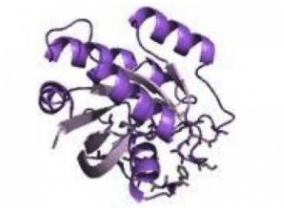
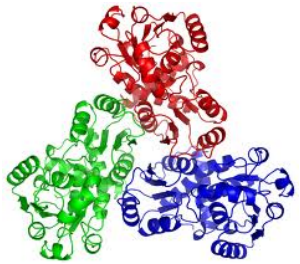
The Origin of Variation in Molecular Complexes

Driven by adaptive processes unique to individual lineages?

Or a consequence of biased mutation pressure and/or random drift?

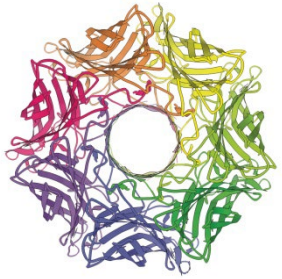


monomer

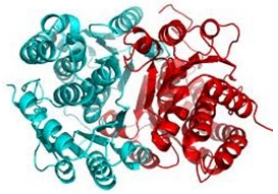


trimer

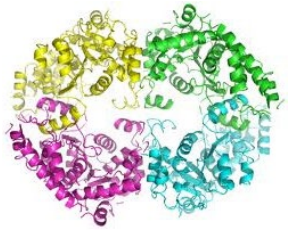
pentamer



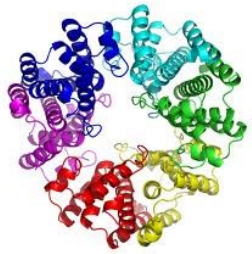
heptamer



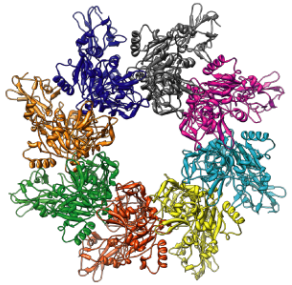
dimer



tetramer



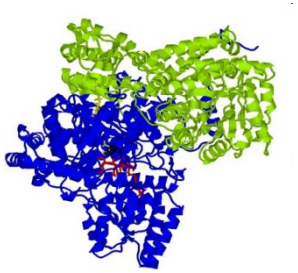
hexamer

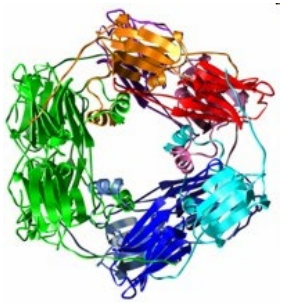


octamer

Most Cellular Components are Assembled from Protein Subunits Derived from the Same Gene or from Related Duplicate Genes

- Metabolic enzymes
- Transcription factors
- Cytoskeleton
- Flagella
- Nuclear pore complexes
- Chaperones and proteasomes
- Transporters and ion channels
- Nucleosomes and chromatin-remodeling complexes





Summary of Comparative Observations on Multimeric Structures

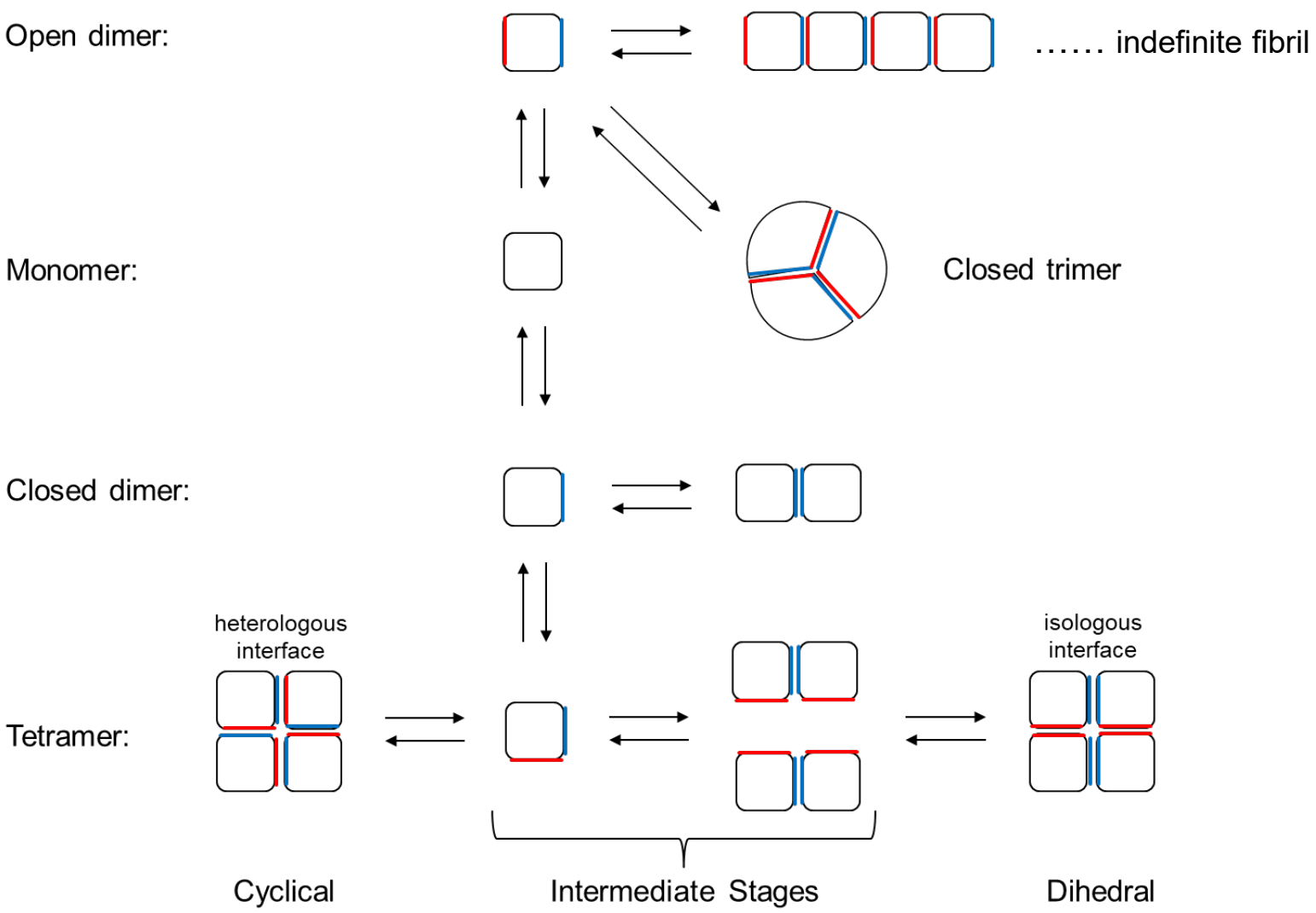
- The majority of proteins operate as multimers, not monomers.

- There is substantial variation in multimeric states both within and among phylogenetic groups.

- No tendency for more complex organisms to harbor more complex molecules – **in striking contrast to what is seen with the complexity of gene structure and genome architecture**.

- No evidence that multimers are generally superior in performance to monomers.

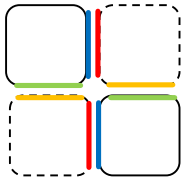
Interfaces and Structures for Homomers



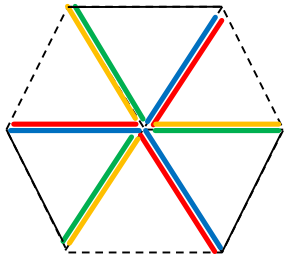
Structural Topologies of Heteromers

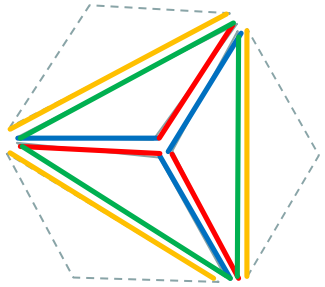
Heterodimer:

Tetramer
(dimer of dimers):

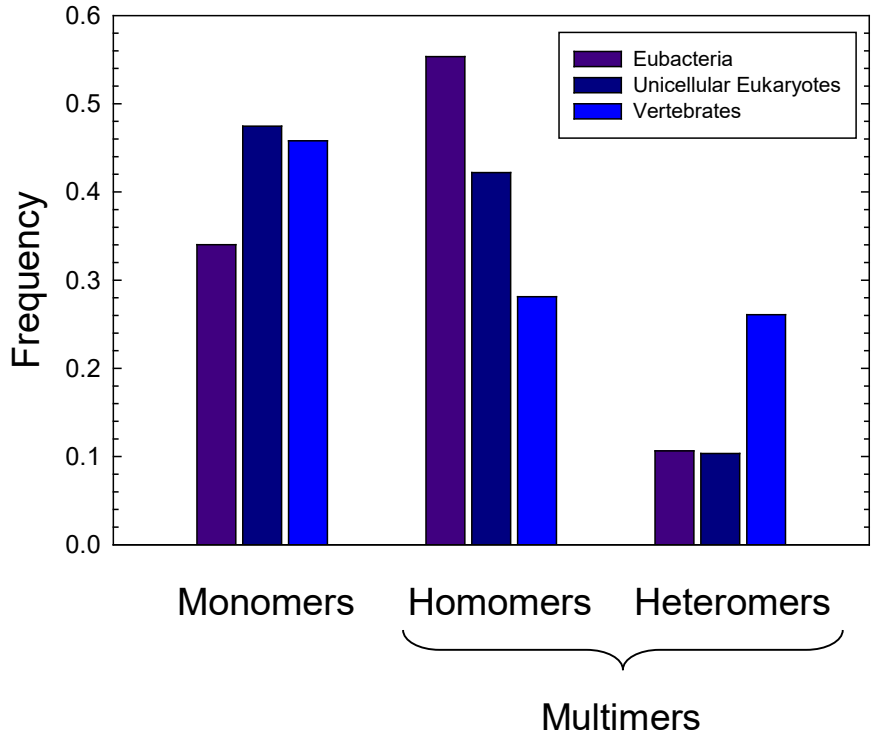


Hexamer
(trimer of dimers):



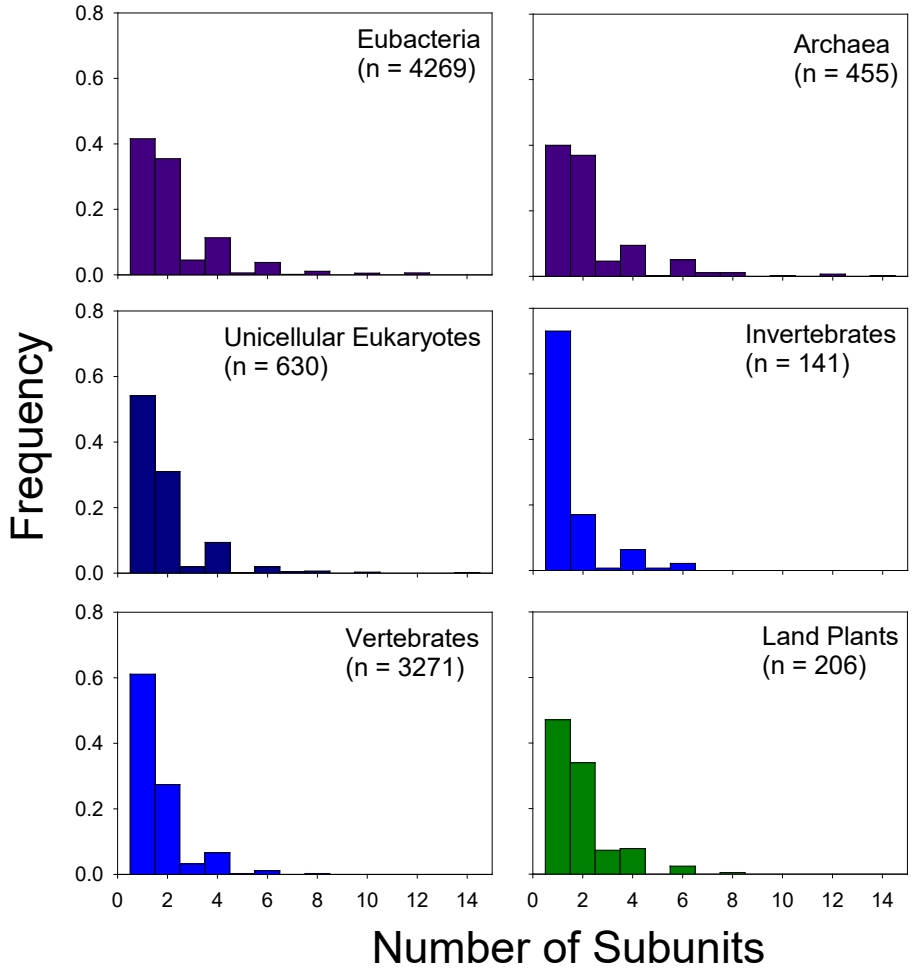


Distribution of Complex Types



- ~40% of proteins with known structures are dimers or higher-order complexes, roughly independent of phylogenetic lineage.
- Most multimers are homomers, with all subunits derived from the same locus.

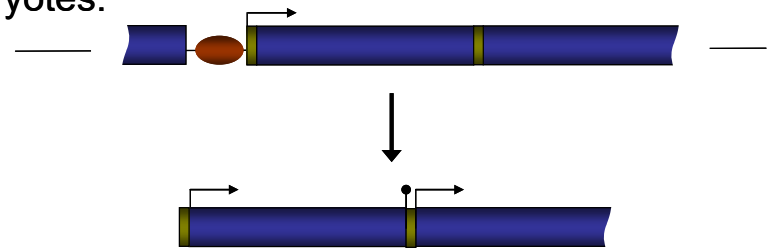
Distribution of Homomeric Types: Approximate Constancy Across the Tree of Life



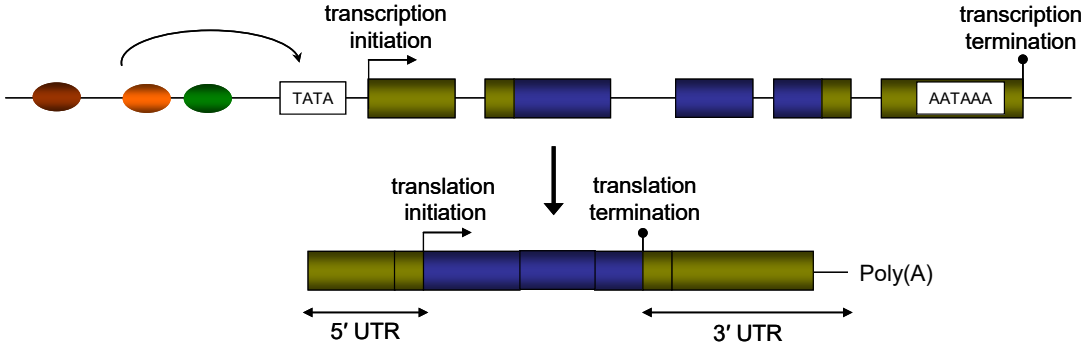
- Distributions are independent of phylogenetic lineage.
- Approximately negative exponential in form.
- Oddmers are underrepresented relative to evenmers
- No gradient in the level of protein architectural complexity with organismal complexity.

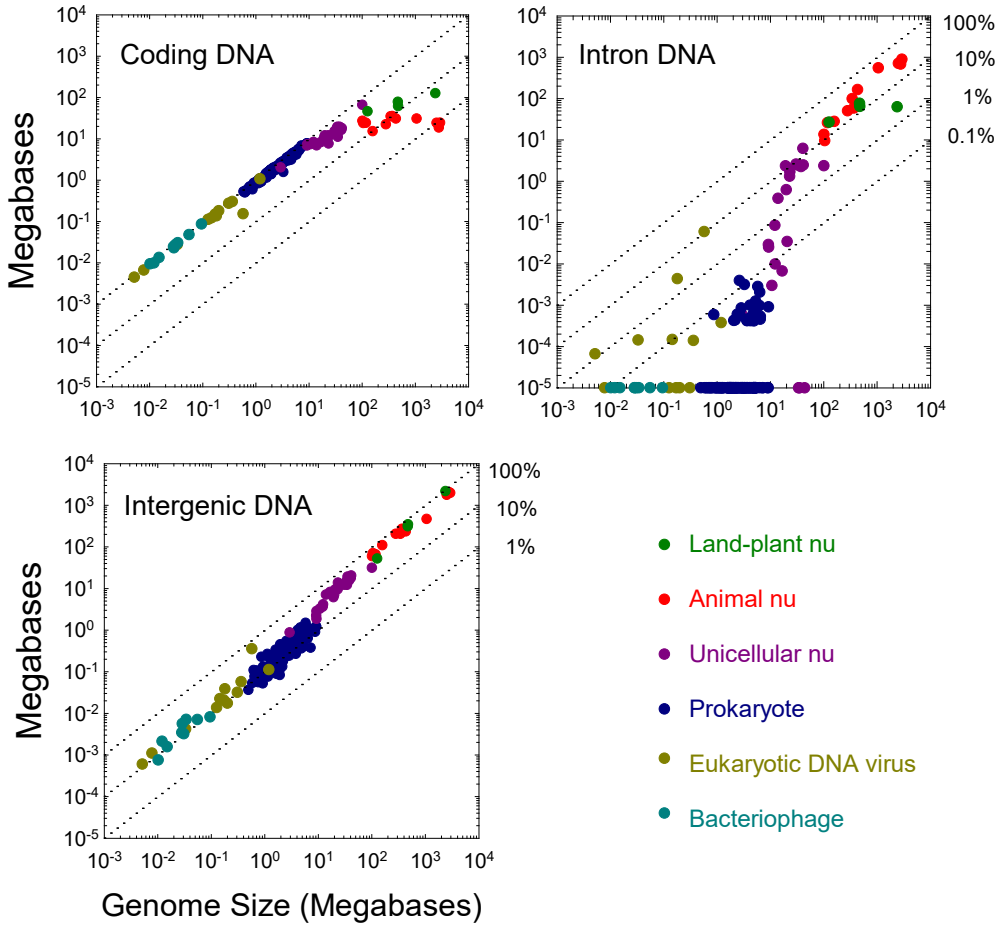
The Expansion of Gene-structure and Genome Complexity with Organismal Complexity

Prokaryotes:

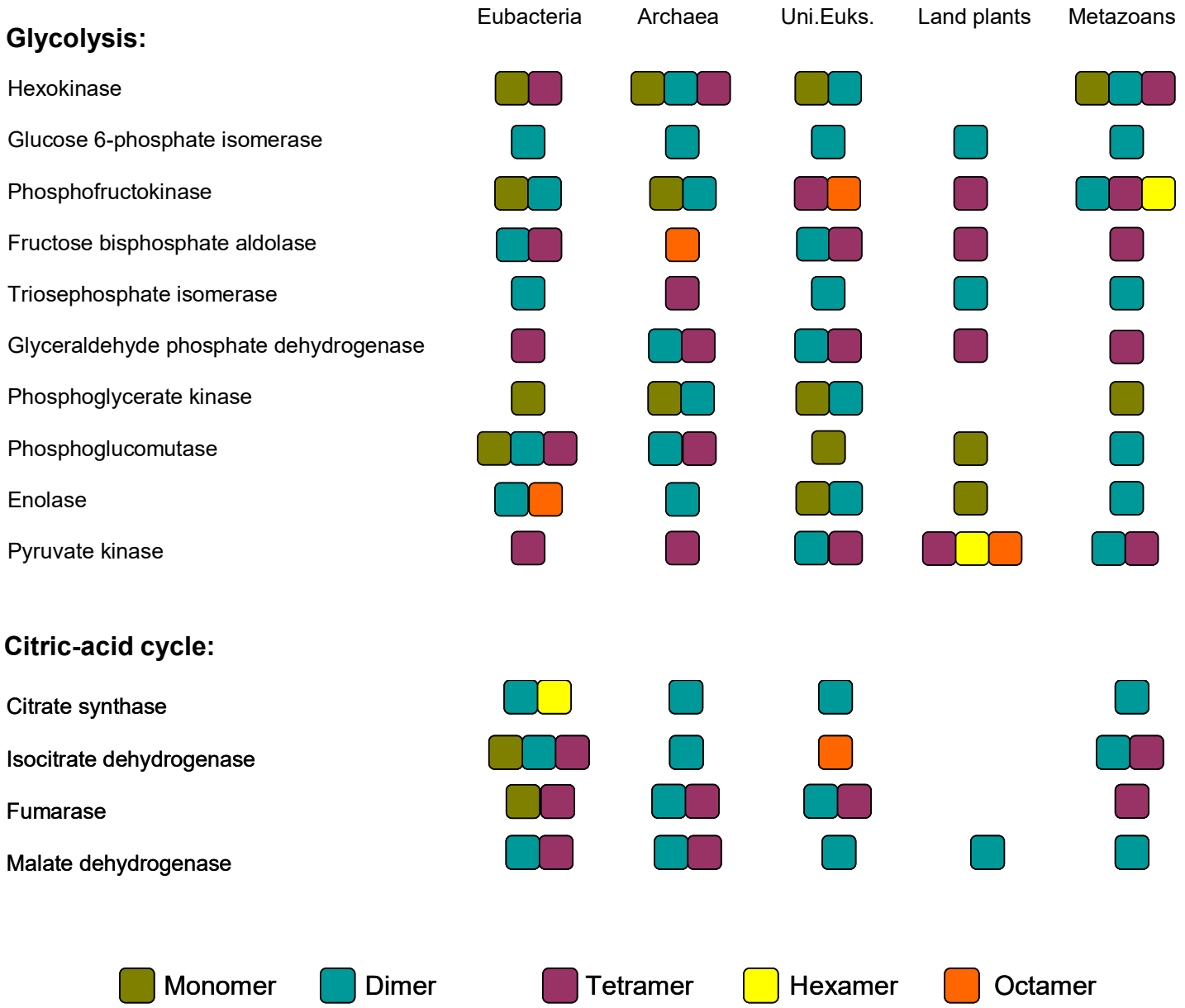


Eukaryotes:





Known Oligomerization Structures for the Enzymes of Central Metabolism



Enzymes with Identical Multimeric States Need Not Have the Same Structural Basis

Dihydrodipicolinate synthase (involved in lysine synthesis)

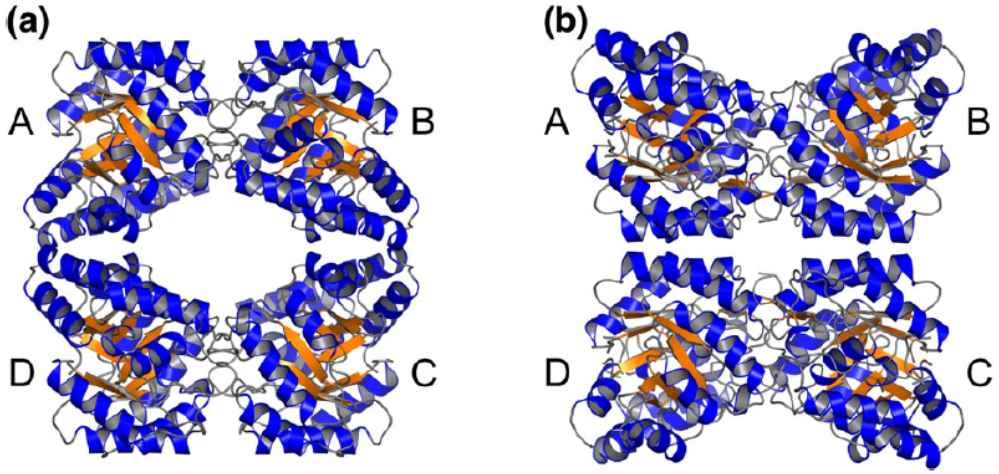


Fig. 1. The X-ray crystal structures of DHDPS from (a) *E. coli*[18,19] and (b) *N. sylvestris*.[23] Each enzyme is a homotetramer of $(\beta/\alpha)_8$-barrels composed of two tight-dimer units (A–B and C–D), but the arrangement of the two dimeric units is different. (Griffin et al. 2008).

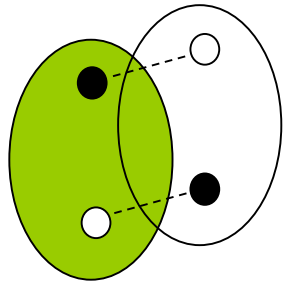
- About 70% of protein families containing homomers exhibit phylogenetic variation in the binding interfaces (Dayhoff et al. 2010).

Some Pure Biophysical Explanations for Frequent Homomers

- To ensure stable complexation, interfaces must overcome the energetic cost of thermal motion.

  Random symmetric interfaces are more likely to generate extremes of binding strength than random asymmetric interfaces (Lukatsky et al. 2007; Andre et al. 2008).

- *Two for the price of one*: any pair of adhesive residues in a symmetric interface must be present twice (Monod et al. 1965).

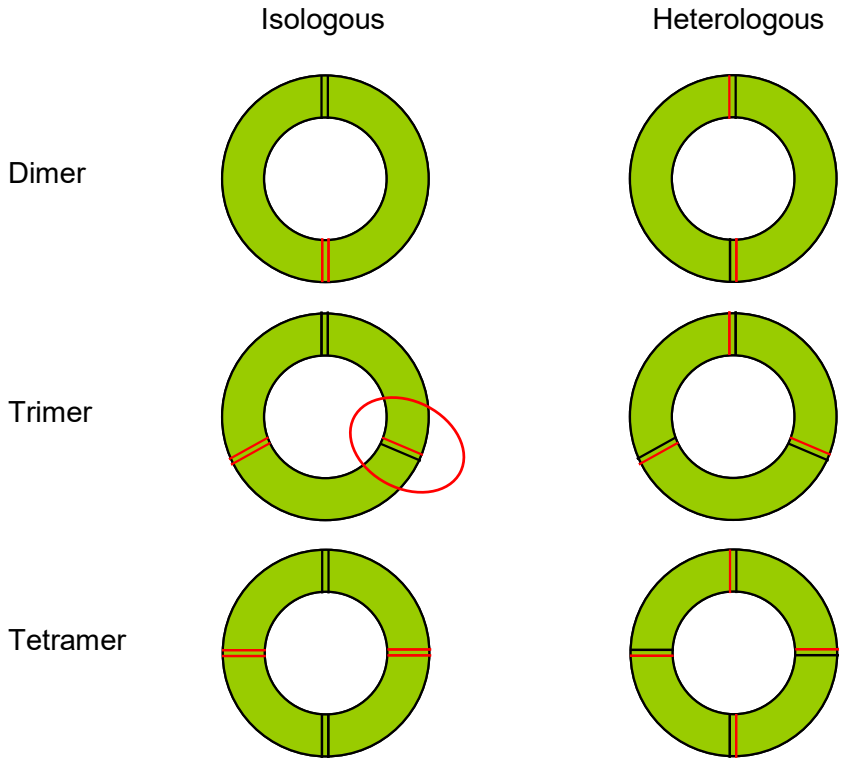


  - However, deleterious AA changes are twice as severe.
  - Heterologous interfaces provide twice the number of opportunities for mutations for adhesion.

- Linkage of sites in the same gene enhances the opportunity for coevolution.

Simple Geometric Limitations on Oddmers

- Can’t take the “two for the price of one” route because an isologous structure cannot be completed.



- Unless equipped with the correct angular orientation for creating a closed loop, heterologous interfaces encourage concatenation into endless fibrils.

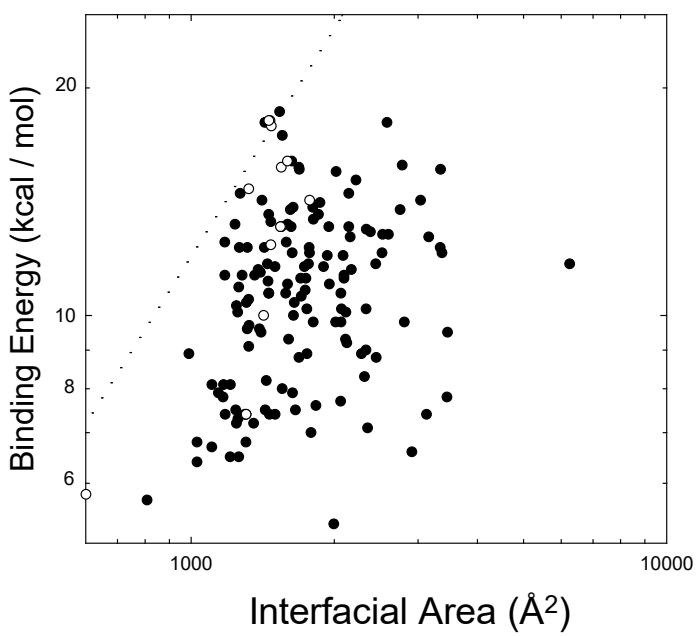
Closed dimer

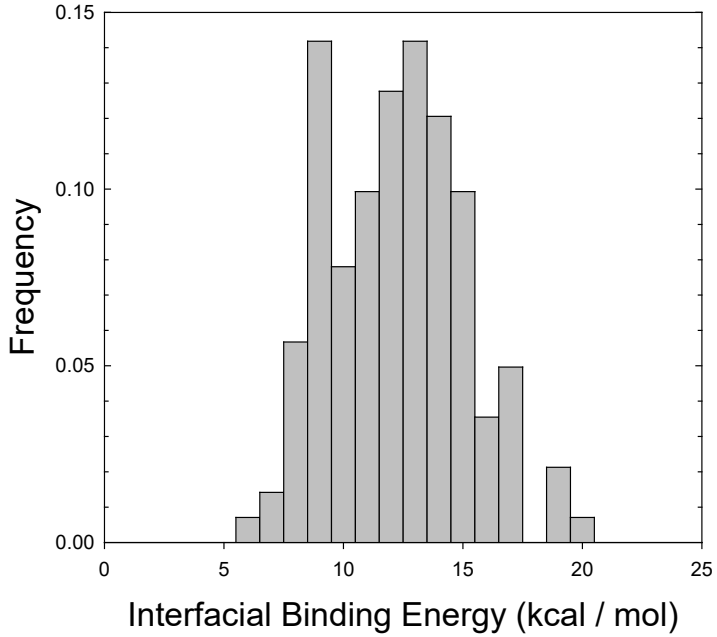
Open-ended with two interfaces

Most Interface Binding Strengths Are on the Margin of Stability

Kastritis et al. (2011)

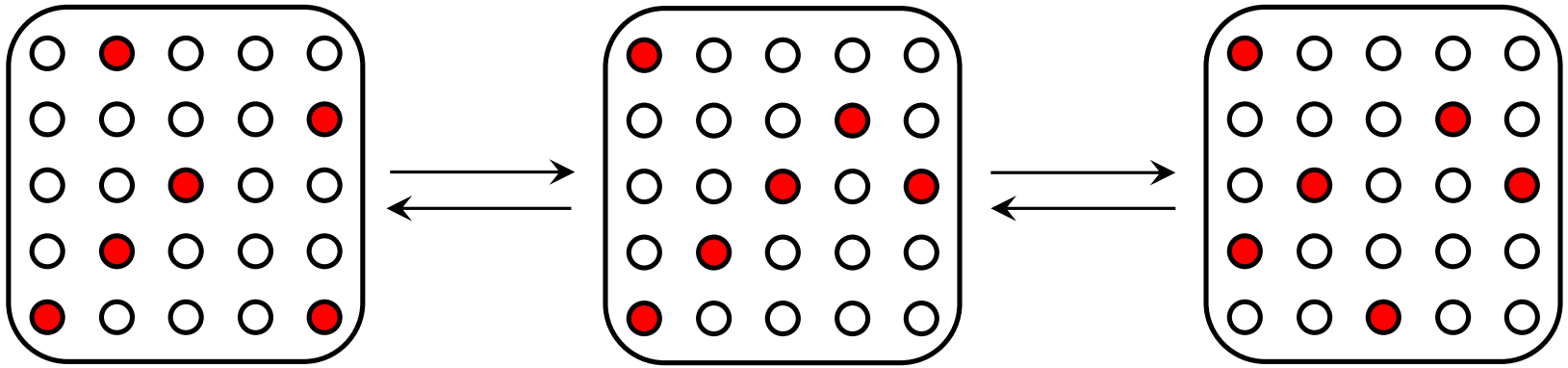
(1 Angstrom = 0.1 nm)





- On average, binding strengths are ~10 to 20x the energy associated with background thermal motion.
- There is only a weak positive relationship between interface size and stability.
- Typically, <10 residues are involved in binding, and removal of 1 or 2 is sufficient to eliminate binding.

Excess Degrees of Freedom in Interface Patches Enables Wandering of Key Residues



- Can lead to passive emergence of interspecies incompatibilities.

Promotion of Multimers by Selection or Effective Neutrality?

- Potential advantages to complex formation:
  - increased structural diversity,
  - increased enzyme size and reduced surface area will increase productive encounter rates with substrate,
  - reduced problems of folding single large proteins,
  - reduced vulnerability to denaturation and/or engagement in promiscuous interactions,
  - reduced molecular motion at the catalytic site increases substrate specificity,
  - increased flexibility for allosteric regulation.

- Potential costs of oligomerization:
  - Elevated production levels necessary for a critical encounter rate for successful multimerization.
  - Concatenation into indefinite filaments – human disorders involving the production of inappropriate protein aggregates include Alzheimer's, Parkinson's, and amyotrophic lateral sclerosis (ALS).

Are Multimeric Molecules Functionally Superior to Monomers?

- The mismatch-repair machinery in eubacteria employs monomers, whereas that in eukaryotes employs dimers.
  Yet, MMR efficiency is greater in bacteria.

- Sliding clamps used in DNA replication are homodimers in bacteria, but homotrimers in eukaryotes.
  Yet, replication-fork progression rates are an order of magnitude greater in eubacteria.

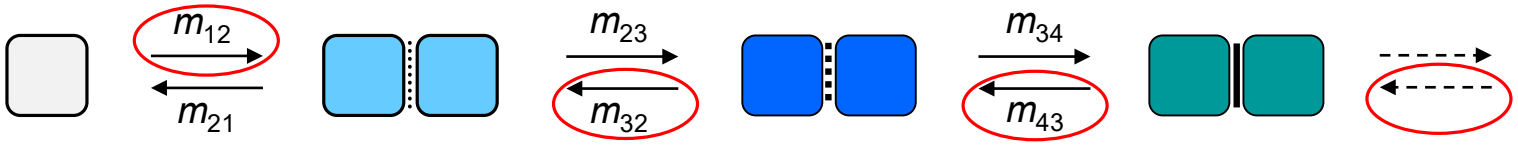
- The protein repertoire of eukaryotic ribosomes is substantially more complex than that in prokaryotes.
  Yet, the level of translation fidelity is no greater (and possibly lower) in eukaryotes.

- Class II amino-acyl tRNA synthetases are dimeric or tetrameric, yet monomeric class I synthetases are much less error-prone with respect to amino-acid charging.

Theory: Evolutionary Determinants of Oligomeric States; Monomers → Strong Dimers

- What are the joint roles of mutation bias, selection, and drift in the relative probabilities of establishment of alternative forms?
    - A/T mutation pressure promotes the evolution of more hydrophobic (and interactive) amino acids.
- Are the most common evolutionary states the optimal states?
- How much variation is expected among species subject to identical population-genetic environments?

Evolution of a Dimeric Interface

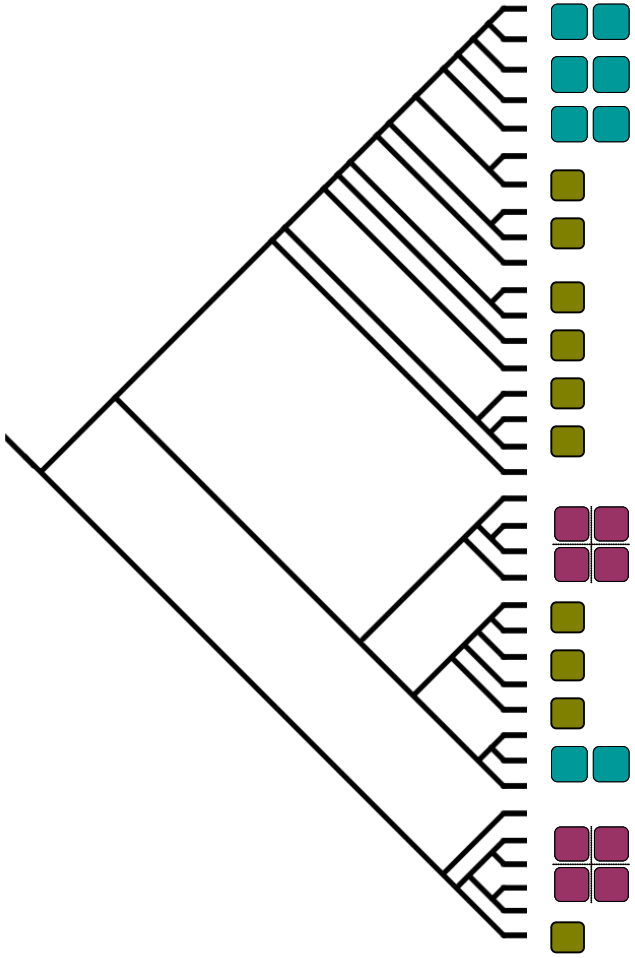


- Each transition rate is equal to the product of the number of relevant mutations arising per generation and the fixation probability.

- At steady state, the flux rate must be equal in both directions. This means that the net rate of establishment of dimers from monomers must equal the reverse rate.

- The equilibrium probability of each state is simply proportional to the product of the total set of transition rates towards the state from both directions.

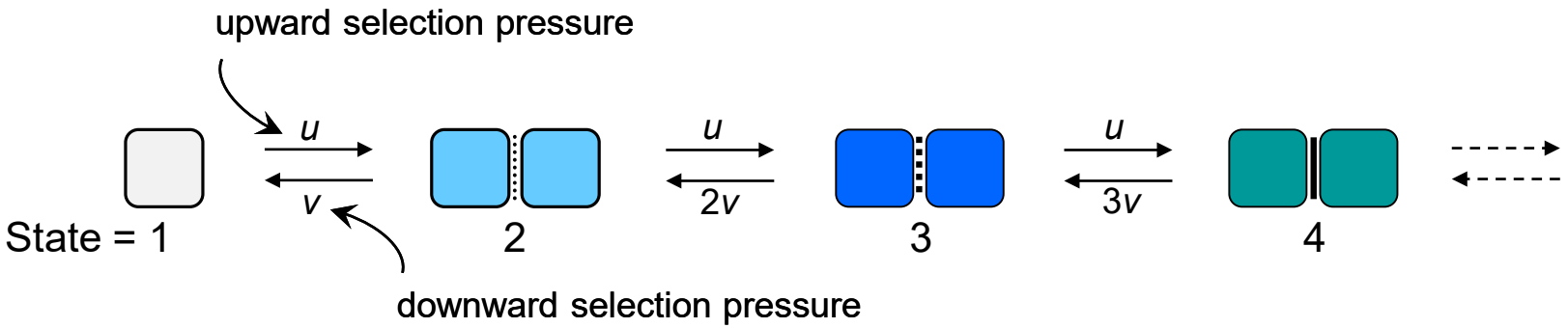
A Hypothetical Distribution of Multimeric Structures on a Phylogeny



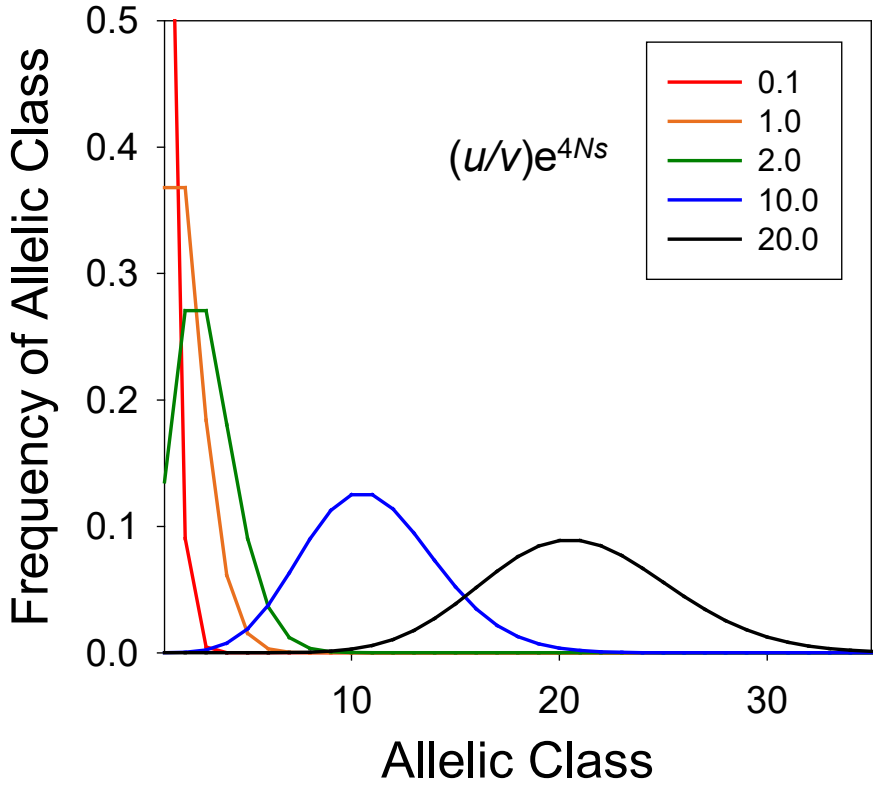
- Stochastic gains and losses of:

  monomers
  dimers
  tetramers

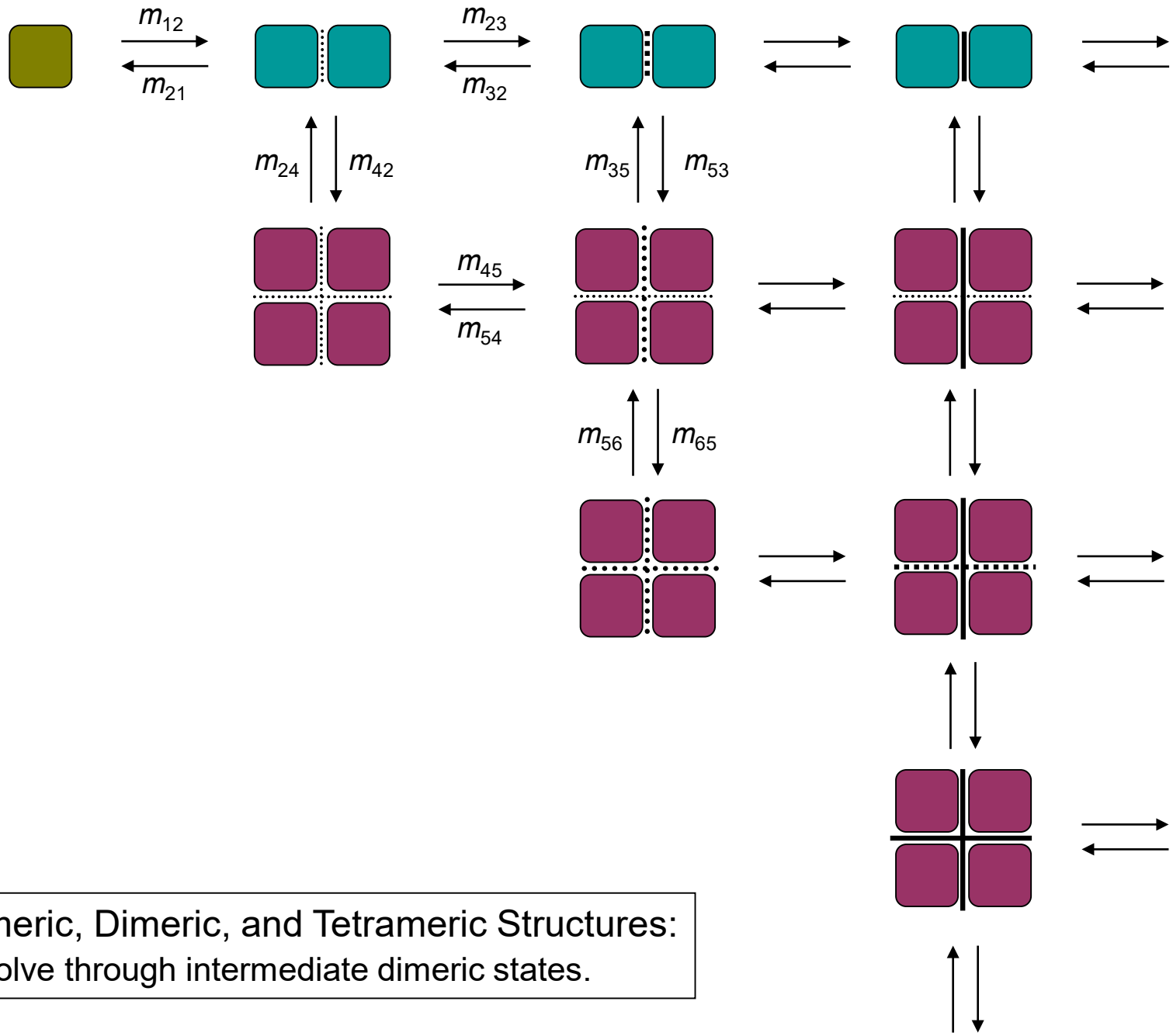
The Steady-State Evolutionary Distribution



- The probability distribution of alternative states is Poisson, with key parameter $(u/v)e^{4Ns}$, where N = effective population size, and s = selective advantage of each additional interface bond.



- Substantial phenotypic variation exists among lineages, even when selection and mutation are operating in identical manners in all lineages.
- The most common state is not necessarily the optimum.
- Under effective neutrality, the distribution is independent of N.



Transitions Between Monomeric, Dimeric, and Tetrameric Structures:
homotetramers generally evolve through intermediate dimeric states.

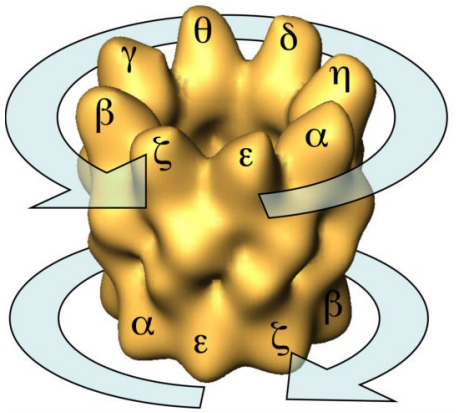
The Eukaryotic CCT Chaperonin Complex

- CCT (chaperonin containing tailless complex) proteins are hetero-hexadecamers (16 proteins in the total barrel; sometimes 18) in eukaryotes.

- The eight components diverged following a series of ancient gene duplications prior to LECA (Archibald et al. 2000).

- Each component is thought to have a specialized binding function, and sites known to be involved in binding seem to be under positive selection in Eukaryotes (Fares and Wolfe 2003).

- A classical case of a homomer becoming a heteromer through duplication, degeneration, and complementation.



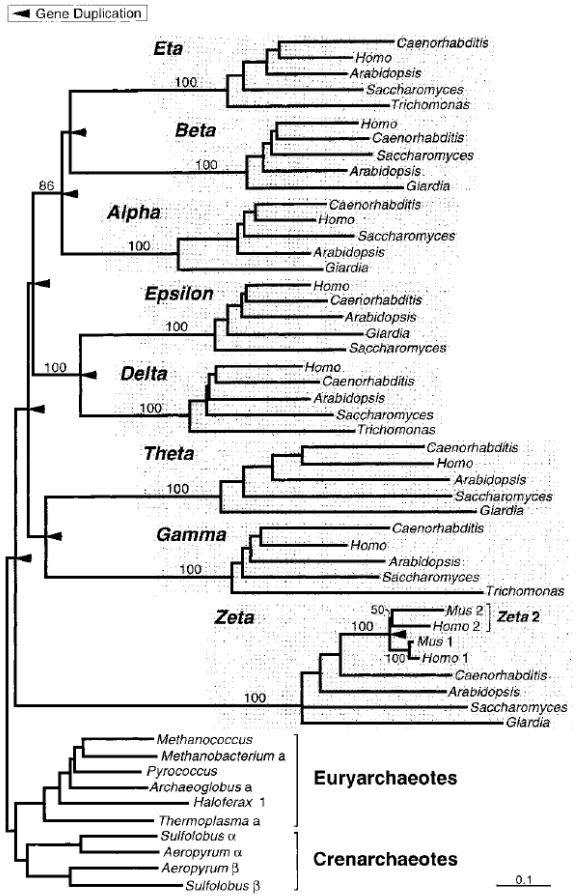


FIG. 2. Phylogeny of eukaryotic CCTs. The tree shown was constructed using the Fitch–Margoliash distance algorithm (Felsenstein, 1995) from an alignment of 53 sequences and 355 unambiguously aligned amino acid sites. The eight different CCT subunit families found in eukaryotes are highlighted in gray. For each CCT subunit family, 5 representative sequences were chosen to represent the full spectrum of eukaryotic diversity: two animals (*Homo* and *Caenorhabditis*) and one from each of fungi (*Saccharomyces*), plants (*Arabidopsis*), and protists (*Giardia* or *Trichomonas*). For CCTζ, the ζ-1 and ζ-2 subunits of *Homo* and *Mus* were also analyzed. The tree is rooted with a representative sample of archaeal chaperonin sequences. Support values for important nodes on the tree are given above the branches and were calculated by bootstrapping with 100 resampling replicates. The scale bar represents the estimated number of substitutions per amino acid site.

Gene Duplication and Chaperonin Evolution

- Archaeal chaperonins have 1 to 3 nonspecialized subunits; whereas eukaryotes have 8 specialized components.

- Parallel duplications leading to heteromeric structures have occurred in the archaea, and reversions to homomers have also occurred.

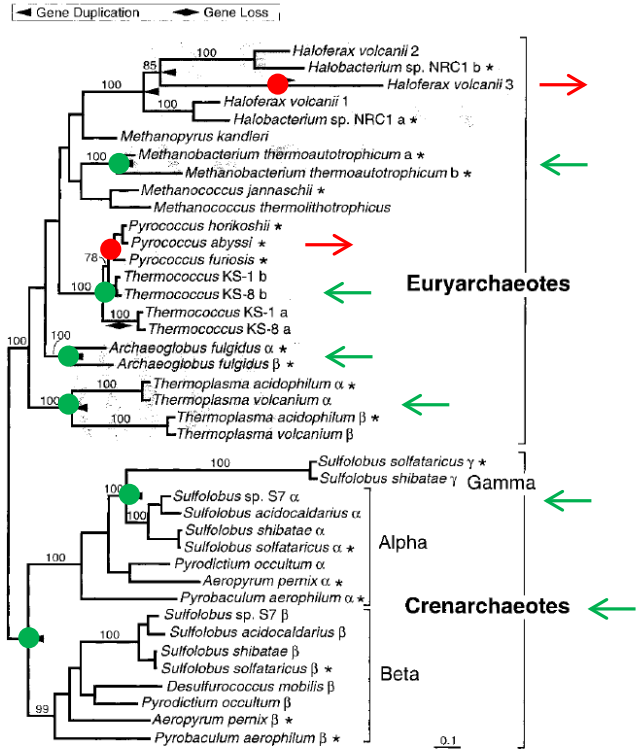


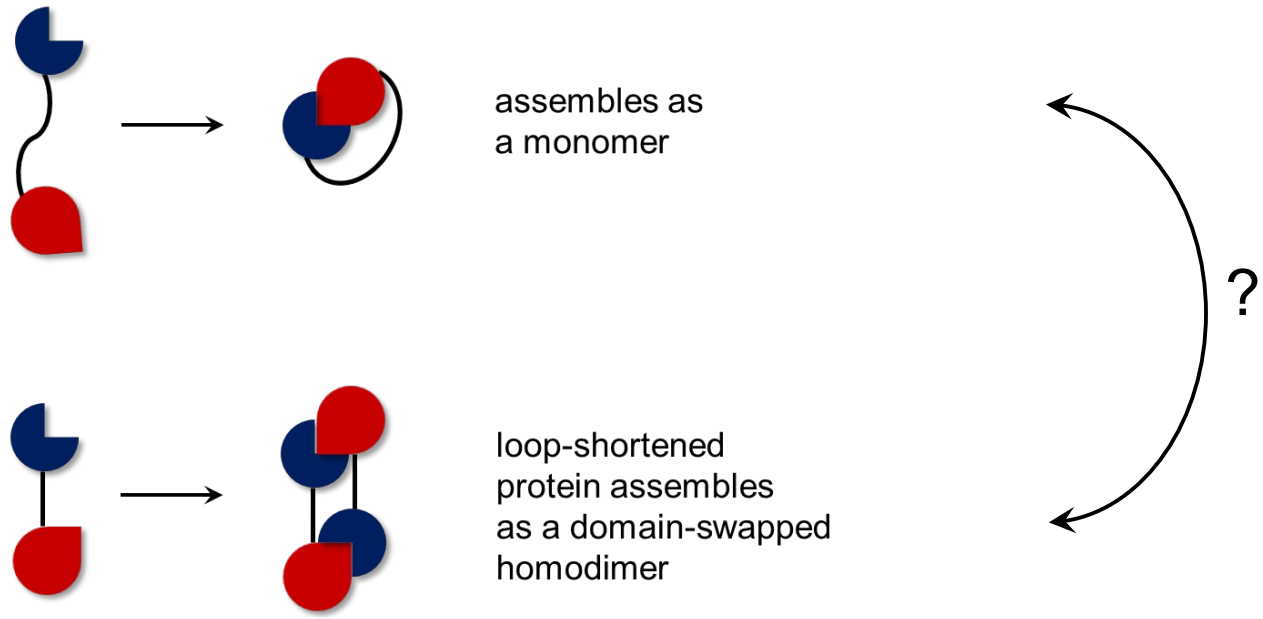
FIG. 1. Phylogenetic analysis of archaeal chaperonins. The tree shown is a ML tree (lnL −15614.72) inferred from a chaperonin protein sequence alignment containing 40 sequences and 452 unambiguously aligned amino acid positions. The two recognized kingdoms within Archaea (euryarchaeotes and crenarchaeotes) are labeled, and inferred gene duplications and gene losses are indicated (see text). Within euryarchaeotes, regions of the tree in which lineage-specific gene duplications have occurred are shaded. For crenarchaeotes, the three different gene/subunit families (α, β, and γ) are indicated. Asterisks appear next to sequences from organisms whose genomes have been completely sequenced. Statistical support values for significant nodes appear above the branches (ML RELL bootstrap values; inferred from a heuristic search of 1000 trees in protML) (Adachi and Hasegawa, 1996) (see text). The scale bar represents the estimated number of amino acid substitutions per site.

Archibald et al. (2001, J. Struct. Biol.)

The origin of protein interactions and allostery in colocalization

John Kuriyan[1,2] & David Eisenberg[3]

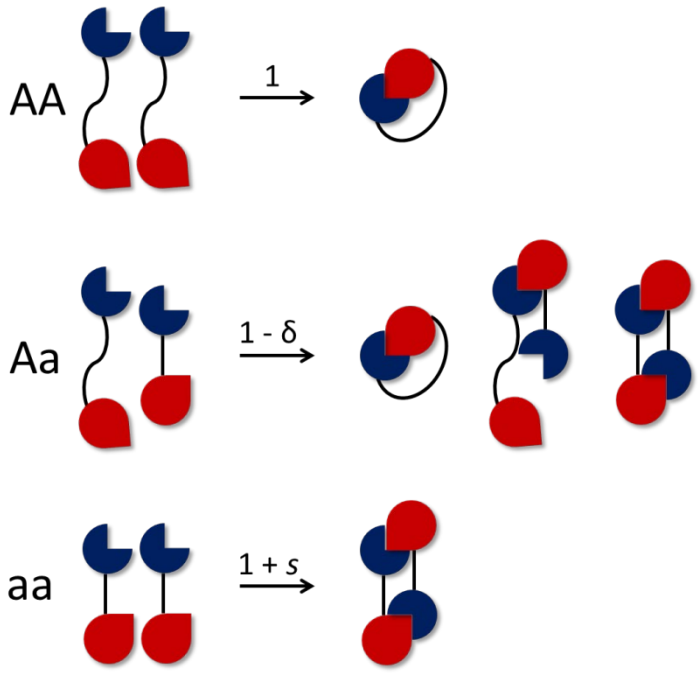
NATURE|Vol 450|13 December 2007|doi:10.1038/nature06524



The Domain-swapping Model

- Interface is preadapted to complexation; and transition requires only a single deletion mutation.

The Population-genetic Conditions for the Origin of Domain Swapping



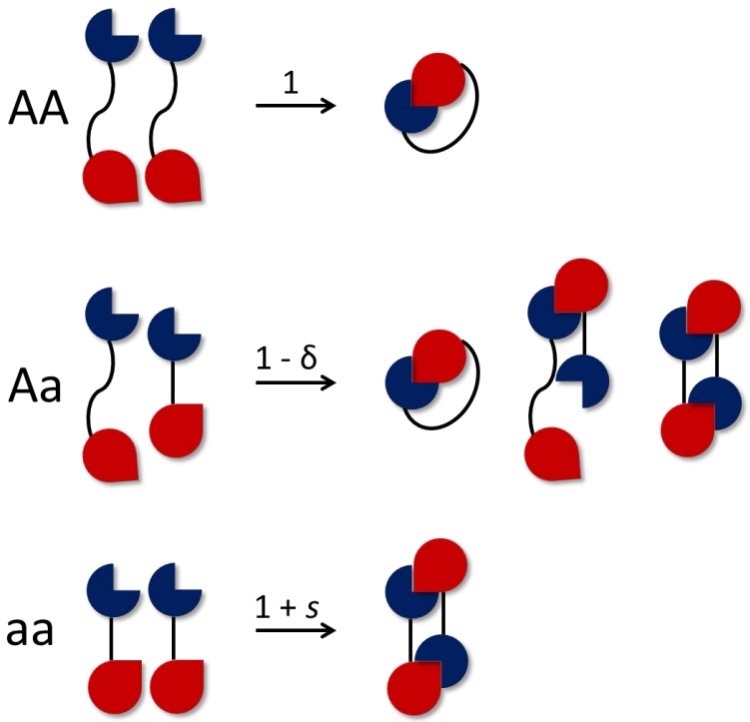
Pure Monomer

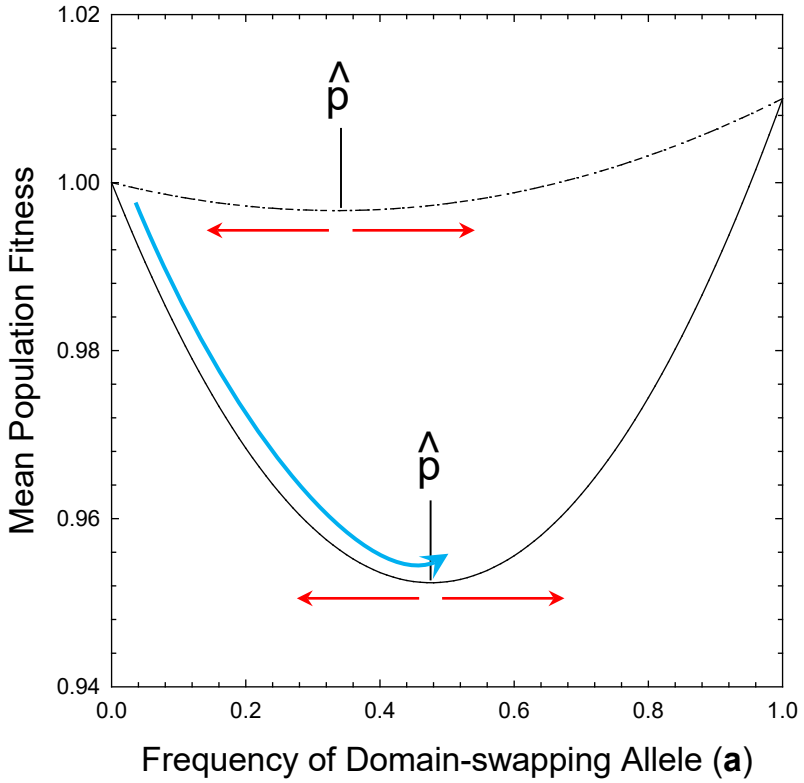
Heterozygote with
mixture of structures; potentially fibrils

Pure Homodimer

- Disadvantage in diploids: reduced heterozygote fitness may impose a strong barrier to fixation;
  the aa homozygote might also be weakly disadvantageous due to the diffusion barrier to assembly.

The Fitness Valley With Heterozygote Disadvantage



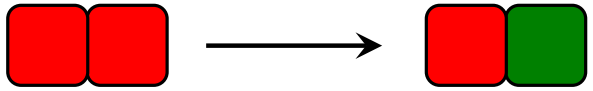


- Evolution of domain-swapping homodimers is strongly inhibited in large diploid populations, unless the heterozygote disadvantage is extremely weak.

mostly AA / Aa

mostly Aa / aa

Transition from Homomers to Heteromers



- Especially commonly observed in eukaryotes, and usually following gene duplication with the sister genes then becoming specialized binding partners.

- Transitions may initiate when there is a balanced ancestral polymorphism at a dimer-forming locus, but fixation of the hetero-complex being impossible until the locus is duplicated, with each daughter locus adopting a particular allelic type.

Predisposition to Transition to Heterodimeric State From Pre-existing Variation

"Neofunctional" alleles segregating in the base population at single-copy loci?

- Maintained, for example, by balancing selection (heterozygote superiority).

Extreme case of homozygote lethality:

Fitness: **AA** = 1 **Aa** = 1 + s **aa** = 0

Spofford's (1969)
Forgotten Insight

Frequency of the **a** allele maintained by selection-mutation balance = s.

"Fixation of heterozygosity" following gene duplication.

Eventual reinforcement of "cross-allelic" binding by secondary mutations.