The Role of Gene Origin in the Evolution of Evolvability

July 24, 2015, EvoEvo Workshop, ECAL 2015, University of York, UK

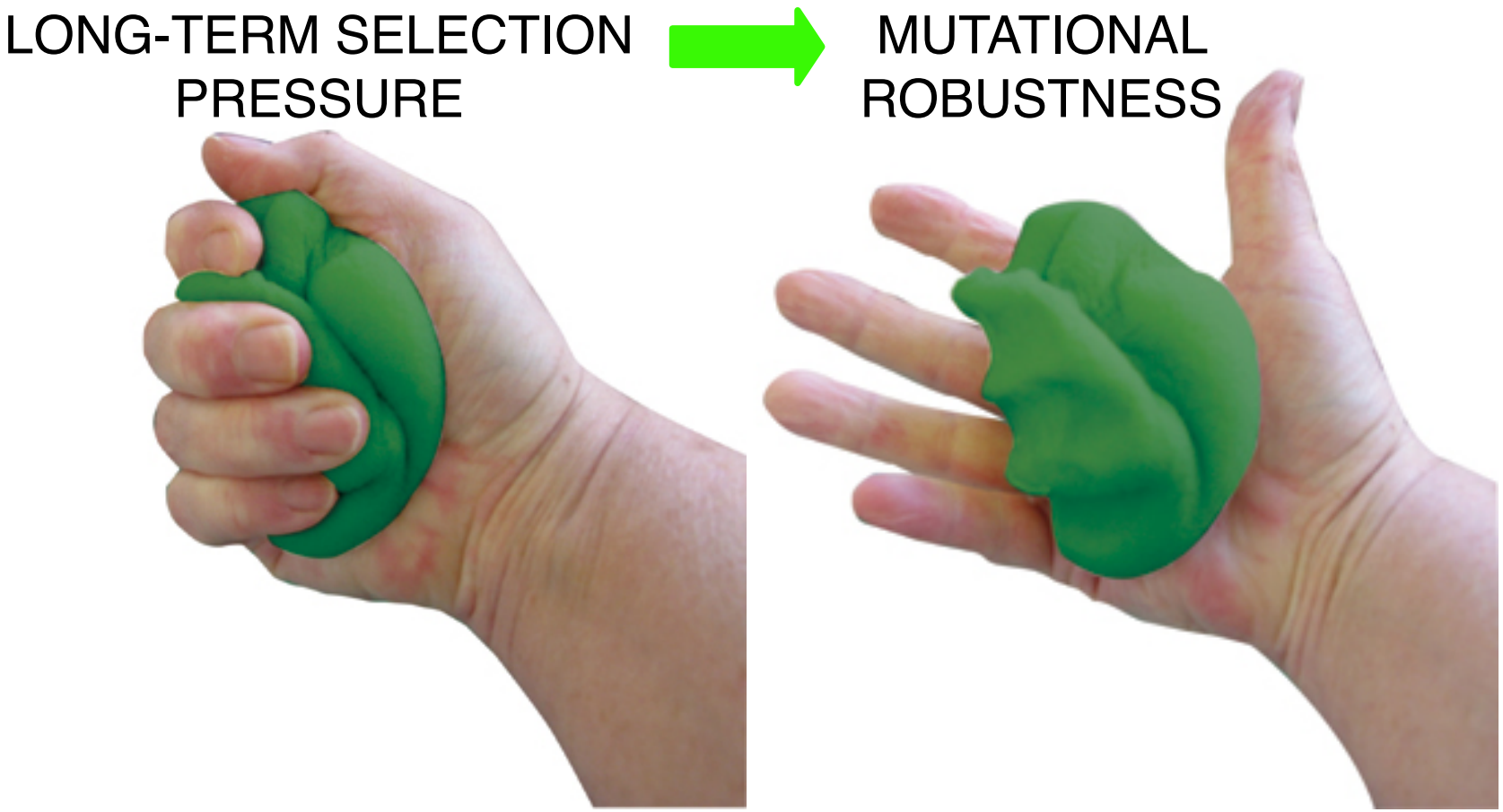
Lee Altenberg
The Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria Lee.Altenberg@kli.ac.at
http://dynamics.org/Altenberg/

Basic Problem of Evolvability

Q. How do you randomly alter a complex organism and improve its function with non-vanishing magnitude at non-vanishing probability?

A. By diverse mechanisms that focus variation in directions with adaptive opportunity while suppressing variation in harmful directions.

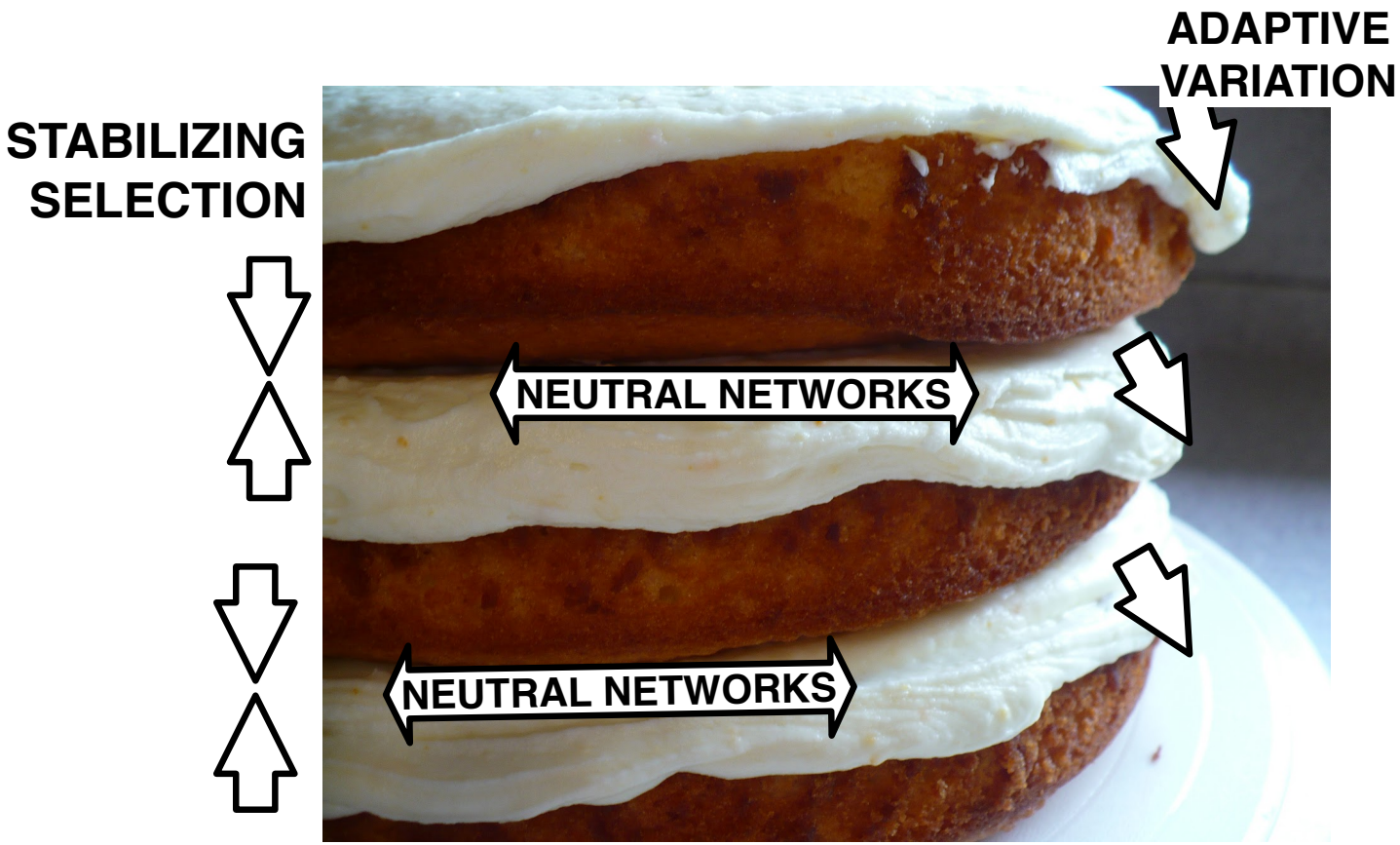
- New theory and some empirical studies reveal variation production of the organism resembling 'viscoelasticity' (suppression of deleterious mutation production)



- Classical quantitative and population genetics models precluded such 'viscoelasticity'

LONG-TERM SELECTION PRESSURE ⇏ NO MUTATIONAL ROBUSTNESS

- Variation can proceed along neutral networks of genotypes to find adaptive variation (Wagner, 2008)



- Varying selection can wear 'trails' in the genotype-phenotype map that enhance re-evolvability:
  - Meyers et al. (2005)
  - Kashtan and Alon (2005)
  - Crombach and Hogeweg (2007, 2008)
  - Draghi and Wagner (2008)
- Needed is a comprehensive understanding of how the variational properties of organisms are shaped in evolution.

Levinton (1988, p. 494) *Genetics, Paleontology, and Macroevolution*

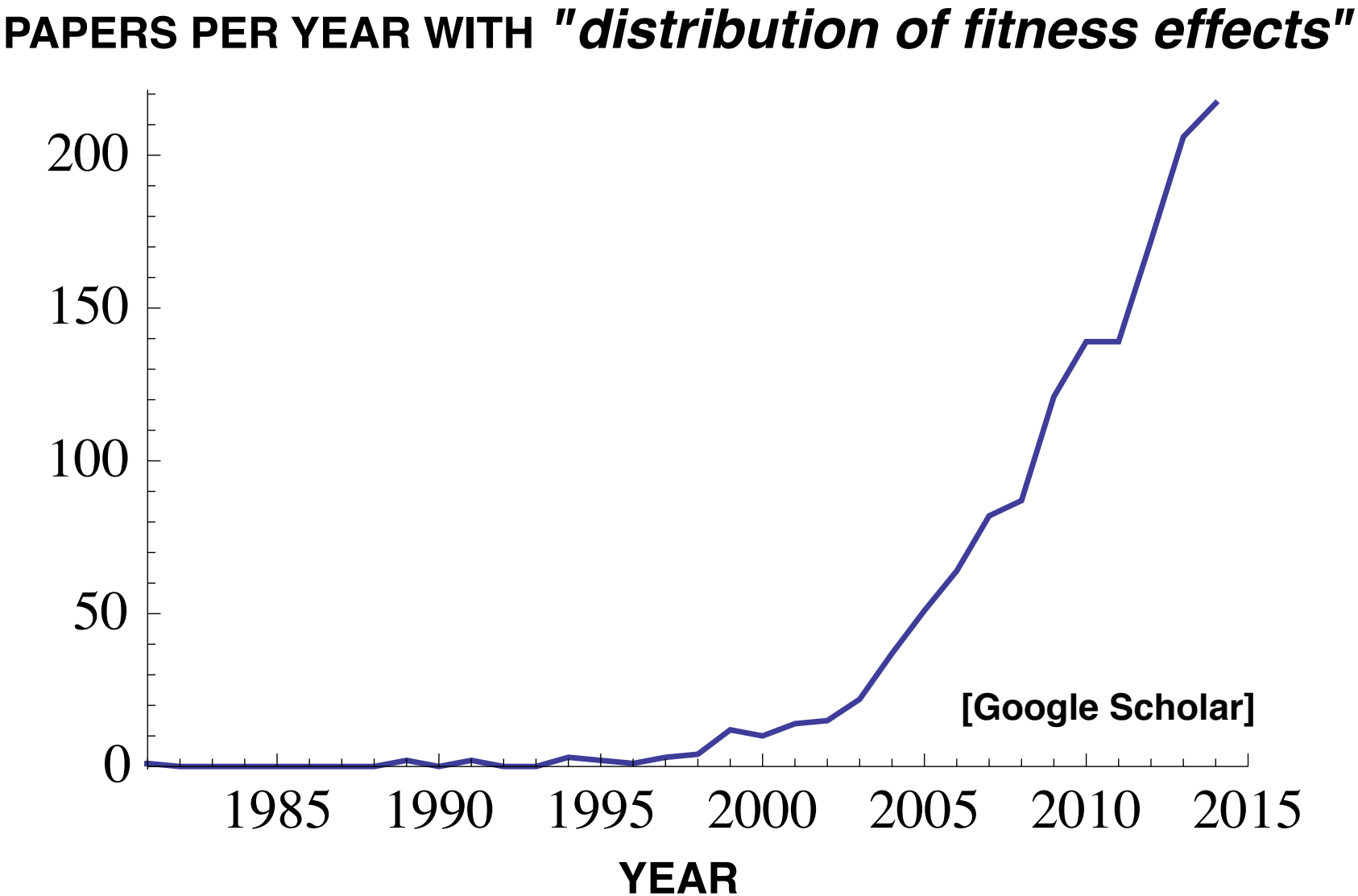
- "Evolutionary biologists have been mainly concerned with the fate of variability in populations, not the generation of variability.
- . . . The genetic and epigenetic factors that generate variability have received relatively little attention.
- This could stem from the dominance of population genetic thinking, or it may be due to a general ignorance of the mechanistic connections between the genes and the phenotype.
- Whatever the reason, the time has come to reemphasize the study of the origin of variation."

Classical and "Extended" Population Genetics

1. The *fate* of variation: The main concern of classical population genetics
2. The *generation* of variation: A central concern of the "extended evolutionary synthesis" (Pigliucci and Müller, 2010)
3. Theoretical approach to 2: Examine the fate of *variation for the generation of variation*.

A Fundamental quantity in the "generation of variability":

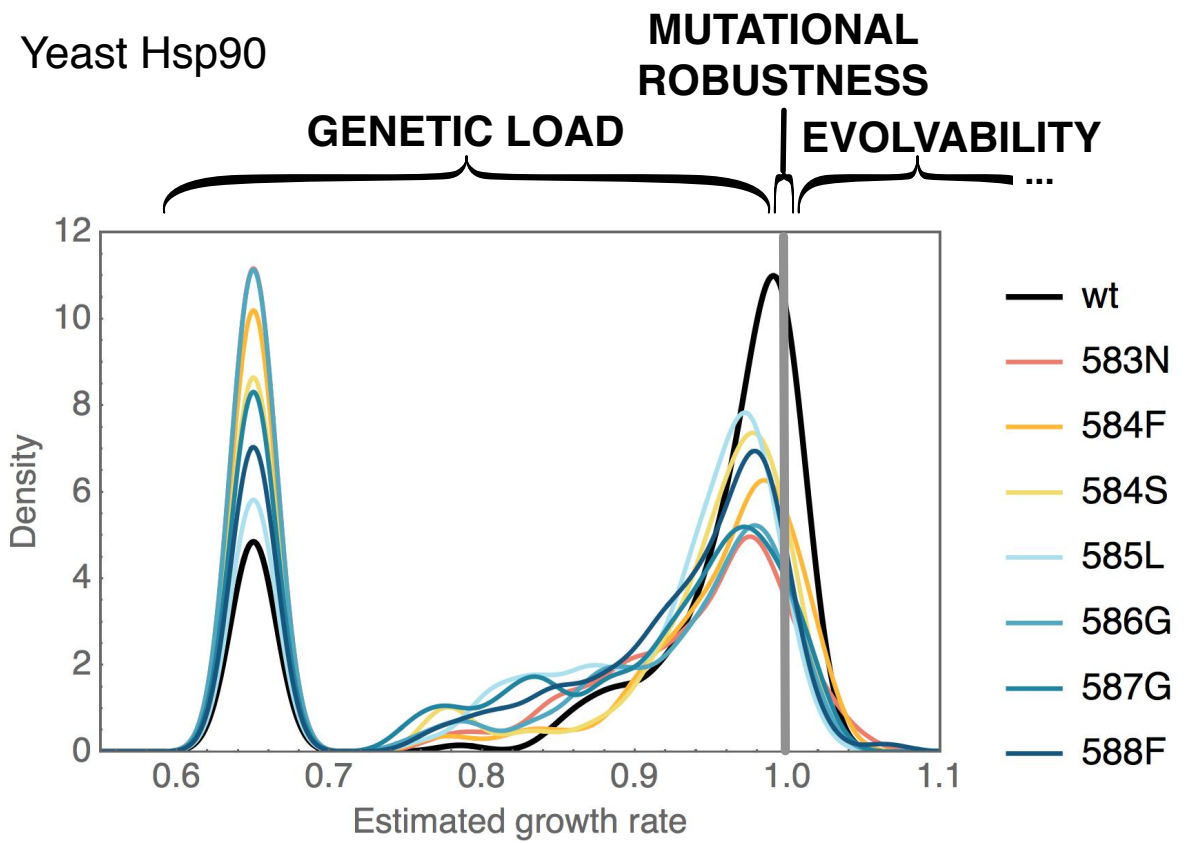
- The Distribution of Fitness Effects of Mutation



Distribution of Fitness Effects of Mutation

Recent talk by Jeffery D. Jensen:

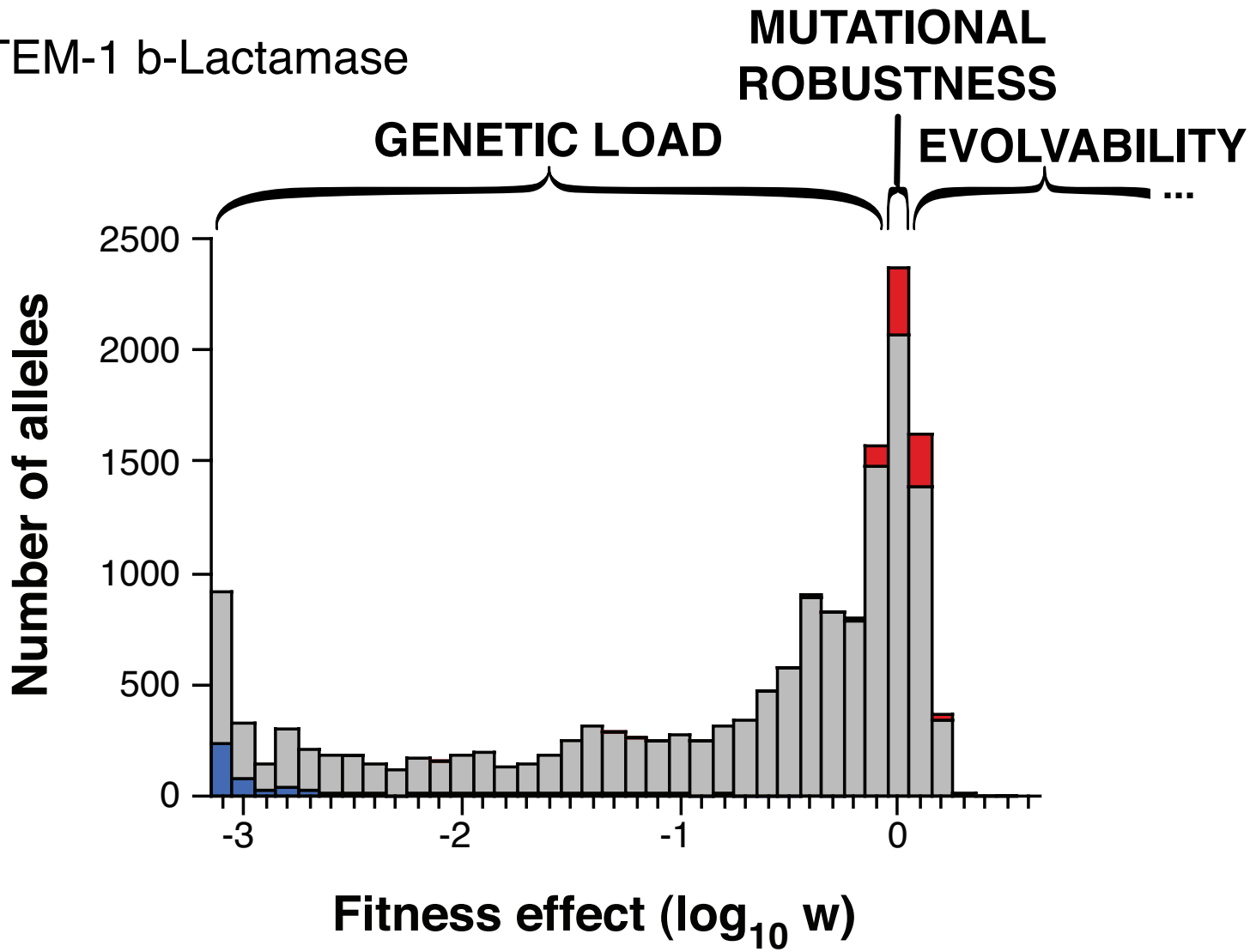
- "On quantifying the distribution of fitness effects of new mutations—an experimental approach".



Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. (2015). A systematic survey of an intragenic epistatic landscape. *Molecular biology and evolution*, *32*(1), 229-238.

Distribution of Fitness Effects of Mutation

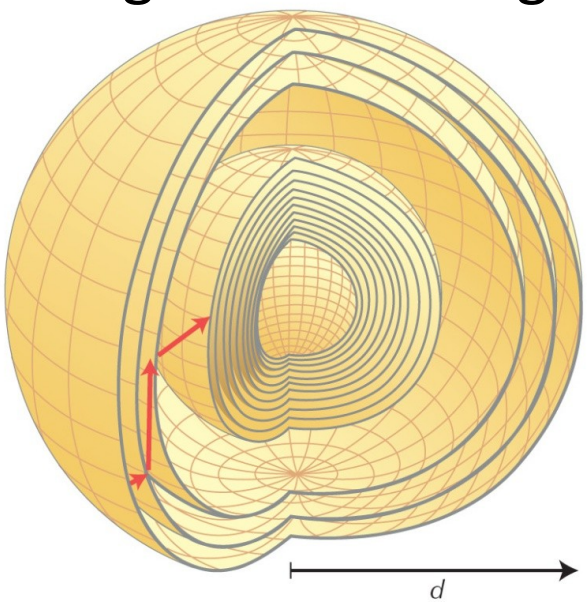
E. coli TEM-1 b-Lactamase



Firnberg, E., Labonte, J. W., Gray, J. J., & Ostermeier, M. (2014). A comprehensive, high-resolution map of a gene's fitness landscape. *Molecular Biology and Evolution*, msu081.

Do distributions of fitness effects evolve?

- Yes, all the time. Described as early as (1930):
  Fisher's geometrical argument

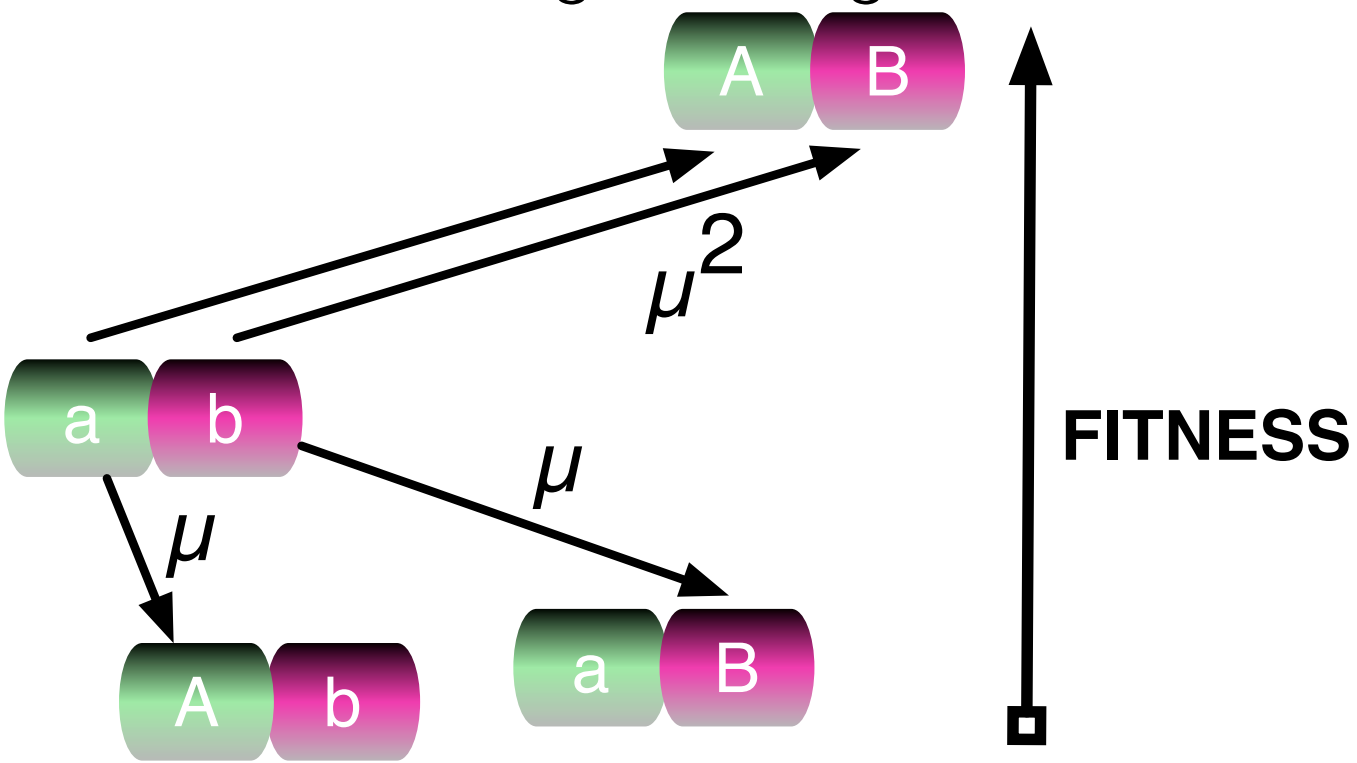


- But in Fisher's model, evolvability always goes down with adaptation.
- How can evolvability ever evolve to increase?

Riedl (1977) A systems-analytical approach to macroevolutionary phenomena. Q. Rev. Biol. 52: 351–370.

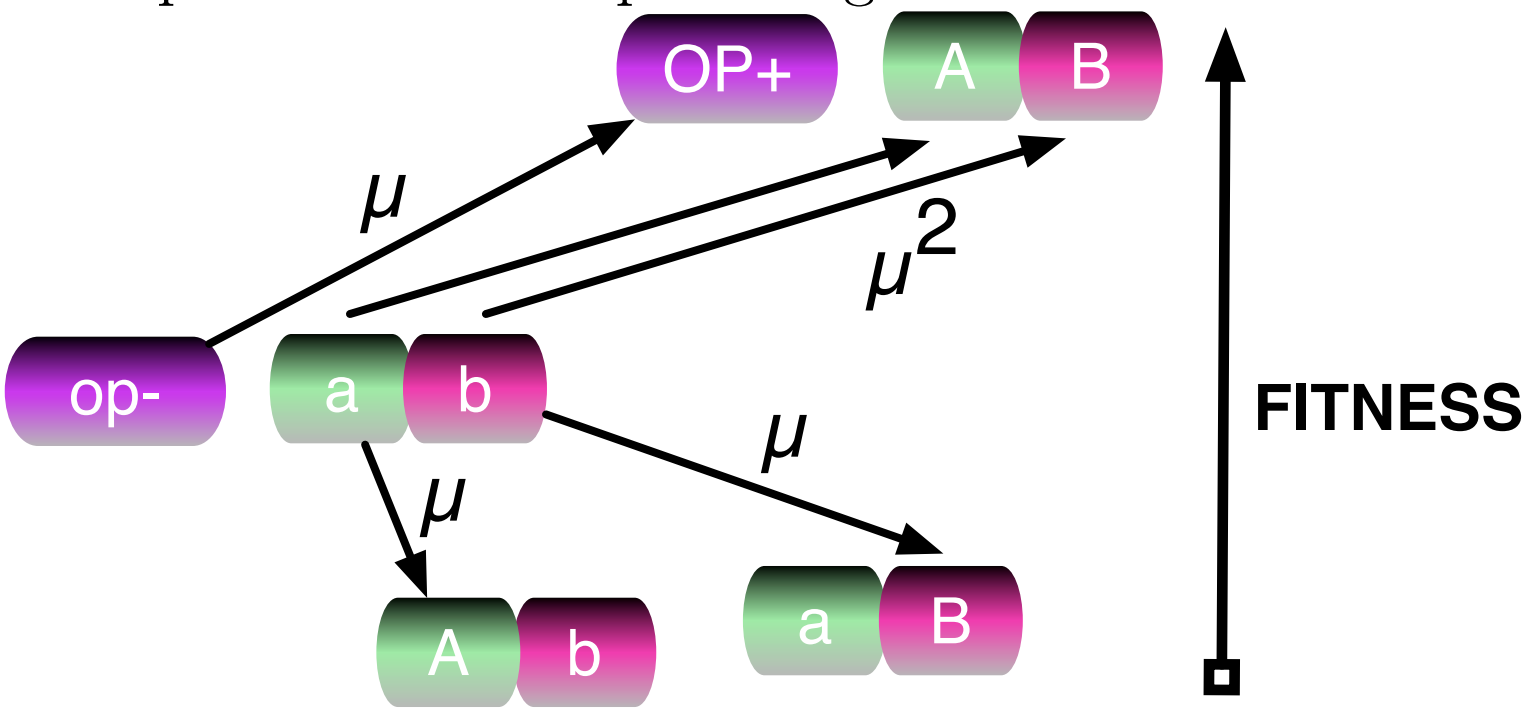
Evolution of evolvability and mutational robustness through new genes.

Riedl considers the following situation, where adaptation requires simultaneous changes in two genes:

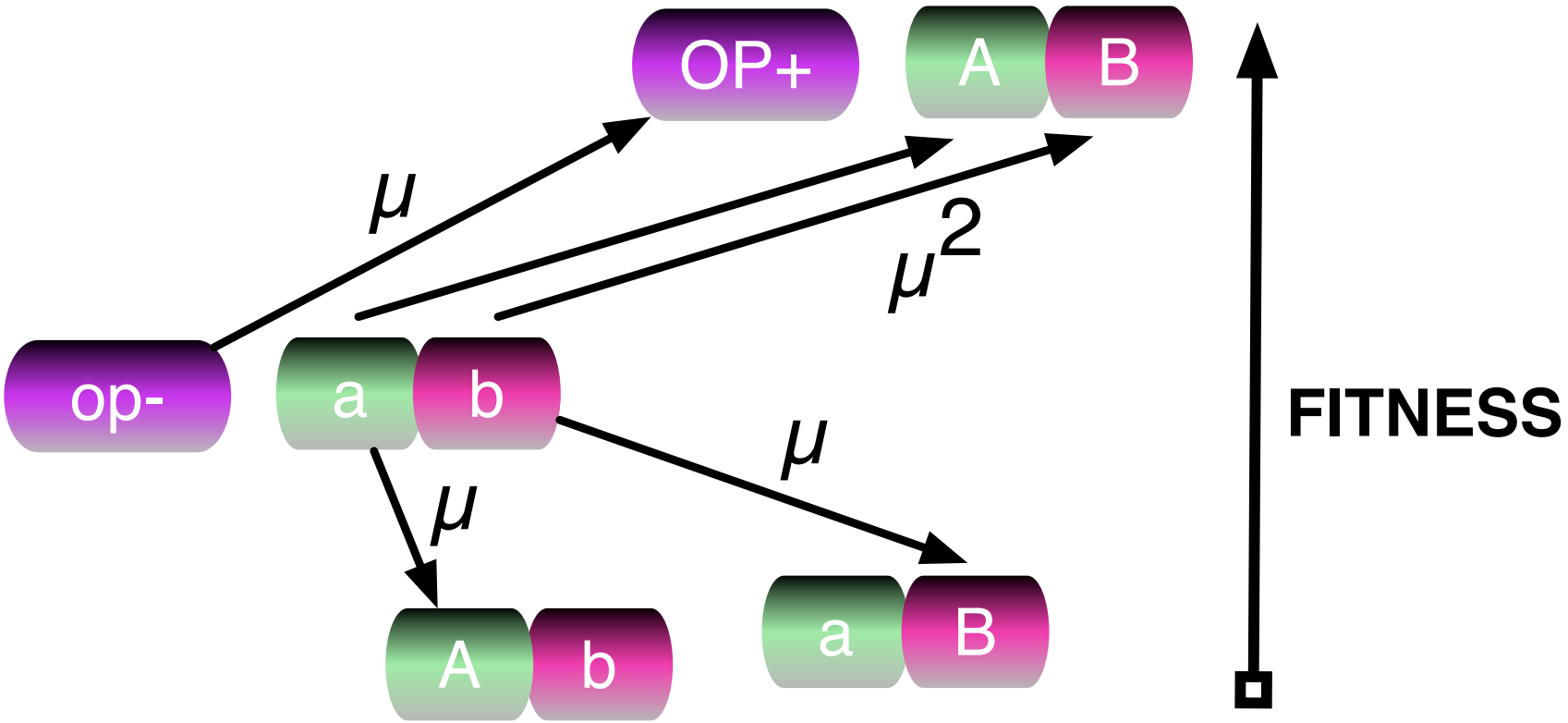


Riedl (1977), cont'd.

"What would happen if independent genetic units, the structural results of which have become functionally dependent, were also to become epigenetically dependent, for example, by adopting a superimposed genetic unit upon which both are dependent, as in the case of two structural genes dependent on an operator gene?



Riedl (1977), cont'd.



A. "The mutation of only one genetic unit, the operator, will result in the change of both. Then the chance of a successful alteration . . . would increase as much as a millionfold."

Riedl (1977), concluded

- "Such adaptive advantages . . . are so tremendous that the invention of a superimposed genetic unit must be expected, even if it would be a millionfold or trillionfold more unlikely than every other alteration within the genome."
- ". . . The chances of successful adaptation increase if the genetic units, by insertion of superimposed genes, copy the functional dependencies of those phene structures for which they code.

A radical idea: variation production records of the history of adaptation.
Riedl's mechanism arises from new gene creation.

Is new gene creation no different from allelic substitution?

Reviewer for *Evolution*:

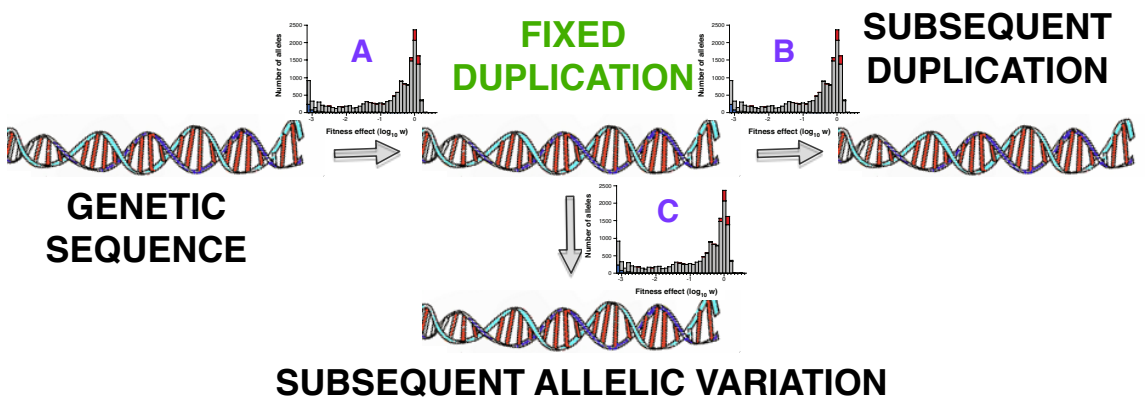
- "The subject of new genes and their role in the evolution is, indeed, interesting.
- The interest, however, is in the mechanisms by which they are introduced rather than in what happens to them after the introduction.
- The moment a new gene is introduced, it becomes totally equivalent from the population genetics point of view to any other gene in the genome.
- A new gene can be regarded as a mutation from a preexisting 'zero' allele in the respective locus."

What did the reviewer miss?

Unique Effects of Gene Duplication:

When new genes are added to the genome, three things are changed:

1. The number of degrees of freedom for genetic variation is increased
2. The probability of allelic variation in that gene family is increased
3. The probability of subsequent gene duplications of that gene family is increased.



New dimensions of variation added by new genes

- Much thought has gone to the causes of gene duplication.
- But what are the cumulative consequences of gene duplication?
- Processes that shape genome growth shape genetic variation production.

Yang et al. (2003) Organismal complexity, protein complexity, and gene duplicability

successful than duplication of a mono-

no monomers were included in their analysis, so the magnitude of difference in survivability between duplication of a monomer and duplication of a protein complex subunit is not known. It is emphasizing that duplication of a monomer

important to include monomers. Indeed, we study relationship between the survivability of a gene duplication and by classifying proteins into monomers ($n = 1$), dimers ($n = 2$), complexes ($3 \leq n \leq 10$), and large complexes ($n >$

midsize complexes ($3 \leq n \leq 10$), and large complexes ($n > 10$). Another factor that may affect the survivability of duplicate genes is organismal complexity. It was suggested that, for transcription factors, dosage imbalance occurs more frequently

types of cells.

Previously we talked about survivability, which may be defined as the probability for a duplicate gene to survive, but adaptive duplication of duplicate genes may also be important. Because

whether a gene has been

Gene "Survivability"

- Gene "survivability" is the language of natural selection.
- But they are not talking about organismal fitness.
- By "gene survivability" they mean long-term preservation of the gene in the genome.
- Differences in gene "survivability" therefore amount to a kind of natural selection in the genome-as-population.
- Suppose we pursue this analogy in full?

Population Analogous Processes

- Eukaryotic species have relatively closed genomes.
- This renders the genome analogous to a population of organisms:
  1. gene duplication $\rightarrow$ reproduction
  2. gene deletion or pseudogenization $\rightarrow$ death
  3. de novo gene creation, horizontal gene transfer $\rightarrow$ immigration
  4. maintenance of gene properties over time $\rightarrow$ heritability

A Novel Level Darwinian Processes: Genome-as-Population

- Lewontin's (1970) sufficient elements for Darwinian evolution:
  1. Heritable
  2. Variation in
  3. Fitness (viability and fecundity)
- Are these elements all present in the genome-as-population?

A Novel Level Darwinian Processes: Genome-as-Population

If there are

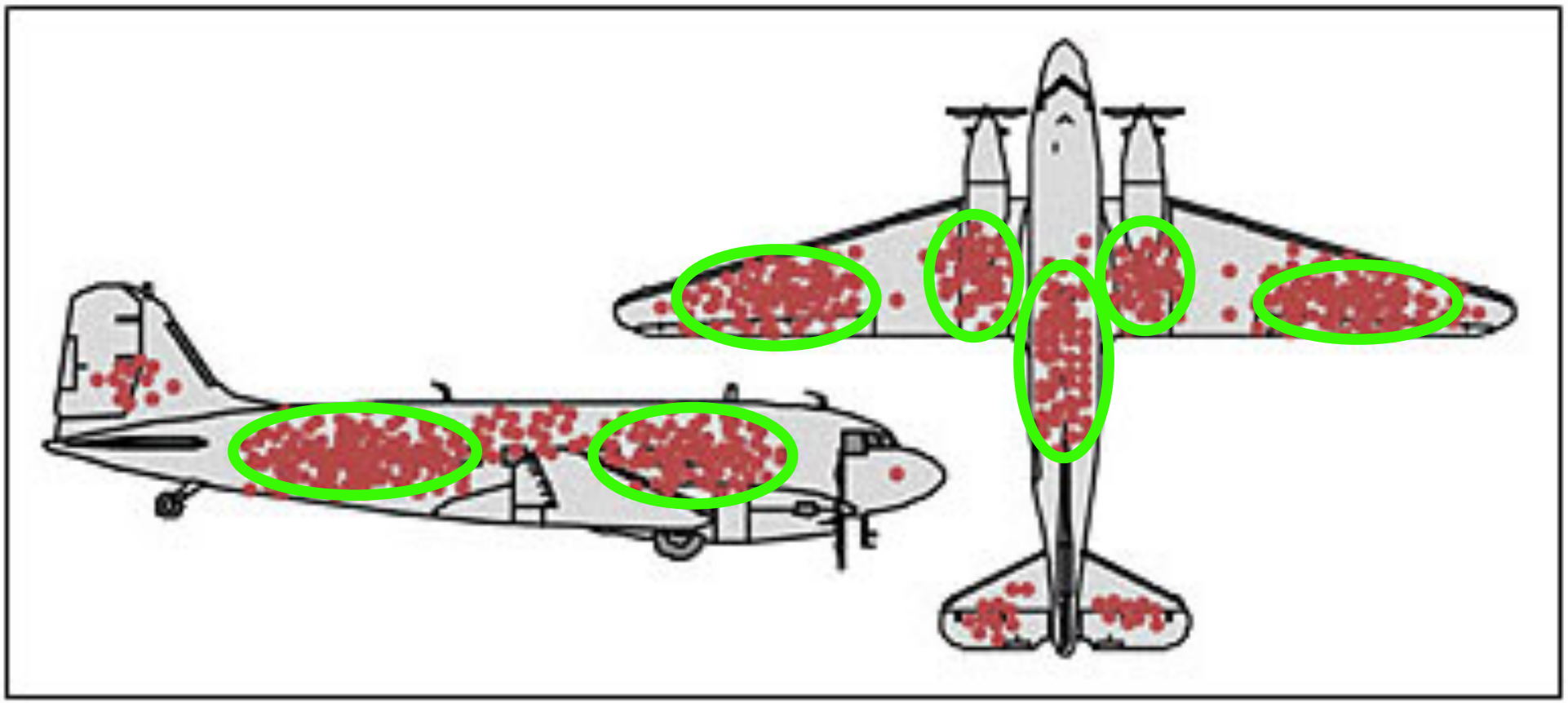
1. gene properties that produce differences in gene survivability and duplicability, and
2. these gene properties persist over time and over gene duplications (heritability),
3. then the genome can become populated by genes with higher gene duplicability and survivability.

That is, evolvability can evolve.

An analogy for selection: Vulnerabilities of WWII bombers

Q. WHERE SHOULD STRATEGIC ARMOR BE ADDED?

DISTRIBUTION OF HOLES ON RETURNING R.A.F. BOMBERS, WWII



Credit: Cameron Moll

An analogy, cont'd

WALD, ABRAHAM. 1943. "A METHOD OF ESTIMATING PLANE VULNERABILITY BASED ON DAMAGE OF SURVIVORS"

Let $\delta(i,j)$ be the conditional probability that part i is hit by gun j knowing that a hit has been scored and the plane survived the hit. Furthermore, let q be the probability that the plane survives a hit (not knowing which part was hit and which gun scored the hit). Then, similar to equation 82, we shall have

$$q(i,j) = \frac{\delta(i,j)}{\gamma(i,j)} q \quad . \tag{101}$$

Let q(j) be the probability that the plane will survive a hit by gun j (not knowing the part hit). Then obviously

$$q(j) = \sum_{1} \gamma(i|j) q(i,j) \quad . \tag{102}$$

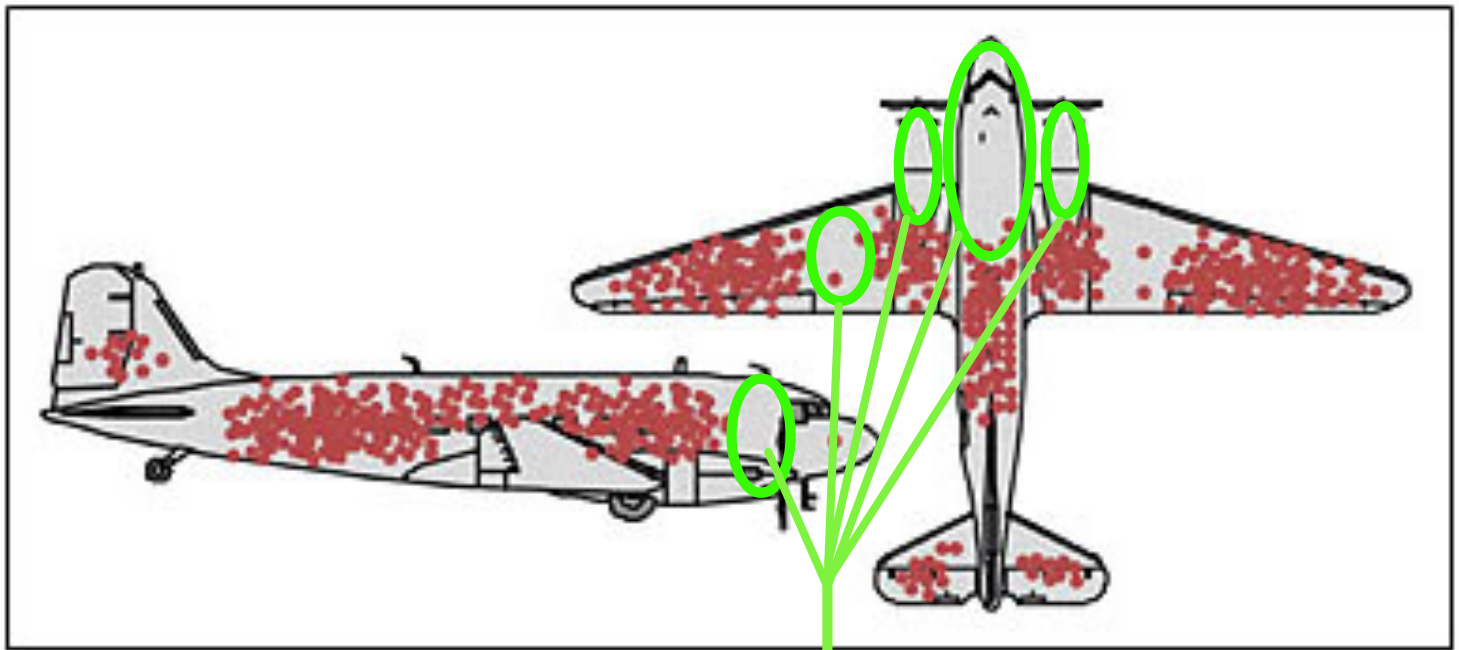
Let $\delta(i|j)$ be the conditional probability that part i is hit by gun j knowing that a hit has been scored by gun j and the plane survived the hit. Clearly

---

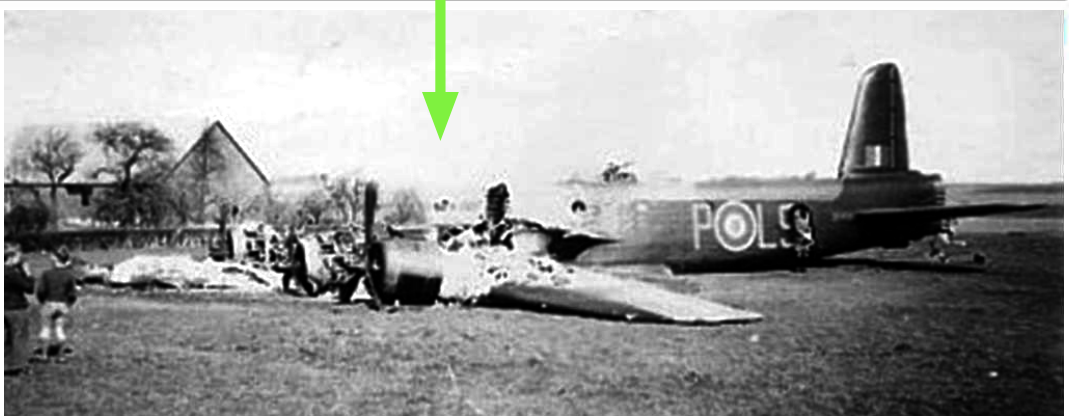
[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 126 and AMP memo 76.8.

An analogy, cont'd

A. ADD ARMOR WHERE THERE AREN'T ANY HOLES!



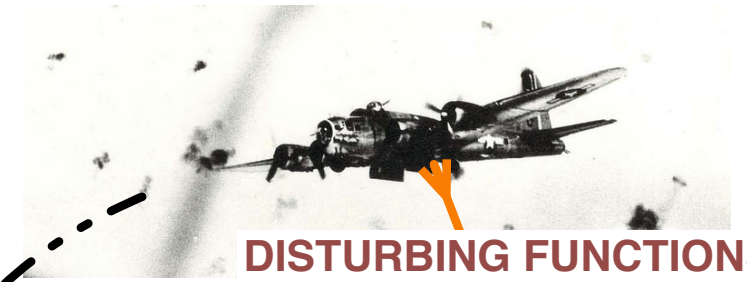
ABRAHAM WALD



An analogy, cont'd

- The sampled planes were all very special—they were the ones that survived.
- In the same way, the genes functioning in the genome are very special—they are the ones that survived.

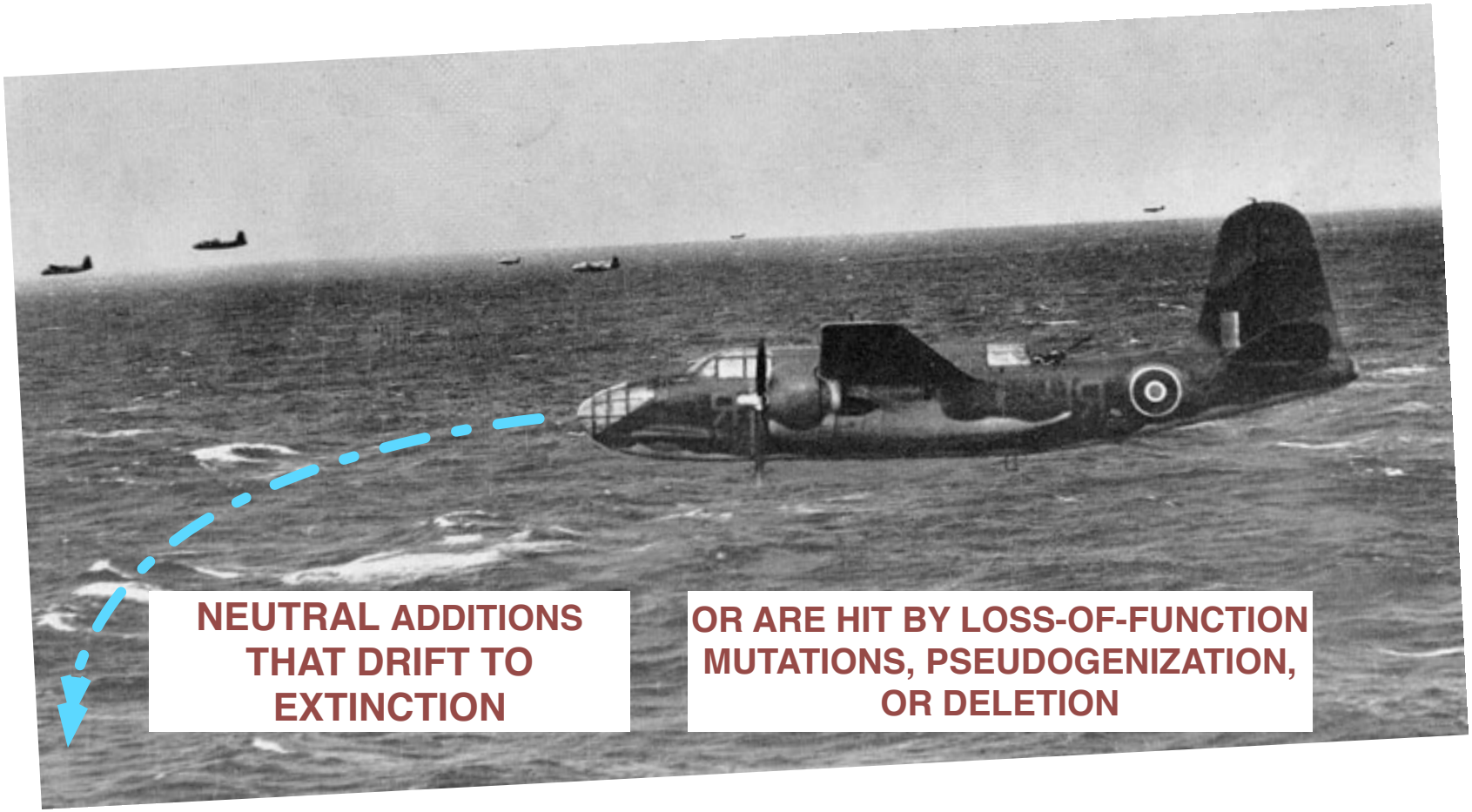
WHAT ARE THE "MISSING HOLES" AMONG NEW GENES?



DISTURBING FUNCTIONS UNDER STABILIZING SELECTION,

e.g. DOSAGE

"MISSING HOLES" AMONG NEW GENES



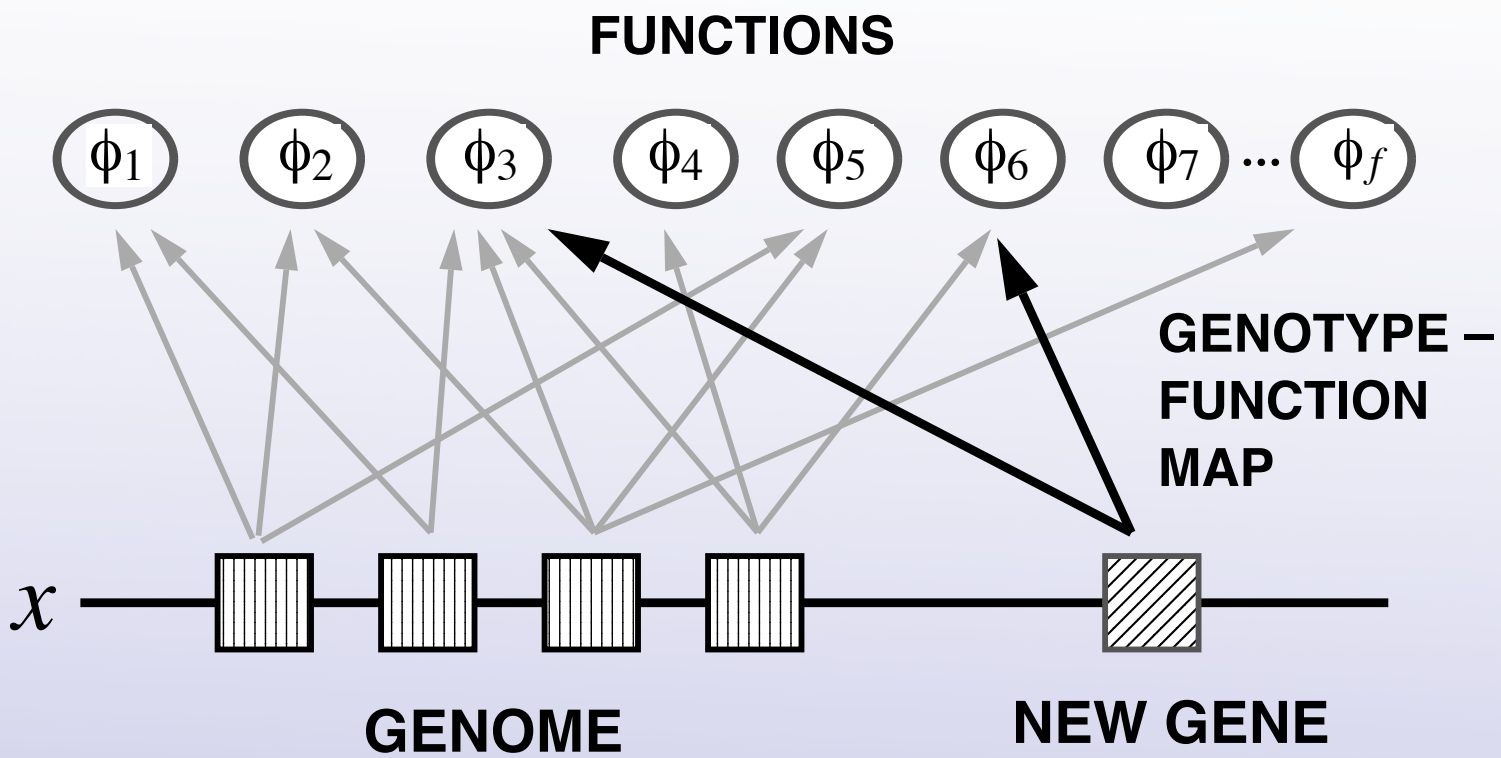
NEW GENES THAT "MAKE IT"

- IMMEDIATE EFFECT IS ADVANTAGEOUS OR NEUTRAL AND GOES TO FIXATION
- SUBSEQUENT MUTATIONS GIVE
  - NEOFUNCTIONALIZATION,
  - SUBFUNCTIONALIZATION, OR
  - ESCAPE FROM ADAPTIVE CONFLICT

An Illustrative Model

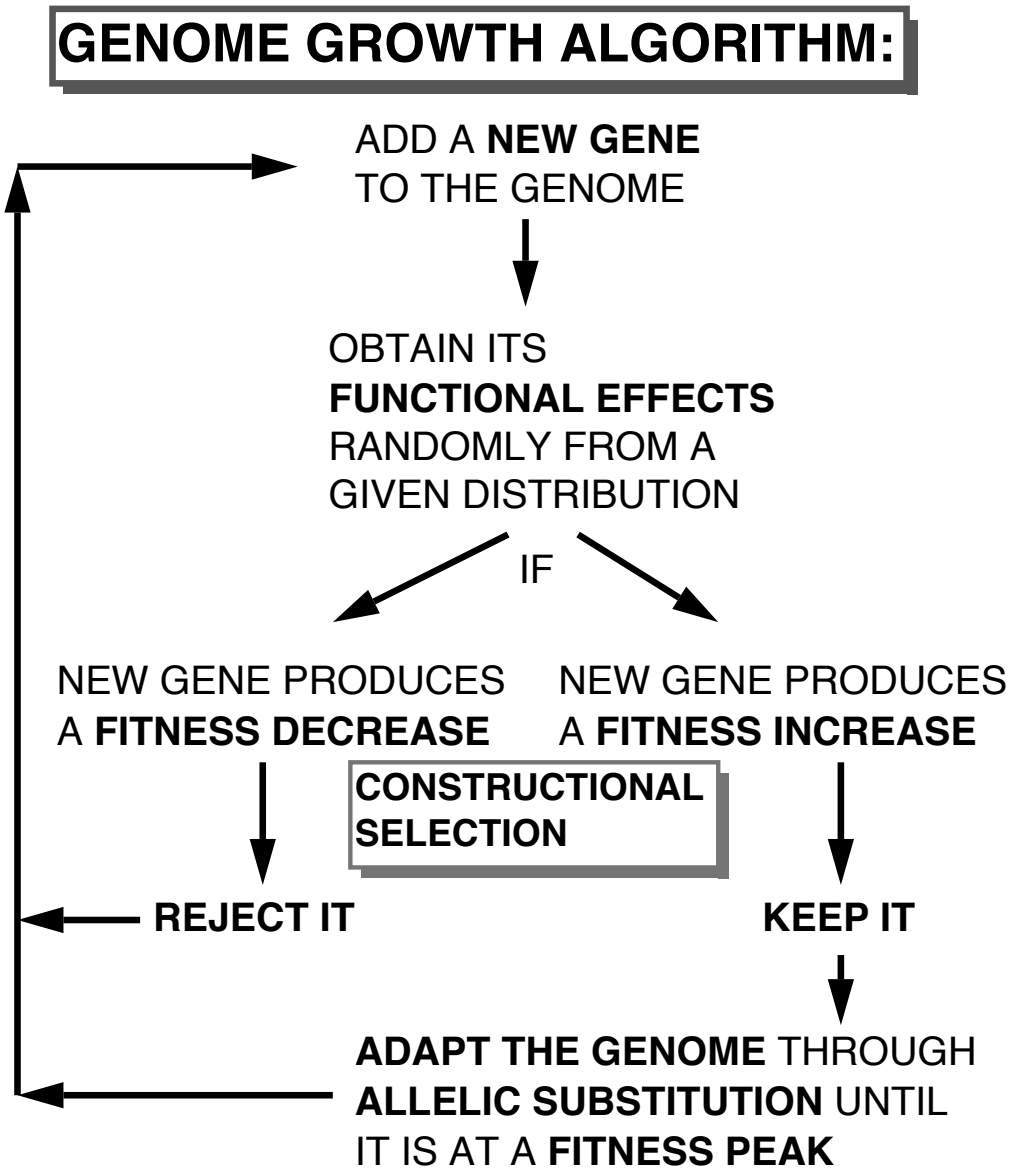
- To illustrate how selection on de novo gene origins can shape the genotype-phenotype map, I will use the NK landscape model of (Kauffman and Levin, 1987).

NK Model (Kauffman and Levin, 1987) for Genome Growth



$$w(x) = \frac{1}{f}\sum_{i=1}^{f}\phi_i(x)$$

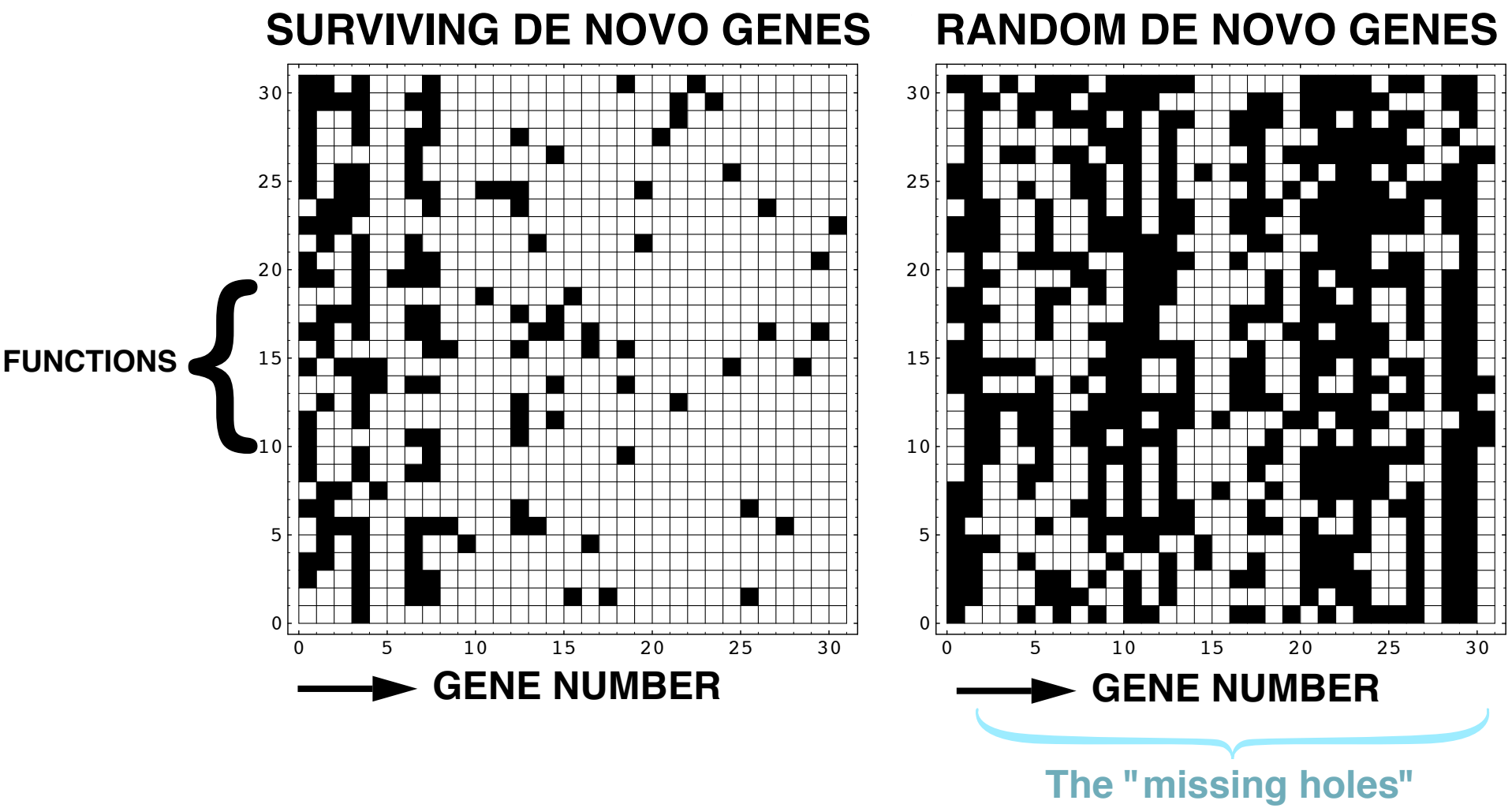
A Model of *De Novo* Gene Origin



Assumptions:

1. Strong selection
2. Weak mutation
3. Rare gene creation

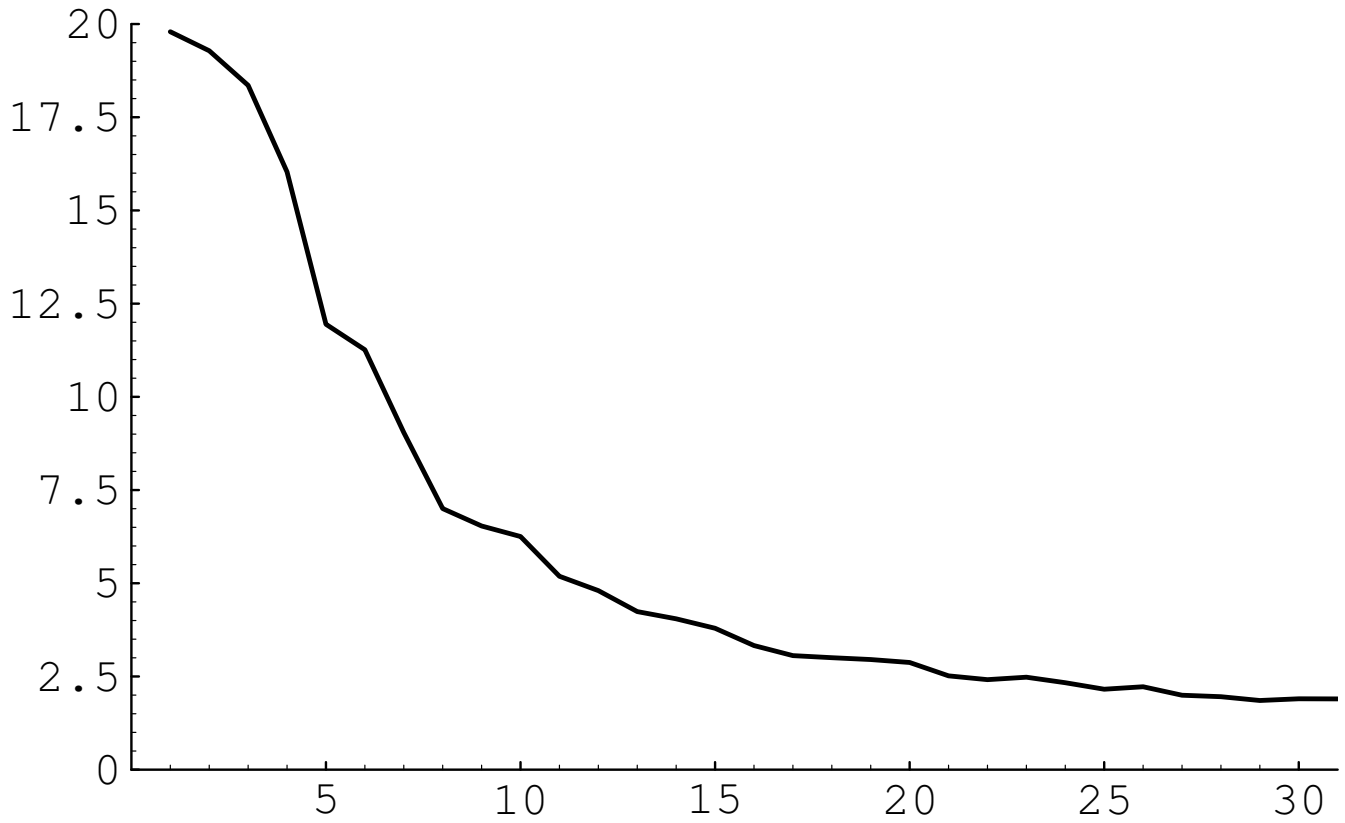
NK Model for De Novo Genes



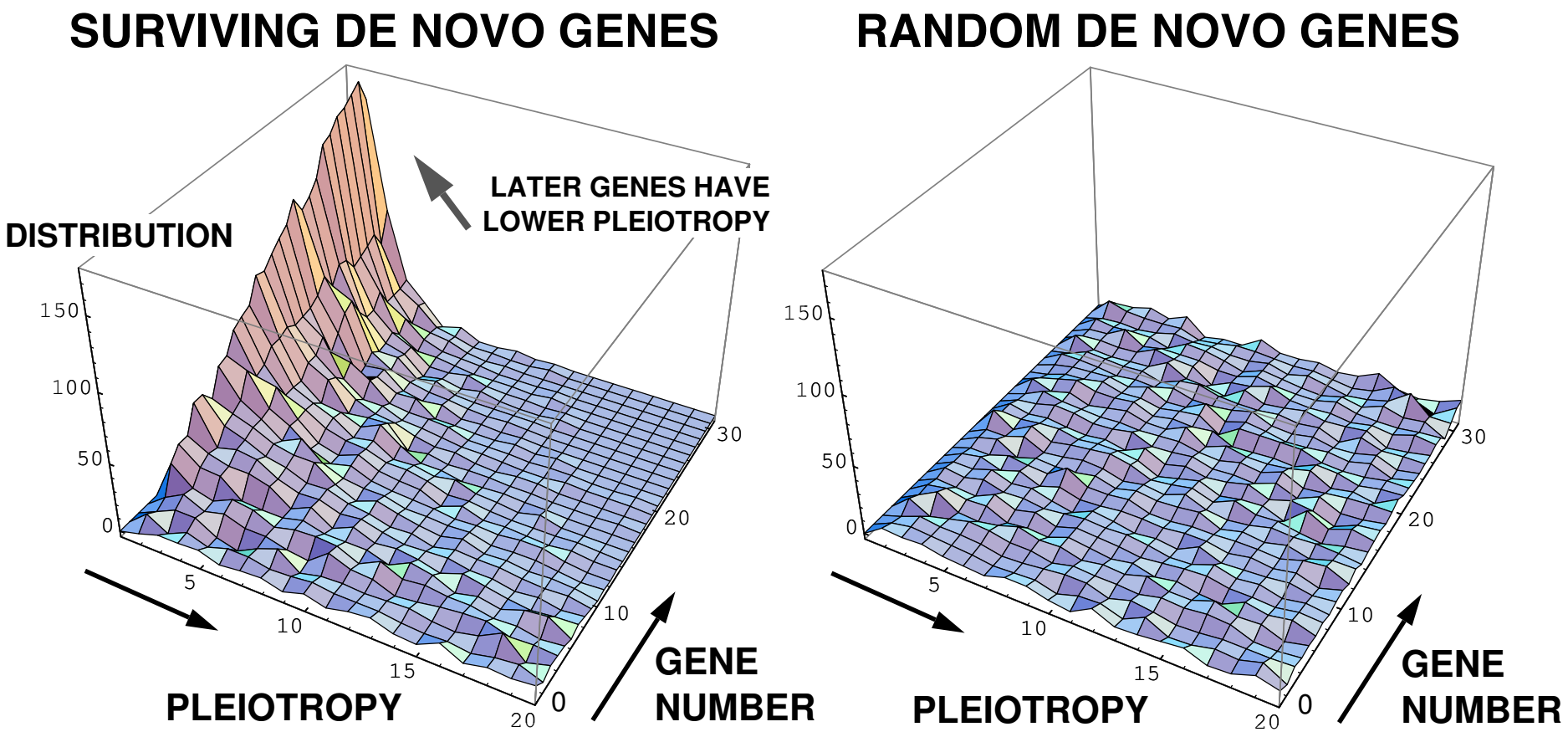
New genes are filtered out to target only phenotypes with adaptive opportunity—i.e. "genes as followers"

NK Model for De Novo Genes

AVERAGE PLEIOTROPY, k_n



NK Model for De Novo Genes



De novo genes:

1. If selection has driven them to fixation in the population, they should be enriched for some properties of gene survivability (in the simulation, enrichment is for low pleiotropy)
2. But de novo genes have not been through iterations of gene duplication
3. They should therefore differ from highly duplicated genes in having
    1. lower duplicability and
    2. lower heritability of their survivability in time.

Key Idea #1: Common Conditioned Histories

The mere existence of a gene tells us something about its history:

- The functioning genes in the genome generally share a common conditioned history that
  1. When created, their effects were neutral or advantageous
  2. The new genes grew in frequency or became fixed in the population [*a step skipped in whole genome duplication]
  3. They came to be preserved by selection before they could be deleted or pseudogenized.

Key Idea #1

Being neutral or advantageous when created (distribution of fitness effects of gene creation) in turn depends upon

- the gene's map to the phenotype

— i.e. its spectrum of phenotypic effects

1. varying traits with adaptive opportunity, while
2. leaving alone functions under stabilizing selection.

- i.e. a kind of modularity.

In a nutshell: genes are born modular, or not born at all.

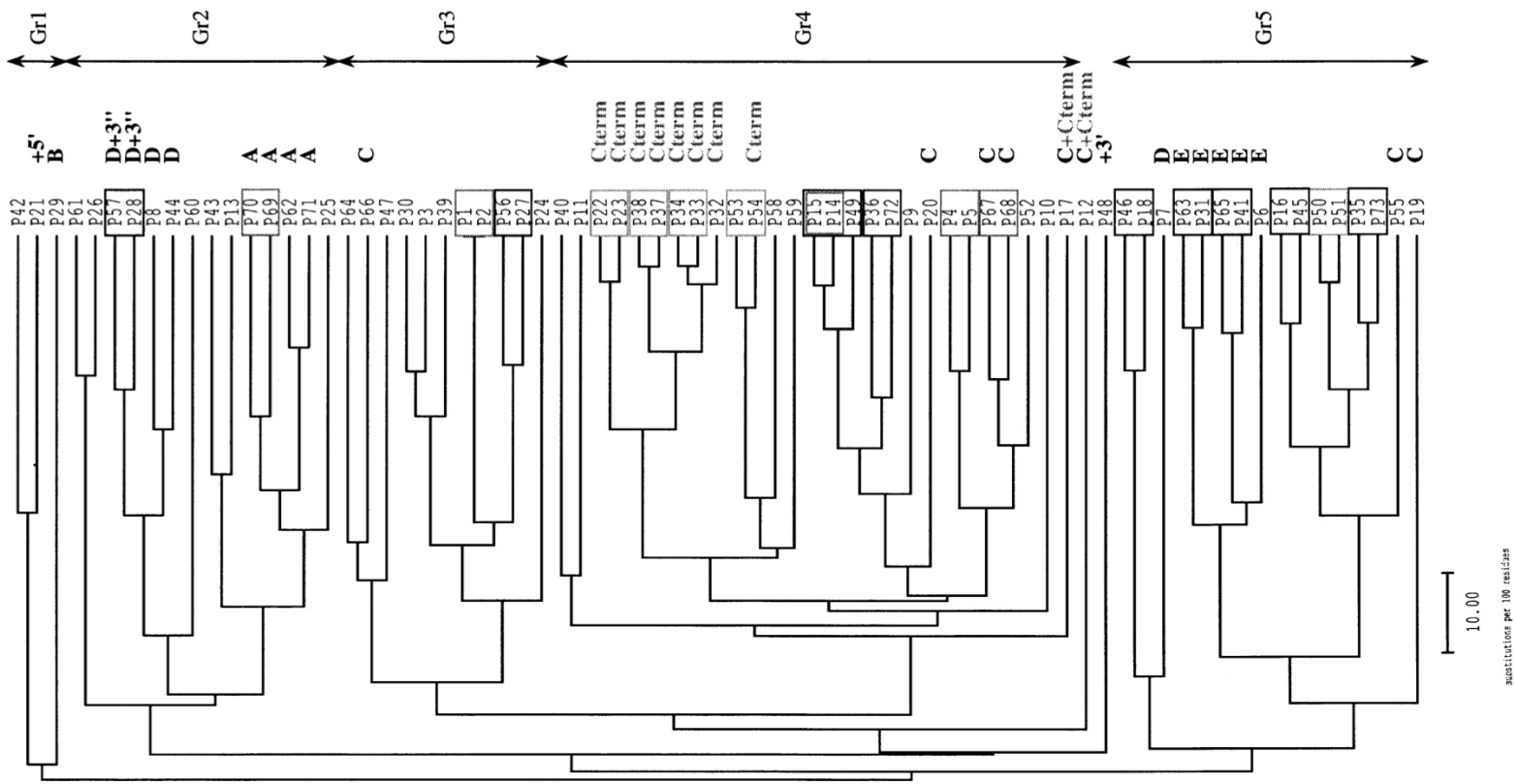
Key Idea #2: The present connects to the past

The spectrum of phenotypic effects is preserved to varying degrees over time

- Those highly preserved over long periods:
  - role in regulatory and interaction networks and development
  - peptide secondary structure
- Less preserved over long periods:
  - Primary amino acid sequences
  - Primary nucleotide sequences
  - Distribution of fitness effects of mutation or duplication
- This question is almost entirely unexplored theoretically and empirically.

Heritability between Duplications

Phylogenetic relationships of the *Arabidopsis* peroxidases (Tognolli, et al (2002)) (duplicates all have peroxidase function):



Singh et al. (2014) Human Dominant Disease Genes Are Enriched in Paralogs Originating from Whole Genome Duplication

Singh et al. (2014) show that conditioning on the gene origin (whole genome duplication of 500 mya, w/ or w/out subsequent small duplication) alters the production of dominant vs. recessive disease mutations.

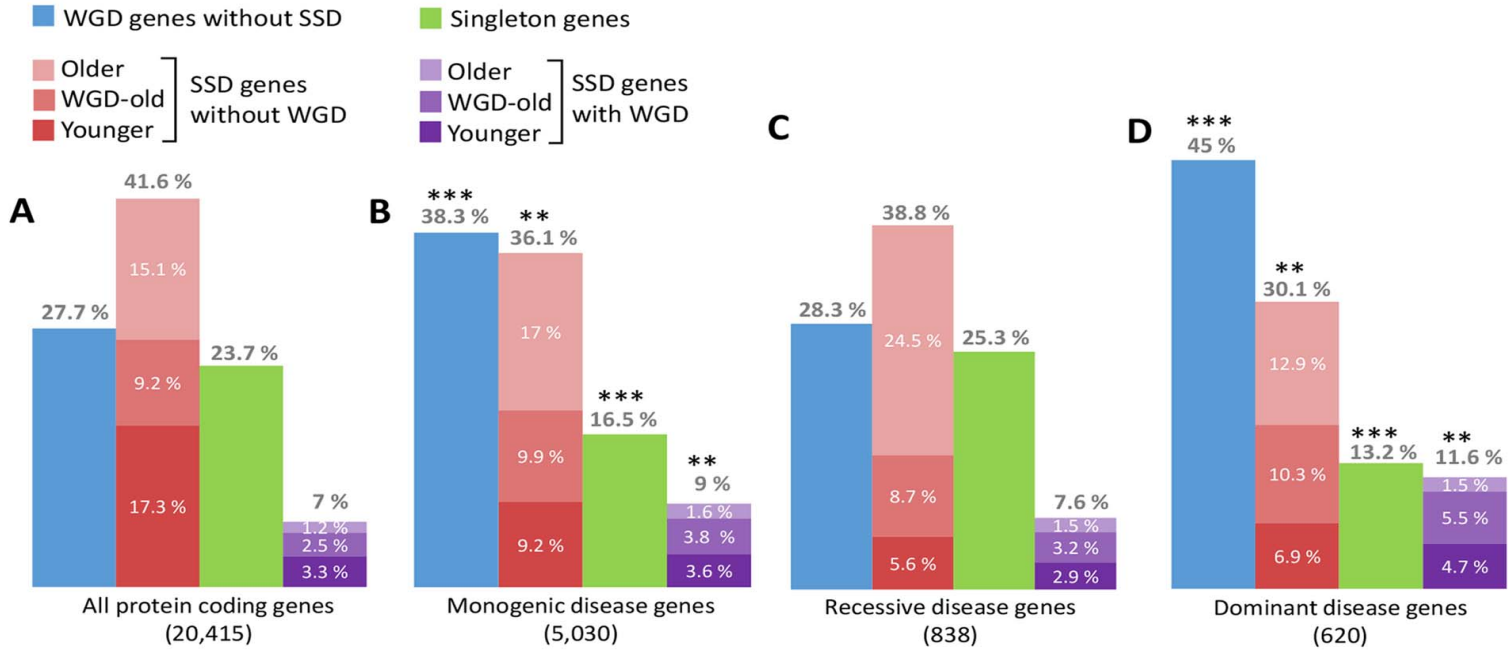


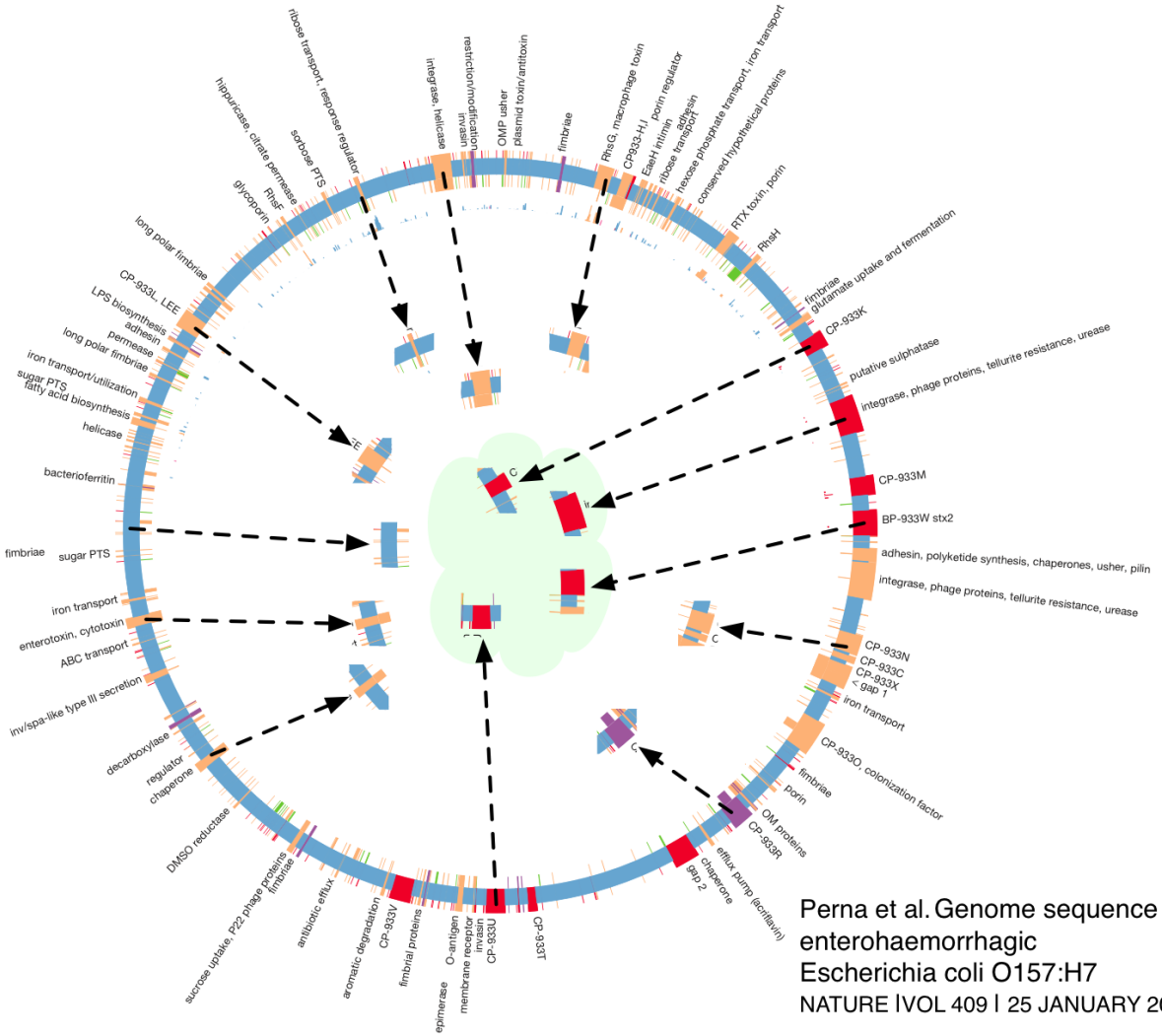
Figure 1. Distributions of WGD, SSD, and singletons in (A) the whole human genome, (B) monogenic disease (MD) genes [1], **(C) recessive MD genes, and (D) dominant MD genes.** (***) corresponds to highly significant deviations ($p<10^{-6}$, FE test) and (**) to significant deviations ($p<10^{-3}$, FE test) from the references in (A).

Thought Experiment

- A Thought Experiment:
  1. Suppose that different genes have different gene duplicabilities.
  2. Suppose also that gene duplicability is preserved after gene duplication.
  3. Then: the genome will come to be populated by genes with high gene duplicability.

Gene duplication 'tournament'

The differential expansion of the genome toward genes more likely to give rise to other genes.



Perna et al. Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
NATURE |VOL 409 | 25 JANUARY 2001

Some equations for this thought experiment

The duplicability of genetic elements of type i is :

$$d_i := \text{Rate}[i \textit{ duplicates}] \times \text{Prob}[i \textit{ fixes}] \times \text{Prob}[i \textit{ maintained}]$$

- Mobile genetic elements have high values of Rate[i duplicates] but low values of $\text{Prob}[i$ *fixes*] and Pr[i *maintained*].
- Genes in large gene families and regulatory elements have high values for $\text{Prob}[i$ fixes] $\times$ $\text{Prob}[i$ maintained]

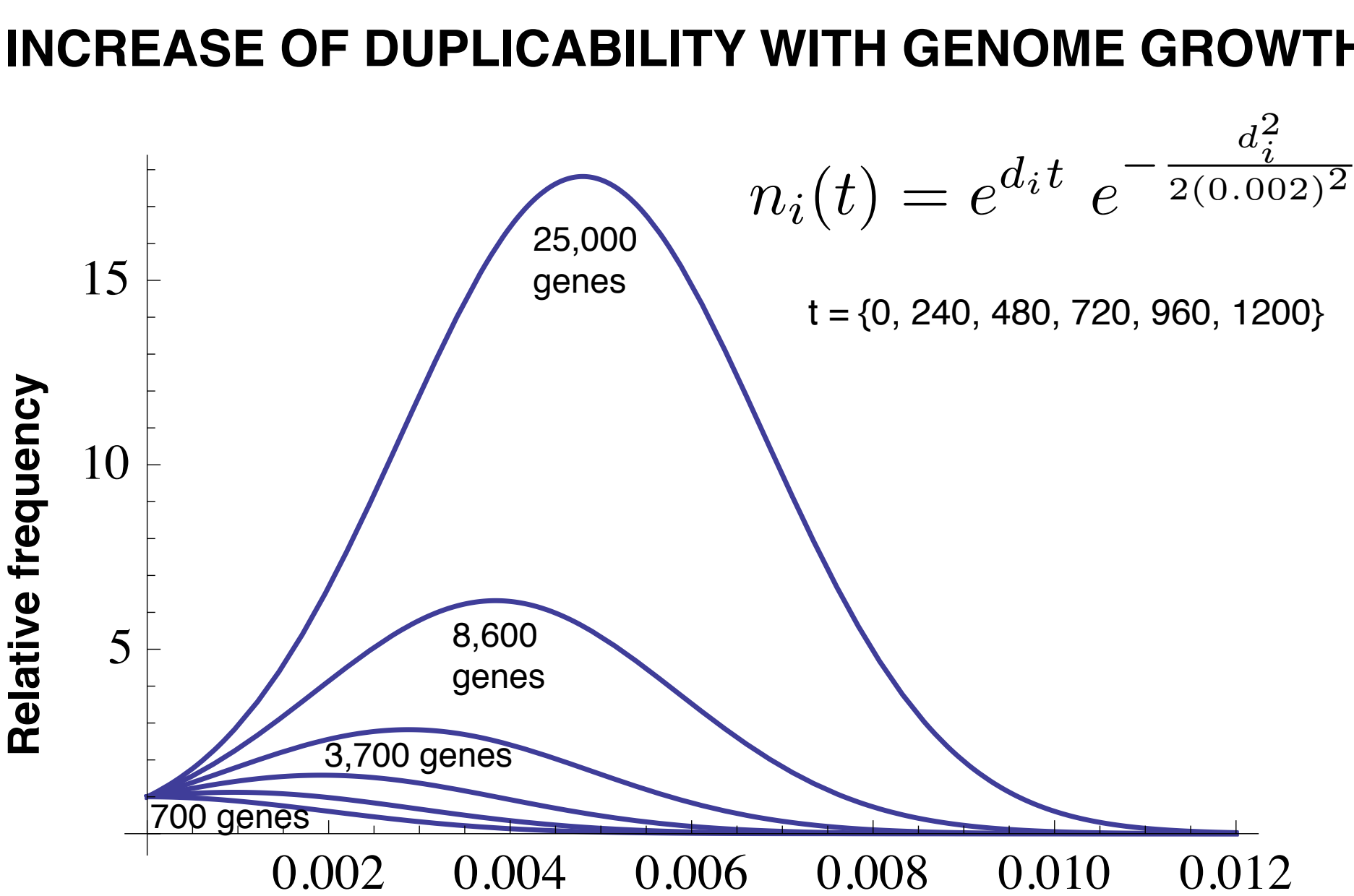
Fundamental Implication

Prob[i **fixes**] —increases with the upper tail of the distribution of fitness effects of duplication

Prob[i **maintained**] —increases with the lower tail of the distribution of fitness effects of deletion or pseudogenization

- Because bigger upper tails of the DFE of duplication confer high gene-duplicability,
- the genome can grow bigger upper tails of the DFE of duplication through the addition of new genes.
- i.e. the genome can grow in evolvability through the addition of new genes.

Thought Experiment Model, cont'd



Thought Experiment Model, concluded

Result (Fisher's Fundamental Theorem of Natural Selection applied to genome growth)

Assuming that gene-duplicability d_i is perfectly transmitted between gene duplications, the genomic rate of production of genes that go to fixation and are maintained,

$$\overline{d}(t) = \sum_{i \in \mathcal{G}} d_i \, \frac{n_i(t)}{N(t)},$$

increases at rate

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{d}(t) = \mathrm{Var}(d_i) \geq 0.$$

Arbitrary heritability of duplicability

Result (Price's Equation in genome expansion)

For a gene of type j, let

- d_j *be j's probability of being stably incorporated in the genome, while*
- ξ_j *be j's offspring's probability of being stably incorporated in the genome:* $\xi_j = \sum_{i \in \mathcal{G}} d_i \; T(i \leftarrow j)$.

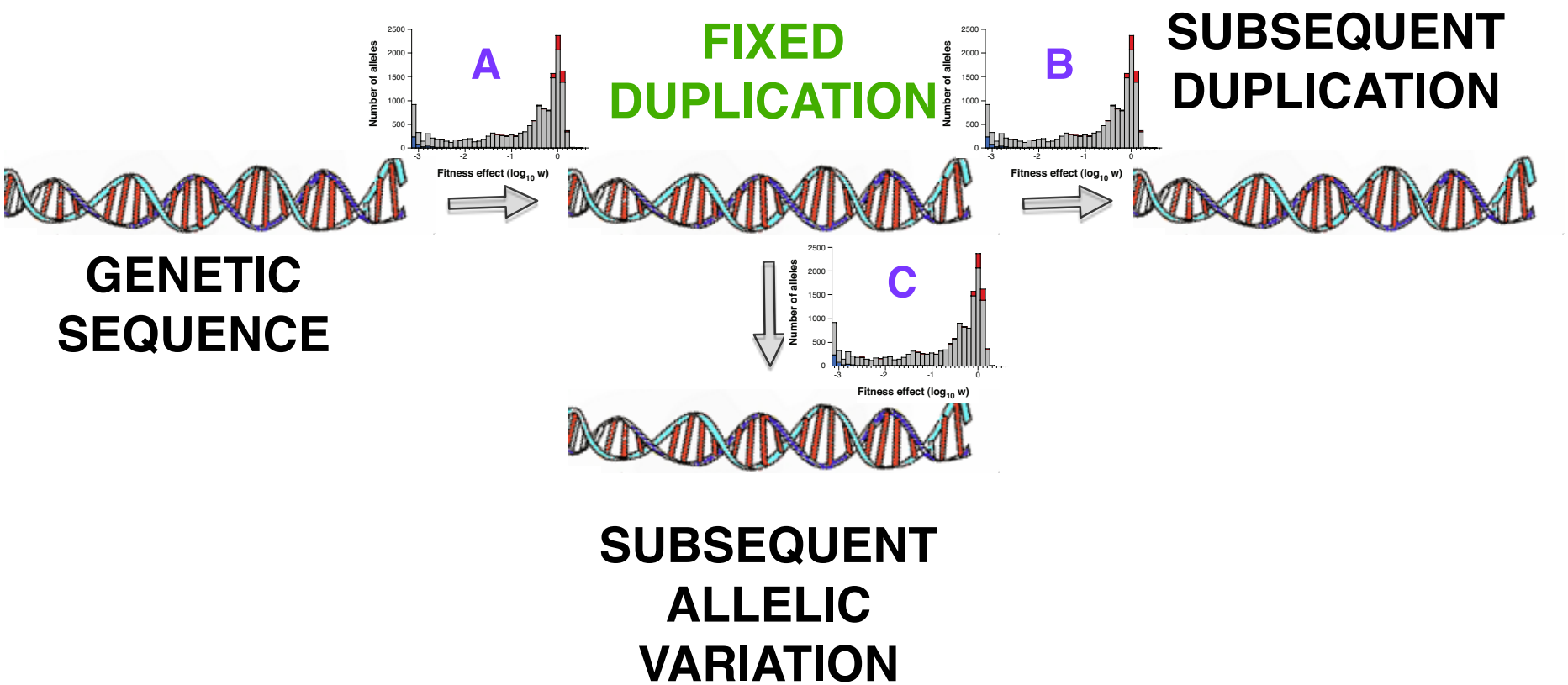
The rate of change in the average d_i of the genome is

$$\frac{d}{dt}\overline{d}(t) = \alpha \left\{ \mathrm{Cov}(\xi, d) + [\overline{\xi}(t) - \overline{d}(t)]\, \overline{d}(t) \right\},$$

where

$$\overline{\xi}(t) = \sum_{i \in \mathcal{G}} \xi_i \; p_i(t), \mathrm{Cov}(\xi, d) = \sum_{i \in \mathcal{G}} \xi_i \; d_i \; p_i(t) - \overline{\xi}(t) \; \overline{d}(t).$$

This was correlation between the upper tails of A and B.
What about between A and C?



Traits that are Correlated with Gene Duplicability

Robertson's Secondary Theorem of Natural Selection for Gene Duplicability:

1. Suppose that different genes have different gene duplicability.
2. Suppose that there are other properties that covary with gene duplicability.
3. Then these covarying properties in the genome will increase with addition of new genes.

Idea in a Nutshell-1st & 2nd

So, on the level of genome-as-population, we have

1. "Fisher's Fundamental Theorem of Natural Selection"
2. The Price equation, and
3. "Robertson's Secondary Theorem of Natural Selection"

where

1. gene duplicability takes on the role of fitness.

Heritability of Duplicability Can Evolve

- Suppose two genes both have high gene duplicability, but one has higher heritability of the duplicability from one gene duplication to the next.
- Then the gene with higher heritability of duplicability will predominate.

This is the Reduction Principle (Feldman, 1972; A., 2012) on the level of genome-as-population.

The Empirical Research Program

Population Genetics of the "Genome-as-population":
Recall the components of gene duplicability:

$$d_i := \text{Rate}[i \text{ duplicates}] \times \Pr[i \text{ fixes}] \times \Pr[i \text{ maintained}].$$

For each possible event that adds or removes genetic material i in the genome: Quantify

1. The rate at which it occurs
2. The probability that it goes to fixation in the population
3. The probability that it becomes stably maintained by selection
4. Properties that are correlated with each of the above:
    1. Effects of dosage change
    2. Positions within gene interaction networks
    3. Effects on phenotypes under directional selection
    4. Effects on phenotypes under stabilizing selection

The Empirical Research Program, cont'd

Quantify the heritability of these properties,

i.e the preservation of these properties:

1. Over time
2. From one duplication to the next
3. With primary sequence evolution of the element itself
4. With primary sequence evolution of other epistatic elements
5. With changes in the environment.

The Empirical Research Program, cont'd

Parts of this program already underway:

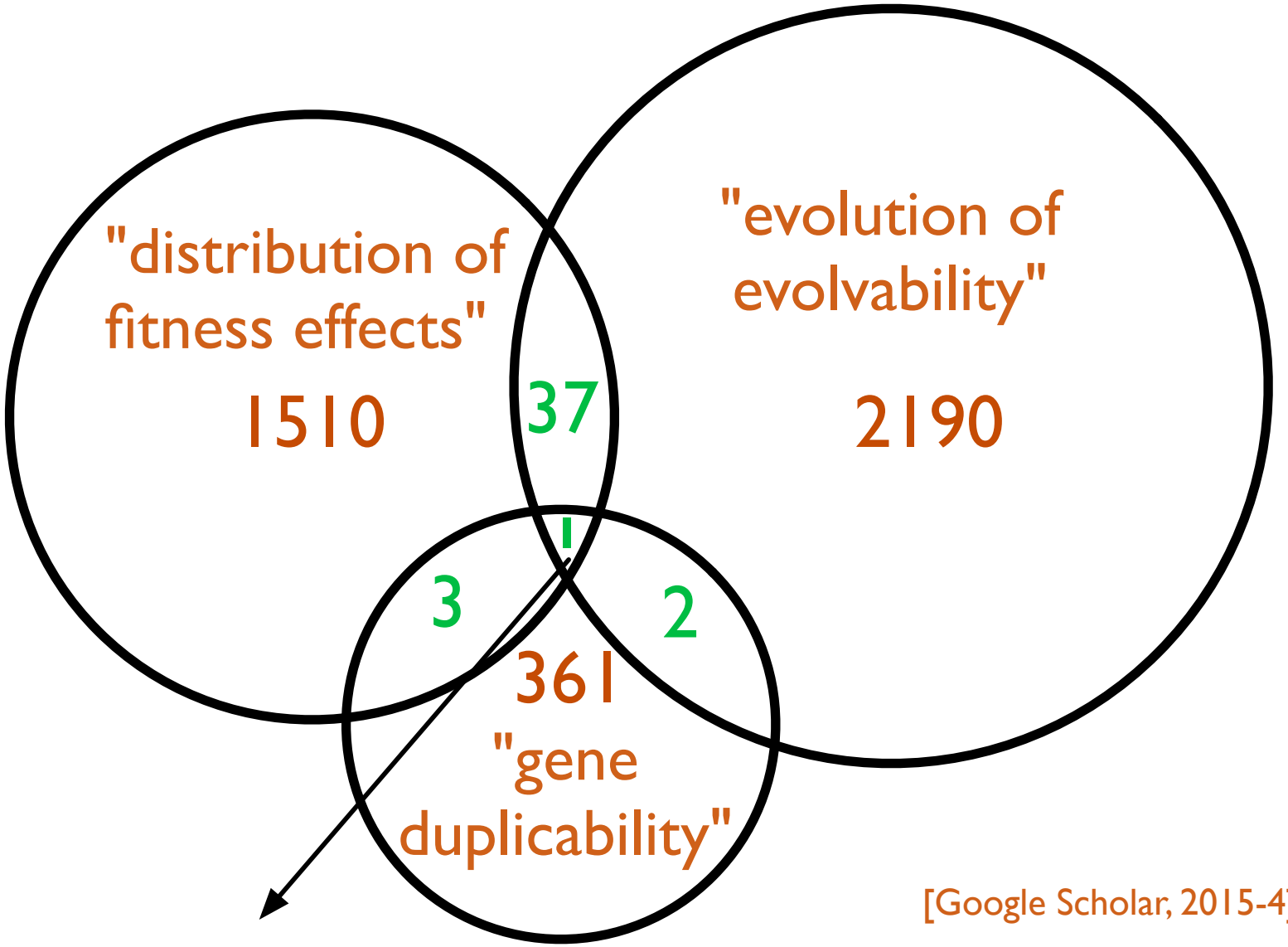
- Quantifying properties correlated with gene duplicability:
  - Connectedness of the protein in the protein interaction network (Guo et al., 2014)
  - Dosage effects of gene duplication (Qian and Zhang, 2008)
  - Complexity of the protein (Yang et al., 2003)
  - Complexity of the organism (Yang et al., 2003)
  - Production of dominant deleterious allelic variation (Singh et al., 2014)
- *In silico* studies (Fischer, Bernard, Beslon, and Knibbe (2014) A Model for Genome Size Evolution)

The Empirical Research Program, cont'd

What is missing from the research program so far:

1. Quantification of the heritability of properties on the level of genome-as-population
2. Examining the consequences of differential gene duplicability for the variational properties of the genome
3. Widespread appreciation that all the diverse events that add and remove genetic material in the genome are part of a single theoretical framework.

NUMBER OF PAPERS WITH:



[Google Scholar, 2015-4]

Systems-biology approaches for predicting genomic evolution
B Papp, RA Notebaart, C Pál - Nature Reviews Genetics, 2011 - nature.com

Conclusion

- The only ways this mechanism will have not left a mark on evolvability and the genotype-phenotype map is:
  1. There are no differences in the duplicability of new genes, or
  2. The erasure of the variation in duplicability happens uniformly and rapidly among all genes
- Neither condition is tenable empirically.
- Therefore, we expect the conditions of gene origin have shaped the genotype-phenotype map and evolvability, and its extent remains an open empirical area for research.

Acknowledgements

This research is supported by:

- The Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria

- The Mathematical Biosciences Institute, Columbus, Ohio under their grant from the National Science Foundation

Further reading: Altenberg (1995) Genome growth and the evolution of the genotype-phenotype map.

Thank you for your attention!

A. 2012. Resolvent positive linear operators exhibit the reduction phenomenon. *Proceedings of the National Academy of Sciences U.S.A.*, 109(10):3705–3710.

Altenberg, L. 1995. Genome growth and the evolution of the genotype-phenotype map. In Banzhaf, W. and Eeckman, F. H., editors, *Evolution and Biocomputation: Computational Models of Evolution*, volume 899 of *Lecture Notes in Computer Science*, pages 205–259. Springer-Verlag, Berlin.

Crombach, A. and Hogeweg, P. 2007. Chromosome rearrangements and the evolution of genome structuring and adaptability. *Mol Biol Evol*, 24:1130–1139.

Crombach, A. and Hogeweg, P. 2008. Evolution of evolvability in gene regulatory networks. *PLoS Computational Biology*, 4(7):e1000112.

Draghi, J. and Wagner, G. P. 2008. Evolution of evolvability in a developmental model. *Evolution*, 62(2):301Ð315.

Feldman, M. W. 1972. Selection for linkage modification: I. Random mating populations. *Theoretical Population Biology*, 3:324–346.

Fischer, S., Bernard, S., Beslon, G., and Knibbe, C. 2014. A model for genome size evolution. *Bulletin of mathematical biology*, 76(9):2249–2291.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Guo, Z., Jiang, W., Lages, N., Borcherds, W., and Wang, D. 2014. Relationship between gene duplicability and diversifiability in the topology of biochemical networks. *BMC genomics*, 15(1):577.

Kashtan, N. and Alon, U. 2005. Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences U.S.A.*, 102:13773–13778.

Kauffman, S. A. and Levin, S. 1987. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128:11–45.

Levinton, J. 1988. *Genetics, Paleontology, and Macroevolution*. Cambridge University Press, Cambridge.

Meyers, L. A., Ancel, F. D., and Lachmann, M. 2005. Evolution of genetic potential. *PLoS Computational Biology*, 1:236–243.

Pigliucci, M. and Müller, G. B., editors 2010. *Evolution — The Extended Synthesis*. MIT Press, Cambridge, MA.

Qian, W. and Zhang, J. 2008. Gene dosage and gene duplicability. *Genetics*, 179(4):2319–2324.

Riedl, R. J. 1977. A systems-analytical approach to macroevolutionary phenomena. *Quarterly Review of Biology*, 52:351–370.

Singh, P. P., Affeldt, S., Malaguti, G., and Isambert, H. 2014. Human dominant disease genes are enriched in paralogs originating from whole genome duplication. *PLoS computational biology*, 10(7):e1003754.

Wagner, A. 2008. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences*, 275(1630):91–100.

Yang, J., Lusk, R., and Li, W.-H. 2003. Organismal complexity, protein complexity, and gene duplicability. *Proceedings of the National Academy of Sciences*, 100(26):15661–15665.