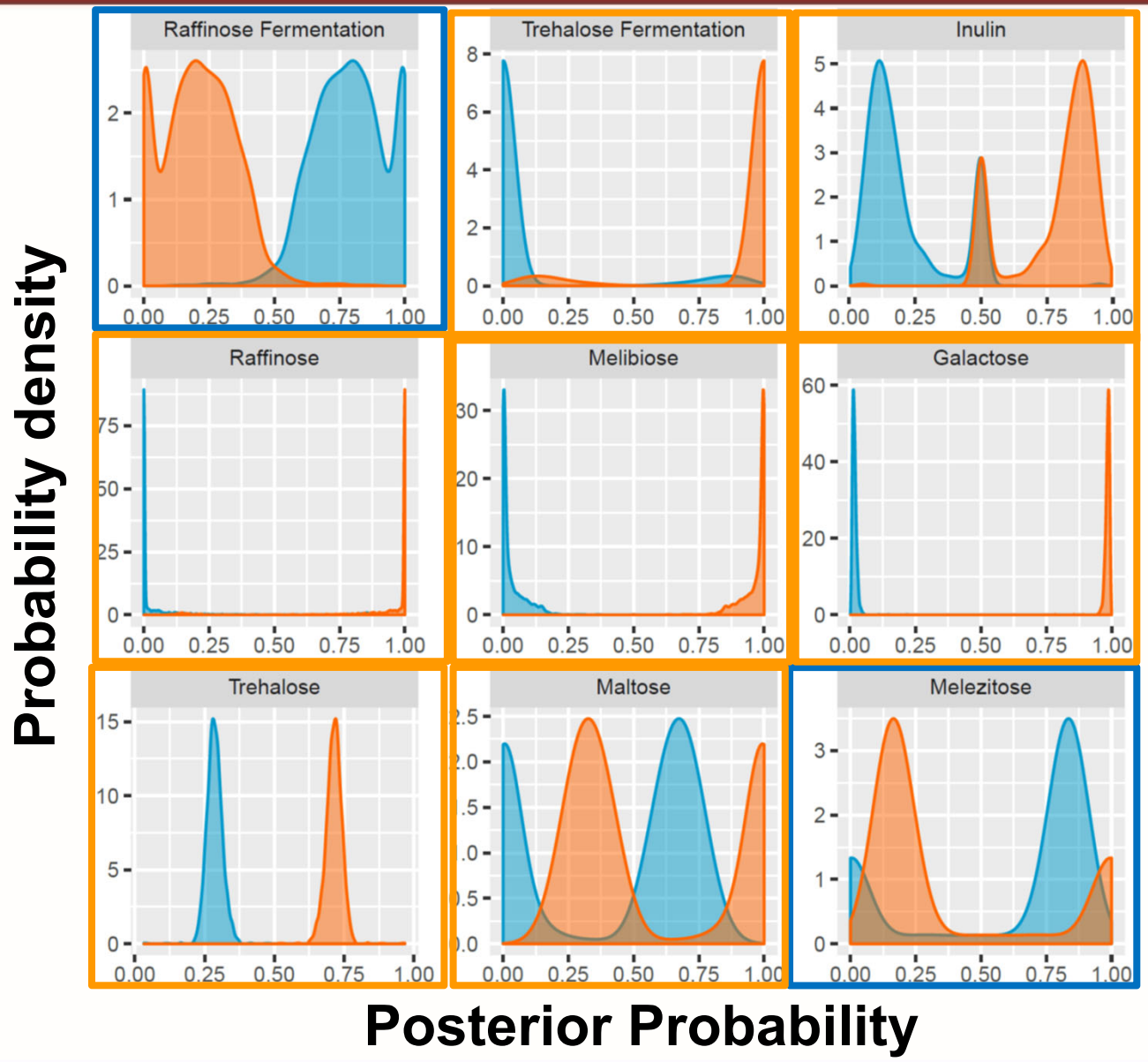
# *BYCA (Budding Yeast Common Ancestor) was a Generalist*

**28 /45 traits likely present in BYCA**

**Prob. of trait presence**

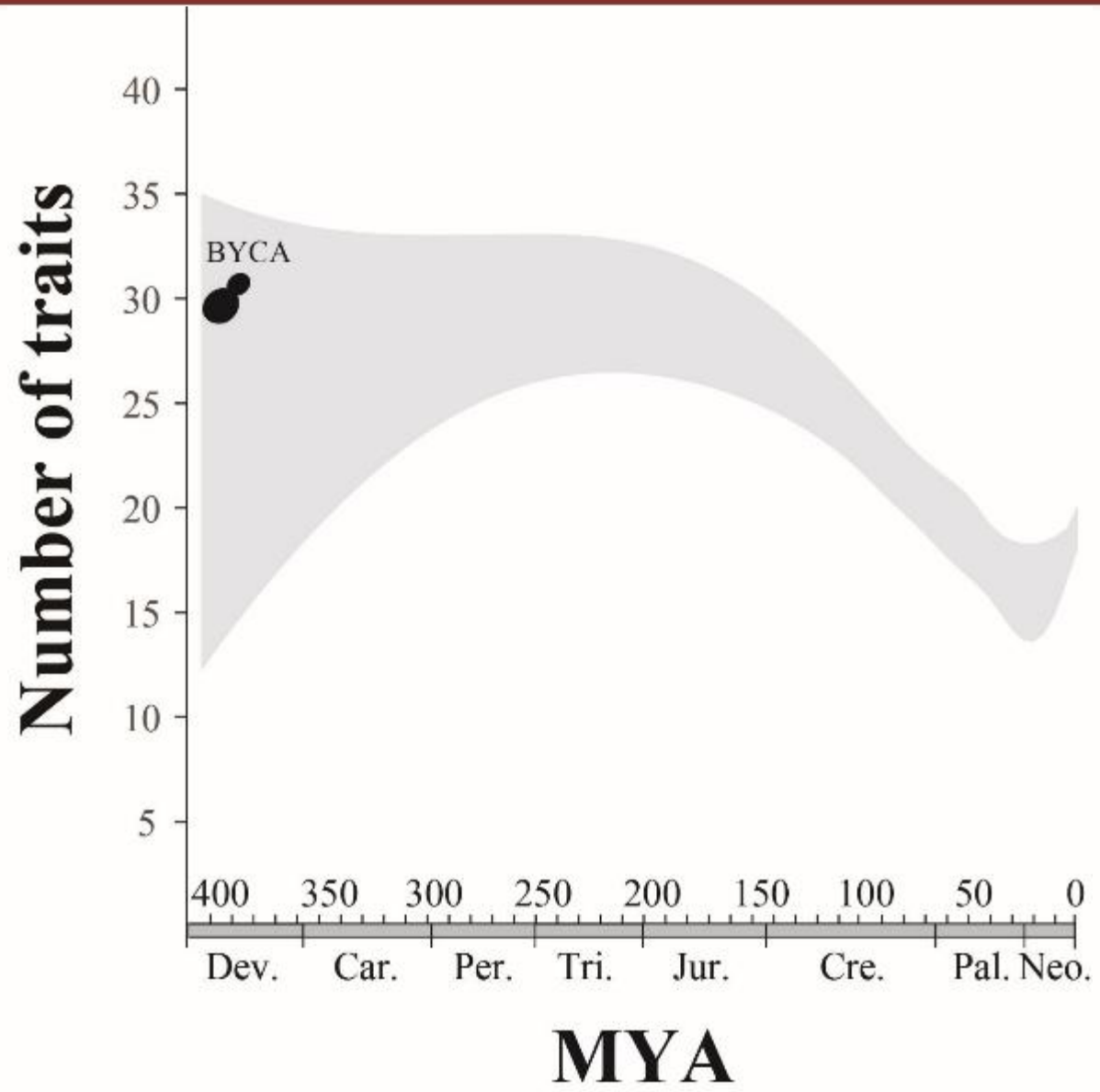
**Prob. of trait absence**

**17 / 45 traits likely absent in BYCA**



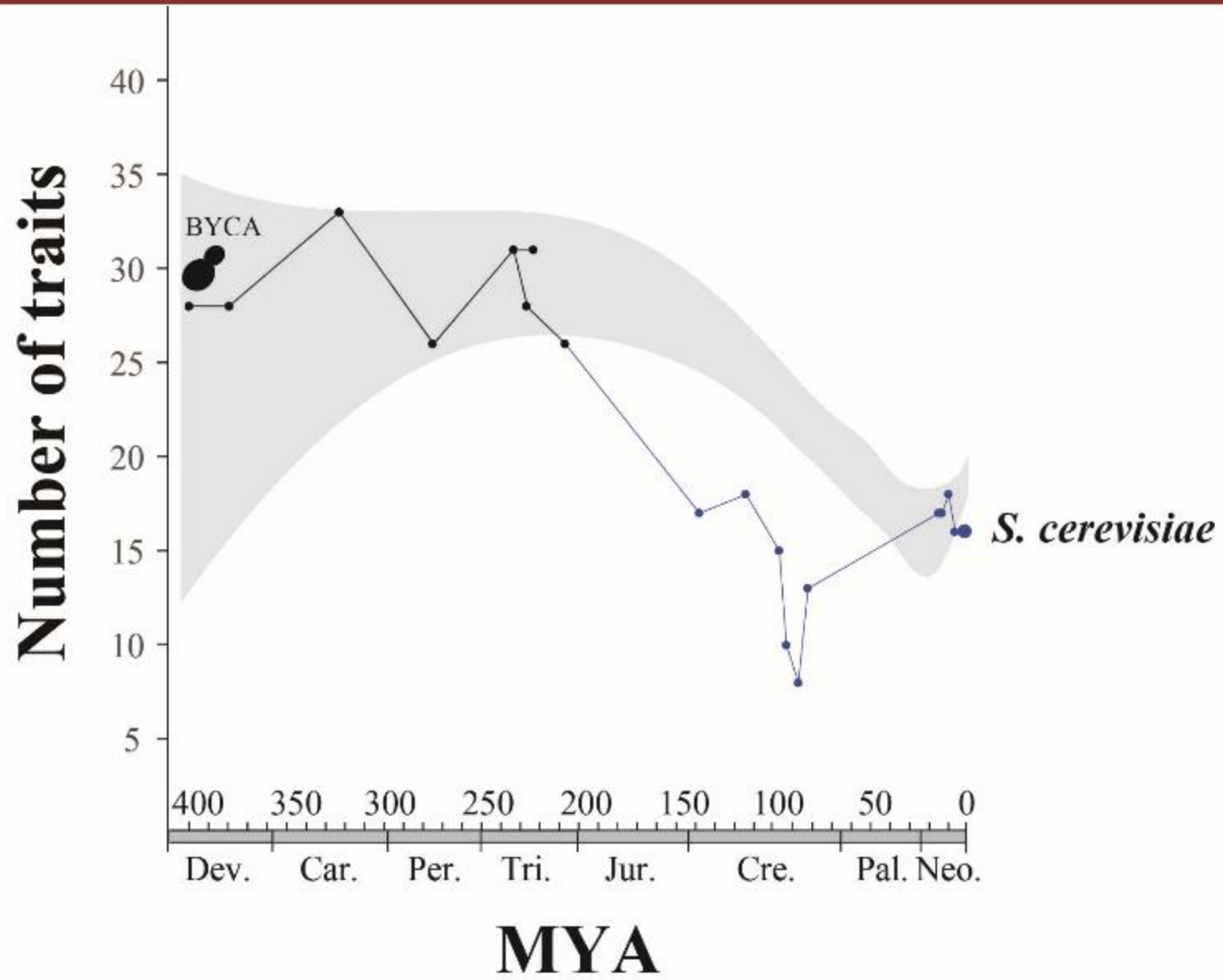

*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

# *Widespread Loss of Traits*



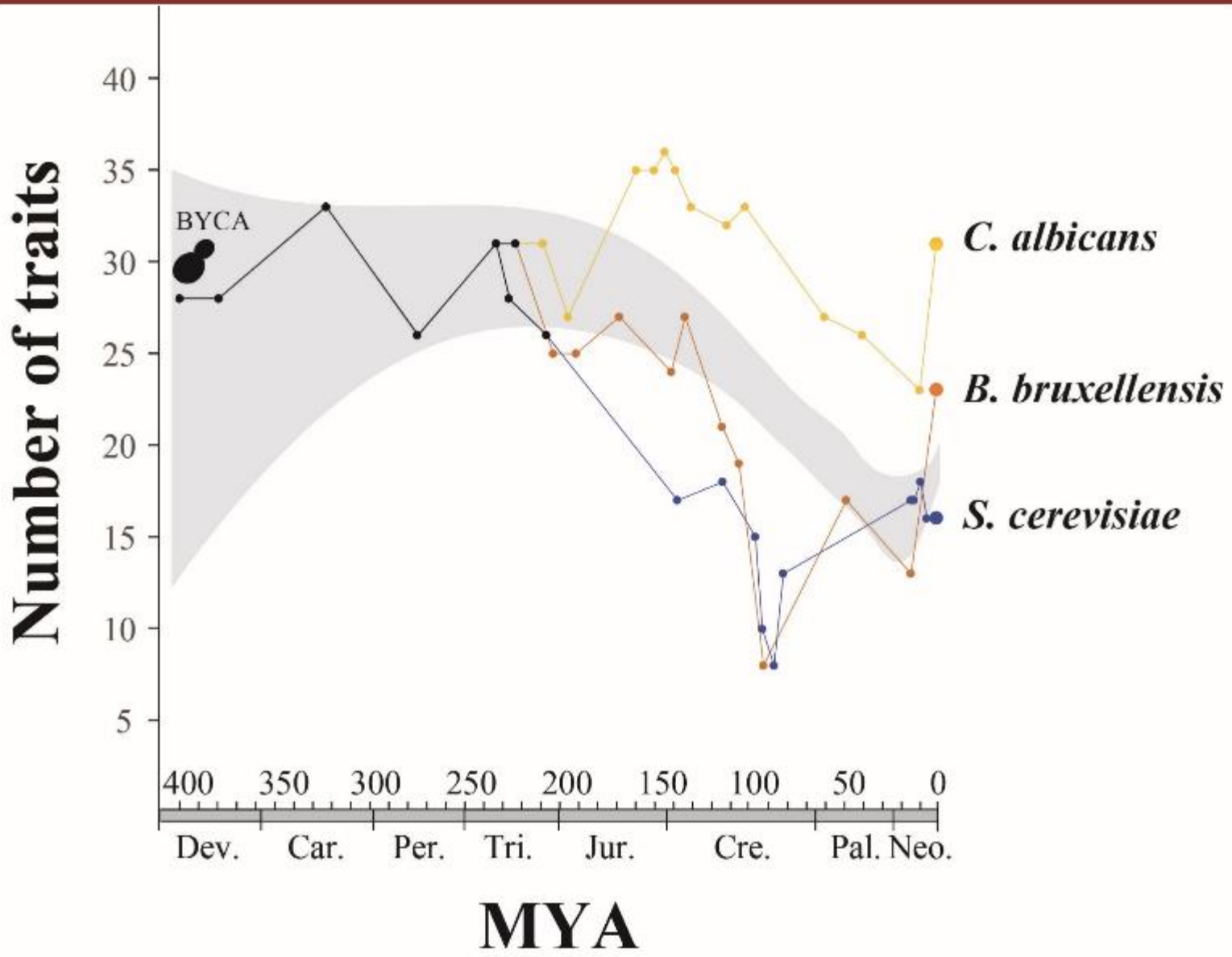

*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

# *Widespread Loss of Traits*



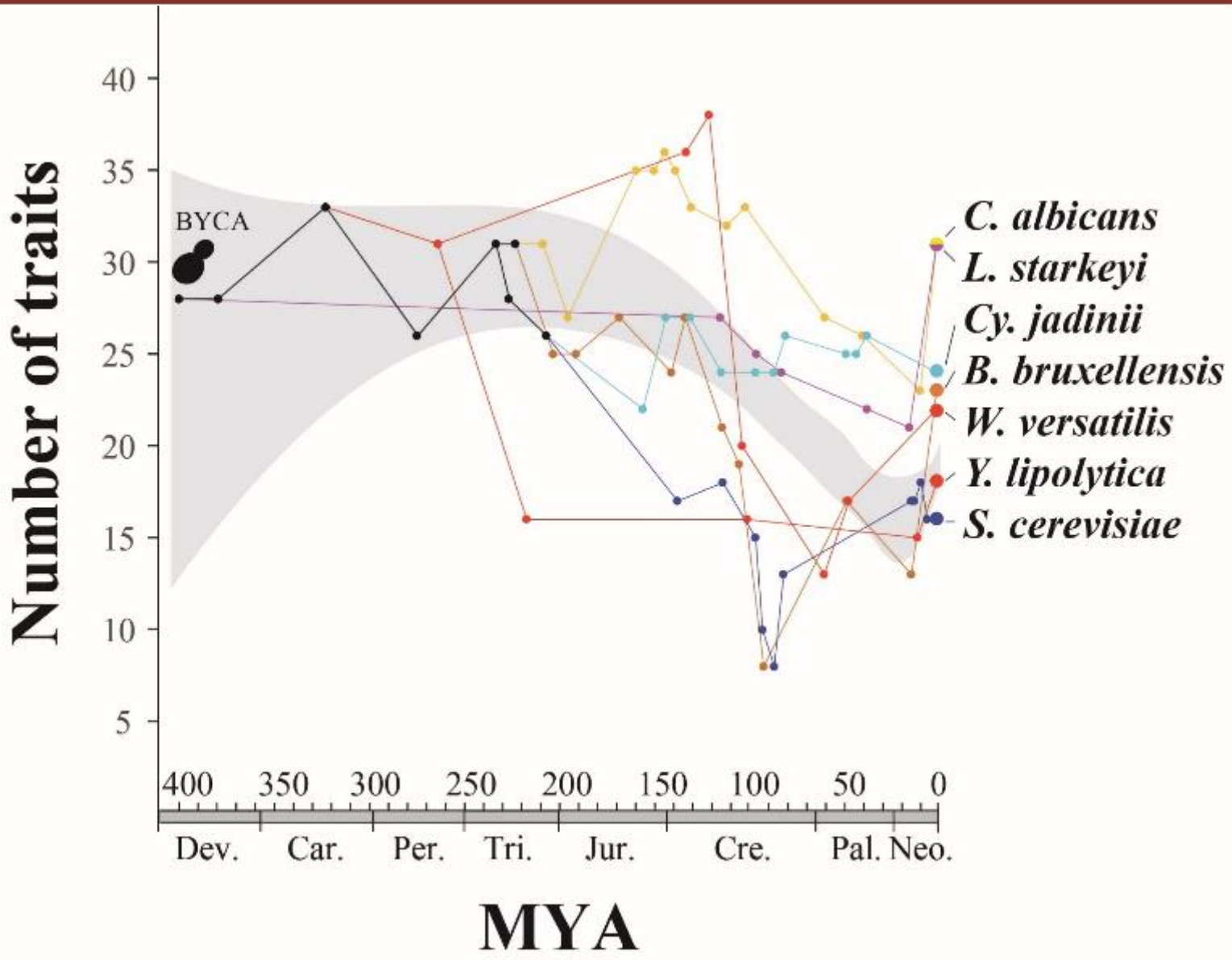

*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

# *Widespread Loss of Traits*




*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

# *Widespread Loss of Traits*



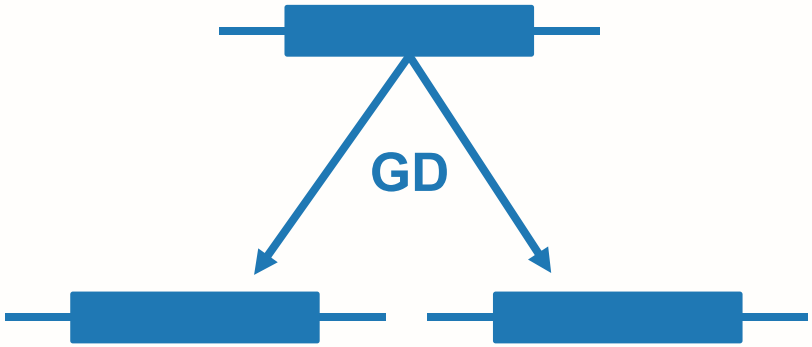

*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

**with so much loss, how did new metabolic traits evolve in the lineage?**

# *Two Major Sources of Gene Innovation*

## Gene duplication (GD)

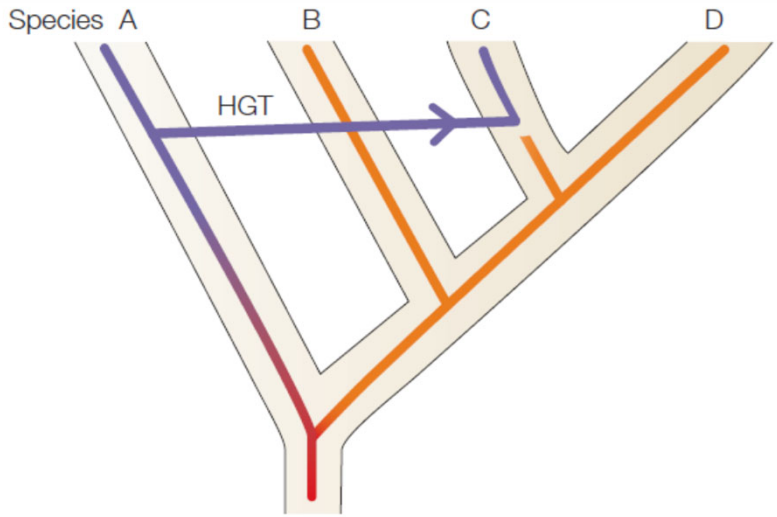
**Any duplication of a region of DNA that contains a gene**




- Plant organic material decay
- Starch catabolism
- Degradation of host tissues
- Toxin production

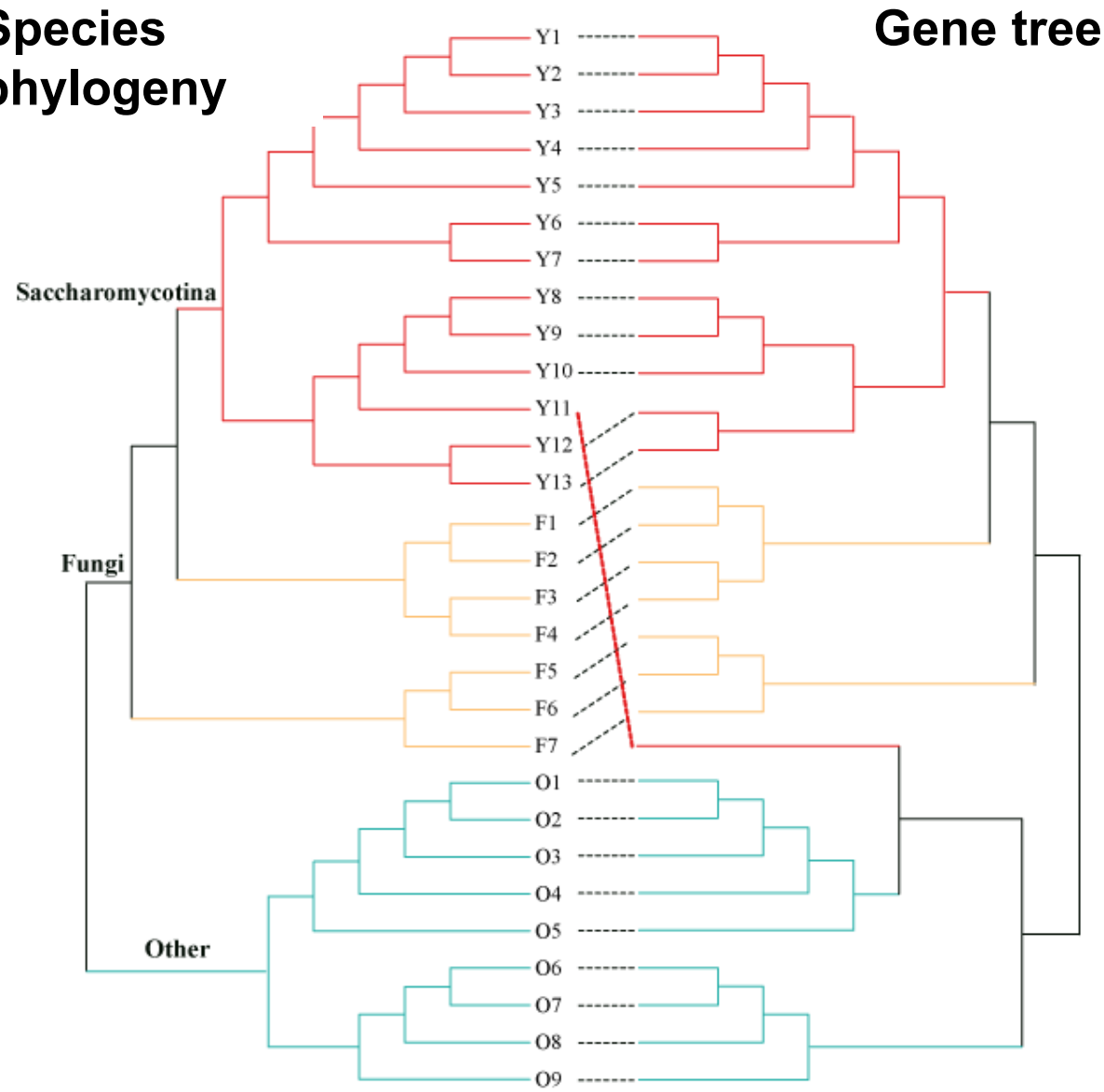
## Horizontal gene transfer (HGT)

**Exchange of genes between organisms other than through reproduction**



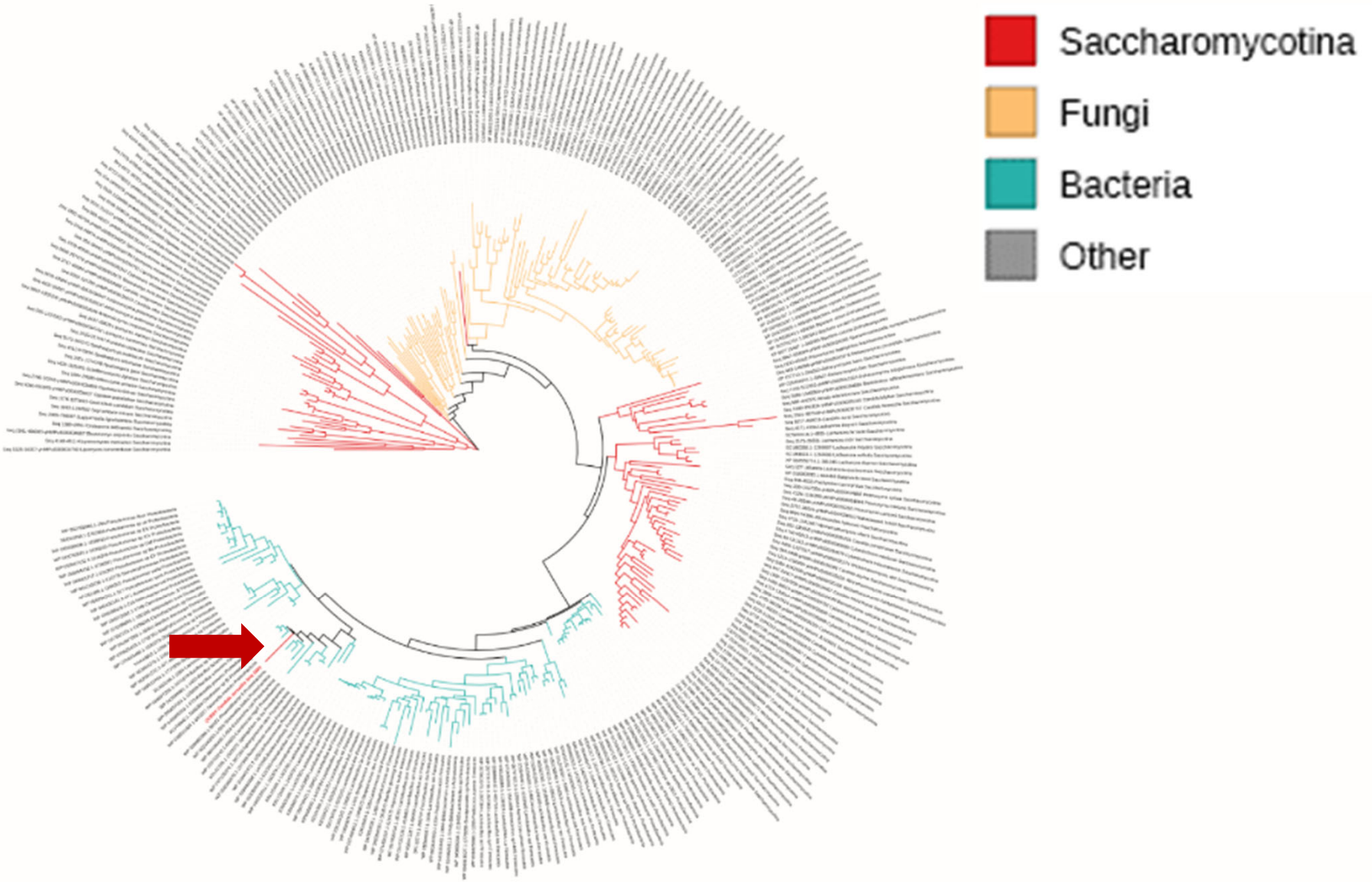

- Xenobiotic catabolism
- Toxin production
- Degradation of plant cell walls
- Wine fermentation

# *How Do We Search for Horizontal Gene Transfer?*



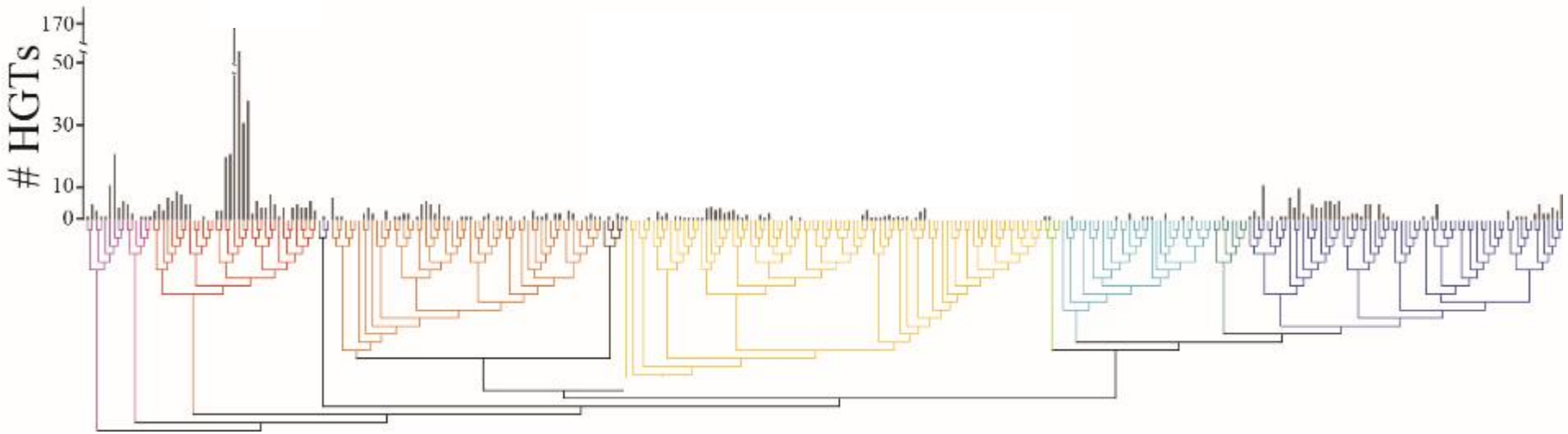

*Fitzpatrick (2011) FEMS Microbiol. Let.*

# *Example Horizontal Gene Transfer*




*Shen, Opulente, Kominek, Zhou et al. (2018) Cell*

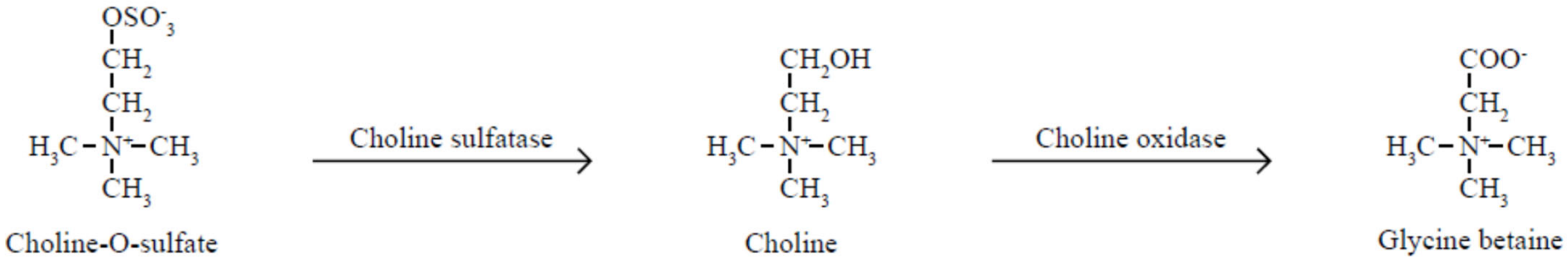
# *Distribution of 878 Horizontally Acquired Genes*

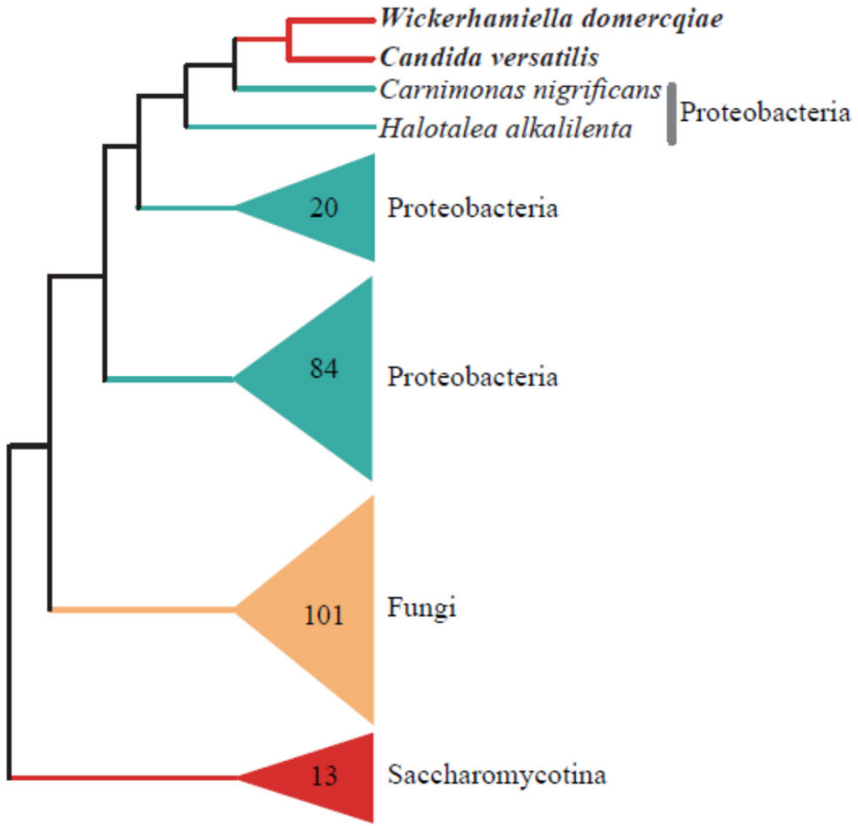


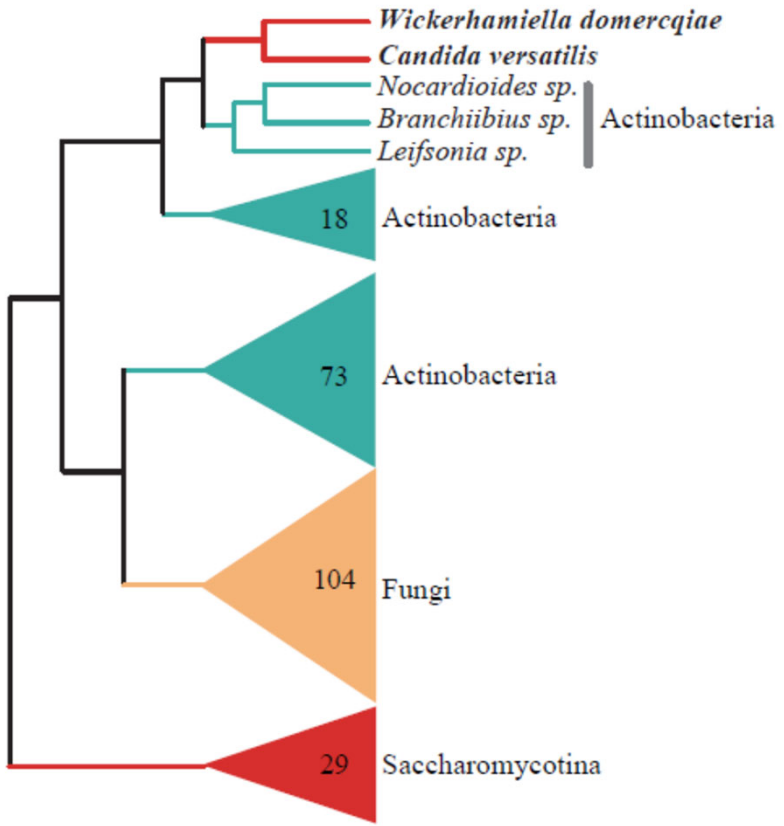

**HGT in 226 yeasts with universal code: 0.071%**

**HGT in 103 yeasts with non-universal code: 0.025%**

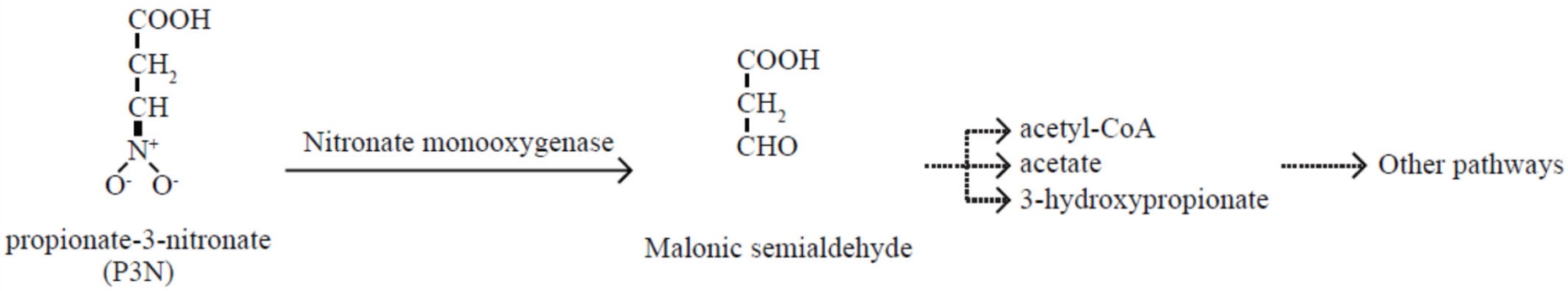
# *Osmotolerant Yeasts Acquired Enzymes for Osmoprotectant Betaine*



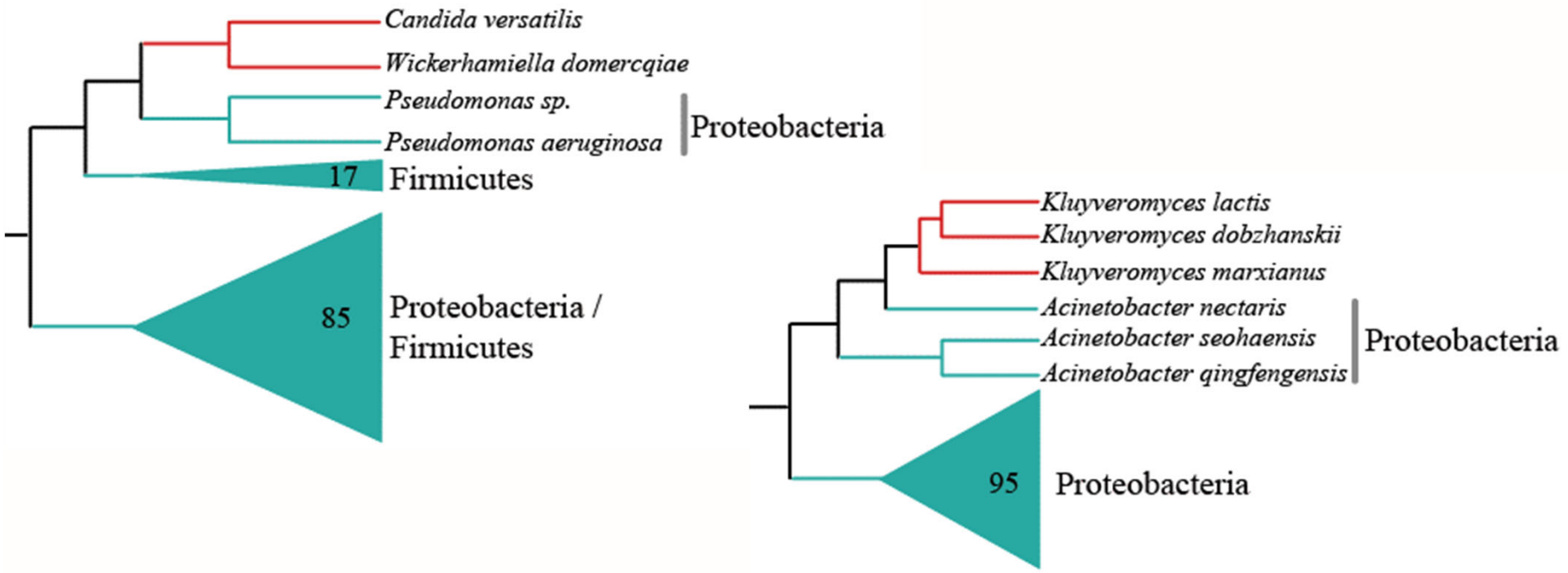
# *Two Yeast Lineages Independently Acquired P3N Detoxifier Genes*




propionate-3-nitronate (P3N)

Malonic semialdehyde

**plant and fungal toxin**



# *Loss is More*

**Specialization is often considered an evolutionary dead end**

**Starting 400 mya, a metabolically complex budding yeast gave (and keeps on giving) rise to organisms with highly diverse metabolisms & ecologies**

- **Differential pathway loss**
- **Gene duplication and functional divergence**
- **HGT**

# *Acknowledgments*


**Hittinger lab – UW**

Jacek Kominek

Dana Opulente

And many others

**Kurtzman lab – USDA**

Jeremy DeVirgilio

http://y100plus.org

**Rokas lab – Vandy**

Xing-Xing Shen

Xiaofan Zhou

Jacob Steenwyk

Leonidas Salichos

And many others

**Budding Yeast Community**

Moriya Ohkuma, Rikiya Endoh, Masako Takashima, Ri-ichiroh Manabe, Neža Čadež, Diego Libkind, Carlos A. Rosa, Marizeth Groenewald