Evolutionary Paths to Complexity and Novel Functions

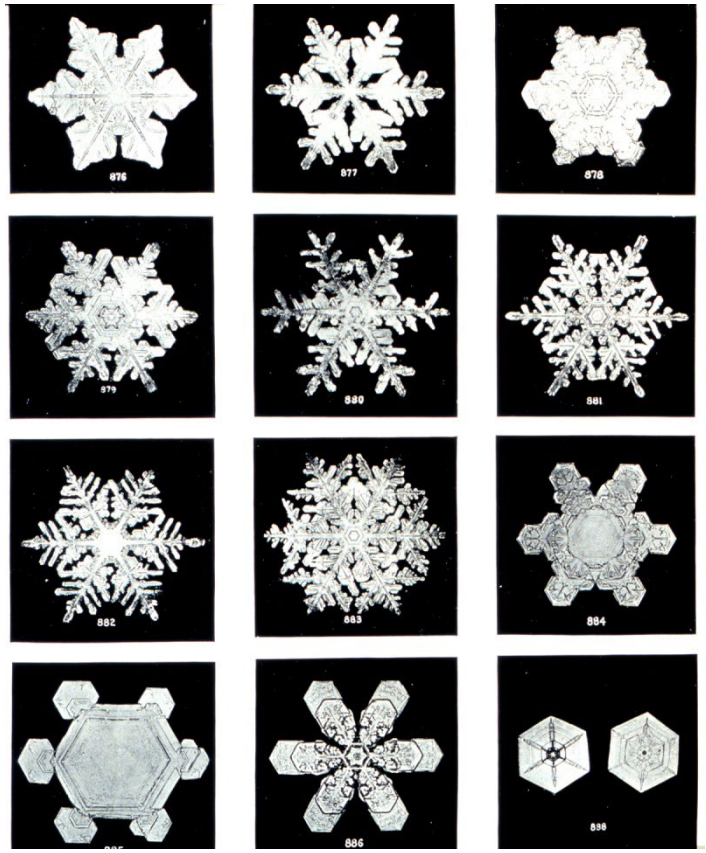
- A common view is that biological complexity represents the crown jewel of the awesome power of natural selection, with animals (humans in particular) representing the pinnacle of what can be achieved.
  - This is a peculiar assumption, as there is no evidence that increases in complexity are intrinsically advantageous.
  - Nor is there any evidence that biology's metabolic, morphological, and behavioral features have reached a maximum level of refinement or ever will.

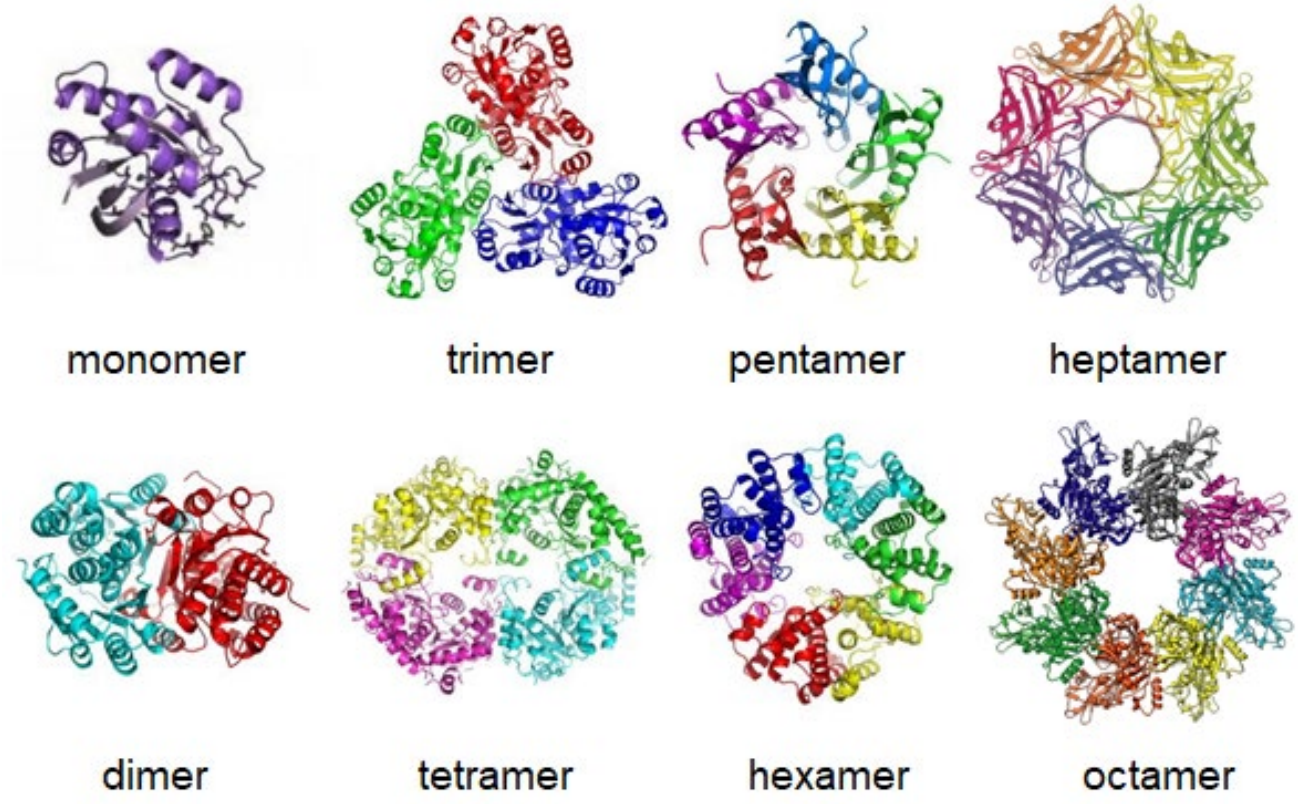
- To minimize energetic costs and mutational vulnerability, all other things being equal, natural selection is expected to always favor simplicity over complexity.

- Yet, many aspects of cell biology are demonstrably “over-designed”, particularly in eukaryotes, and most notably in multicellular species.

- The structural features of biology, combined with fundamental population-genetic processes, result in a complexity ratchet in certain phylogenetic lineages.

Mesmerizing Beauty, Diversity, and the Adaptationist Paradigm

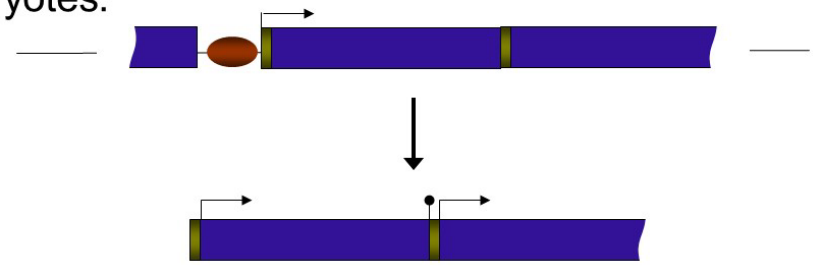




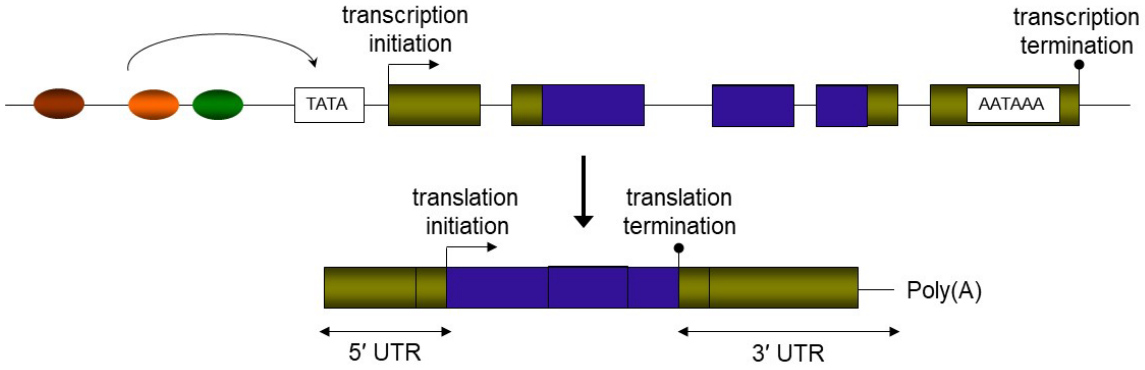
"….. from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."
Charles Darwin

Origin of Gene-structural and Genome Complexity by Nonadaptive Mechanisms

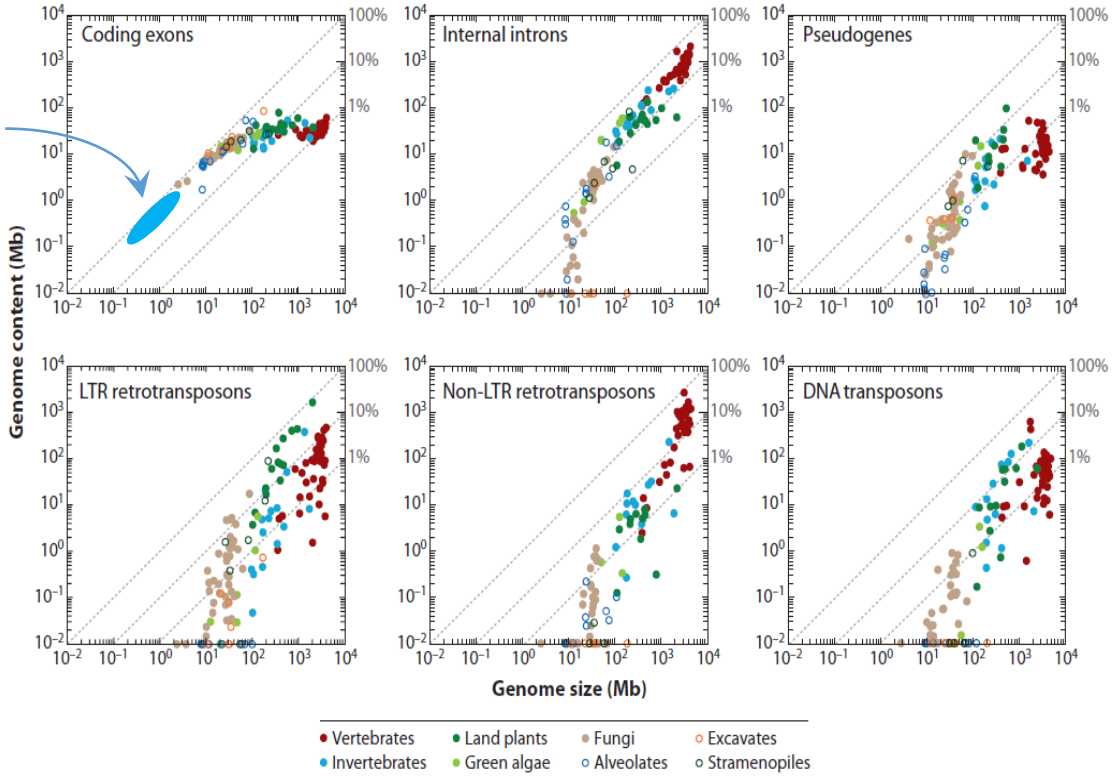
Prokaryotes:



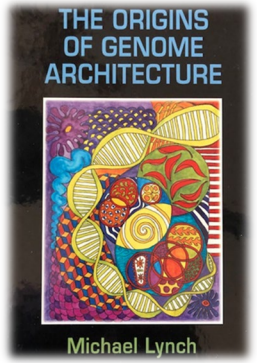
Eukaryotes:



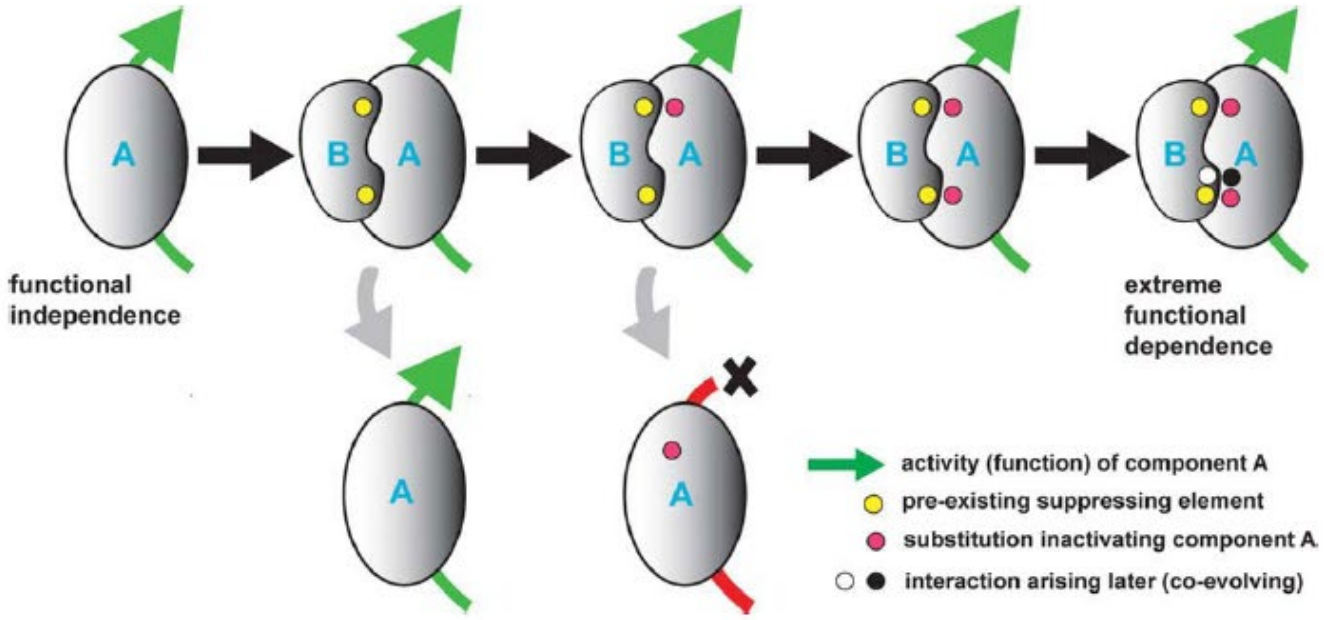
prokaryotes



- All embellishments to gene structure impose weak mutational and bioenergetic disadvantages.
- Efficiently resisted by selection in prokaryotes with large N_e, but can accumulate in an effectively neutral fashion in eukaryotes experiencing relatively high levels of random genetic drift.



Constructive Neutral Evolution: a pathway to the emergence of organismal complexity by nonadaptive mechanisms?



- The fortuitous development of initially neutral interactions between different gene products can alter the selective environment in ways that enable the fixation of previously forbidden mutations, thereby leading to potentially permanent mutual dependence.
- Assumes “excess capacity” in the system to allow diversion of component B.
- Although the formalities of the theory remain to be worked out, the model provides a plausible explanation for the origin of a wide variety of cellular features.

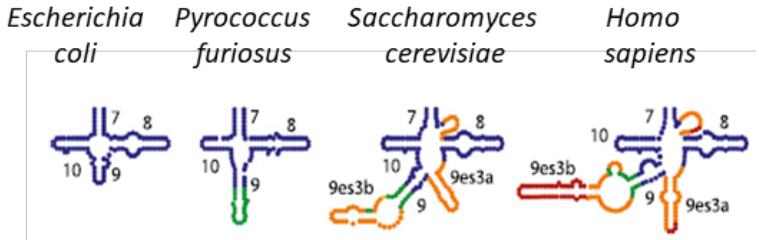
Ribosomes: a Potential Poster-Child for Constructive Neutral Evolution

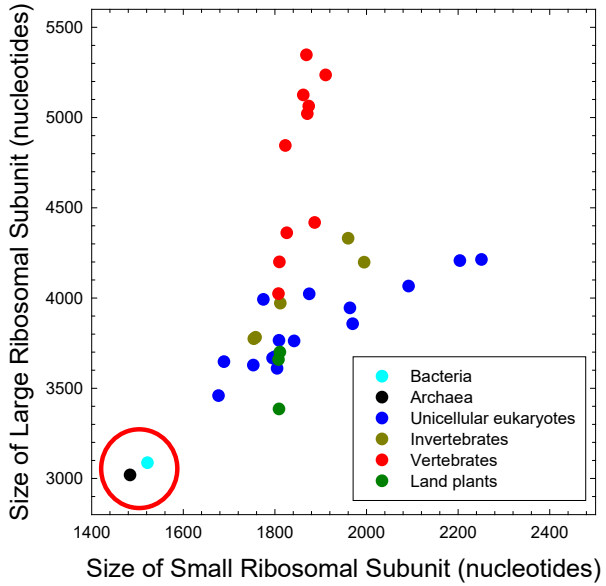
- A complex molecular machine:
  - Catalytic core of 3 to 4 ribosomal RNAs that operate as a ribozyme.
  - Many dozens of surrounding proteins assemble in a sort of “onion-skin” layering.
  - Common core of 34 proteins used in all eukaryotes and prokaryotes.

- Many lineage-specific proteins, and variation in numbers of proteins among lineages.
  - Expansion of the numbers of numbers of proteins in the small / large subunits from 21 / 33 in bacteria to 33 / 46 in eukaryotes, and there is further expansion in some mitochondrial ribosomes.

- 50% expansion in the size of the small and large ribosomal RNAs in eukaryotes relative to bacteria, mostly by expansion segments.

- In contrast, in some mitochondrial ribosomal RNAs, the size is reduced by 50%.





- The machinery for ribosome biogenesis is even more expanded, from just a few proteins in bacteria to ~200 in eukaryotes.

- Despite the massive increase in investment in eukaryotic ribosomes, there is no evidence of improved function.

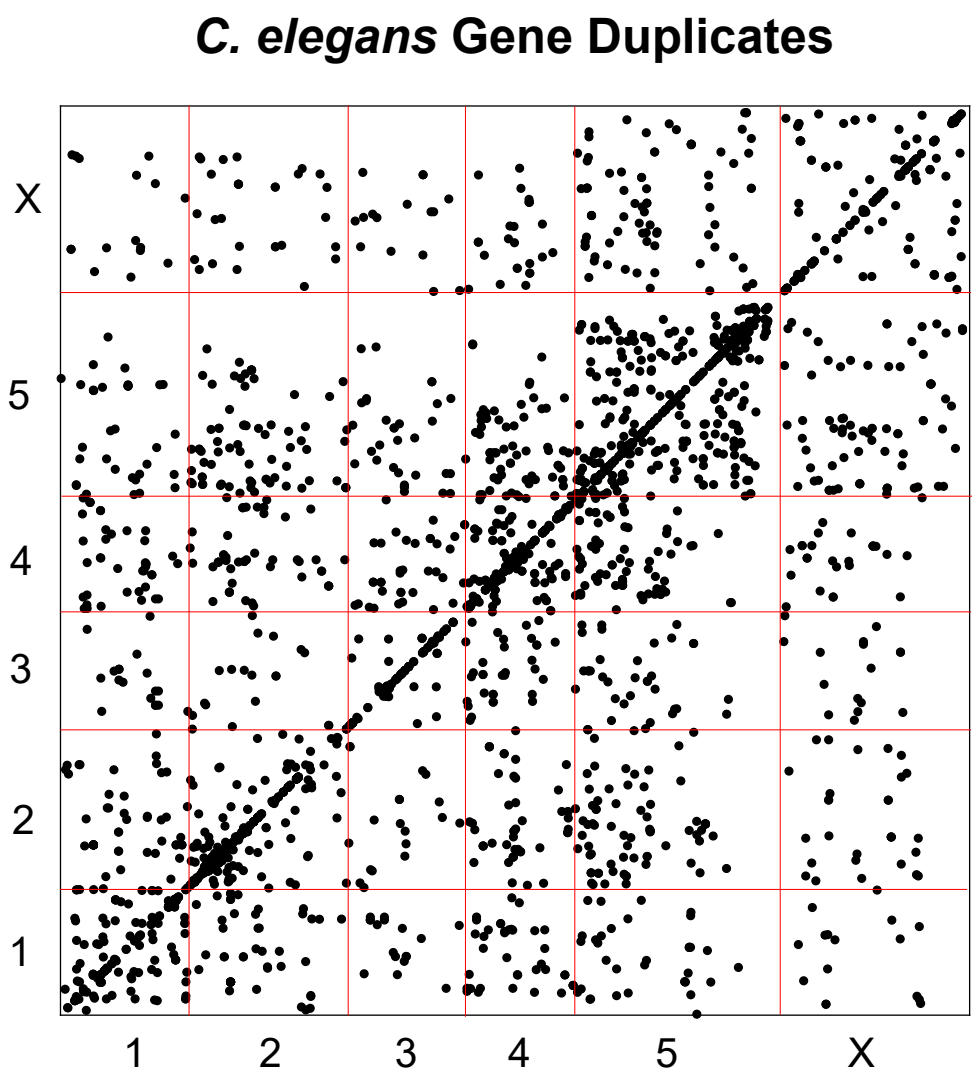
- In bacteria and eukaryotes:

  Translation rate ≈ 5 to 20 amino acids / second.
  Translation-error rate ≈ 0.001 to 0.01 per codon.

Sources of Novel Gene Functions

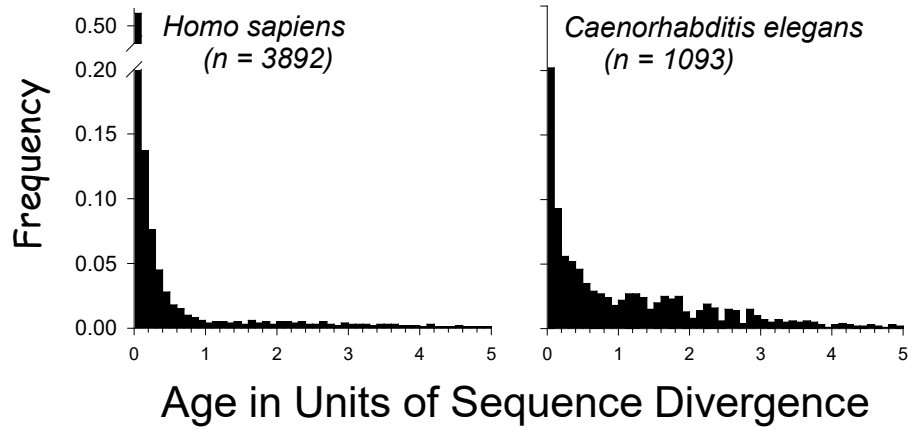
- *De novo* origin from novel open reading frames.
- Protein promiscuity provides latent potential for functional diversification.
- The origin of multimeric protein complexes may lead to novel functions at binding interfaces.
- Gene duplication: neofunctionalization and subfunctionalization.

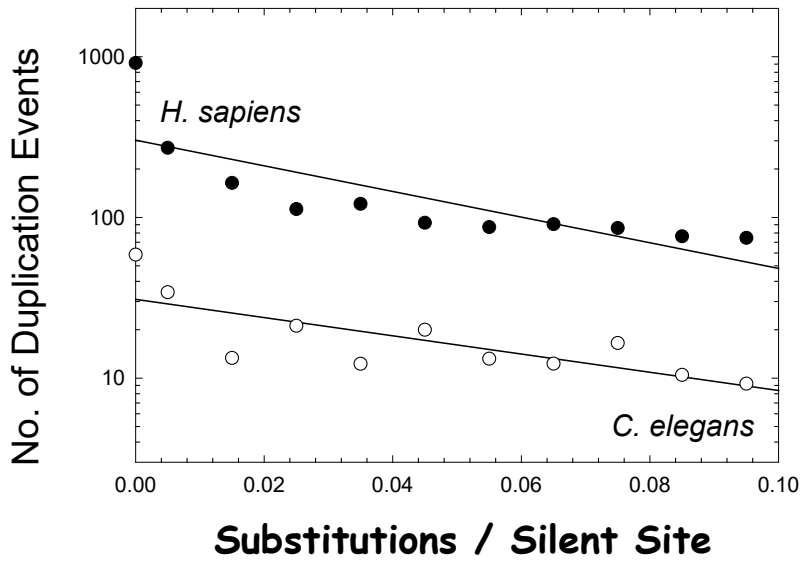
The Pervasiveness of Duplicate Genes



- Each dot denotes a pair of gene duplicates.
- Off-diagonals are on different chromosomes.

The Evolutionary Demography of Duplicate Genes





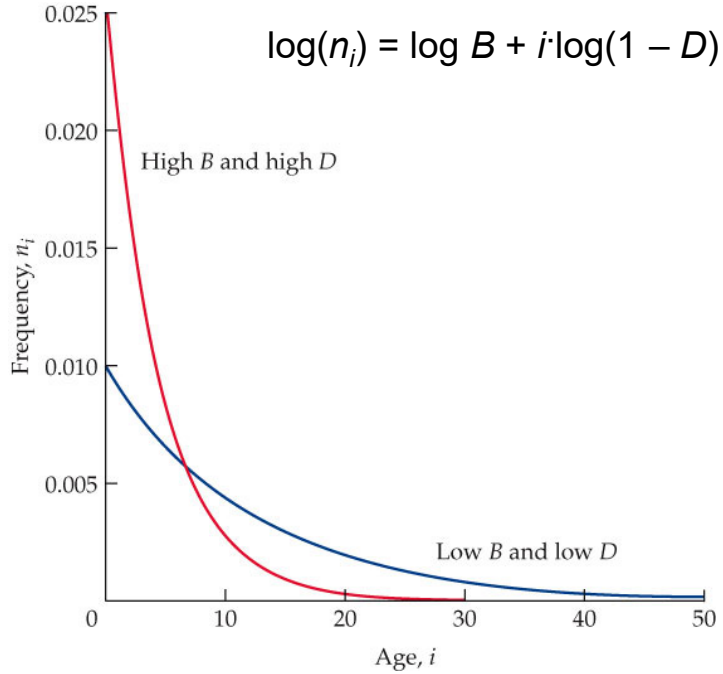
$n_t = n_{t-1} + B(1 + n_{t-1}) - Dn_{t-1}$

where n = no. of excess copies

Stable Age Distribution Under a Steady-state Birth-death Process

$n_i = B(1 - D)^i$,

$\log(n_i) = \log B + i \cdot \log(1 - D)$



Birth Rates of Duplicate Genes (time scale = 1% divergence at silent sites)

Homo sapiens	**0.0049**
Mus musculus	**0.0030**
Fugu rubripes	**0.0043**
Drosophila melanogaster	**0.0011**
Anopheles gambiae	**0.0062**
Caenorhabditis elegans	**0.0028**
Plasmodium falciparum	**0.0003**
Saccharomyces cerevisiae	**0.0025**
Schizosaccharomyces pombe	**0.0016**
Encephalitozoon cuniculi	**0.0118**
Average:	**0.0037 (0.0007)**

- The rate of duplication per gene is nearly as high as the rate of mutation per nucleotide site.
- Incremental gene duplication is sufficient to duplicate an entire genome on a time scale of ~100 MYs.

Ancient Polyploidization (Whole-Genome Duplication) Events Are Common

- *Saccharomyces cerevisiae* (Wolfe and Shields 1997, Nature) – 8% of duplicates still survive.

- The *Paramecium "aurelia"* complex (Aury et al. 2006, Nature) – contains ~40,000 genes.

- *Arabidopsis thaliana* (Bowers et al. 2003, Nature) – three polyploidization events: the first prior to the monocot-dicot split (Simillion et al. 2002; Bowers et al. 2003); the second between 160 and 230 MYA (after the divergence of monocots and dicots); the third between 20 and 85 MYA.

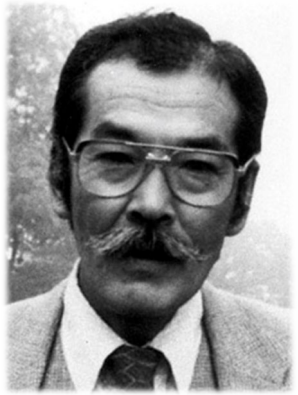
- *Xenopus laevis*, the African clawed frog, is a tetraploid.

- Polyploidization is very common in fish lineages – an ancient event occurred deep in the ray-finned fish lineage; more recent secondary events in salmon, suckers, carp, etc.

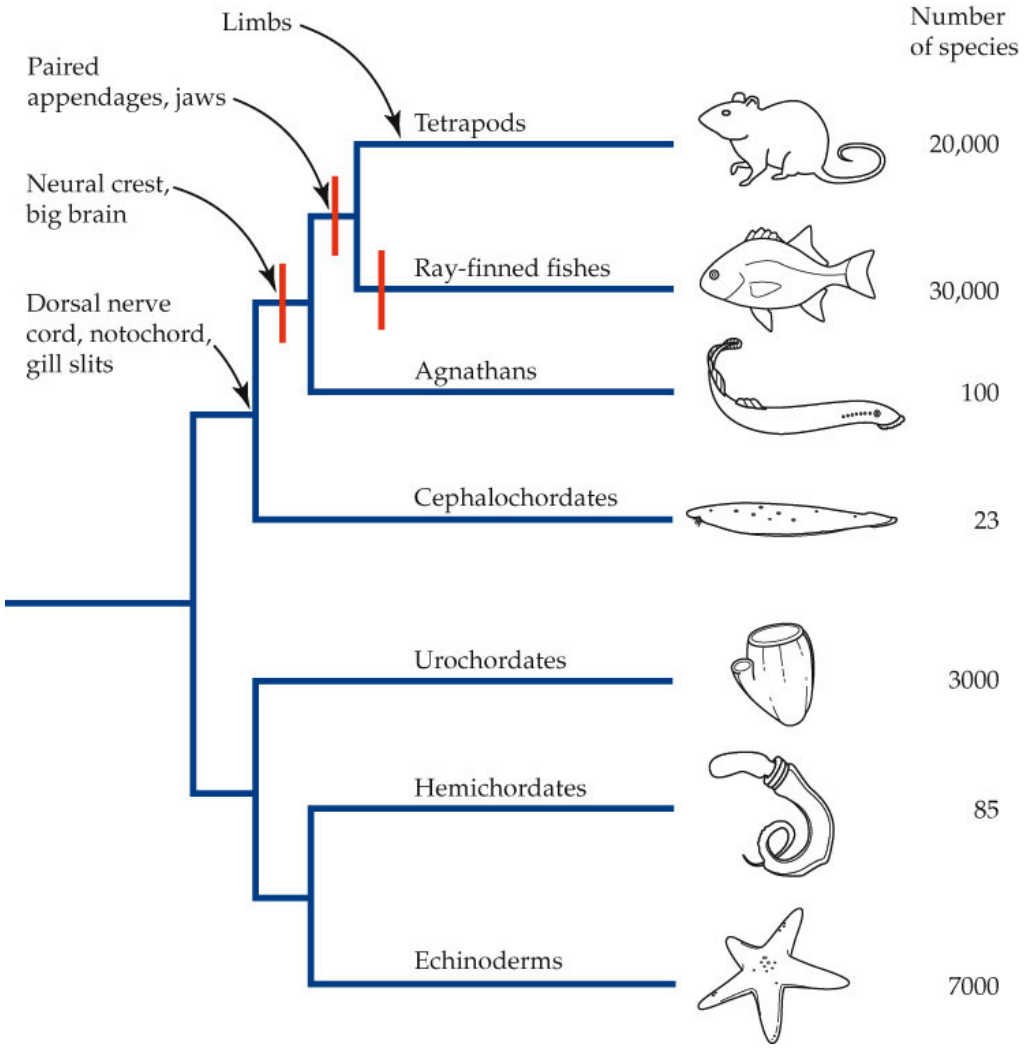
  Zebrafish and pufferfish are tetraploids (in comparison with tetrapods).

- No known prokaryote has experienced whole-genome duplication.

Ohno's 2R Hypothesis and the Origin of Vertebrate Innovations

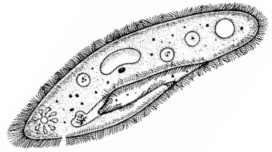


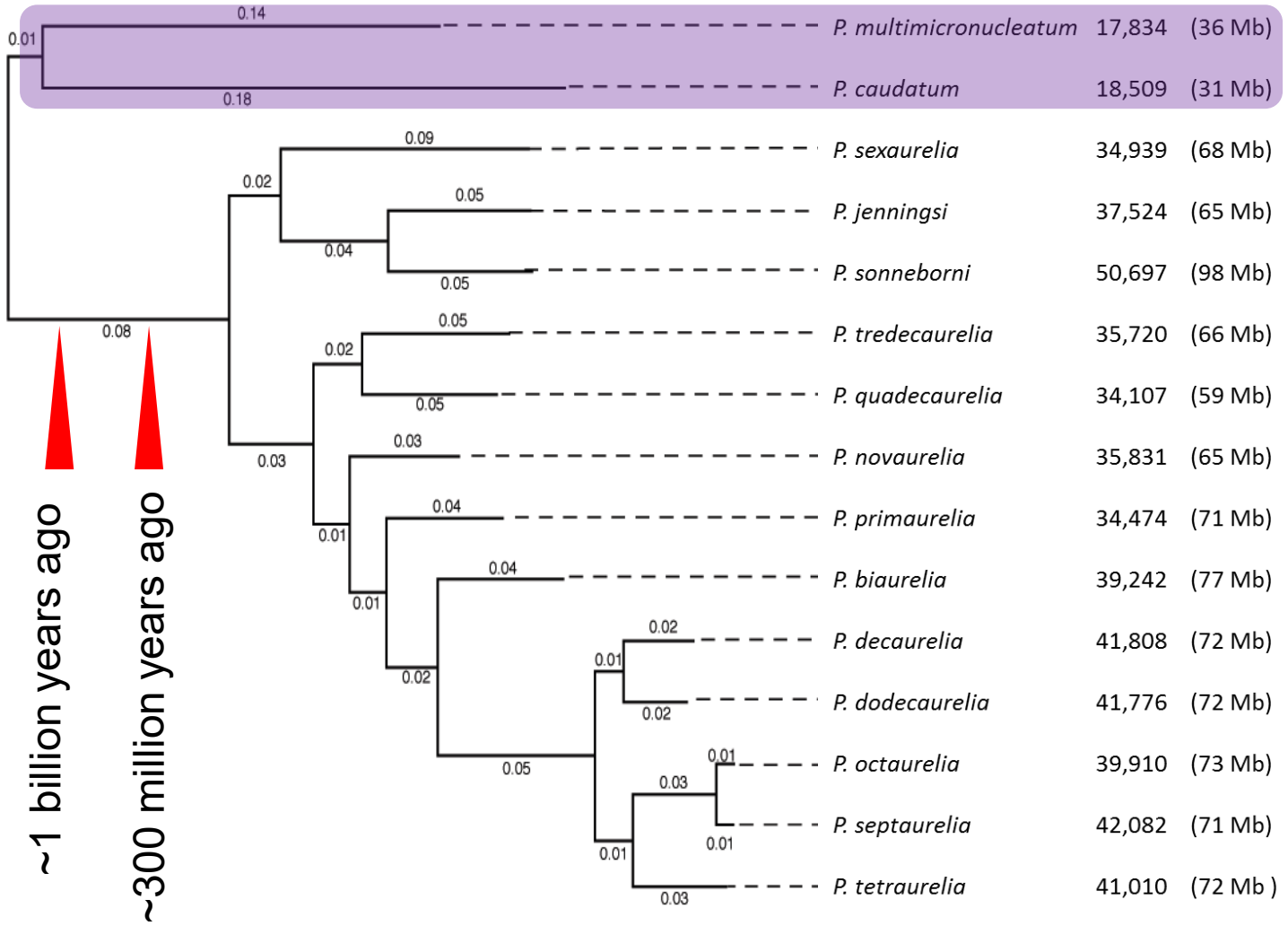
Susumu Ohno
(1928-2000)



www.youtube.com/watch?v=9Q1EkWtff2I

Two Whole-genome Events Preceded the Origin of the *Paramecium aurelia* Group: a cryptic species complex with no discernable morphological differentiation after 100s of millions of years.





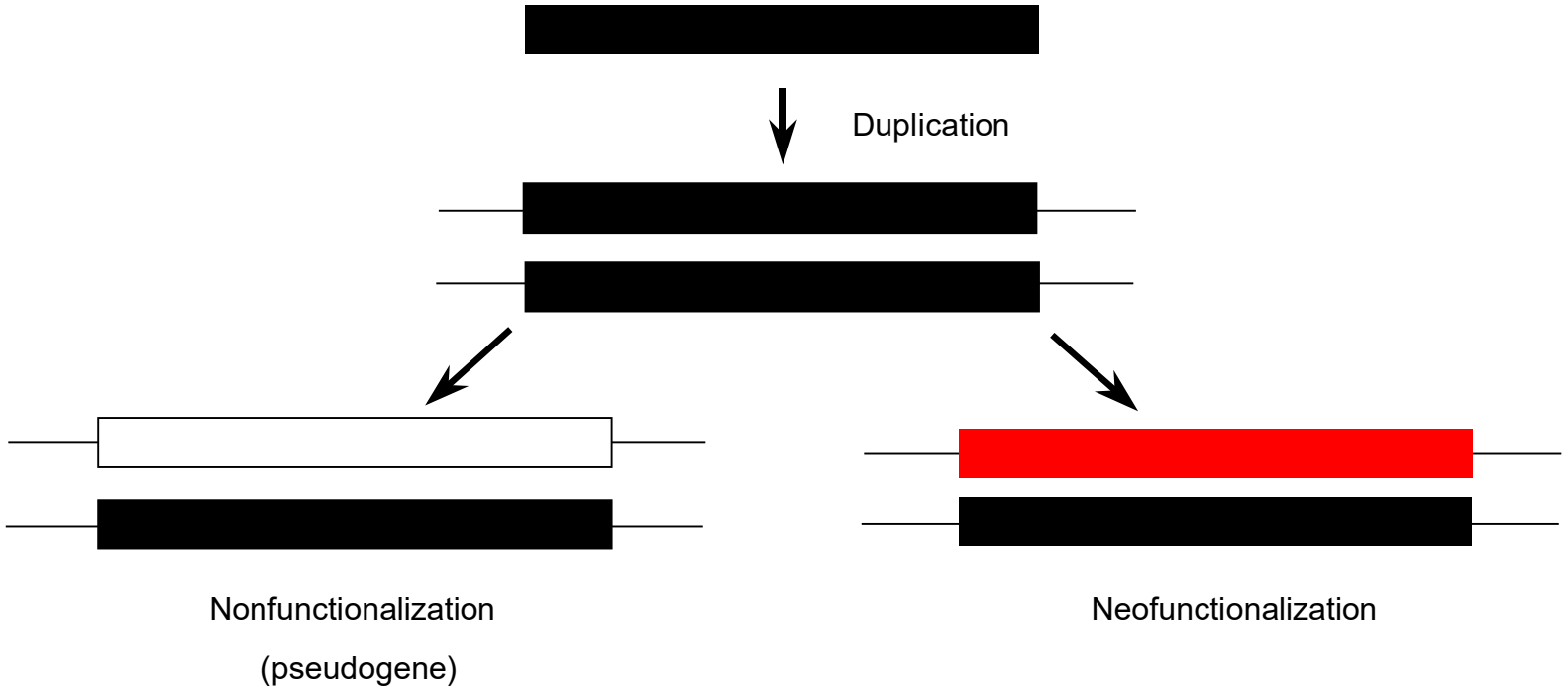
	Gene Number	Genome Size
P. multimicronucleatum	17,834	(36 Mb)
P. caudatum	18,509	(31 Mb)
P. sexaurelia	34,939	(68 Mb)
P. jenningsi	37,524	(65 Mb)
P. sonneborni	50,697	(98 Mb)
P. tredecaurelia	35,720	(66 Mb)
P. quadecaurelia	34,107	(59 Mb)
P. novaurelia	35,831	(65 Mb)
P. primaurelia	34,474	(71 Mb)
P. biaurelia	39,242	(77 Mb)
P. decaurelia	41,808	(72 Mb)
P. dodecaurelia	41,776	(72 Mb)
P. octaurelia	39,910	(73 Mb)
P. septaurelia	42,082	(71 Mb)
P. tetraurelia	41,010	(72 Mb)

Tracy Sonneborn
(1905 – 1981)

Casey “Bombshell Shock” McGrath

The Classical Model for the Fates of Duplicate Genes

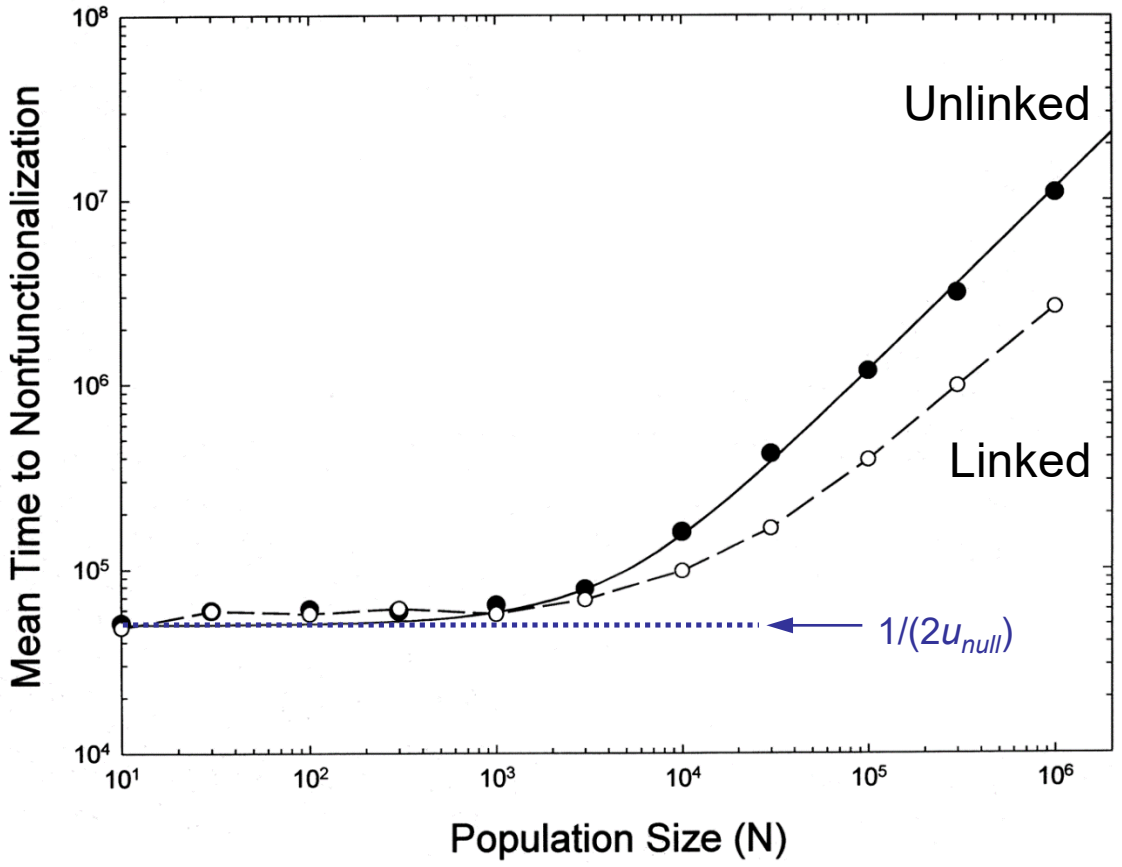
J. B. S. Haldane



H. J. Muller

Under the classical model, the mean time to duplicate-gene loss:

- is no more than a few million generations (the reciprocal of the mutation rate to nulls),
- increases with increasing population size.

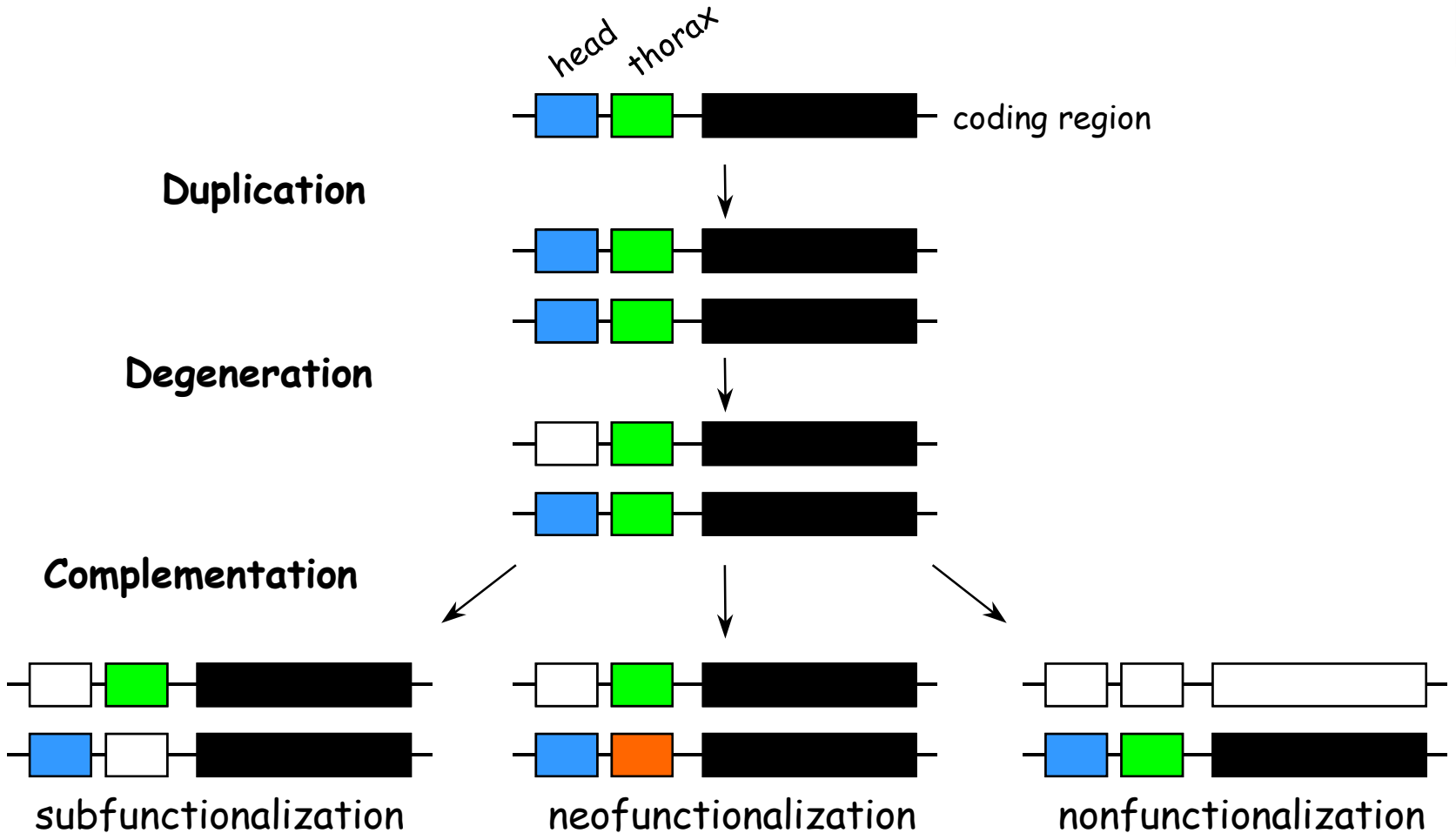


Long Half-Lives of Duplicate Genes in Polyploid Species

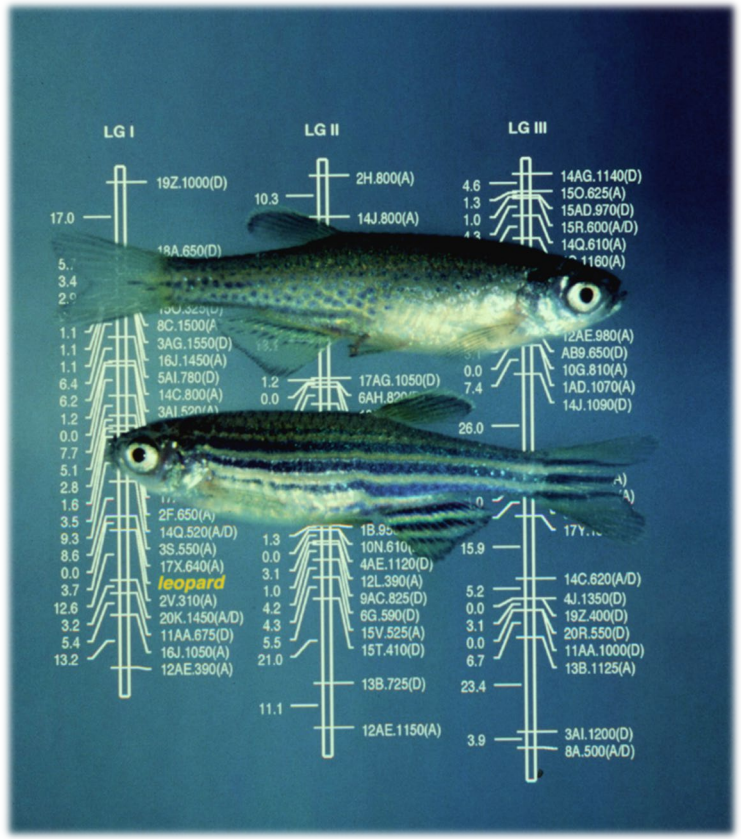
Phylogenetic group	Fraction Surviving	Half-life of Duplicates (millions of years)	
Salmonids	0.50	100	Allendorf et al. (1975)
Catastomid fish(suckers)	0.50	50	Ferris and Whitt (1979)
Cyprinus carpio (carp)	0.60	16	David et al. (2003)
Loaches	0.25	14	Ferris and Whitt (1977)
Arabidopsis thaliana	0.33	31	Ermolaeva et al. (2003)
Oryza sativa (rice)	0.34	35	Vandepoele et al. (2003)
Saccharomyces cerevisiae	0.08	27	Wolfe and Shields (1997)

The Duplication / Degeneration / Complementation Model

Force et al.
(Genetics, 1999)

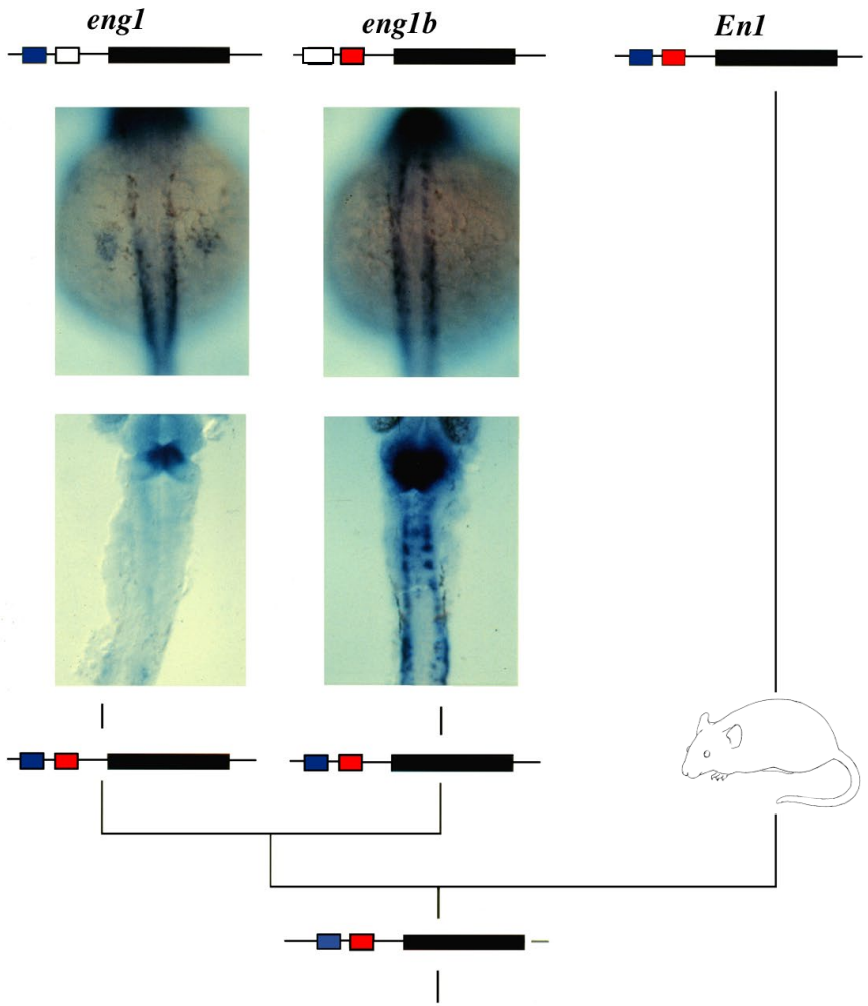


soundcloud.com/bendingspoons

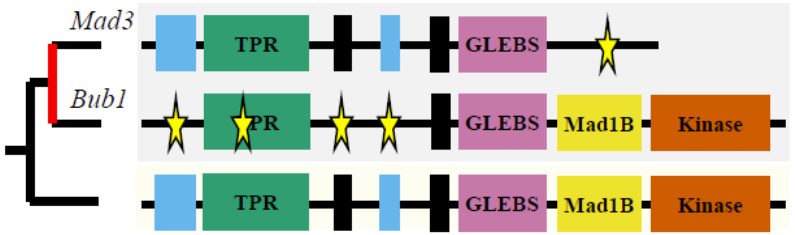


Duplicate *engrailed* genes in zebrafish

Single copy in tetrapods



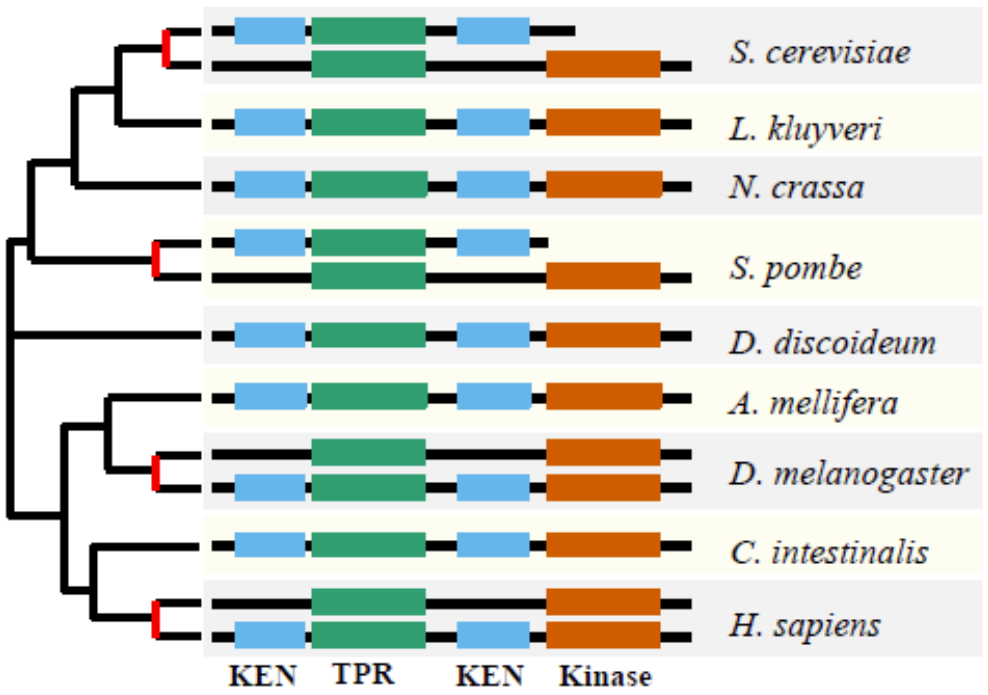
An Example of Parallel Subfunctionalization: the spindle-checkpoint pathway



Has an APC inhibitor domain, and is constitutively expressed in nuclei.

Has kinase domain, and is localized to kinetochores

Single-copy gene in *Lk*



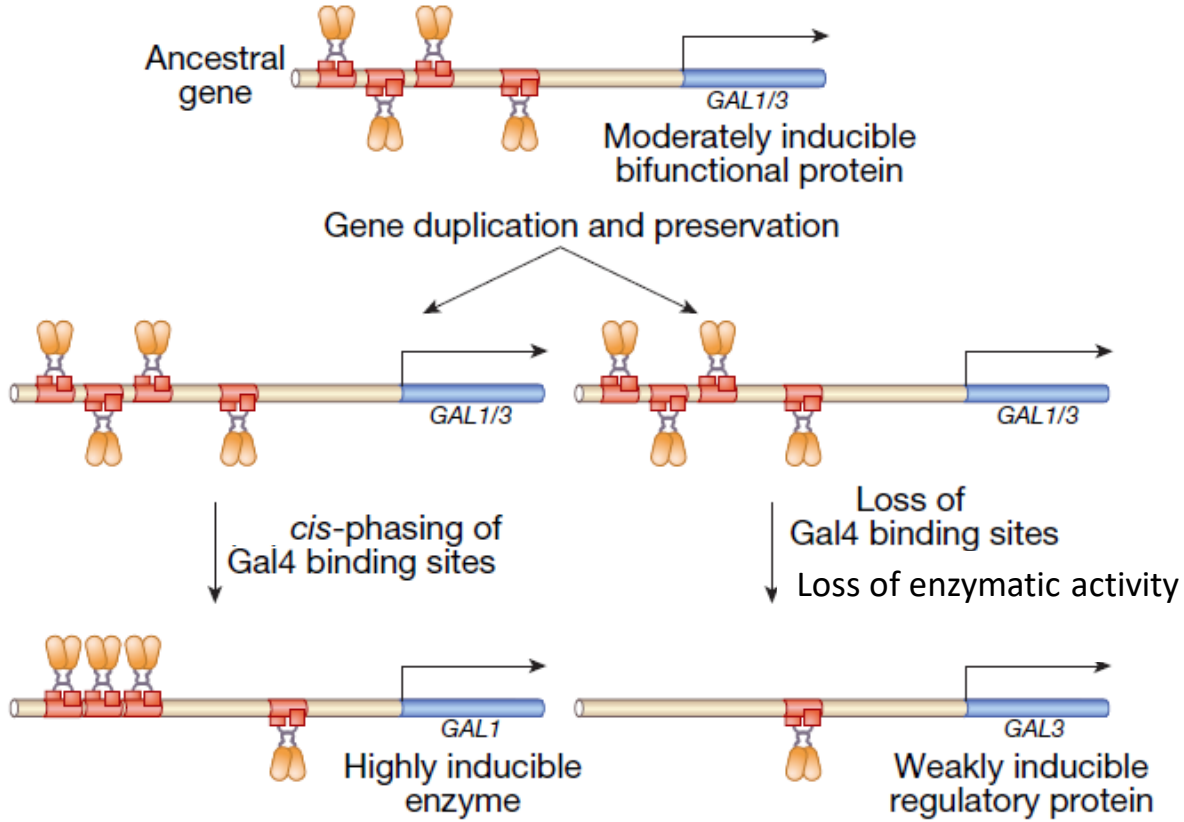
When the duplicate copies of *Sc* are replaced with the single-copy gene from *Lk*, full functionality is retained.

Parallel duplications in diverse phylogenetic lineages.

(Nguyen et al. 2017, PNAS)

Preservation by Adaptive-Conflict Resolution

- Two sister genes are involved in galactose utilization, one (Gal3) playing a regulatory role in pathway induction and the other (Gal1) serving as a galactokinase.
- By reference to an outgroup species, it was determined that the ancestral single-copy gene served both functions.
- Gene duplication then allowed the refinement of regulatory-site configurations that had previously been constrained in the ancestral gene, enabling the emergence of a more tightly regulated system.

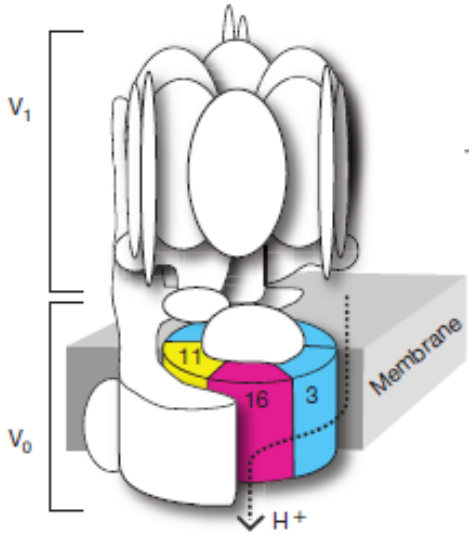


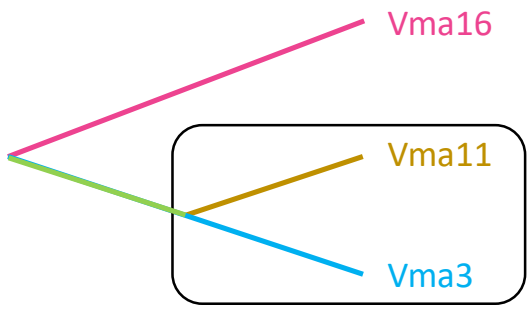
Modern *S. cerevisiae* Genes

Hittinger and Carroll (2007)

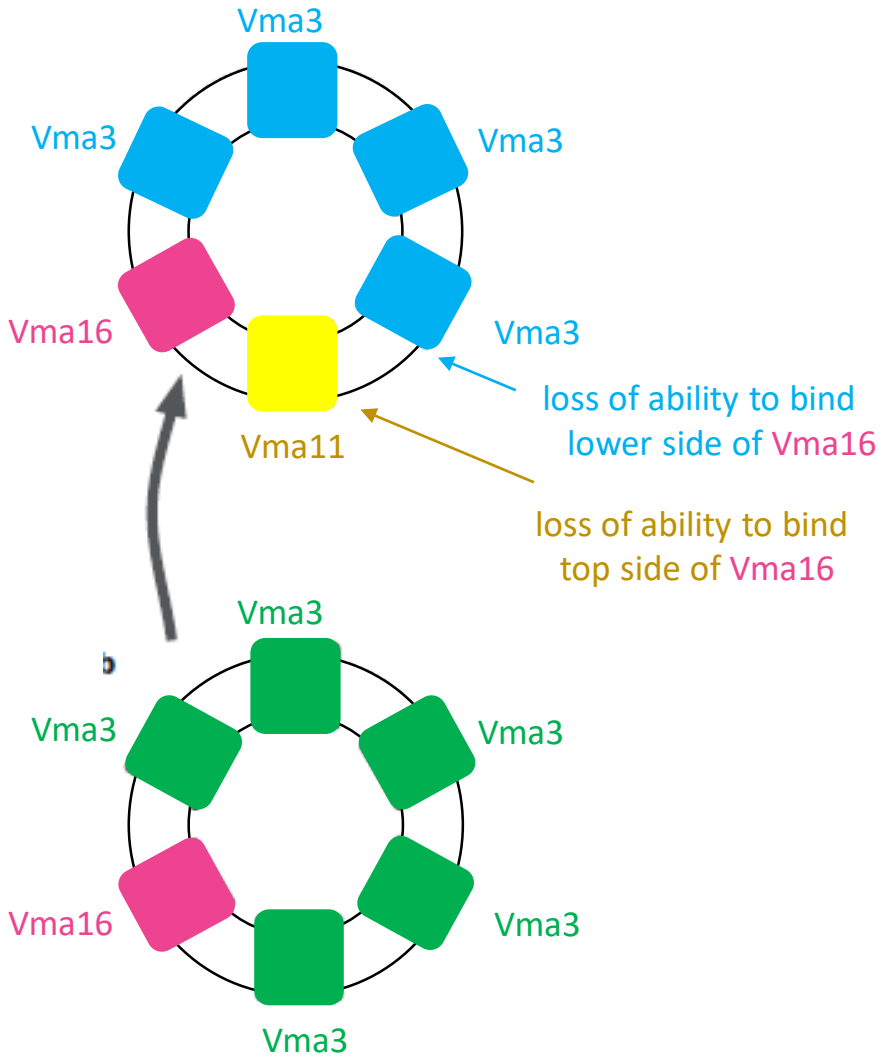
An Example of Subfunctionalization by Structural Alterations: vacuolar ATP synthase.

- In most eukaryotes, the ring consists of five copies of one protein (Vma16) and one of another (Vma3), both of which arose from an ancient gene duplication.
- In fungi, a third duplicate (Vma11) that arose from Vma3 replaces one subunit of Vma16, specifically resides between Vma16 and Vma3.
- One side of Vma3 lost the ability to bind to one side of Vma16, whereas the other side of Vma11 lost the ability to bind to Vma3.
- No evidence that this increase in complexity has endowed yeast with increased fitness.



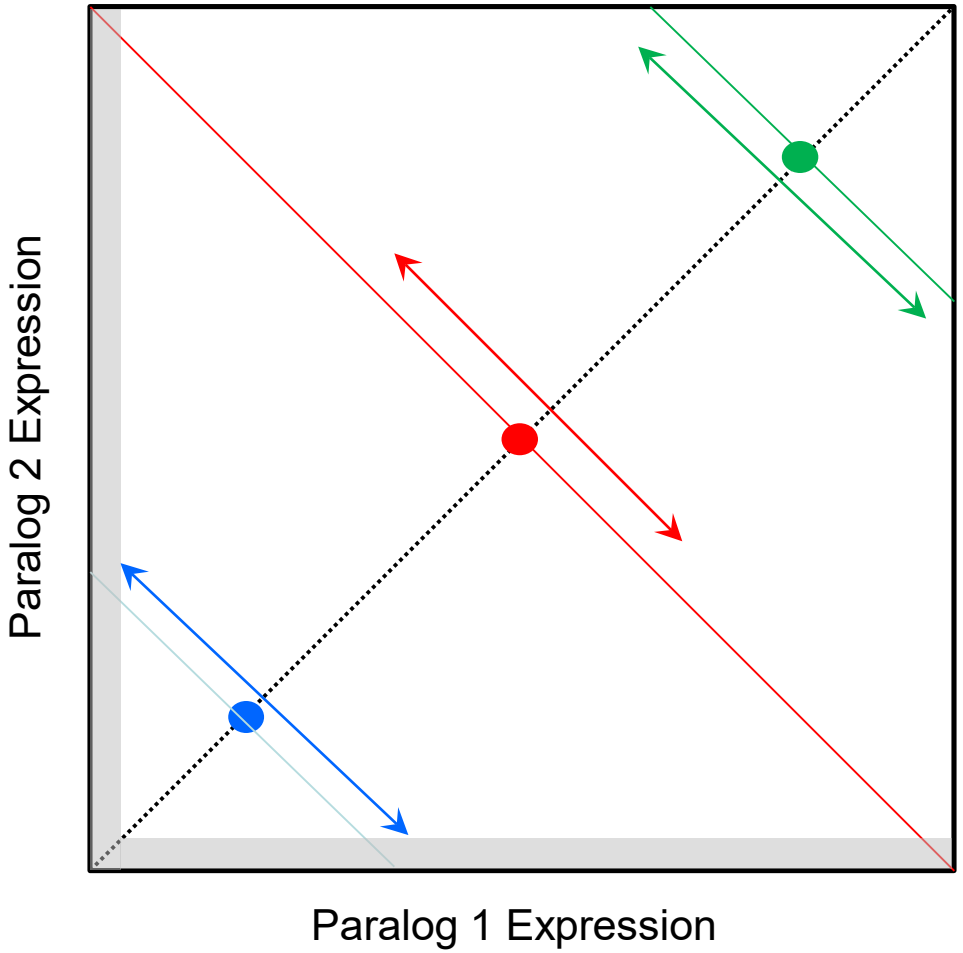


fungal-specific duplication

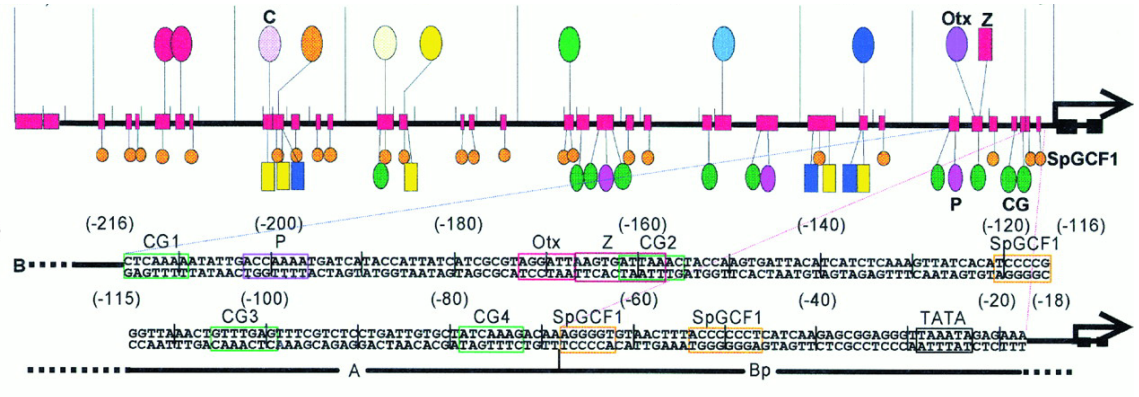


Finnigan et al. (2012)

Quantitative Subfunctionalization: Joint Meandering of Paralog Expression Along the Drift Barrier



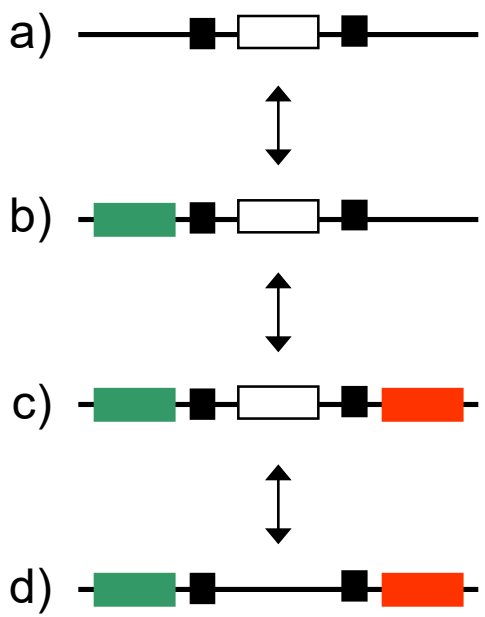
How do independently mutable subfunctions originate?



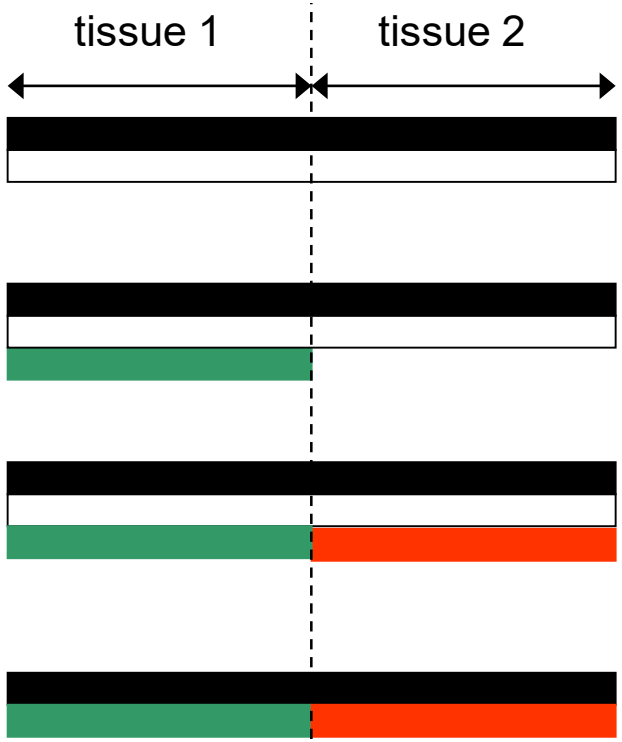
Endo16 cis-Regulatory Structure in Sea Urchin

From: Yuh, Bolouri, and Davidson (1998, Science).

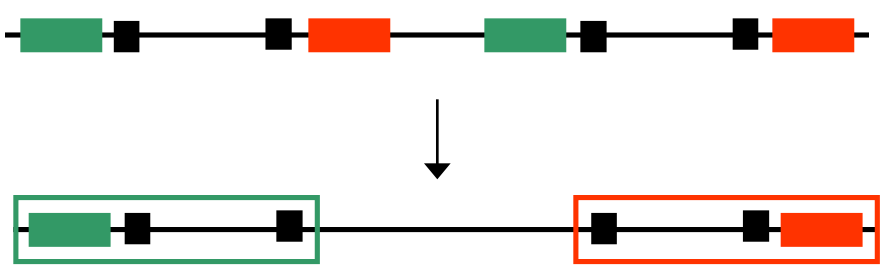
Accretion, degeneration, and partial replacement of TFB sites



Transcription-factor utilization



Duplication, degeneration, and complementation



Evolutionary Novelty Out of Degenerative Mutations and Random Genetic Drift

- Gene duplication is a primary mechanism for the origin of organismal complexity, with neofunctionalization of one member of a pair providing a potential route to the origin of novel gene features.

- More commonly, duplicate genes are preserved by the partitioning of ancestral gene functions via complementary degenerative mutations. Subfunctionalization is facilitated in populations with small effective sizes.

- The same processes that lead to subfunctionalization of duplicate genes promote the evolution of modular forms of gene structure upon which the process of subfunctionalization depends.

- Thus, by facilitating the recurrent emergence and partitioning of gene subfunctions, reduced effective population sizes can lead to the passive increase in organismal complexity without any direct selection for such changes.

- Subfunctionalization eliminates pleiotropic constraints unique to single-copy genes, and as a preservational process, buys time for the neofunctionalization process.

Stochastic, Divergent Losses of Duplicate Genes Lead to the Passive Origin of Reproductive Isolating Barriers

- Did whole-genome duplication prior to the last eukaryotic common ancestor passively promote the explosive establishment of reproductively isolated lineages?