

Lecture Outline

- ❖ Introduction to evolutionary genomics
- ❖ Brief review of next-generation DNA sequencing & its applications
- ❖ Benchmarking transcriptome sequencing for functional and evolutionary genomics
- ❖ Transcriptome sequencing for molecular phylogenetics

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

- ❖ **Genomics within species (using next-generation DNA sequencing)**
- ❖ **Genomics between species (using publicly available genome data)**



Fungi: a Model for Comparative Functional Genomics

Medicine & Industry

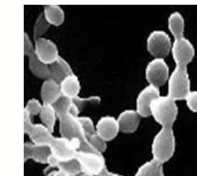
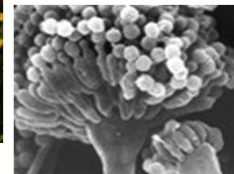
- ❖ Candidiasis (yeast infections)
- ❖ Aspergillosis
- ❖ Antibiotics (e.g., penicillin, cyclosporin)
- ❖ Food (e.g., beer, wine, cheese)

Model Organisms

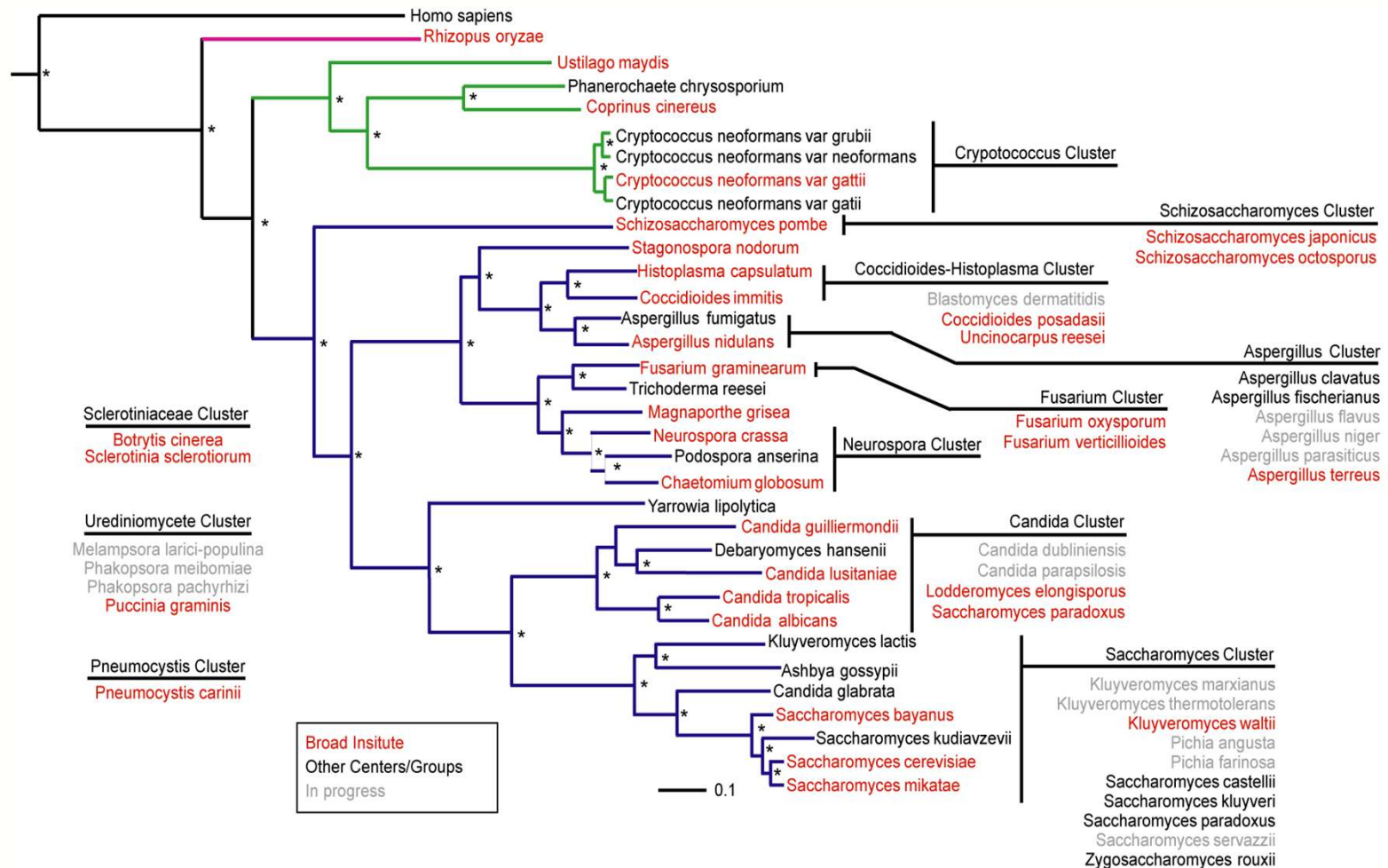
- ❖ Small genome size & high gene density
- ❖ Tractable systems for molecular and genomic studies

Evolution & Ecology

- ❖ Rich ecology and evolutionary history
- ❖ Vital in function of ecosystems (as decomposers)

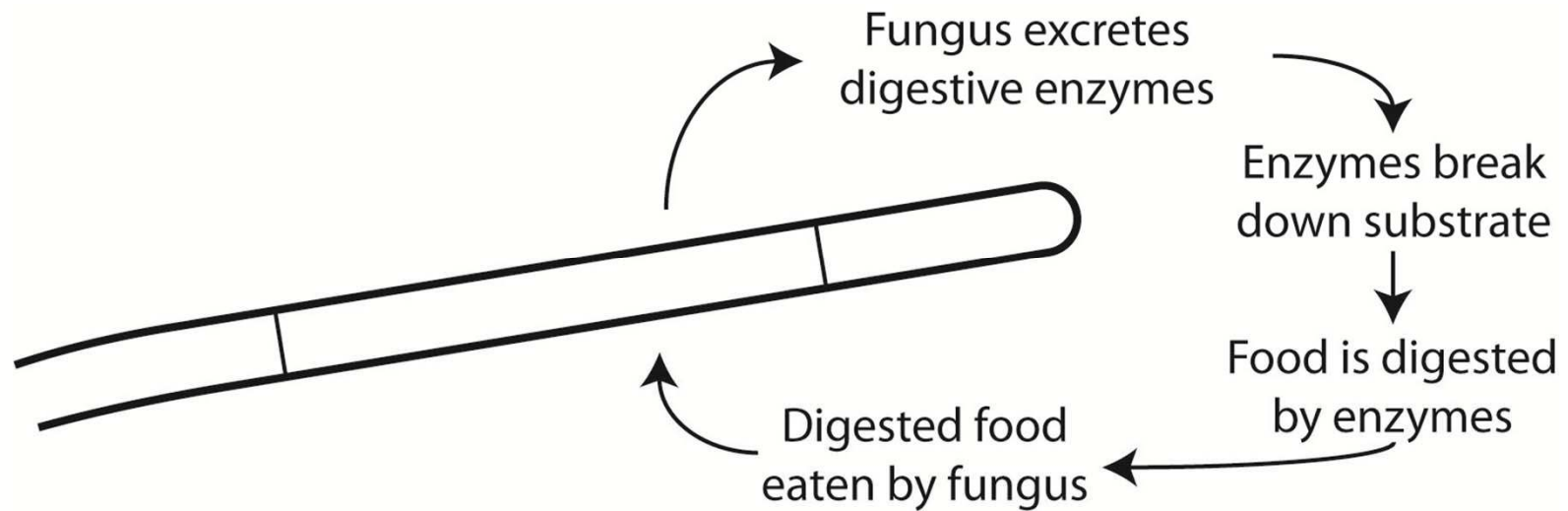


Fungi: the Eukaryotic Clade with the Most Genomes



Galagan et al. (2005) Genome Res.

Fungi Eat by Absorption



- ❖ Different fungi secrete different enzymes and can only “eat” certain materials
- ❖ Fungi exhibit remarkable versatility on what materials they can “eat”: leather, cloth, wood, manure, animal carcasses, live hosts, ink, syrup, paint, glue, hair, etc.



Fungi Eat by Absorption

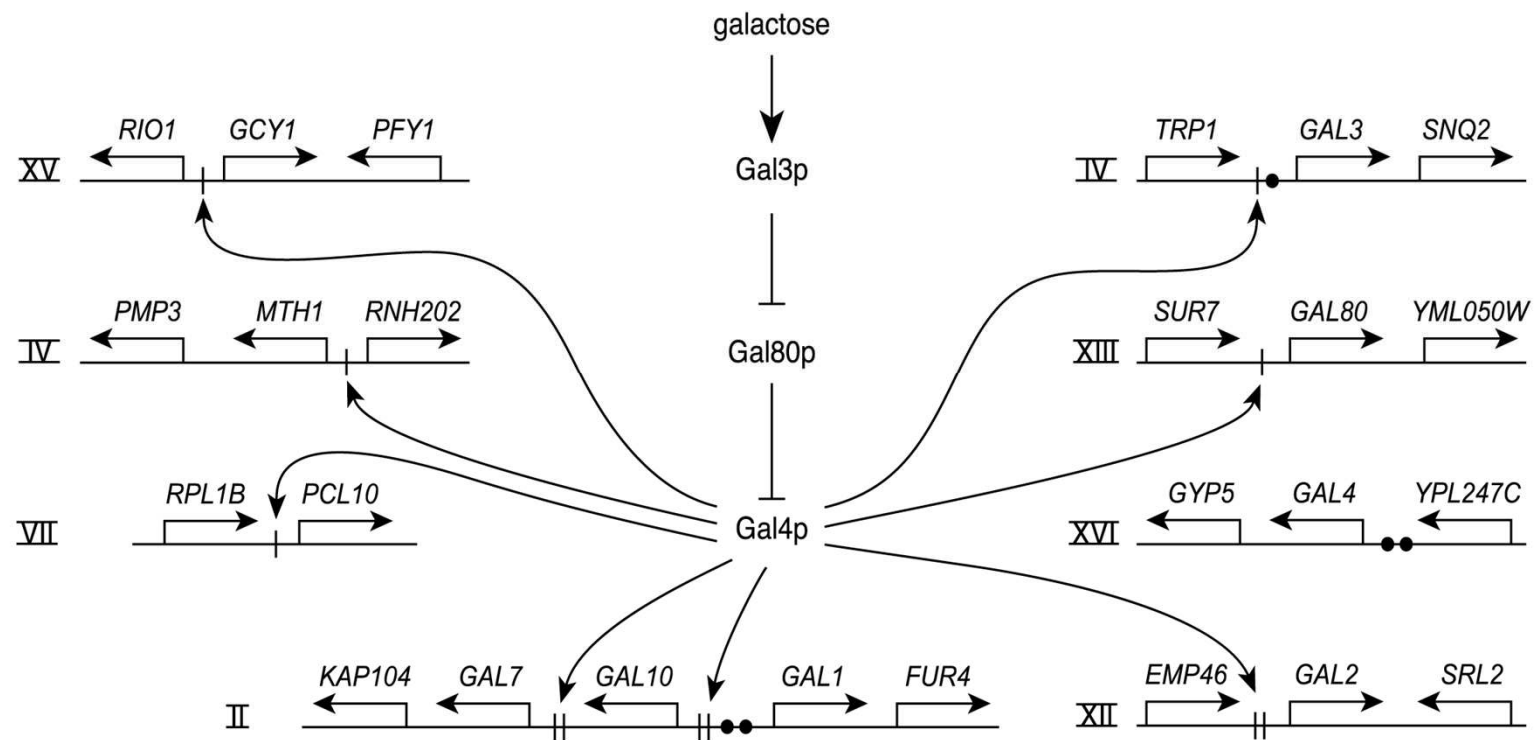


Diner fries

McDonald's fries



The GAL Pathway in the Baker's Yeast *S. cerevisiae*



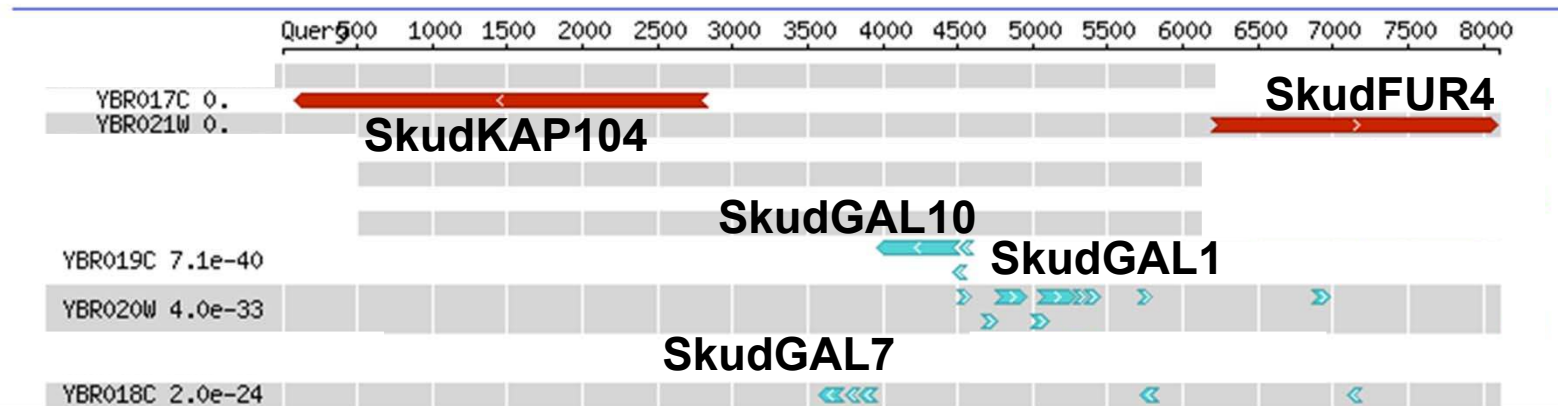
S. kudriavzevii GAL Genes are Syntenic Pseudogenes

Skud: 5026 SPSNFK-L*SYTRDQNMPTQSLREF*NGL*MTFTVKFSNDEEFFEQFGALKNES*VSCDK 5202
 SP F+ L Y R +++ ++SLR + + T F+ DE+FF+QFGAL NES SCDK

Scer: 381 SPVRFQVLKLYQRAKHVYSESLRVL-KAVKLMTTASFTADEDEFFKQFGALMNESQASCDK 439

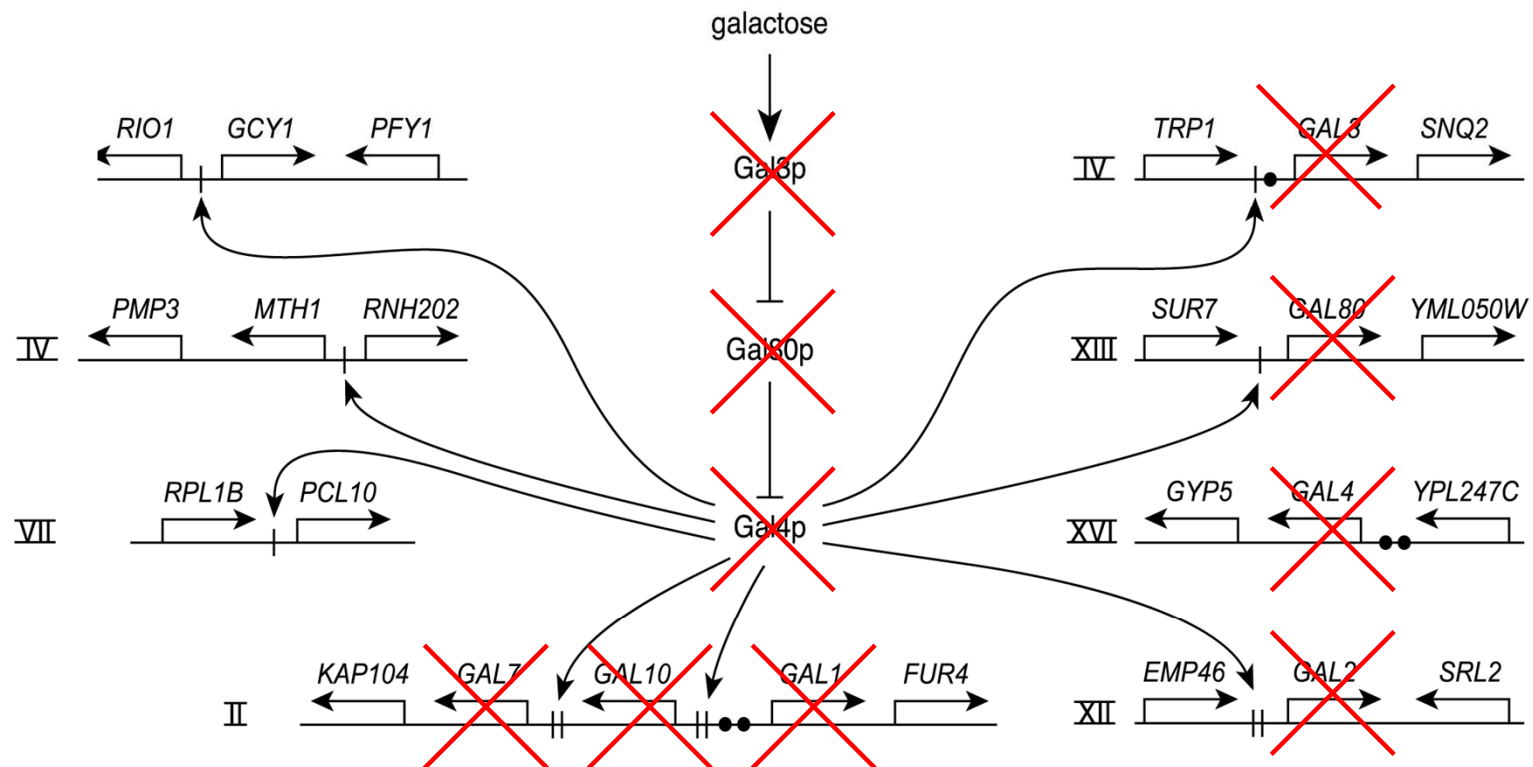
Skud: 5203 LYGCSCAETDTVCSIALLSGSYNSRL 5280
 LY CSC E D +CSIAL +GSY SRL

Scer: 440 LYECSCPEIDKICSIALSNGSYGSRL 465



Hittinger, Rokas & Carroll (2004) PNAS

The Entire GAL Pathway is Degenerate



Hittinger, Rokas & Carroll (2004) PNAS

But S. kudriavzevii is More Diverse!

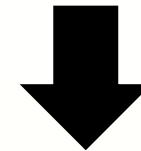
Strain	Nationality
IFO1802	Japan
IFO10990	Japan
IFO10991	Japan
IFO1803	Japan
ZP513	Portugal
ZP537	Portugal
ZP542	Portugal
ZP591	Portugal
ZP594	Portugal
ZP595	Portugal
ZP620	Portugal
ZP621	Portugal
ZP623	Portugal
ZP625	Portugal
ZP627	Portugal
ZP629	Portugal
ZP630	Portugal
ZP634	Portugal

**2004: 4
Japanese
GAL-
strains**

**2007: 14
Portuguese
GAL+
strains
discovered**



**Illumina
sequencing**

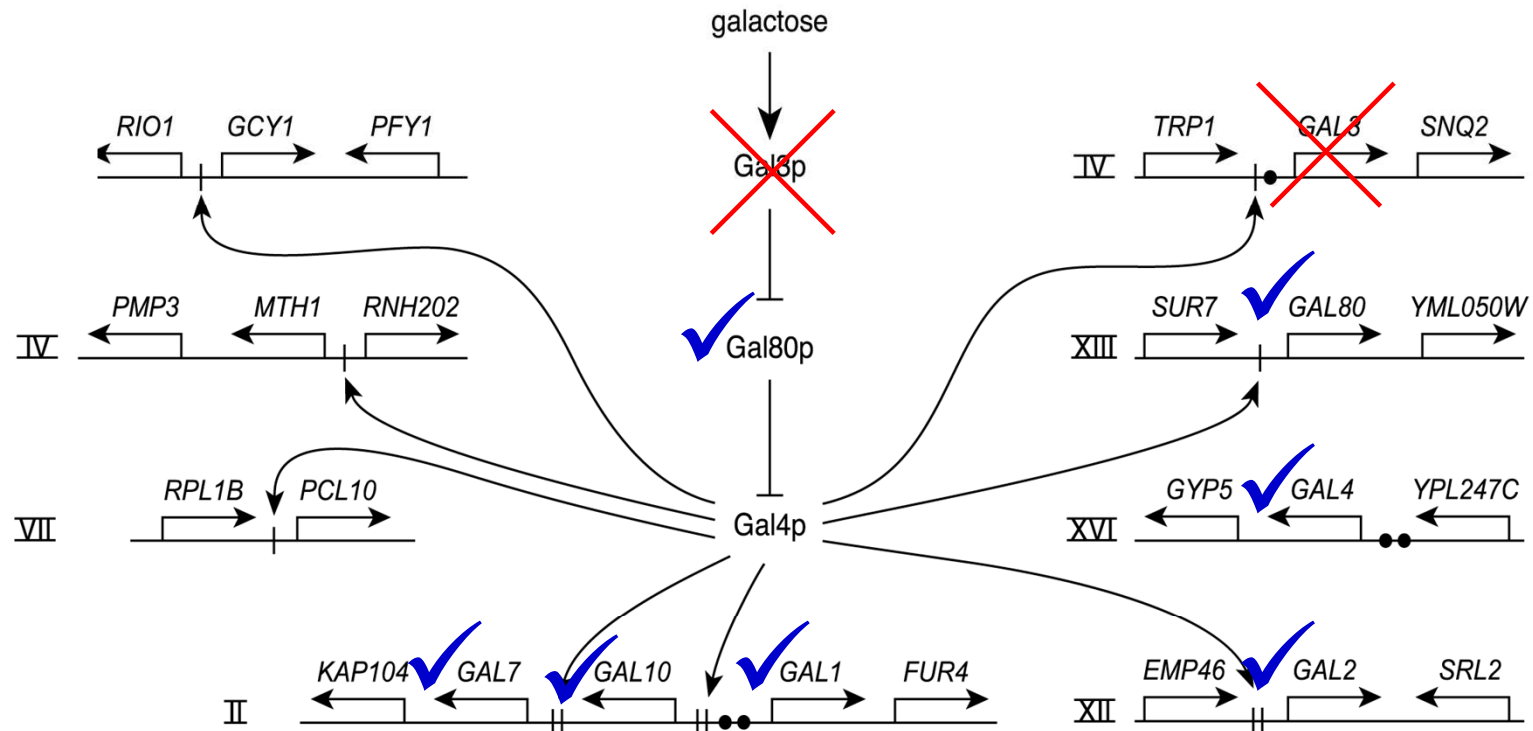


**80,000,000 reads
2,800,000,000 bp**



Sampaio & Gonçalves (2008) Appl. Environ. Microbiol.

The Reduced GAL⁺ Network of *S. kudriavzevii*



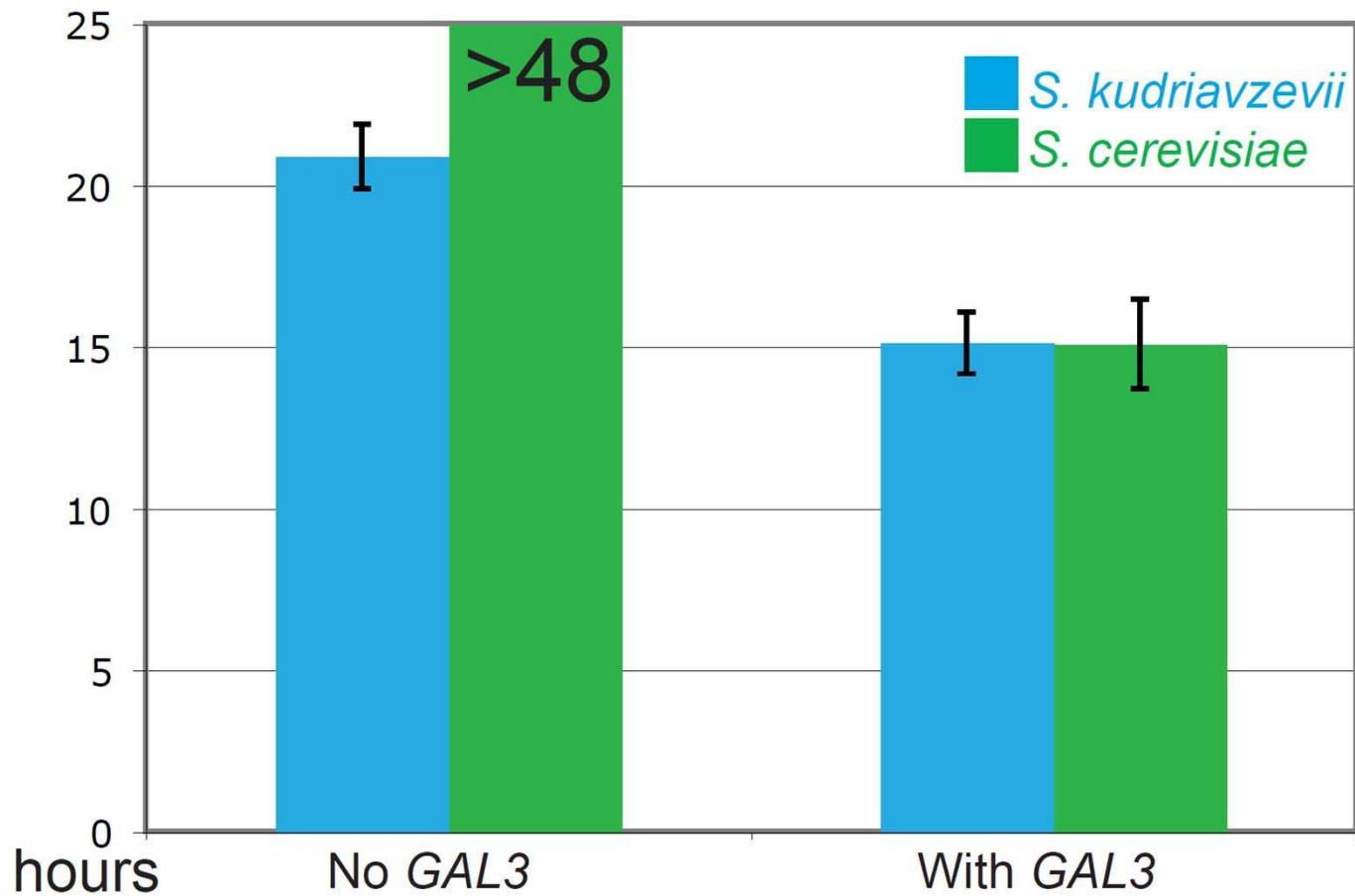
How can we explain such striking polymorphism in a multi-locus gene network?



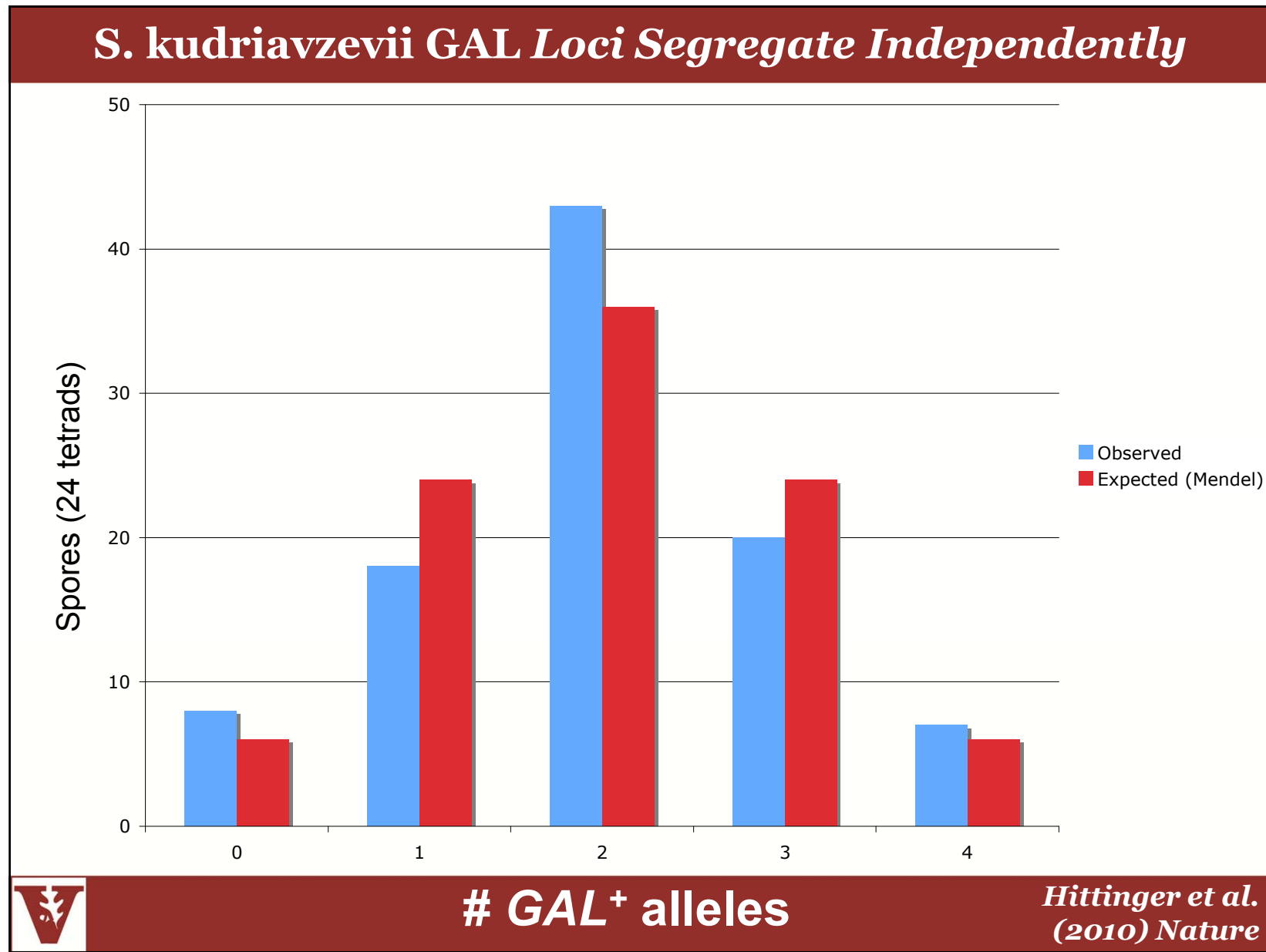
de novo assembly from ~30x Illumina of
ZP591; confirmed by Sanger

Hittinger et al. (2010) Nature

The *S. kudriavzevii* GAL⁺ Network Can Function Without GAL3



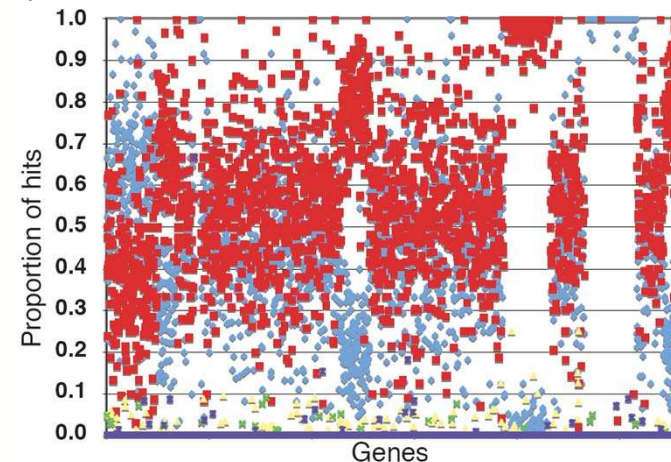
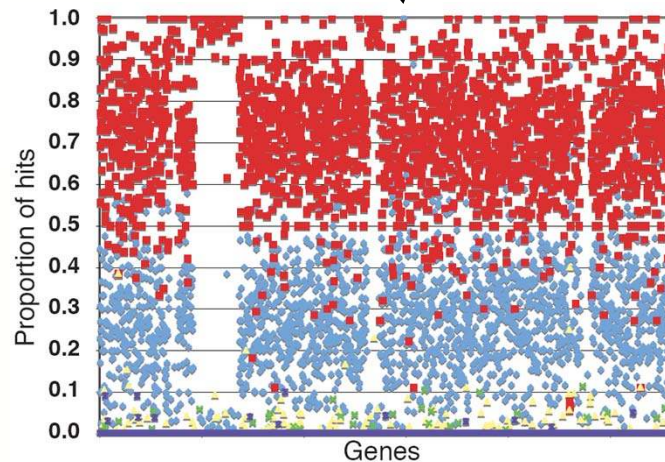
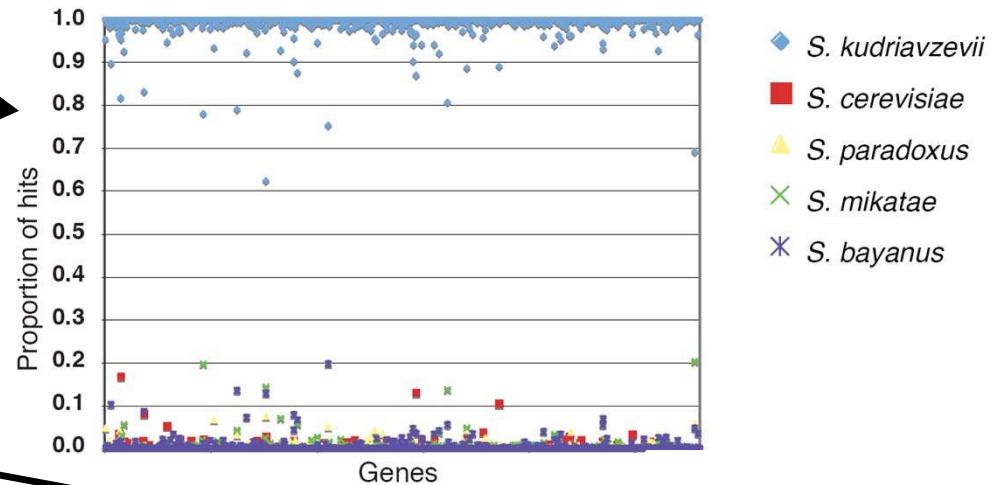
Hittinger et al. (2010) Nature



The GAL⁺ Isolates of *S. kudriavzevii* are Not Hybrids

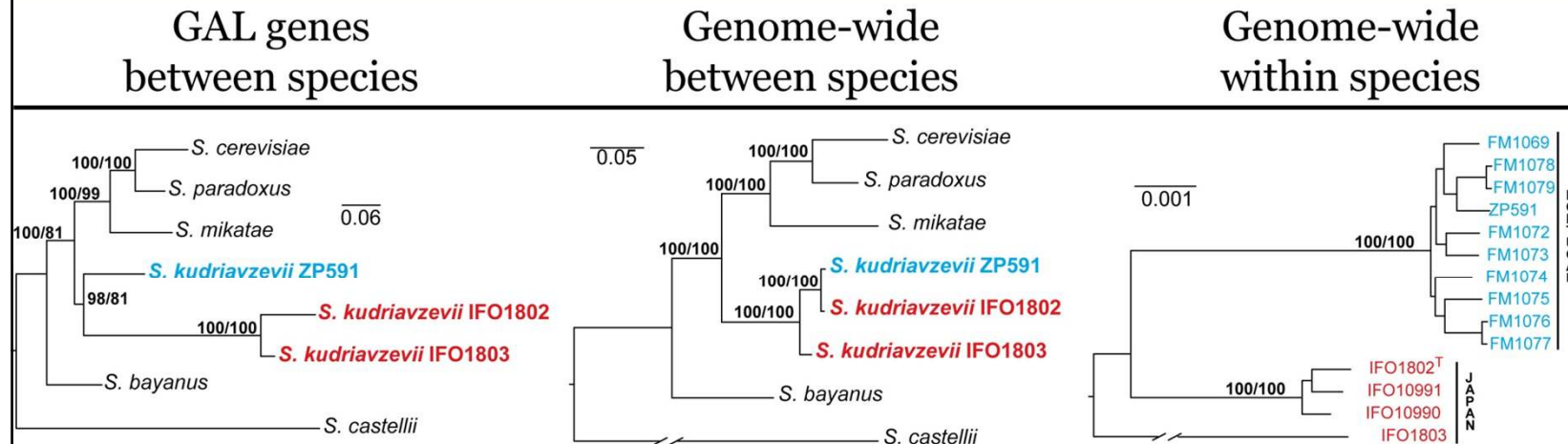
Portuguese
S. kudriavzevii

Known and wild-caught
S. cerevisiae X *S. kudriavzevii*
hybrids



Hittinger et al. (2010) Nature

The Two GAL Network States Evolved Within *S. kudriavzevii*



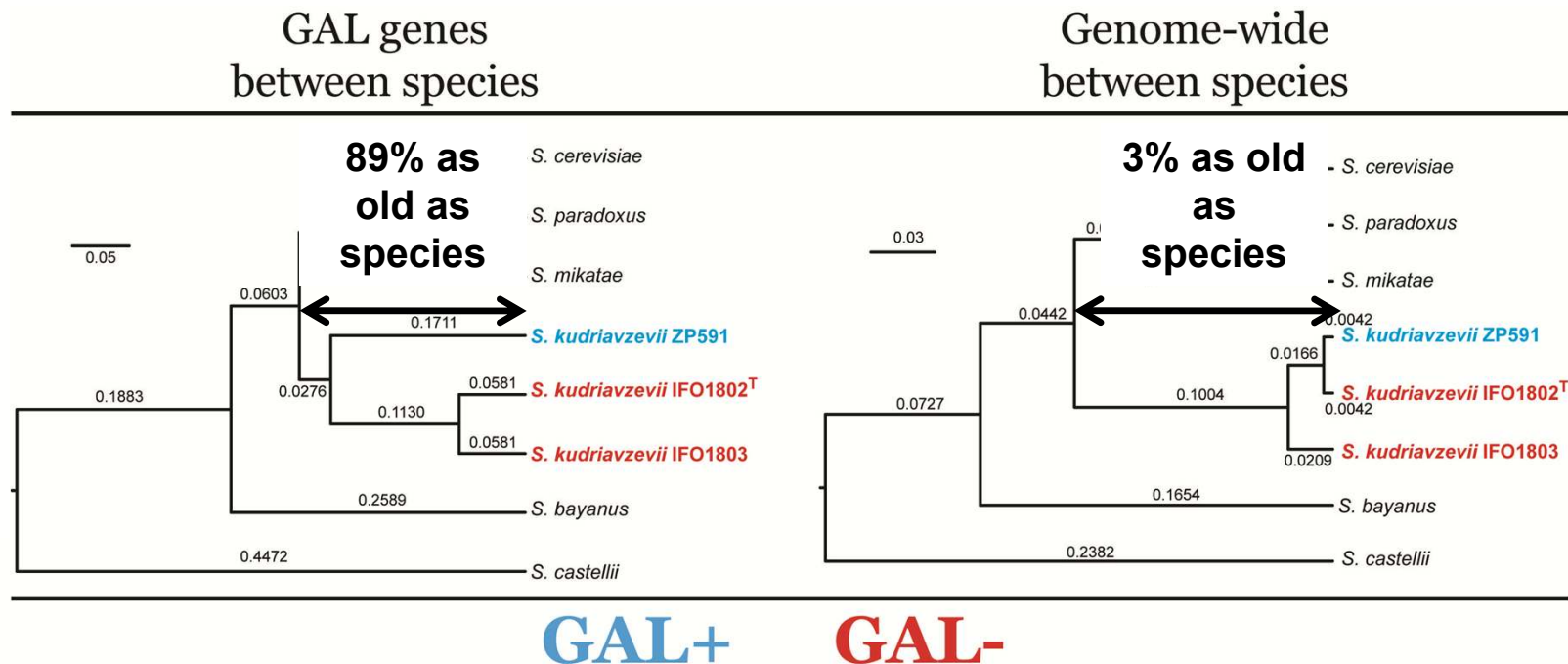
GAL+ **GAL-**

Only two non-GAL genes in the *S. kudriavzevii* genome significantly reject the lineage phylogeny in favor of the GAL gene phylogeny – both flank GAL loci



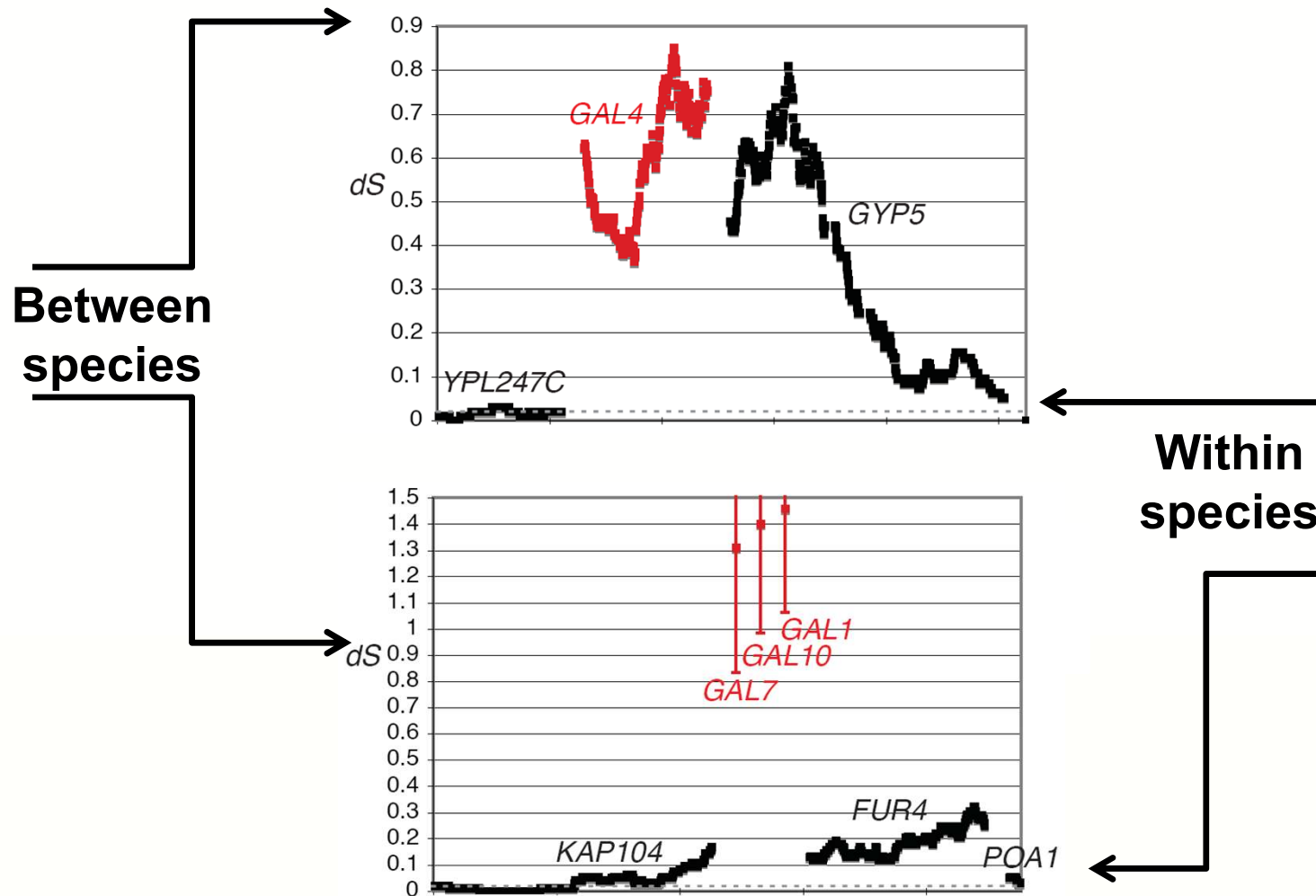
Clade Support Values: Posterior Probability / ML Bootstrap

Ancient GAL Divergence Despite Recent Genome Divergence



Ultrametric trees constructed using BEAST

The *S. kudriavzevii* GAL Loci Are Remarkably Divergent



Hittinger et al. (2010) Nature

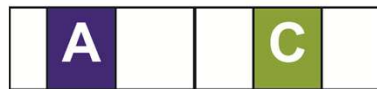
Types of Genetic Variation Within Species



**Genes in
Reference Genome**



Nucleotide Substitutions



Deletions



Insertions



Inversions



Copy Number Differences



Segmental Duplications



Balanced unlinked Gene Network Polymorphisms (BuGNPs)



How can we explain such striking polymorphism in a multi-locus gene network?



Hittinger et al. (2010) Nature

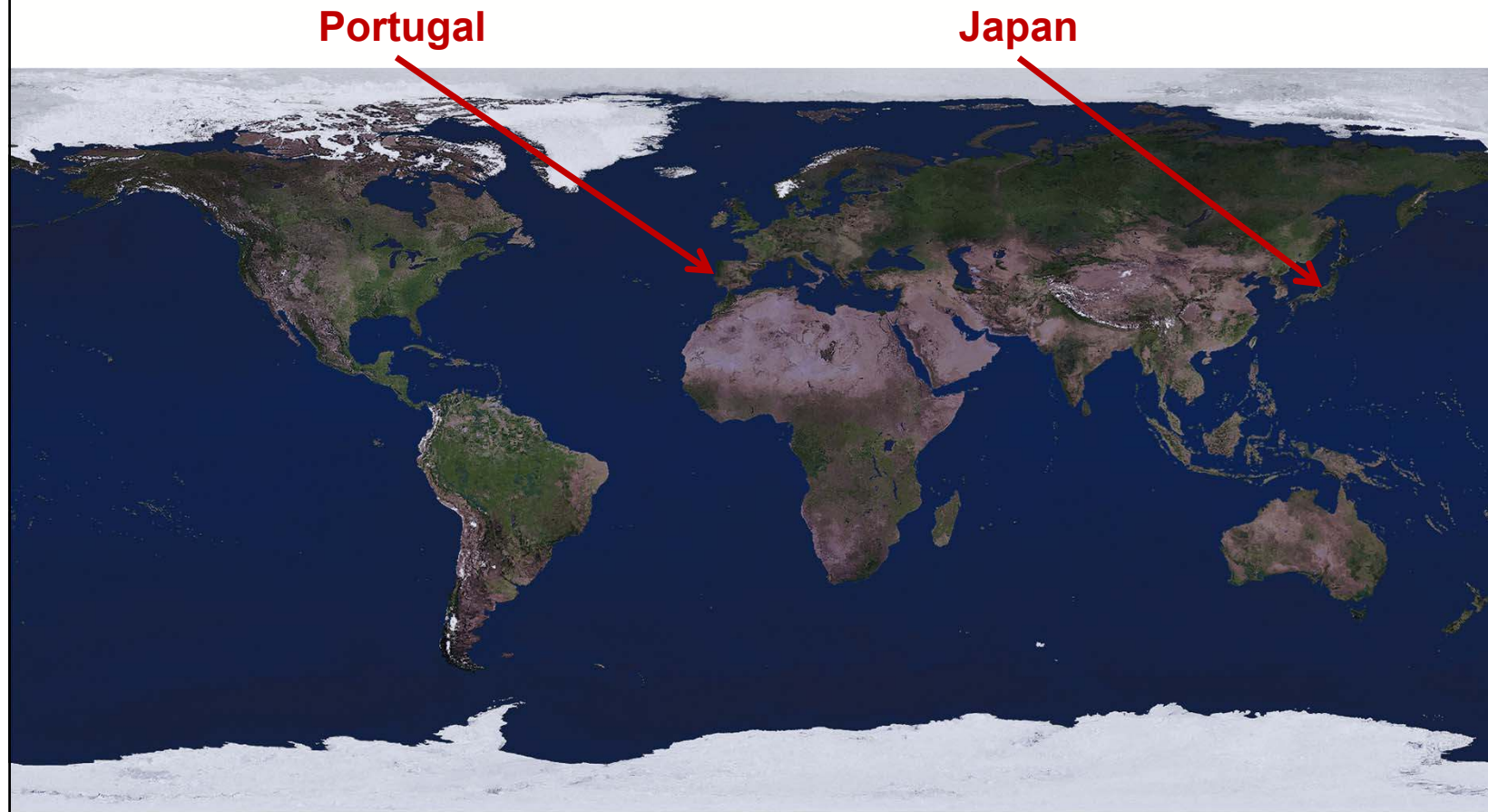
How Can We Explain BuGNPs?

“It is now generally understood that [...] alleles at different loci may not be randomly associated with each other in a population. While this effect is generally regarded as a consequence of linkage, even **genes on different chromosomes** may be held temporarily or permanently out of random association by forces of **selection**, **drift** and **non-random mating**.”



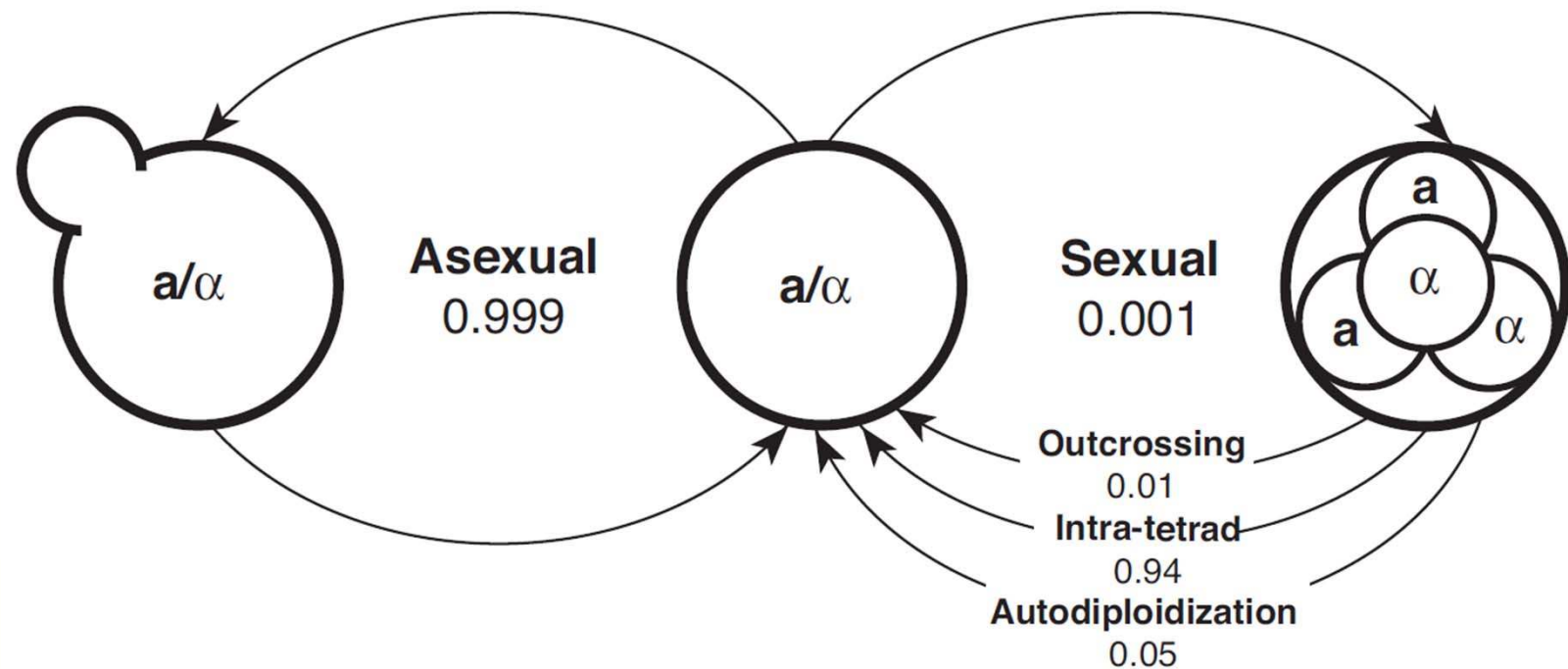
Lewontin (1998) Genetics

Maintaining BuGNPs Through Geographic Isolation



http://nssdc.gsfc.nasa.gov/planetary/image/earth_day.jpg

Maintaining BuGNPs by Selfing

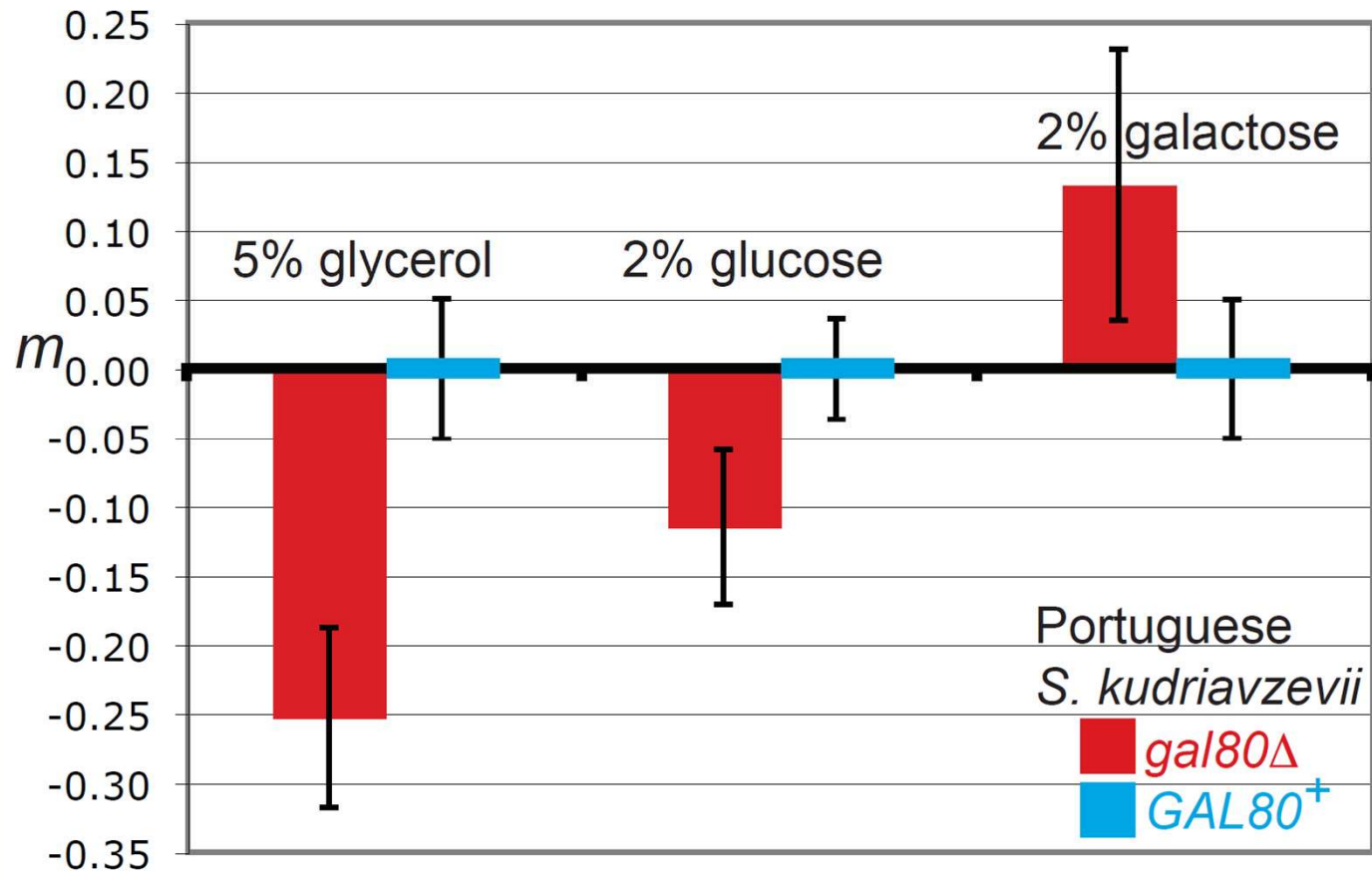


The reproductive lifestyle of *Saccharomyces paradoxus*



Tsai et al. (2008) PNAS

Maintaining BuGNPs by Selecting Against Intermediates



Hittinger et al. (2010) Nature

Lecture Outline

- ❖ Introduction to evolutionary genomics
- ❖ Brief review of next-generation DNA sequencing & its applications
- ❖ Benchmarking transcriptome sequencing for functional and evolutionary genomics
- ❖ Transcriptome sequencing for molecular phylogenetics

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

- ❖ Genomics within species (using next-generation DNA sequencing)
- ❖ **Genomics between species (using publicly available genome data)**



Fungal Metabolic Genes Are Often Clustered

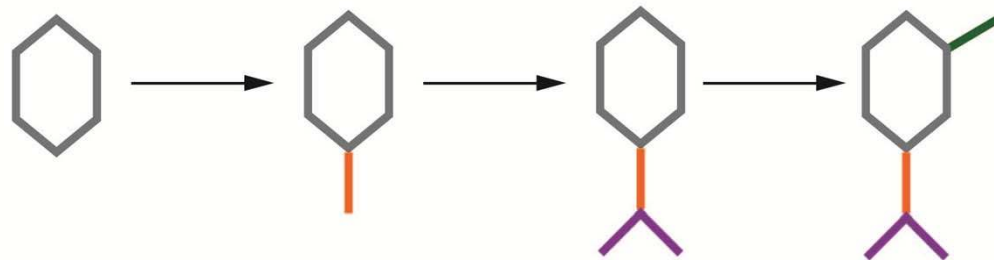
Metabolic gene cluster



Enzymes

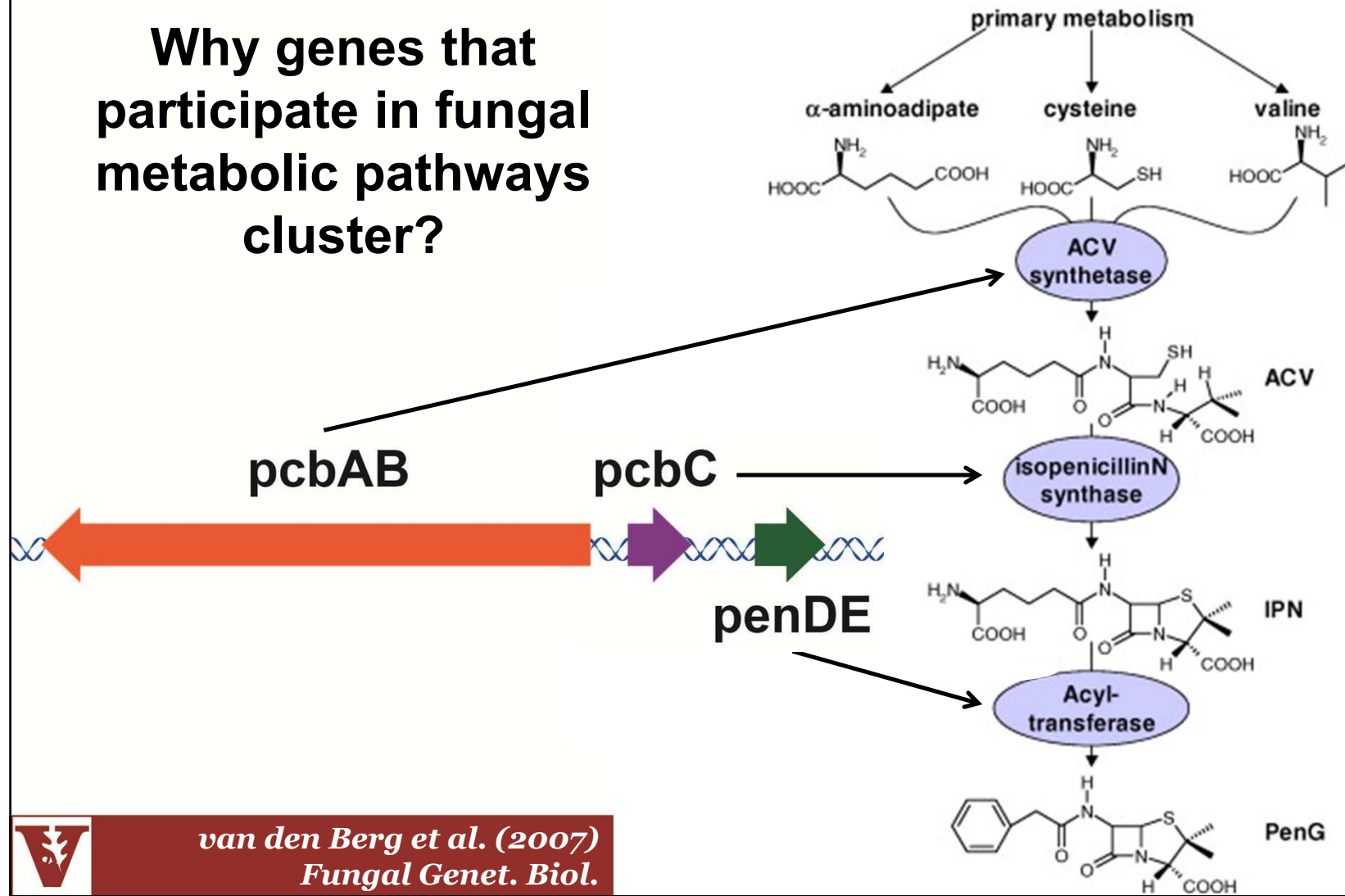


Metabolic pathway



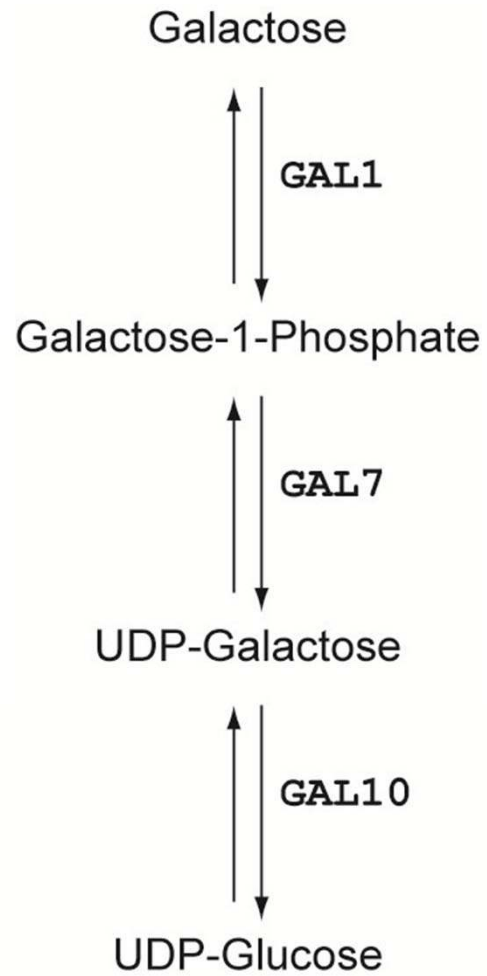
Penicillin is Synthesized by a Gene Cluster

Why genes that participate in fungal metabolic pathways cluster?



van den Berg et al. (2007)
Fungal Genet. Biol.

The GAL Network In Fungi



Jason Slot

Variation in GAL Gene Clustering Across Fungi

Saccharomyces (Ascomycota: Saccharomycetes)



Aspergillus (Ascomycota: Eurotiomycetes)



Cryptococcus (Basidiomycota: Tremellomycetes)



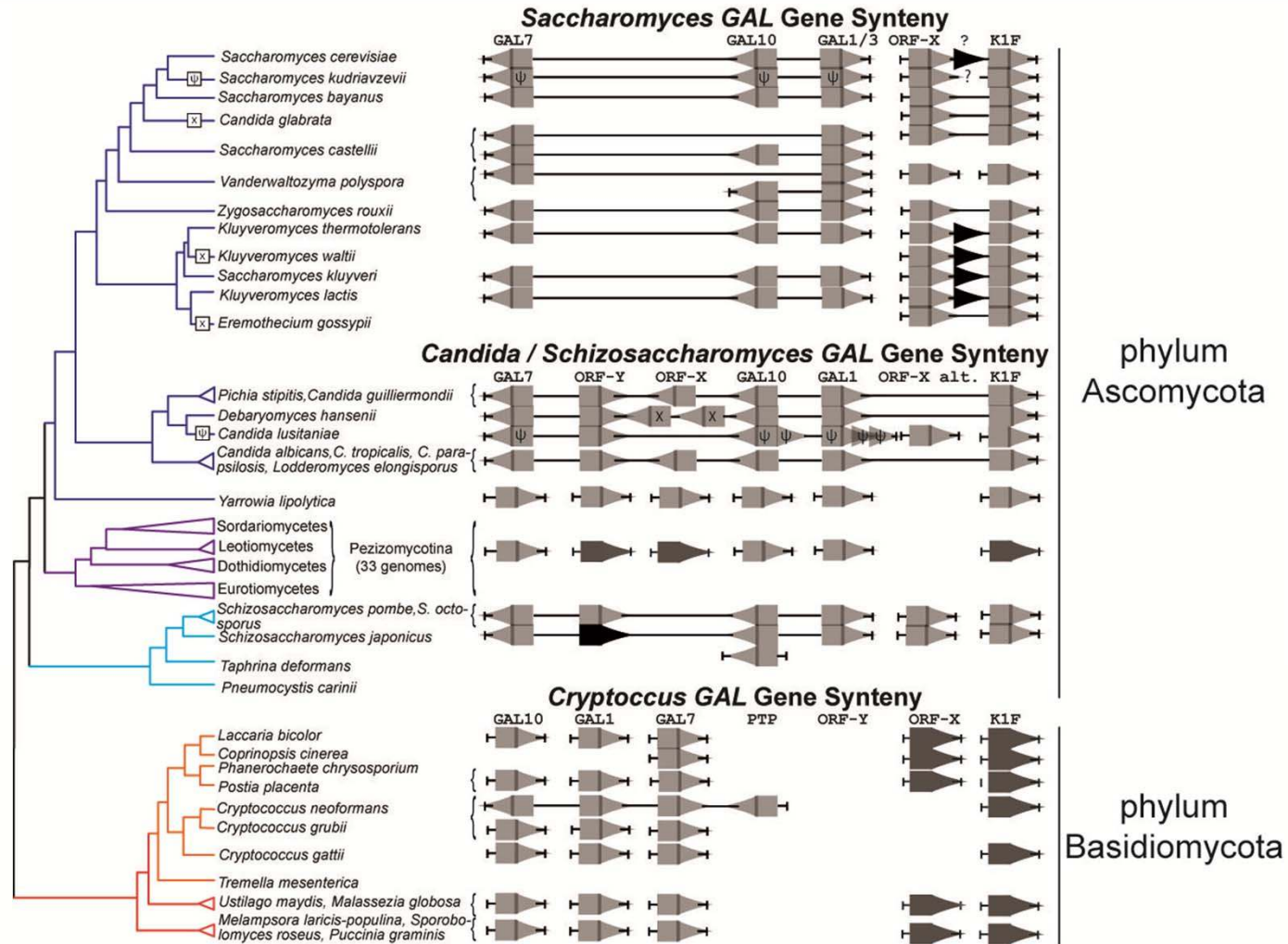
Ustilago (Basidiomycota: Ustilaginomycetes)



Batrachochytrium (Chytridiomycota: Chytridiomycetes)



The Distribution of GAL Gene Clusters Across Fungi



Slot & Rokas (2010) PNAS

Three Distinct Types of Fungal GAL Gene Clusters

Saccharomyces



Candida / Schizosaccharomyces

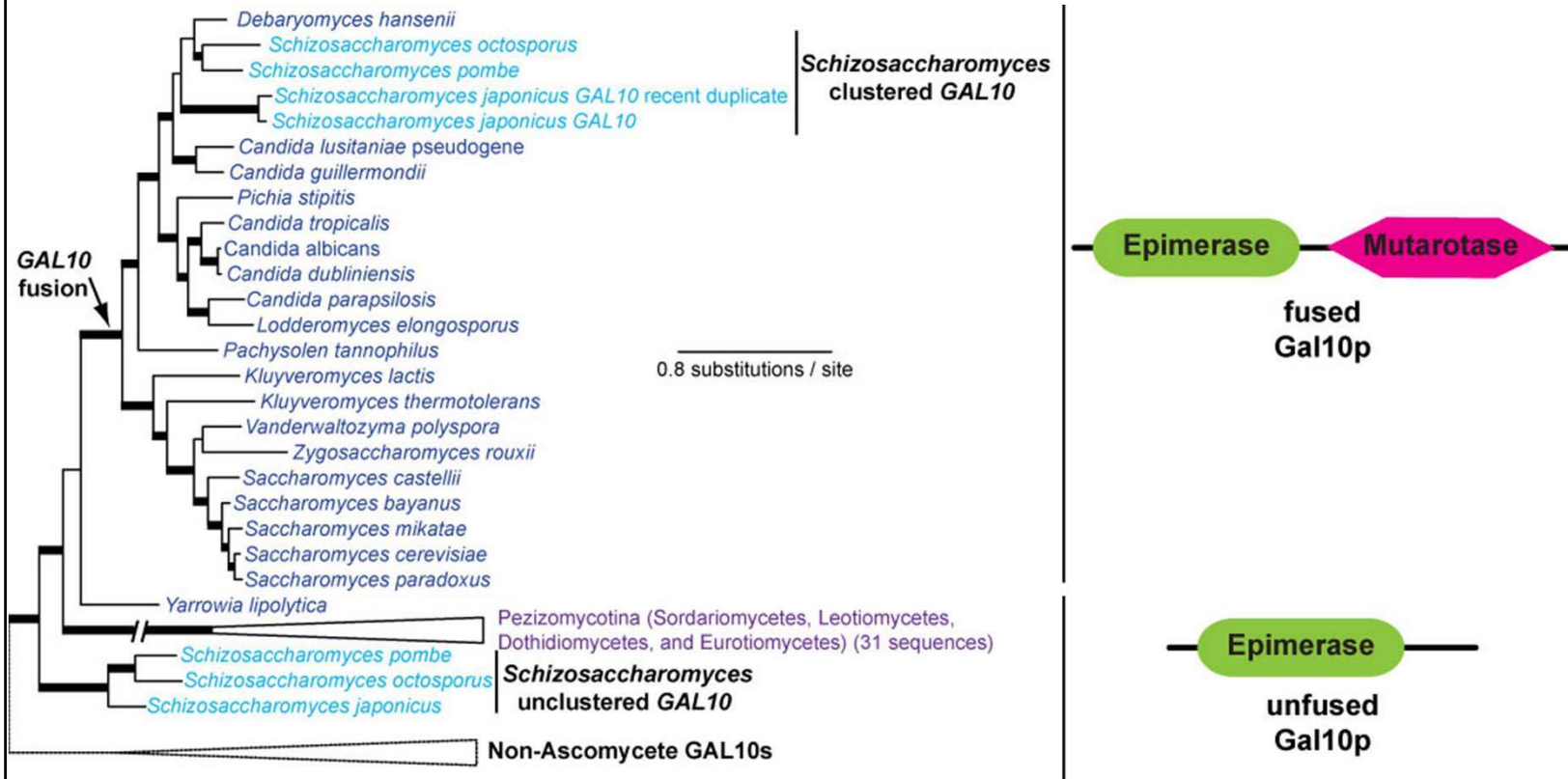


Cryptococcus



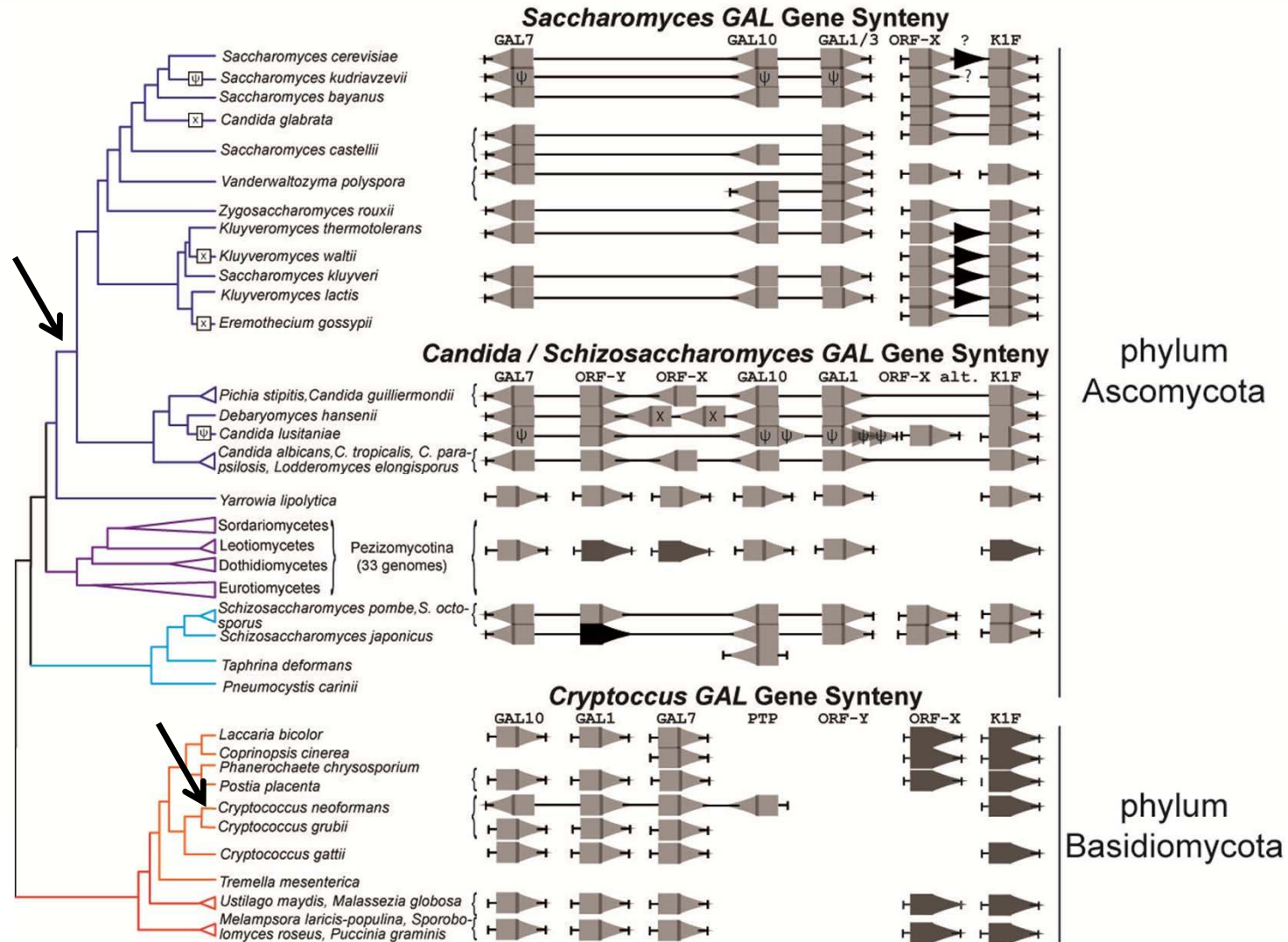
Slot & Rokas (2010) PNAS

Schizosaccharomyces Gained Its GAL Cluster by HGT from Candida

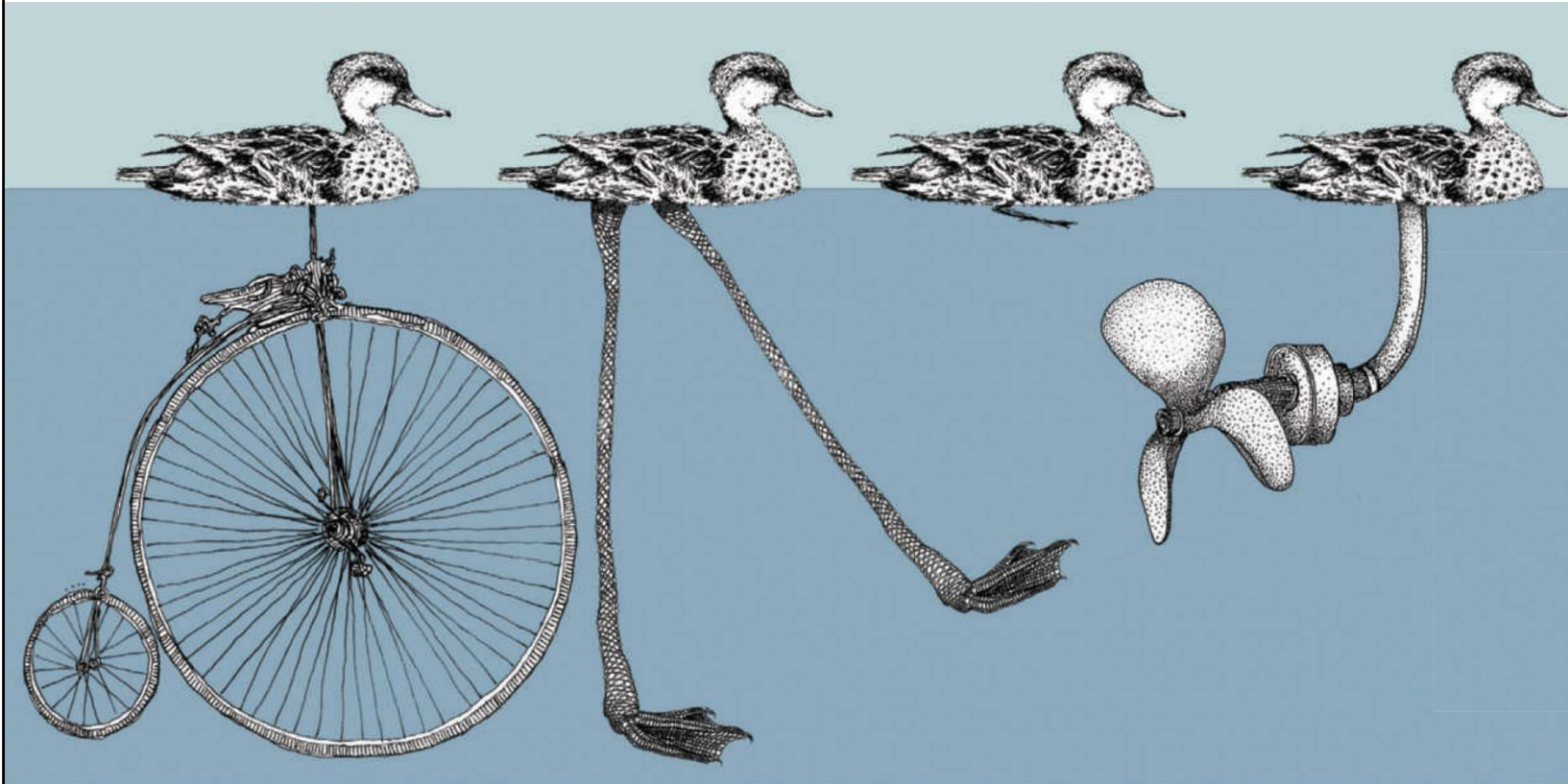


Slot & Rokas (2010) PNAS

The *Saccharomyces* & *Cryptococcus* Clusters Evolved Independently



Analogy At the Phenotypic Level



Chouard (2008) Nature

Analogy At the Genomic Level

Saccharomyces



Candida / Schizosaccharomyces



Cryptococcus



Why Cluster?

Clustering allows more precise coordination of pathway gene regulation

- Sharing of regulatory element binding sites**
- Deleterious metabolic intermediates**

Clustering is a selfish property of the pathway

- Horizontal gene transfer**
- No physiological benefit to the host**



GAL10 ←

Scer TTATATTGAATTTTCAAAAATCTTACTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCATTACCACCATATACA
 Spar CTATGTTGATCTTTTCAGAAATTTT-CACTATATTAAGATGGGTGCAAAGAAGTGTGATTATTATATTACATCGCTTTCCTATCATACACA
 Smik GTATATTGAATTTTTCAGTTTTTTTTCACATCTTCAAGGTTATGTAAAAAA-TGTCAAGATAATATTACATTTTCGTTACTATCATACACA
 Sbay TTTTTTGGATTTCTTTAGTTTTCTTTCTTTAACTTCAAAATTATAAAAGAAAGTGTAGTCACATCATGCTATCT-GTCACTATCATATATA
 * * * * *

TATA

Scer TATCCATATCTAATCTTACTTATATGTTGT-GGAAAT-GTAAAGAGCCCCATTATCTTAGCCTAAAAAACC--TTCTCTTGGAACTTTCAGTAATACG
 Spar TATCCATATCTAGTCTTACTTATATGTTGT-GAGAGT-GTTGATAACCCAGTATCTTAACCCAAGAAAGCC--TT-TCTATGAAACTTGAACGT-TACG
 Smik TACCGATGTCTAGTCTTACTTATATGTTAC-GGGAATTGTTGGTAATCCCAGTCTCCAGATCAAAAAGGT--CTTTCTATGGAGCTTTG-CTA-TATG
 Sbay TAGATATTTCTGATCTTTCTTATATATTATAGAGAGATGCCAATAAACGTGCTACCTCGAACAAAAGAGGGGATTTTCTGTAGGGCTTTCCCTATTTG
 * * * * *

Gal4 **Gal4** **Gal4**

Scer CTTAACTGCTCATTGC-----TATATTGAAGTACCGGATTAGAAGCCGCCGAGCGGCGACAGCCCTCCGACCGGAAGACTCTCCTCCGTGCGTCTCGTCT
 Spar CTTAACTGCTCATTGC-----AATATTGAAGTACCGGATCAGAAGCCGCCGAGCGGACGACAGCCCTCCGACCGGAATATTCCCCTCCGTGCGTCCGCTCT
 Smik TTTAGCTGTTCAAG-----ATATTGAATAACCGATGAGAAGCCGCCGAGCGGACGACAATCCCGACCGGAACATTCTCCTCCGCGGGCGTCTCT
 Sbay TCTTATTGTCCATTACTTCGCAATGTTGAAATAACCGATCAGAAGCTGCCGACCGGATGACAGTACTCCGCGGAAAACGTCTCTCCGTGCGAAGTCTCT
 * * * * *

Gal4

Scer TCACCGG-TCGCGTTCCTGAAACGCAGATGTGCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAA-----TACTAGCTTTT--ATGGTTATGAA
 Spar TCGTCGGGTTGTGTCCCTTAA-CATCGATGTACTCGCGCCGCCCTGCTCCGAACAATAAAGATTCTACAAAGAA-TACTTGTTTTTTTATGGTTATGAC
 Smik ACCTTGG-TCGCGTCCCTGAA-CATAGGTACCGCTCGCACCCGCTGGTCCGAACATATACTTGCATAAAGAGGTACTAATTTCT--ACGGTGTATGCC
 Sbay GTG-CGGATCACGTCCCTGAT-TACTGAAGCGTCTCGCCCCCATACCCCGAACAATGCAATGCAAGAACAAA-TGCCTGTAGTG--GCAGTTATGGT
 * * * * *

Mig1

Scer GAGGA-AAAATTGGCAGTAA---CCTGGCCCCACAAACCTT-CAAAATTAACGAATCAAATTAACAACCATA-GGATGATAATGCGA-----TTAG--T
 Spar AGGAACAAAATAAGCAGCCC---ACTGACCCCATATACCTTTCAAATATTGAATCAAATTTGGCAGCATA-TGGTAATAGTACAG-----TTAG--G
 Smik CAACGCAAAATAAACAGTCC---CCCGCCCCACATACCTT-CAAATCGATGCGTAAACTGGCTAGCATA-GAATTTTGGTAGCAA-AATATTAG--G
 Sbay GAACGTGAAATGACAATTCTTTGCCCT-CCCCAATATACCTTGTTCCTGTACAGCAGTGGATAGAACAATGATGGGGTTGCGGTCAGCCTACTCG
 * * * * *

Mig1 **TATA**

Scer TTTTGTAGCTTTATTTCTGGGGTAATTAATCAGCGAAGCG--ATGATTTTT-GATCTATTAACAGATAATAAAATGAAAAGCTGCATAACCAC-----TT
 Spar GTTTT--CTTTATTCCTGAGACAATTCATCCGCAAAAAATAATGGTTTTT-GGTCTATTAGCAAACATATAAATGCAAAGTTGCATAGCCAC-----TT
 Smik TTCTCA--CTTTCTCTGTGATAATTCATCACCAGAAATG--ATGGTTTA--GGACTATTAGCAAACATATAAATGCAAAGTCGCAGAGATCA-----AT
 Sbay TTTTCCGTTTACTTCTGTAGTGGCTCAT--GCAGAAAGTAATGGTTTTCTGTTCTTTTGCAAACATATAAATATGAAAGTAAGATCGCCTCAATTGTA
 * * * * *

Scer TAACTAATACTTTCAACATTTTCAGT--TTGTATTACTT-CTTATTCAAAT---GTCATAAAAGTATCAACA-AAAAATTGTTAATATACCTCTATACT
 Spar TAAATAC-ATTTGCTCCTCCAAGATT--TTTAATTTTCGT-TTTGTTTTATT---GTCATGGAAATATTAACA-ACAAGTAGTTAATATACATCTATACT
 Smik TCATTTCC-ATTCGAACCTTTGAGACTAATTATATTAGTACTAGTTTTCTTTGGAGTTATAGAAATACCAAAA-AAAAATAGTCAGTATCTATACATACA
 Sbay TAGTTTTTCTTTATTCCGTTTGTACTTCTTAGATTTGTTATTCCGGTTTTTACTTTGTCTCCAATTATCAAACATCAATAACAAGTATTCAACATTTGT
 * * * * *

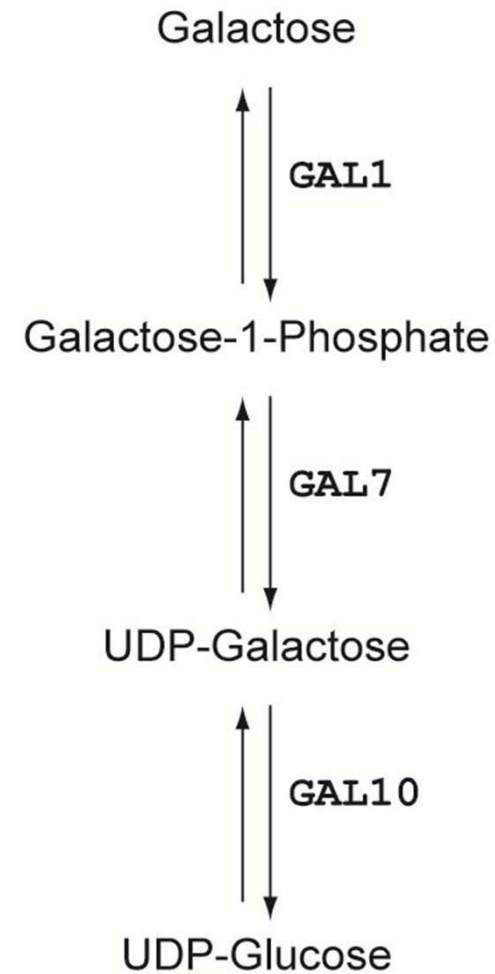
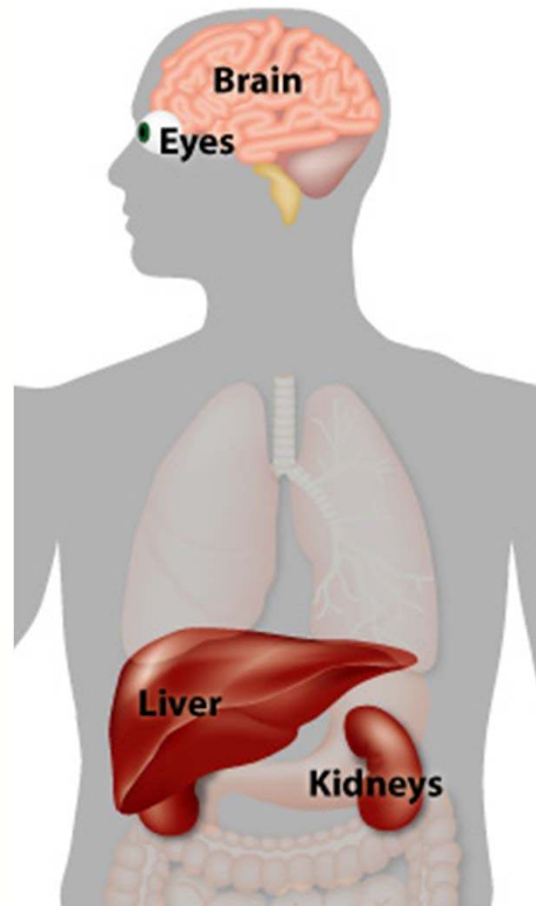
Scer TTAA-CGTCAAGGA---GAAAAAACTATA
 Spar TTAT-CGTCAAGGAAA-GAACAACTATA
 Smik TCGTTTCATCAAGAA---AAAAAACTA..
 Sbay TTATCCCAAAAAACACAACAACATATA
 * * * * *

GAL1 →

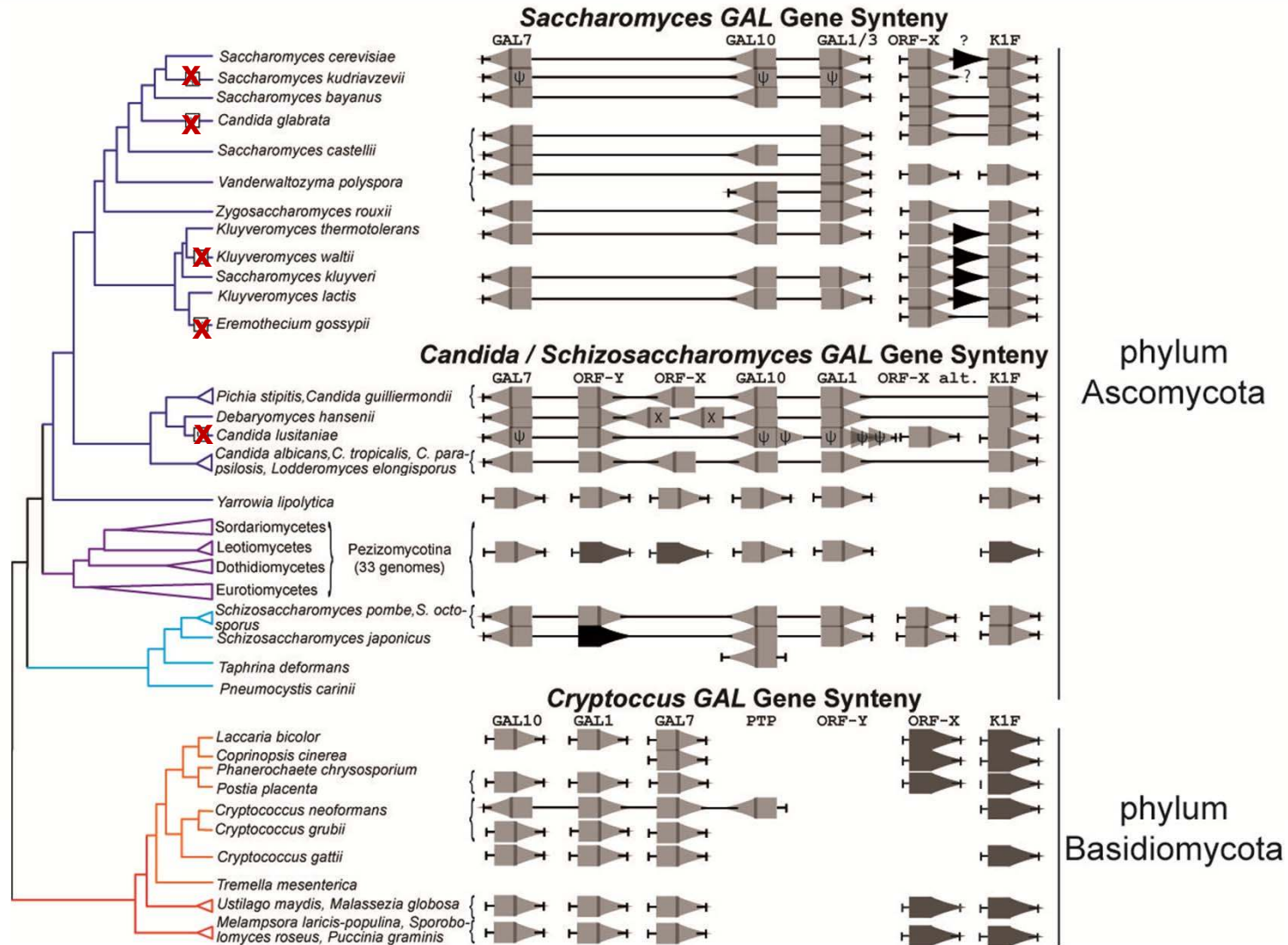
Kellis et al. (2003) Nature

Mutations in GAL7 Lead to Galactosemia

Organs Affected by Galactosemia



Why Cluster?



Are GAL Genes in Clusters Lost More Frequently?

Two potentially functional *GAL* states:

clustered – C

unclustered – U

One nonfunctional *GAL* state:

nonfunctional – N

Rate Restriction	LnL	2ΔL	df	p-value	Mean Transition Rate
none	-28.82				qCN = 30.9 qUN = 1.0 qNC = 77.8 qNU = 2.2
qNC = qNU	-29.64	1.62	1	0.25	qNC = qNU = 12.9
qCN = qUN	-32.95	8.24	2	0.02	qCN = qUN = 2.4

Species with Clusters Have Higher Rates of *GAL* Pathway Loss



Analyses done with BayesTraits (by Pagel & Meade);
<http://www.evolution.rdg.ac.uk/SoftwareMain.html>

A Hybrid Model for the Evolution of Gene Clustering

Clustering originates because it allows more precise coordination of pathway gene regulation and deleterious metabolite avoidance

Clustering makes horizontal gene transfer easier

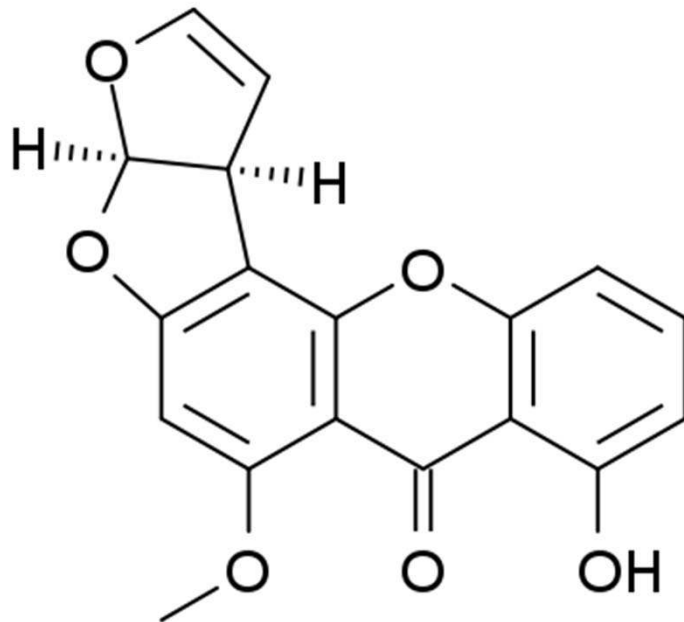
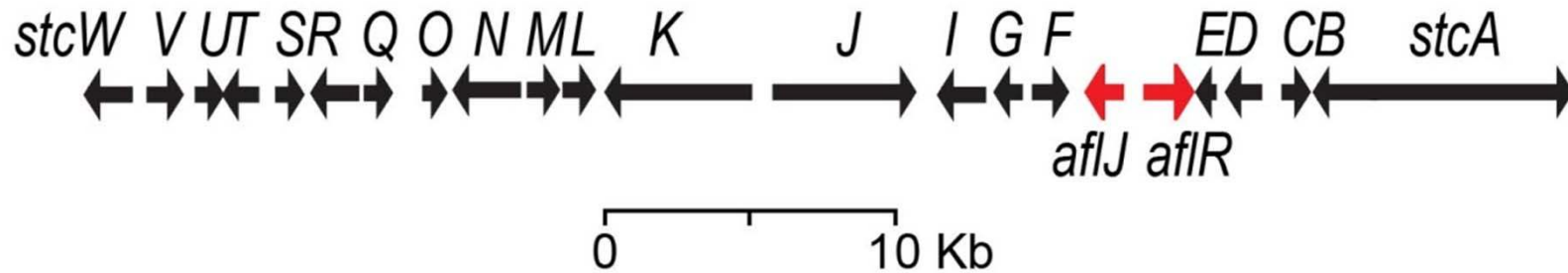
Clustering makes pathway loss easier

=> Clustering facilitates fungal adaptation to diverse and changing nutritional environments



<http://www.genome.jp/kegg/>

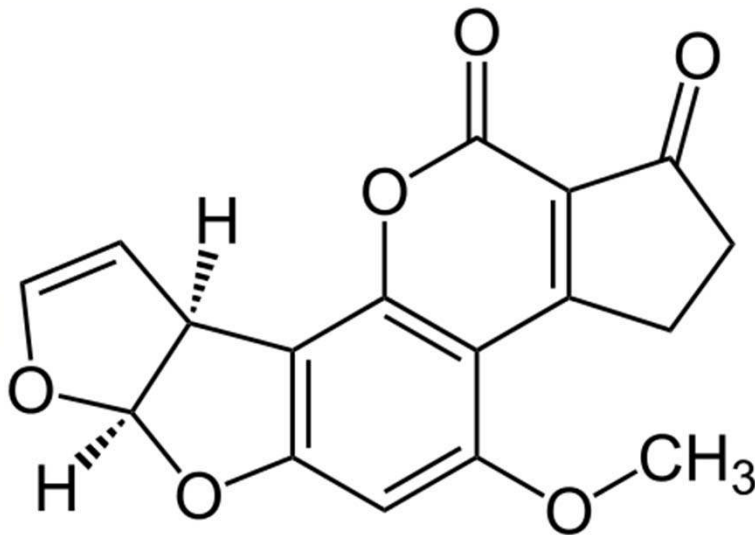
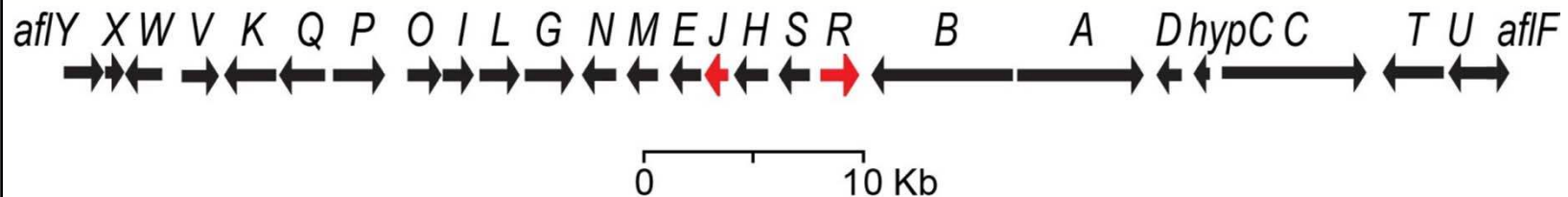
The Sterigmatocystin Gene Cluster in *Aspergillus nidulans*



wheat, maize,
animal feed,
hard cheese,
pecan nuts
and green
coffee beans



The Aflatoxin Gene Cluster in *Aspergillus flavus*

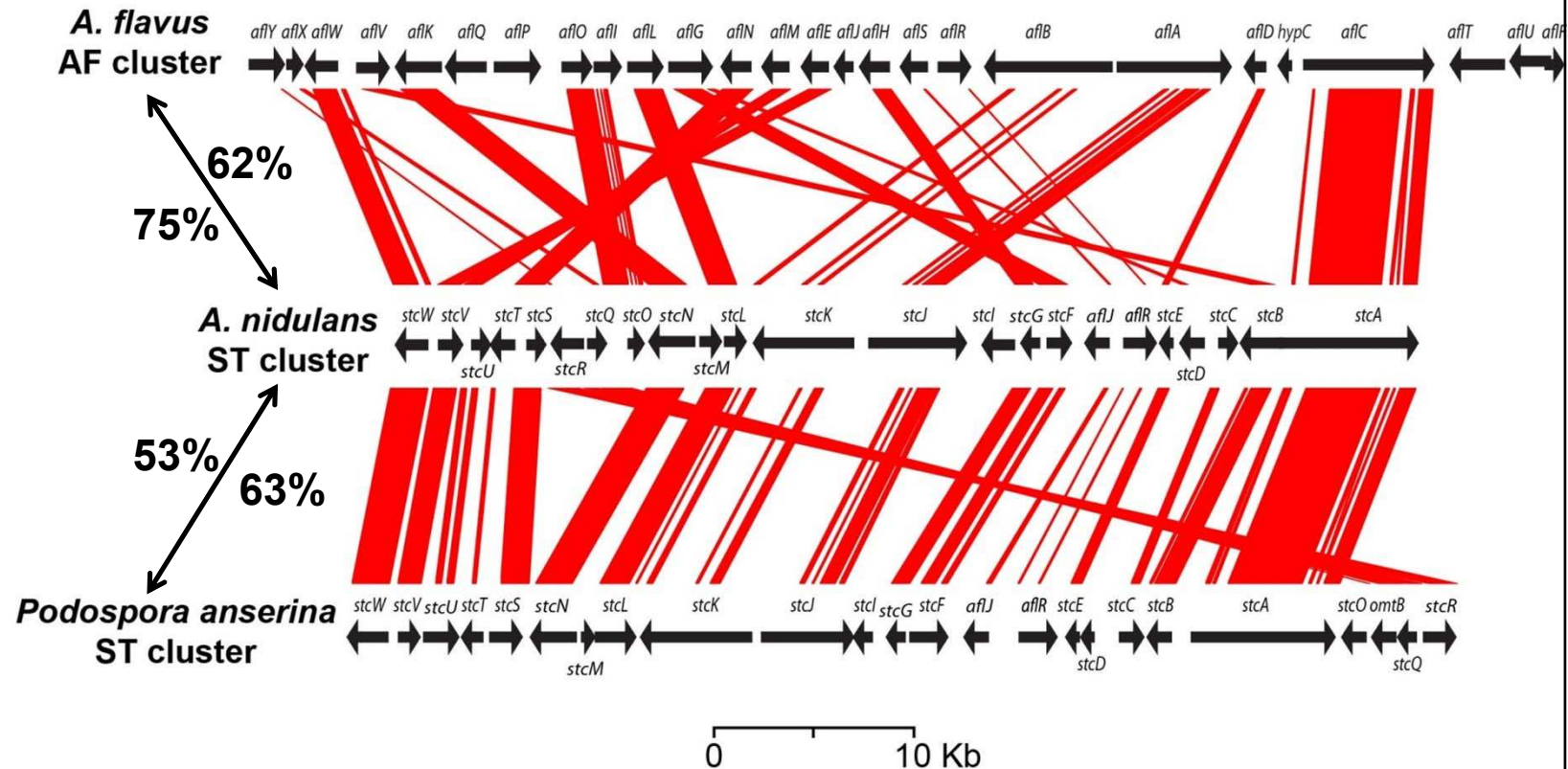


Cereals (maize, sorghum, rice, wheat)
 Oilseeds (peanut, soybean, sunflower)
 Spices (chilli pepper, black pepper)
 tree nuts (almond, pistachio, walnut)
 Milk of animals fed contaminated feed

Virtually all sources of commercial
 peanut butter in the United States
 contain minute quantities of aflatoxin

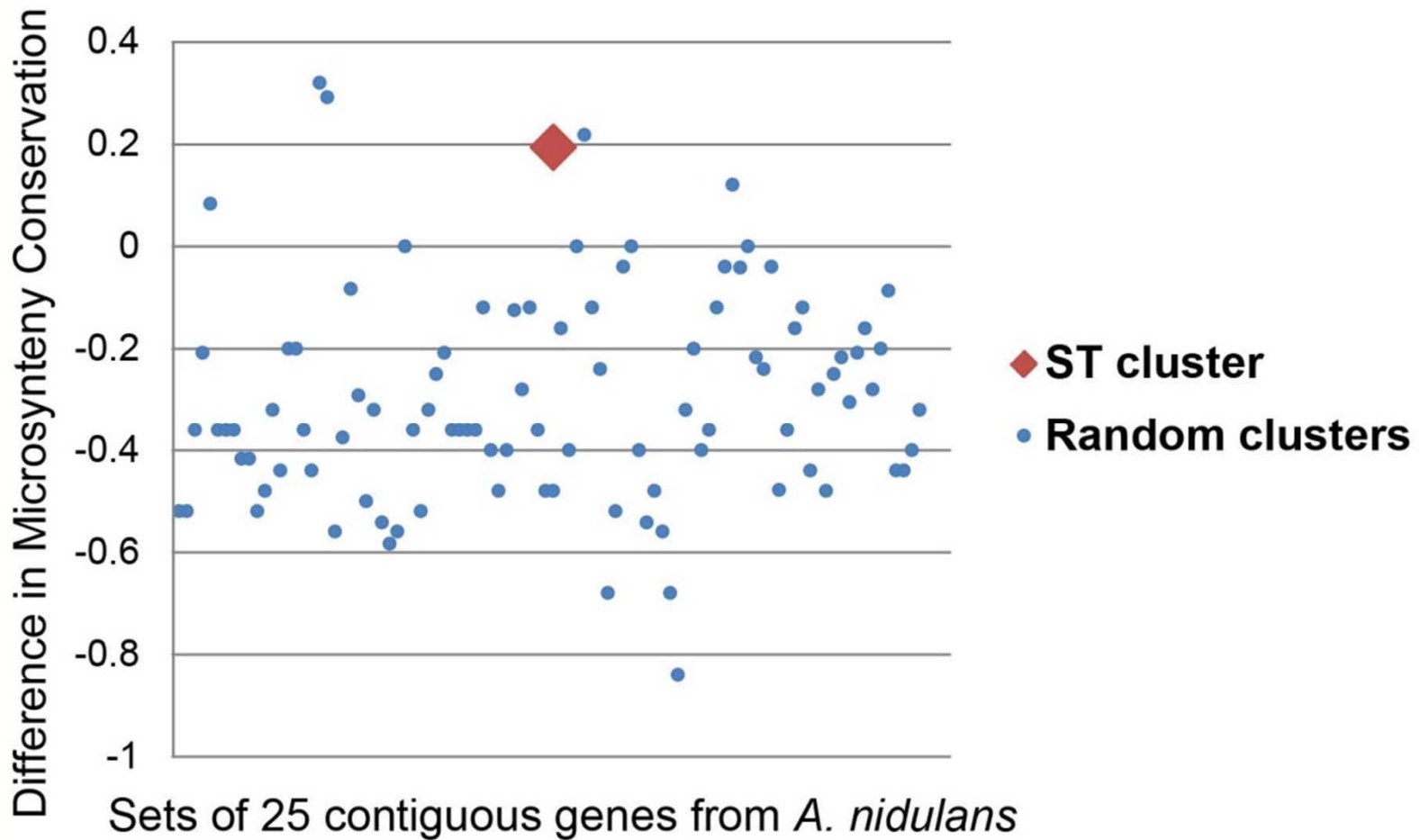
In the United States, agricultural
 economic losses due to AF
 contamination of food and feed amount
 to \$270 M annually

A Near-Identical ST gene cluster in *Podospora anserina*



Slot & Rokas (2011) *Curr. Biol.*

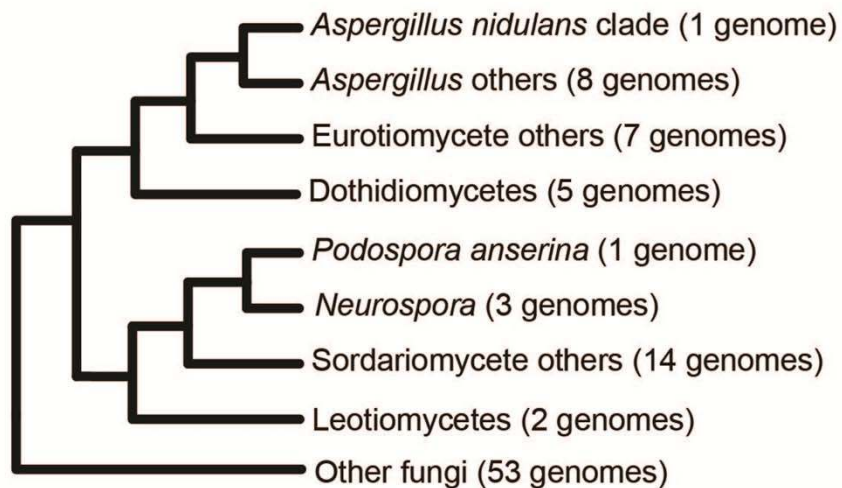
P. anserina ST Cluster Conservation Higher Than Background



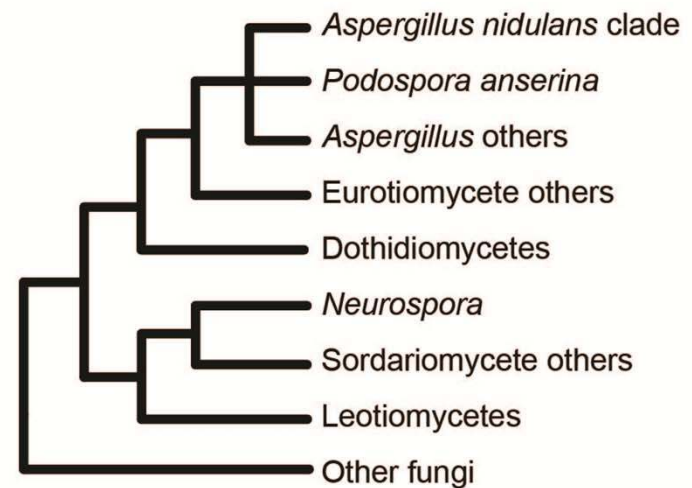
Slot & Rokas (2011) Curr. Biol.

Did P. anserina Acquire the ST Cluster via HGT?

Gene tree expected from species phylogeny

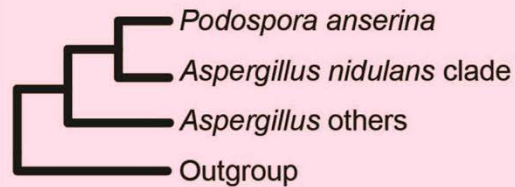


Gene tree expected from HGT



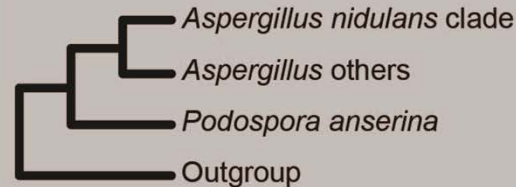
Slot & Rokas (2011) Curr. Biol.

The *P. anserina* ST Cluster was Acquired via HGT



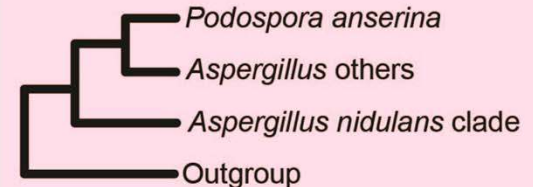
stcW (Figure S1A)
stcV (Figure S1B)
stcR (Figure S1W)
stcQ (Figure S1V)
afIJ (Figure S1N)
stcE (Figure S1P)

1



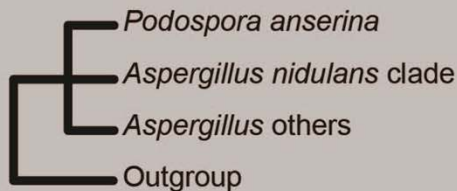
stcU (Figure S1C)*
stcS (Figure S1E)*
stcL (Figure S1H)
stcI (Figure S1K)
stcG (Figure S1L)
stcF (Figure S1M)
stcD (Figure S1Q)

2



stcK (Figure S1I)

3



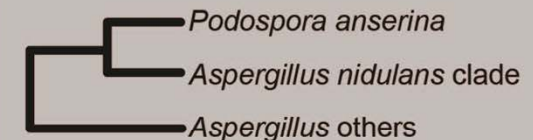
stcN (Figure S1F)
stcJ (Figure S1J)
stcA (Figure S1T)
stcM (Figure S1G)**
stcB (Figure S1S)

4



stcT (Figure S1D)
stcC (Figure S1R)

5



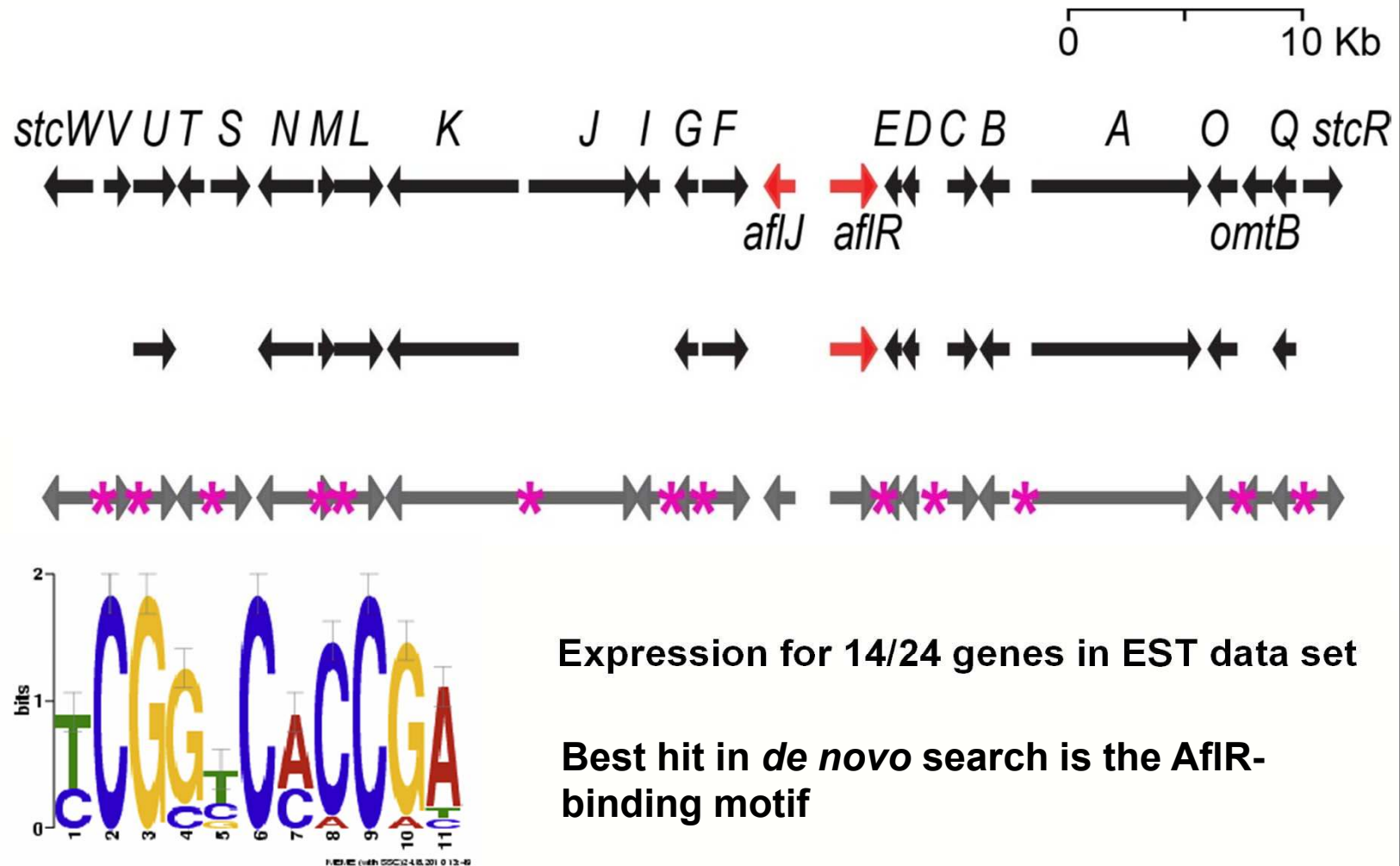
afIR (Figure S1O)
stcO (Figure S1U)

6



Slot & Rokas (2011) Curr. Biol.

Is the *P. anserina* ST Cluster Functional?



Slot & Rokas (2011) *Curr. Biol.*

ORIGINAL PAPER

Larvicidal activity of metabolites from the endophytic *Podospora* sp. against the malaria vector *Anopheles gambiae*

Josphat C. Matasyoh • Birger Dittrich
Anja Schueffler • Hartmut Laatsch

Abstract In a screening for natural products with mosquito larvicidal activities, the endophytic fungus *Podospora* sp. isolated from the plant *Laggera alata* (Asteraceae) was conspicuous. Two xanthones, sterigmatocystin (1) and secosterigmatocystin (2), and an anthraquinone derivative (3) 13-hydroxyversicolorin B were isolated after fermentation on M₂ medium. These compounds were characterised using spectroscopic and X-ray analysis and examined against third instar larvae of *Anopheles gambiae*. The results demonstrated that compound 1 was the most potent one with LC₅₀ and LC₉₀ values of 13.3 and 73.5 ppm, respectively. Over 95% mortality was observed at a concentration 100 ppm after 24 h. These results compared favourably with the commercial larvicide pylarvex® that showed 100% mortality at the same concentration. Com-

Ion Torrent Personal Genome Machine Sequencer

Illumina MiSeq
\$125,000

2 x 150 bp 27 hours 1 Gb



454 / Roche GS Junior
\$100,000

2 x 400 bp 10 hours 35 Mb



Ion Personal Genome Sequencer
\$49,500

1 x 100 bp 2 hours 10 Mb



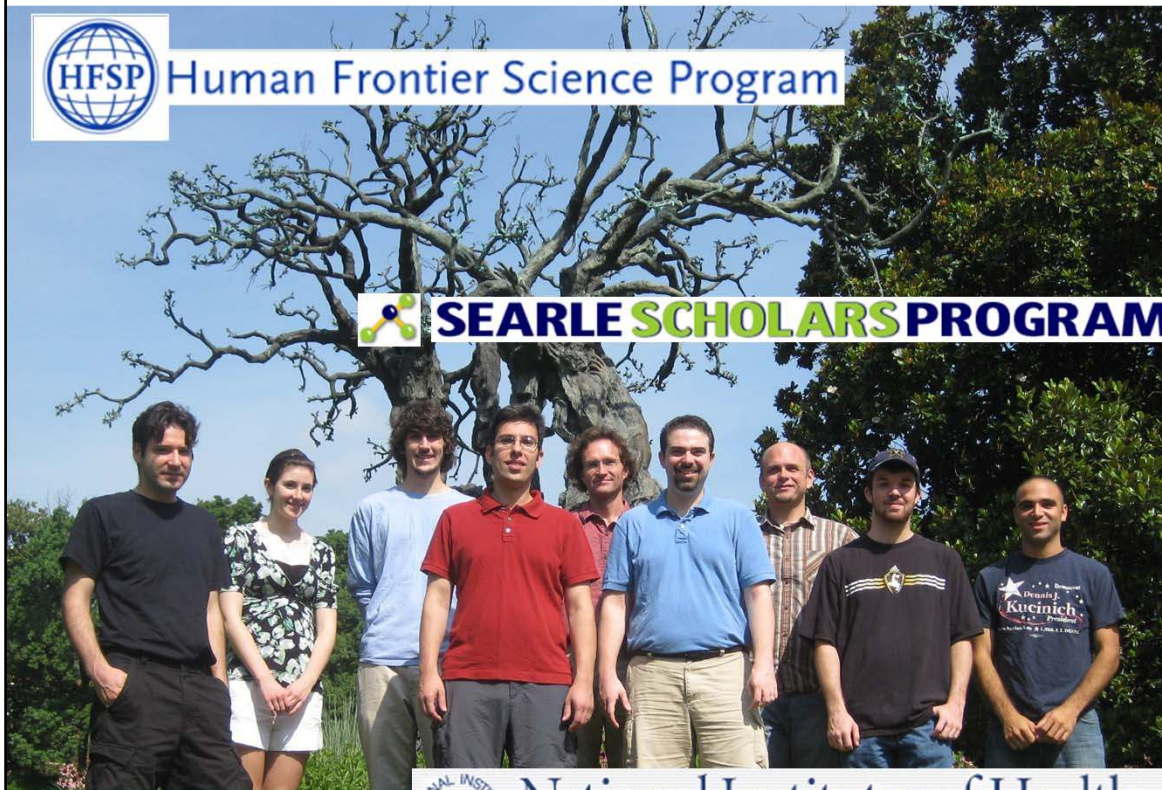
Acknowledgements



Human Frontier Science Program



SEARLE SCHOLARS PROGRAM



Chris Hittinger



Mark Johnston

HHMI
HOWARD HUGHES
MEDICAL INSTITUTE



National Institutes of Health
The Nation's Medical Research Agency



National Science Foundation
WHERE DISCOVERIES BEGIN



<http://people.vanderbilt.edu/~antonis.rokas/>

Literature

1. Mardis ER (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:387-402.
2. Kahvejian A, Quackenbush J, & Thompson JF (2008) What would you do if you could sequence everything? *Nat. Biotechnol.* 26:1125-1133.
3. Rokas A & Abbot P (2009) Harnessing genomics for evolutionary insights. *TREE* 24:192-200.
4. Hittinger et al. (2010) Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *PNAS* 107: 1476-1481.
5. Hittinger et al. (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464: 54-58.
6. Slot JC & Rokas A (2010) Multiple *GAL* pathway gene clusters evolved independently and by different mechanisms in fungi. *PNAS* 107: 10136-10141.
7. Slot JC & Rokas A (2011). Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr. Biol.* 21: *in press*.