The Evolution of Evolvability

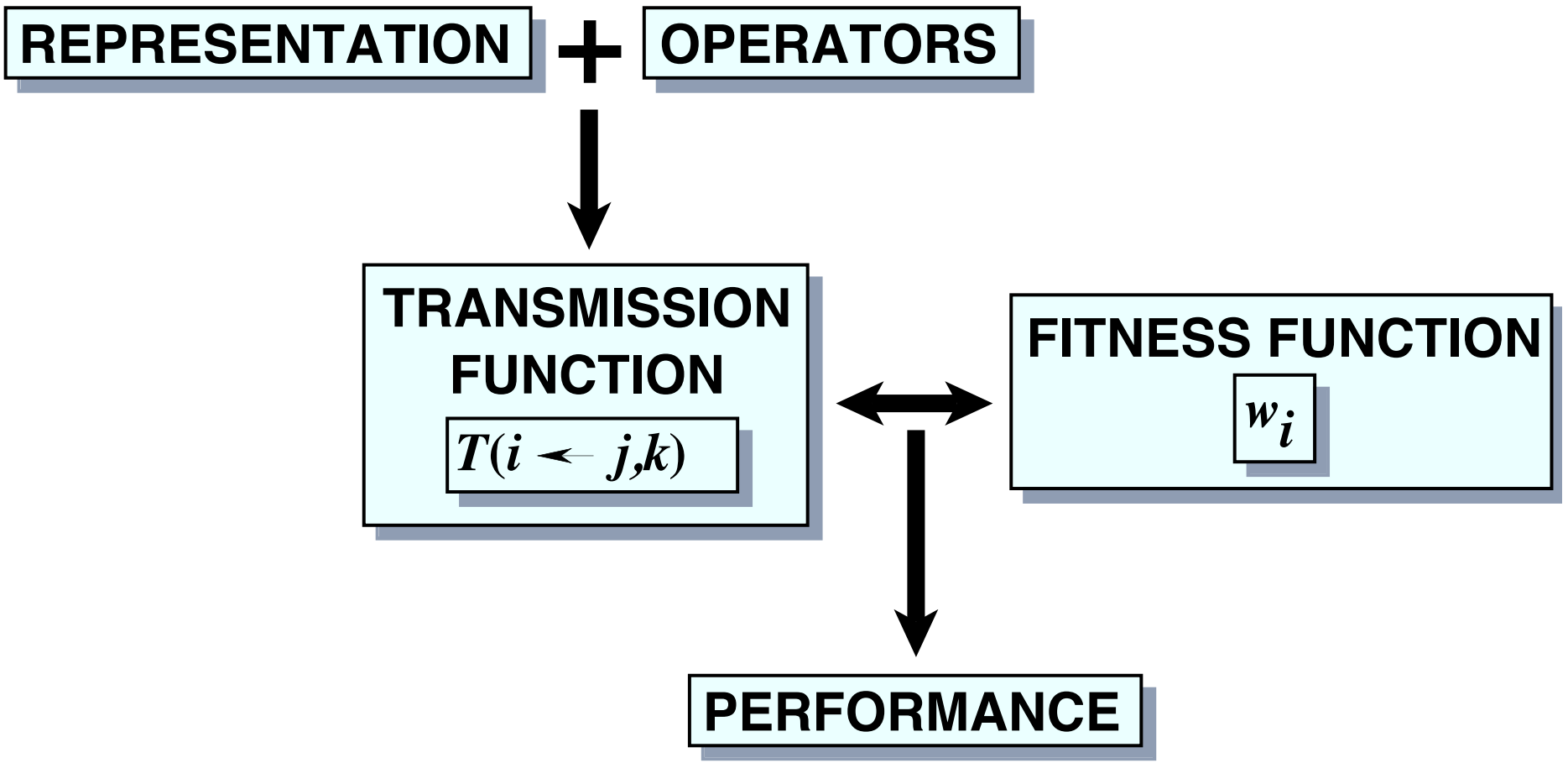
March 17, 2014, Simons Institute

Lee Altenberg
altenber@hawaii.edu
http://dynamics.org/Altenberg/

3 Goals for this talk

1. Review the early work on the evolution of evolvability
2. Dispel the myth that it requires group selection
3. Describe one mechanism for the evolution of evolvability due to individual selection.

A Framework for Evolutionary Computation



'Hard problems' for Evolutionary Algorithms

"Massive Multimodality"

$$F(x,y) = 21.5 + x\sin(4\pi x) + y\sin(20\pi y)$$

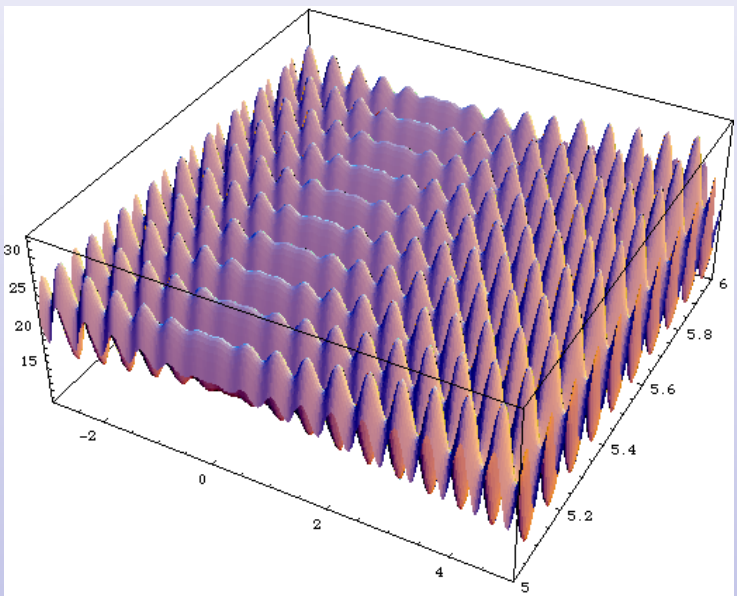
Mutation:

$$(x,y) \to (x+\varepsilon, y+\xi)$$

where mutation is produced by random variables distributed as:

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\varepsilon^2}{2\sigma^2}}$$

$$f(\xi) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\xi^2}{2\sigma^2}}$$

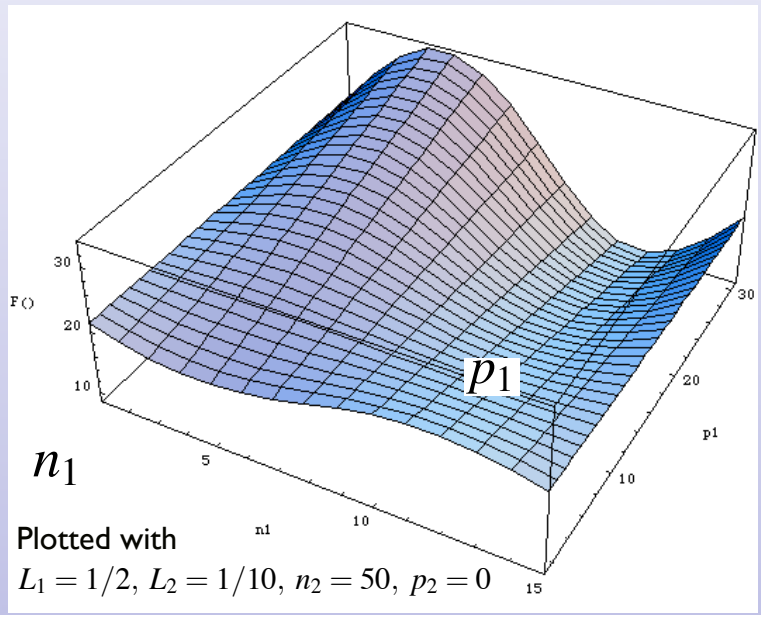


20

Changes in Representation can Solve 'EA Hardness'

- There is no reason the representation has to be the "natural" one.
- Rewrite the representation in terms of **phase** and **wave number**: $x = n_1 L_1 + p_1, \quad y = n_2 L_2 + p_2,$
  - where $L_1 = 1/2,\ L_2 = 1/10,\ n_1, n_2 \in Z,\ p_1, p_2 \in \Re \mod 1$

The adaptive landscape becomes smooth.



Plotted with
$L_1 = 1/2,\ L_2 = 1/10,\ n_2 = 50,\ p_2 = 0$

Can be Equivalent to a Change in Variation Operators

...which is equivalent to a change in the mutation operator:

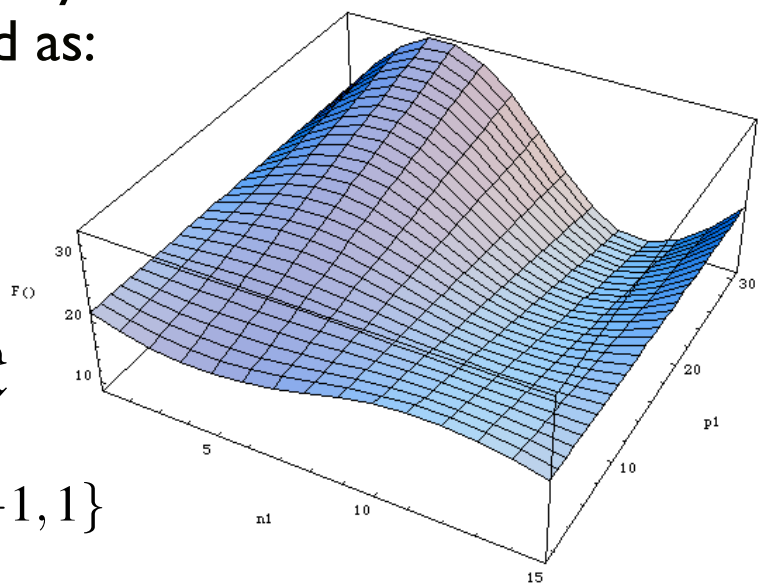
- From: $(x, y) \rightarrow (x + \varepsilon, y + \xi)$
- To: $(x, y) \rightarrow (x + \varepsilon + \nu,\ y + \xi + \mu)$

where mutation is produced by random variables distributed as:

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\varepsilon^2}{2\sigma^2}}$$

$$f(\xi) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\xi^2}{2\sigma^2}} \quad \varepsilon, \xi \in \Re$$

$$f(\nu) = f(\mu) = 1/2 \ \text{ for } \nu, \mu \in \{-1, 1\}$$



Operator/Representation Duality:

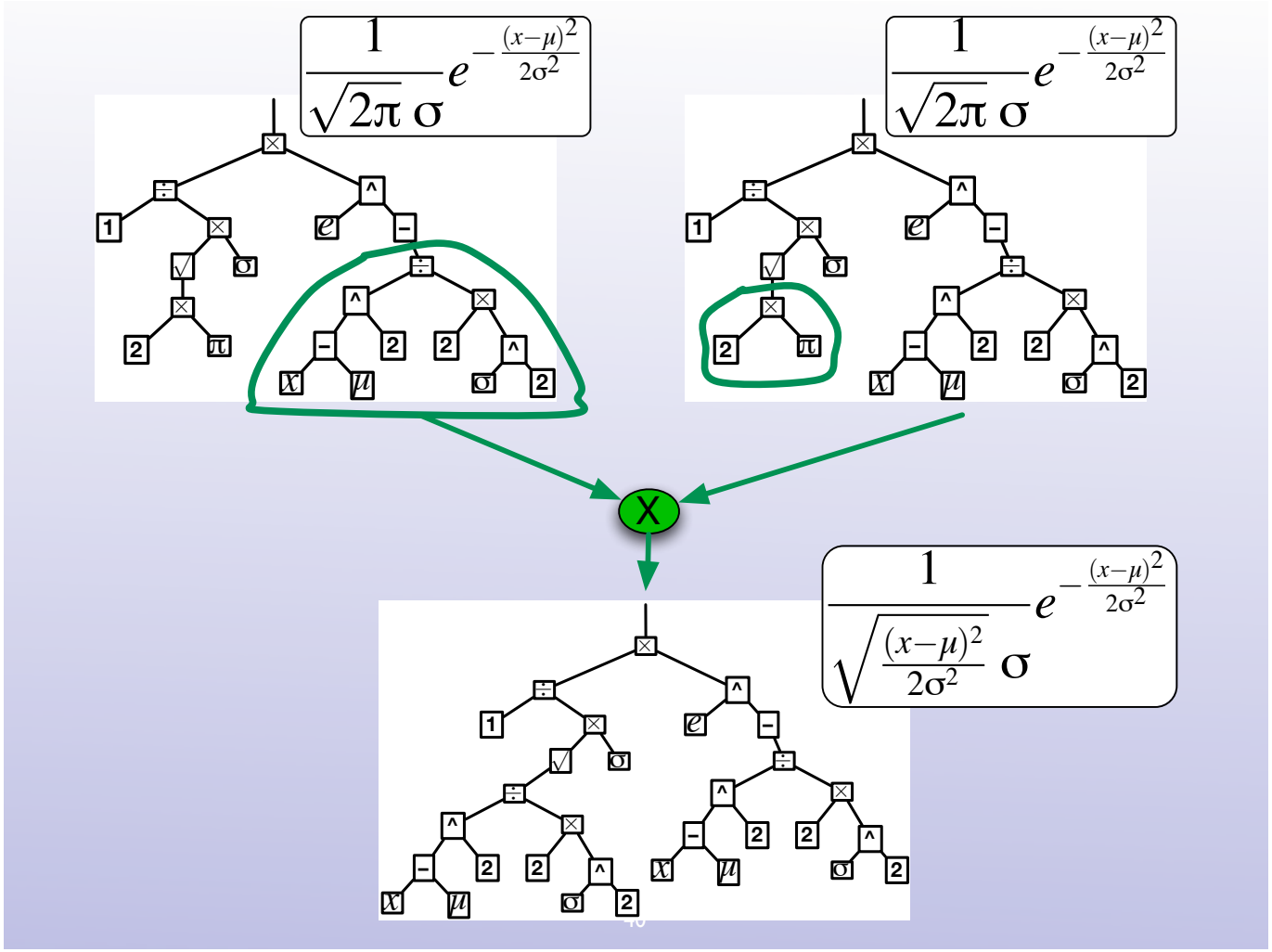
- Changes in representations

may be equivalent to

- changes in genetic operators

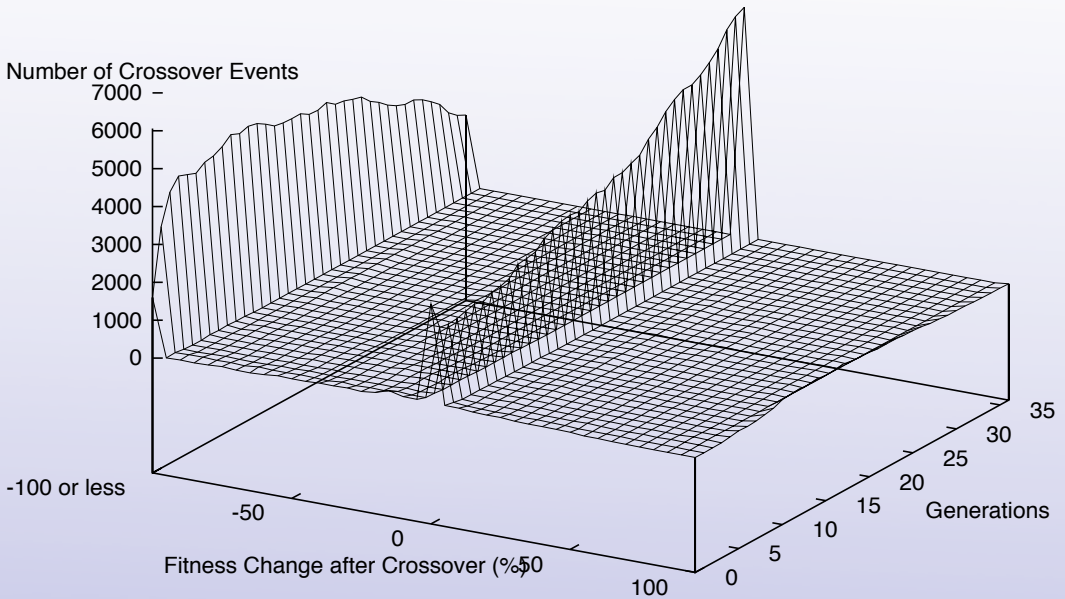
in producing the same new transmission function, $T(i \leftarrow j, k)$.

Emergent Representations: Genetic Programming



An intriguing result:

- Evolution of robustness to crossover in genetic programming



Nordin & Banzhaf, 1995. Complexity compression and evolution. ICGA95. 26

There's that bi-modality that Joanna Masel described in her talk.

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."*

The evolution of evolvability is precisely one of the subjects among "the study of the origin of variation."

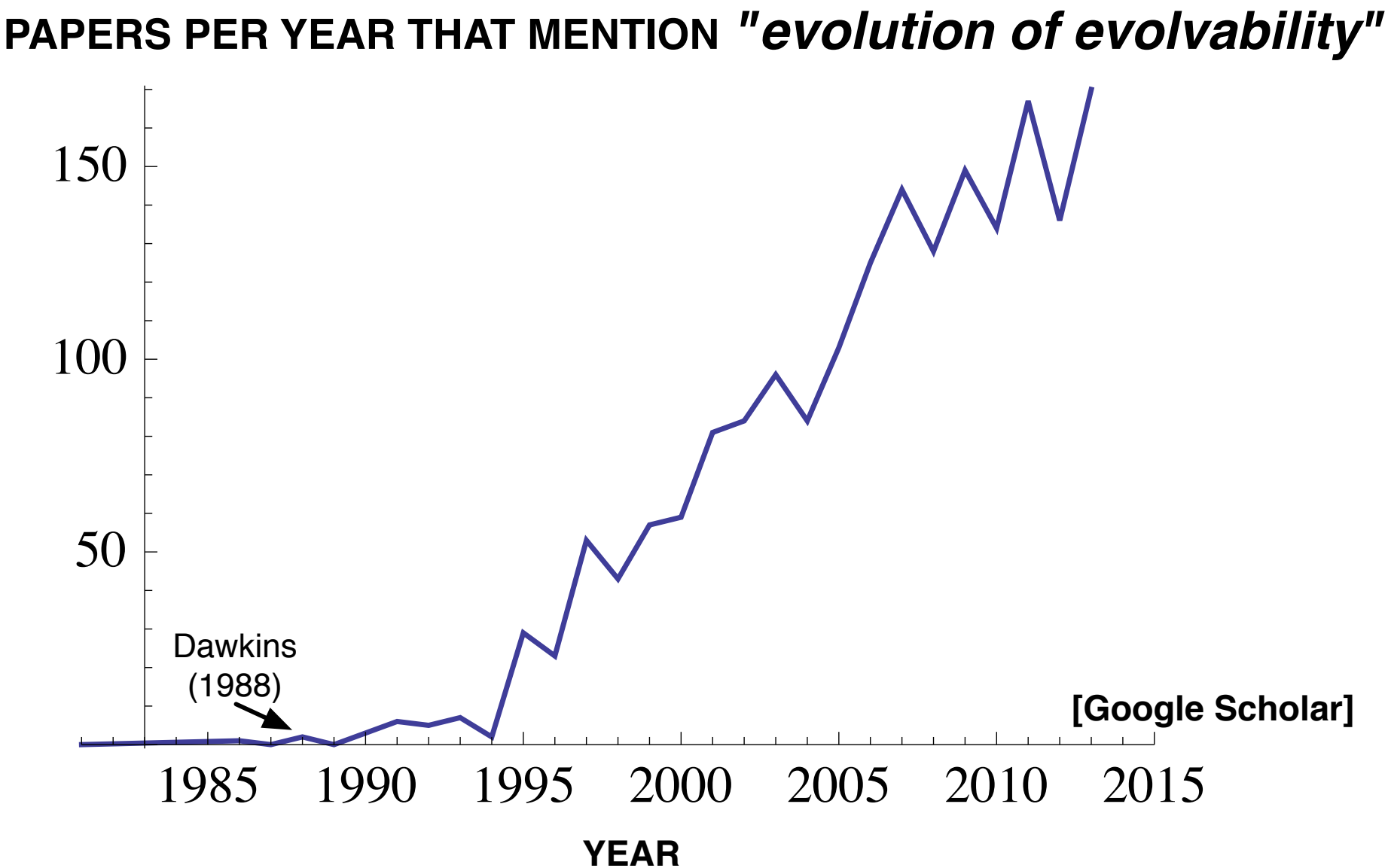
Origins

- The idea of 'evolvability' and its evolution — under different names — goes back at least to
  1. Riedl (1975), *Die Ordnung des Lebendigen: Systembedingungen der Evolution*: "**increase in the probability of a successful adaptation**";
  2. Conrad (1977), "**efficient evolutionary behavior**";
  3. Conrad (1979), "**increase evolutionary amenability**";
  4. Conrad and Volkenstein (1981), "Replaceability of amino acids and the **self-facilitation of evolution**".

"The Evolution of Evolvability"

- Dawkins (1988) coined the catchy phrase 'evolution of evolvability'
- Subsequently adopted by several researchers working on this area but using other terms:
  - Conrad (1990)
  - Kauffman (1990)
  - Alberch (1991)
  - A. (1993)

Adoption of the phrase "evolution of evolvability"



Synopsis

- Over 120 papers per year now refer to the 'evolution of evolvability'
- Many ideas and mechanisms now populate this phrase
- I won't attempt here to review the whole field of work on the evolution of evolvability
- Rather, I want to describe one specific mechanism — the role of gene origin in the evolution of evolvability
- I won't even make claims about its relative importance — that is an empirical question.
- Rather, I want to show how different phenomena are connected theoretically.

'Anathema': Dawkins (1988) The Evolution of Evolvability

- *"A title like The Evolution of Evolvability ought to be anathema to a dyed-in-the-wool, radical neo-Darwinian like me. . . .*
- *As the ages go by, changes in embryology that increase evolutionary richness tend to be self-perpetuating.*
- *I am talking about a kind of higher-level selection, a selection not for survivability but for evolvability. . . .*
- *It now seems to me that an embryology that is pregnant with evolutionary potential is a good candidate for a higher-level property of just the kind that we must have before we allow ourselves to speak of species or higher-level selection."*

'Controversial': Conrad (1990) The Geometry of Evolution

- *"Let us now recall why the concept of evolvability is controversial.*
- *Some evolutionists argue that 'natural selection can act only on properties that are advantageous to the individual. Evolvability is advantageous to the species. Do not, therefore, let the concept of evolvability mix into biological thinking.'*

'Controversial': Conrad (1990), cont'd

1. *This dictum is wrong on two counts:*
2. *Some mutation buffering redundancies are in fact advantageous to the individual organism. Some of the redundancies that confer stability on the phenotypic dynamics also serve to buffer the effect of genetic change.*
3. *Mutation buffering and other evolution-facilitating mechanisms can accumulate even if they are a tax from the standpoint of the individual organism. When they occur, the evolution-facilitating redundancies will hitchhike along with the advantageous traits whose appearance they facilitate."*

'Controversial': Sniegowski and Murphy (2006)

- *"Because populations, not individuals, evolve and adapt, it follows that evolvability-as-adaptation must be the consequence of selection among populations rather than selection among individuals.*
- *Selection among populations is possible, in principle, but it is a very weak force compared with individual-level selection."*

'Group Selection': Lynch (2007)

- *"First, evolution is a population-level feature.*
- *Thus, if an organismal feature that modifies the ability to evolve is to be advanced directly by adaptive mechanisms, selection must operate efficiently at a higher level of organization than the individual.*
- *This requires a significantly subdivided population structure, with levels of evolvability being positively correlated with population longevity and/or productivity.*
- *Because populations survive longer than individuals, such group selection is expected to be a much weaker force than individual selection, and necessarily operates on much longer time scales."*

'Controversial': Zhong and Priest (2011)

- *"One of the reasons the concept of evolvability is controversial is that models of evolvability usually invoke levels of selection above the individual.*
- *Because natural selection lacks foresight and tends to fix alleles that maximise current fitness regardless of the consequences for future evolutionary potential of the population, evolvability is generally not expected to be selected at the level of individuals."*

'Controversial': Graves et al. (2013)

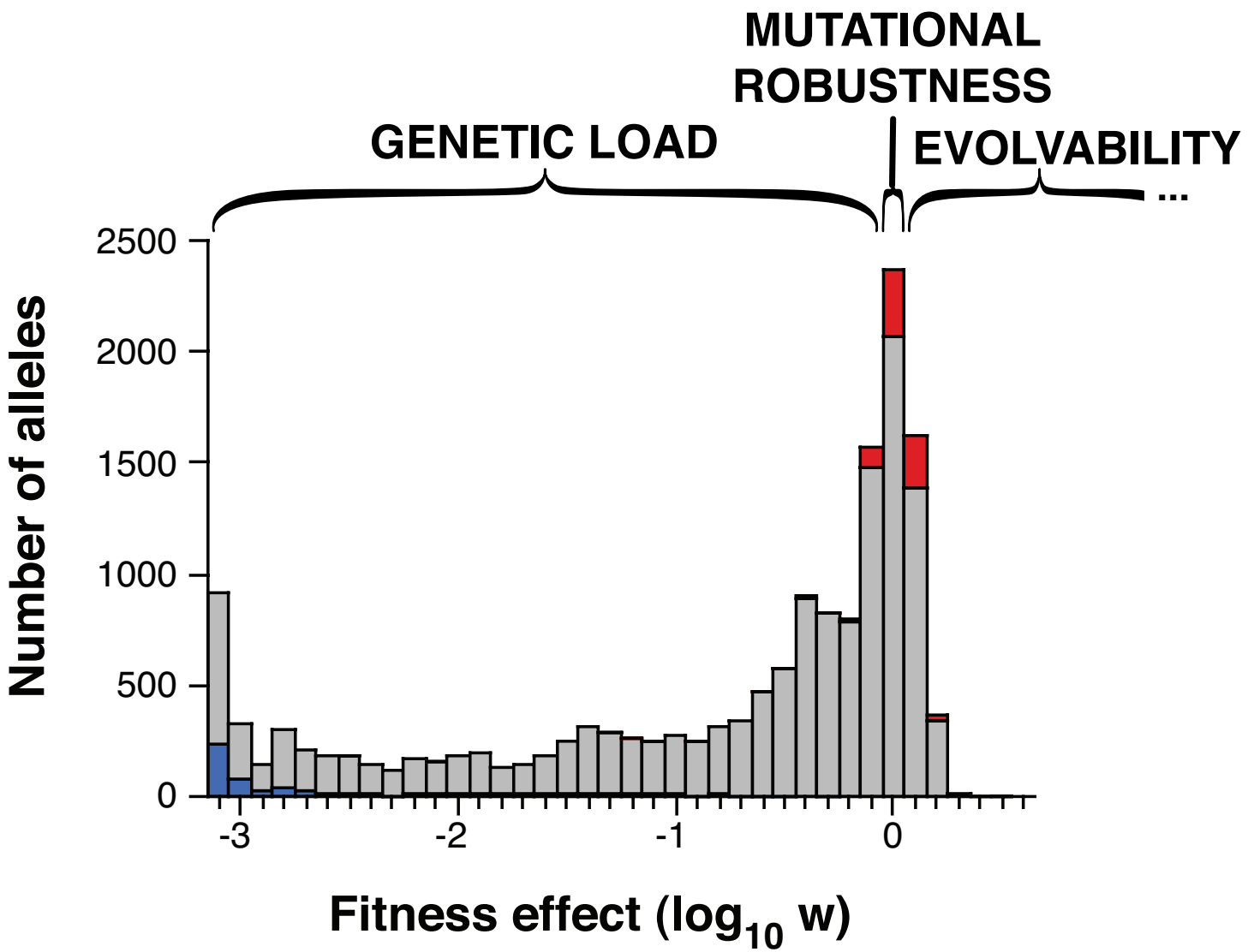
- *"The hypothesis that differences in evolvability result from past natural selection acting on the ability to evolve, however, remains highly controversial for two primary reasons.*
  1. *Evolvability is a population-level phenotype and thus must be favored by the relatively weak forces generated by natural selection at the population level.*
  2. *Selection on evolvability suggests the unlikely scenario that natural selection has the evolutionary foresight to adapt a population to future environmental contingencies."*

Despite the early literature explaining how individual selection can act on evolvability, the myth persists that group selection is required.

Definition: Probability of Producing Adaptive Variants

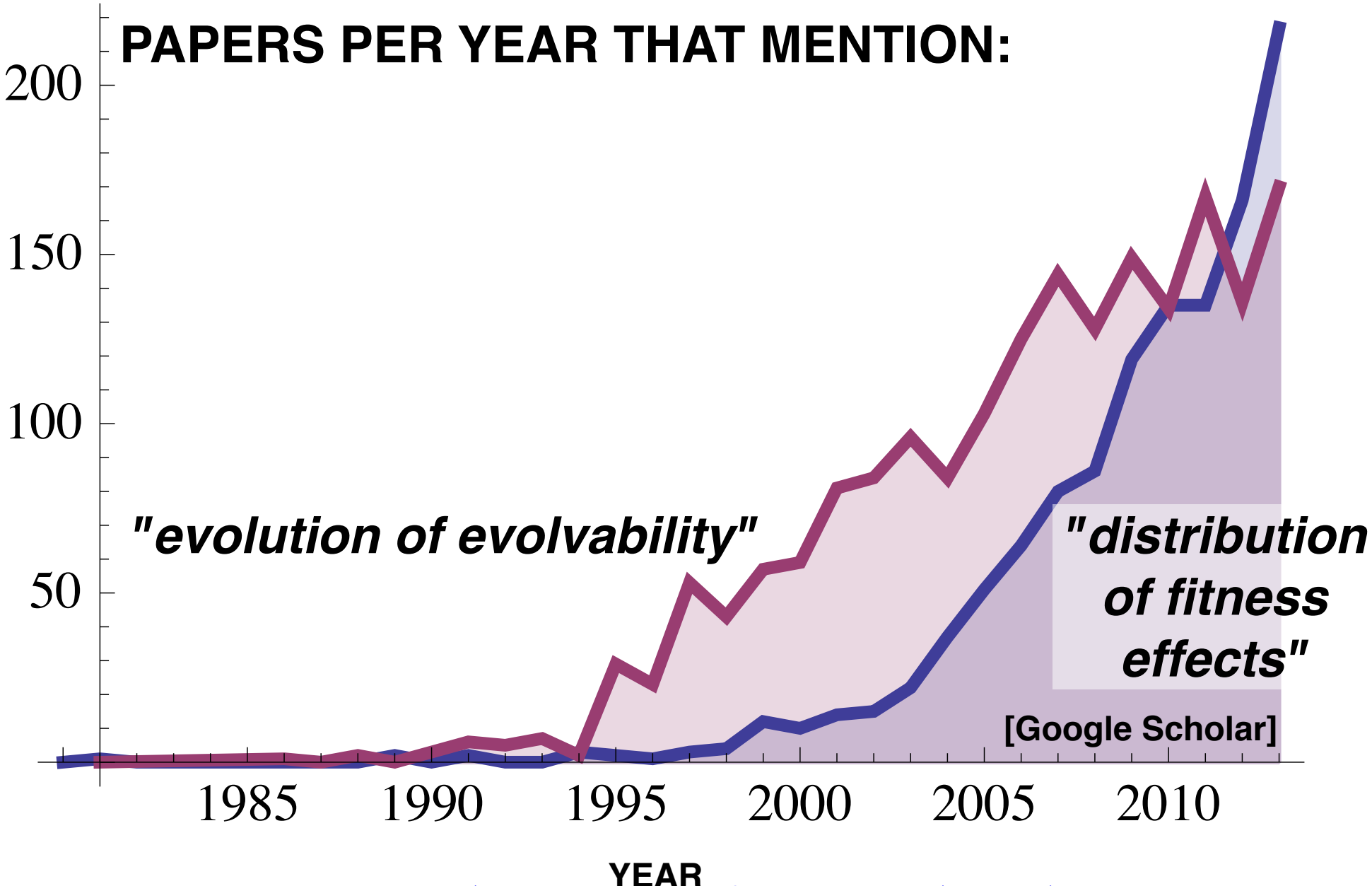
- The property I will refer to as "evolvability" is the probability that an organism generates adaptive genetic variation
- i.e. offspring with higher fitness than the parents.
- This is clearly not a population property but a property of organisms.
- It is the upper tail of the fitness distribution of an organism's offspring.
- This 'evolvability' then is a sub-property of the distribution of fitness effects of genetic variation.

3 Parts of the "Distribution of Fitness Effects"

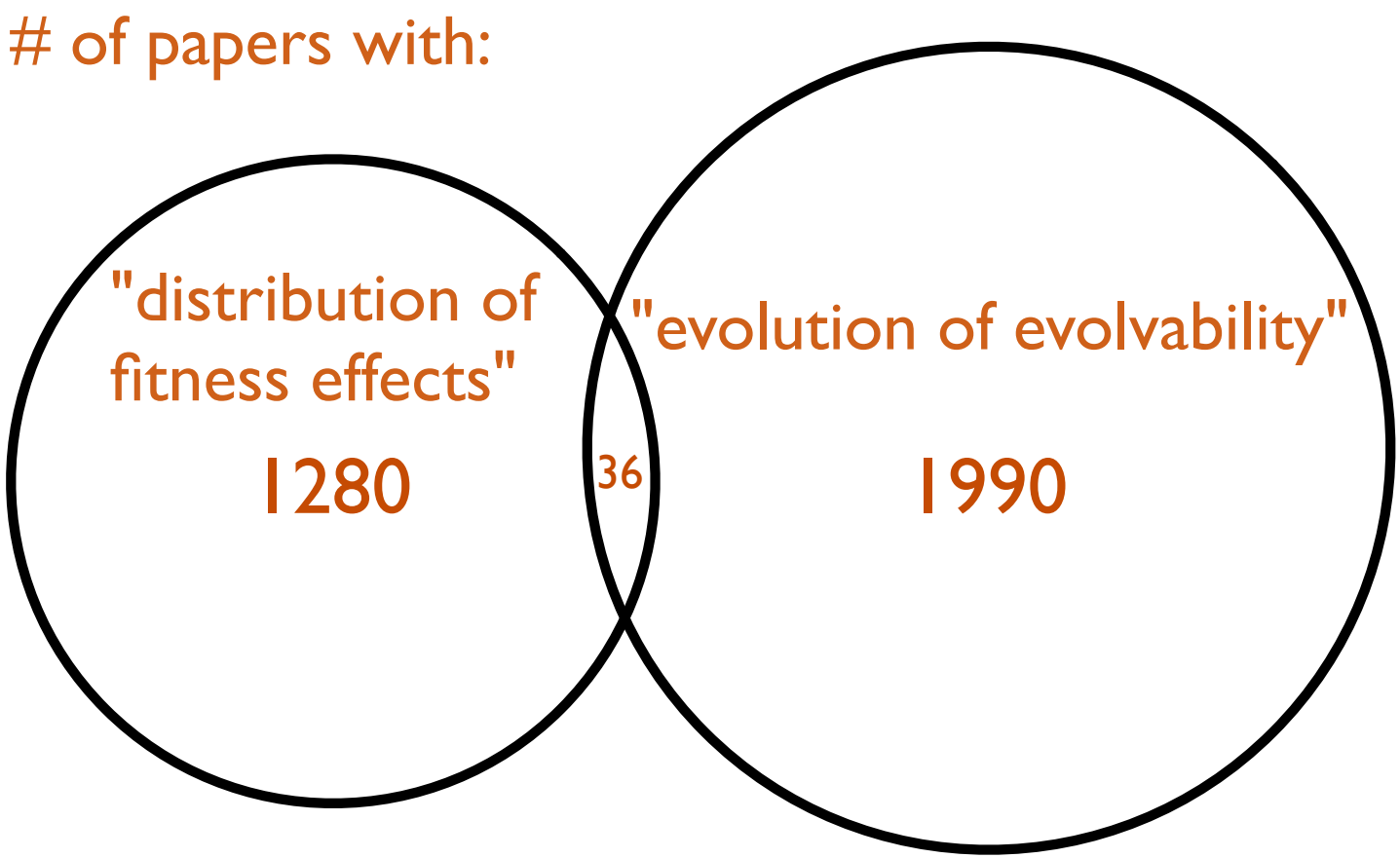


Firnberg, E., Labonte, J. W., Gray, J. J., & Ostermeier, M. (2014). A comprehensive, high-resolution map of a gene's fitness landscape. *Molecular Biology and Evolution*, msu081.

Adoption of the phrase "distribution of fitness effects"



Intersection of "evolution of evolvability" and "distribution of fitness effects"



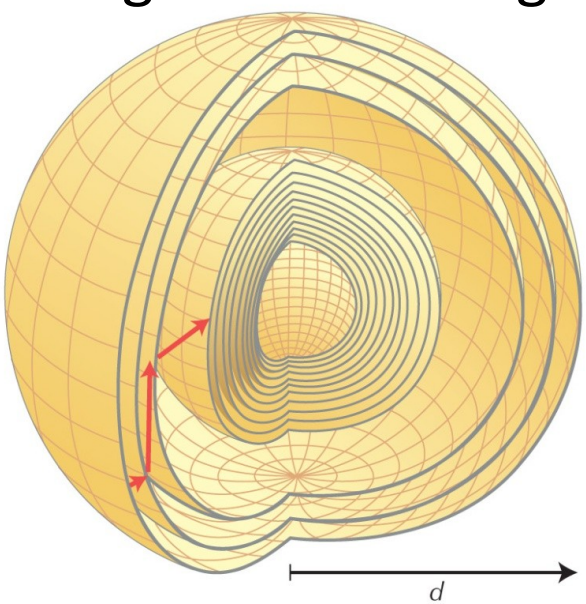
[Google Scholar, 2014-3]

One goal of future work: get these two communities talking to each other.

Evolvability as a distribution of fitness effects

- Defined as a distribution of fitness effects of mutation, evolvability a well-known entity.
- Evolvability changes all the time.
- Described as early as Fisher's (1930) geometric argument.

Fisher's geometrical argument



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

- Q. *"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?*
- A. *"The mutation of only one genetic unit, the operator, will result in the change of both. If the probability of mutation (P_m) and the probability of the success (P_s) of a superimposed gene do not differ greatly from those of structural genes, . . . then the chance of a successful alteration would rise from $(P_m P_s)^2$ to $(P_m P_s)$ or from 10^{-12} to 10^{-6}, i.e., would increase as much as a millionfold."*

Riedl (1977), cont'd

- *"Such a millionfold increase in the adaptive chances can therefore be achieved if the* **unnecessary, genetically redundant or risky independence or adaptive freedom of a single genetic unit is avoided** *by expedient* **dependency**.
- *"Such adaptive advantages by systemization 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."*
- *If a system like the operon were not known to exist, we could have predicted that it must exist."*

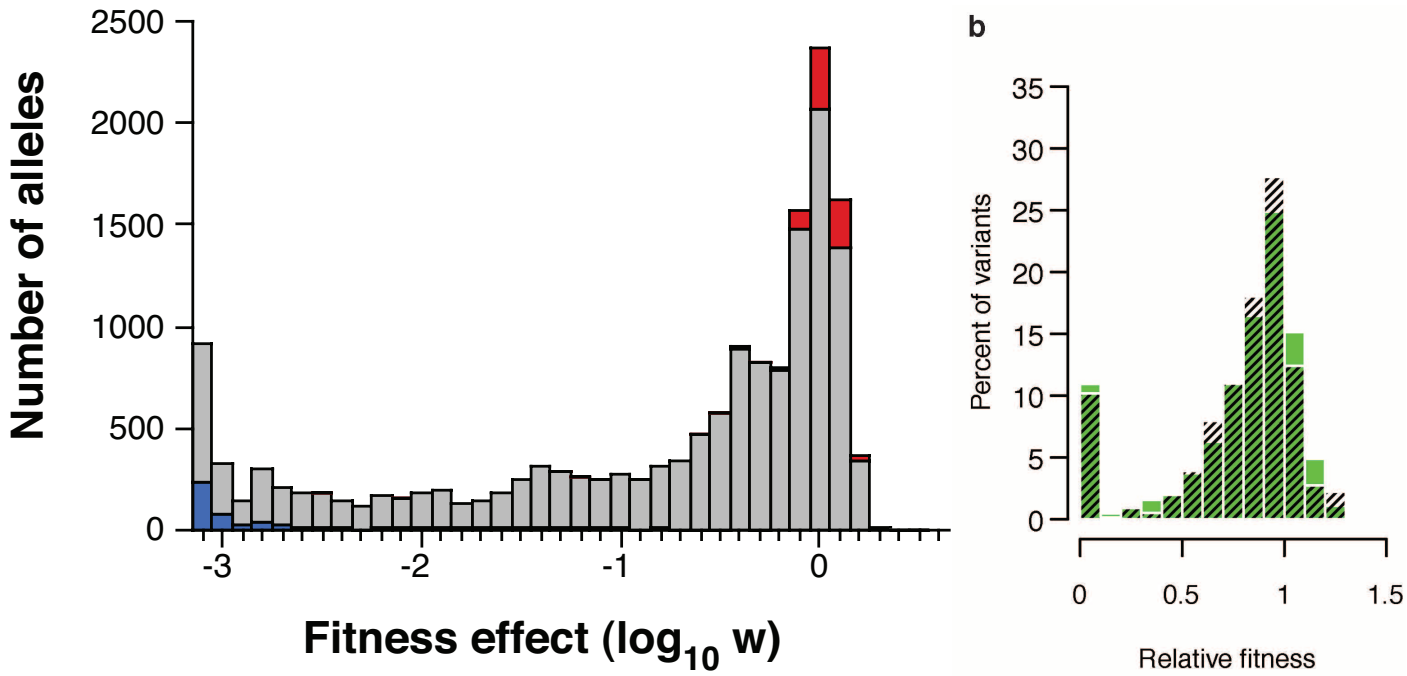
Riedl (1977), concluded

- *"... 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.*
- *This positive feedback of the* **adaptive speed (or probability)** [i.e. evolvability] *within a single adaptive direction is compensated by negative feedback in most of the alternative directions* [i.e. mutational robustness]*."*

'Constructional Selection' for Evolvability

- Different genetic transformations each have their own distribution of fitness effects:
  1. Allelic Mutations
  2. Deletions
  3. Recombination
  4. Gene duplications
  5. *De novo* gene origin
- Here I will be concerned chiefly with 1 and 4, 5 — allelic mutations and gene origins — and the relationship between them.

The Evolution of Evolvability



Basic questions:

- Q. What determines the distribution of fitness effects of mutation, or of gene duplication?

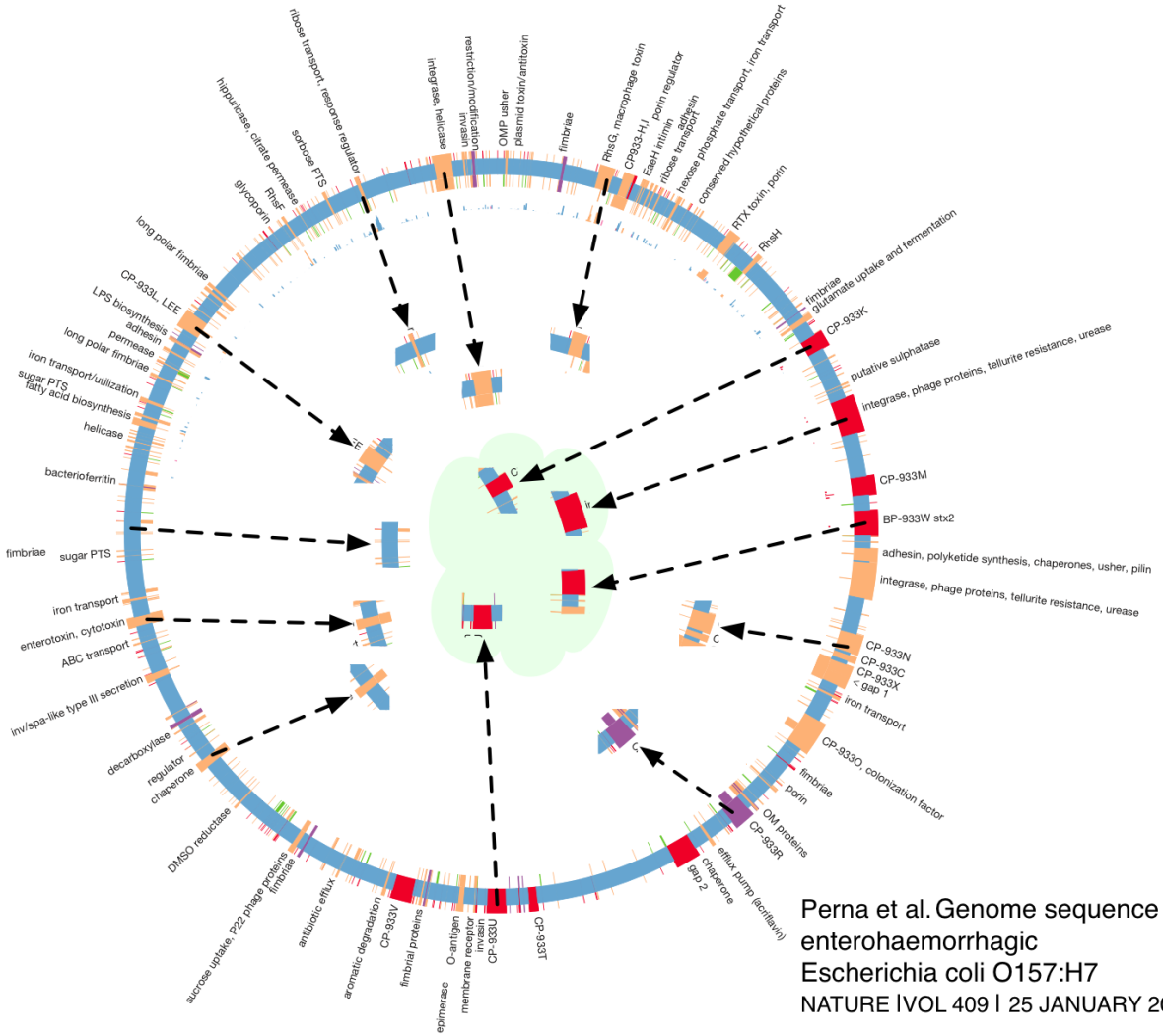
Determinants of the Probability of Adaptive Variants

- Adaptive opportunity due to current imperfection or environment change
- Probability of improving imperfect functions while not disturbing adapted functions
- The way that genetic variation maps to phenotypic variation is fundamental to whether or not that variation has the possibility of producing adaptive change.

- Even when strong opportunity exists for new adaptations in an organism, many of its previously evolved functions will remain under stabilizing selection.
- Adaptation requires variation that be able to move the organismal phenotype toward traits under directional selection without greatly disturbing traits remaining under stabilizing selection.
- Variation that disturbs existing adaptations as it produces new adaptations — i.e. variation which is *pleiotropic* — will have difficulty producing an overall fitness advantage.

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

'Constructional selection':

Selection during the origin of genes

- provides a filter on the construction of the genotype-phenotype map
- which naturally produces evolvability.

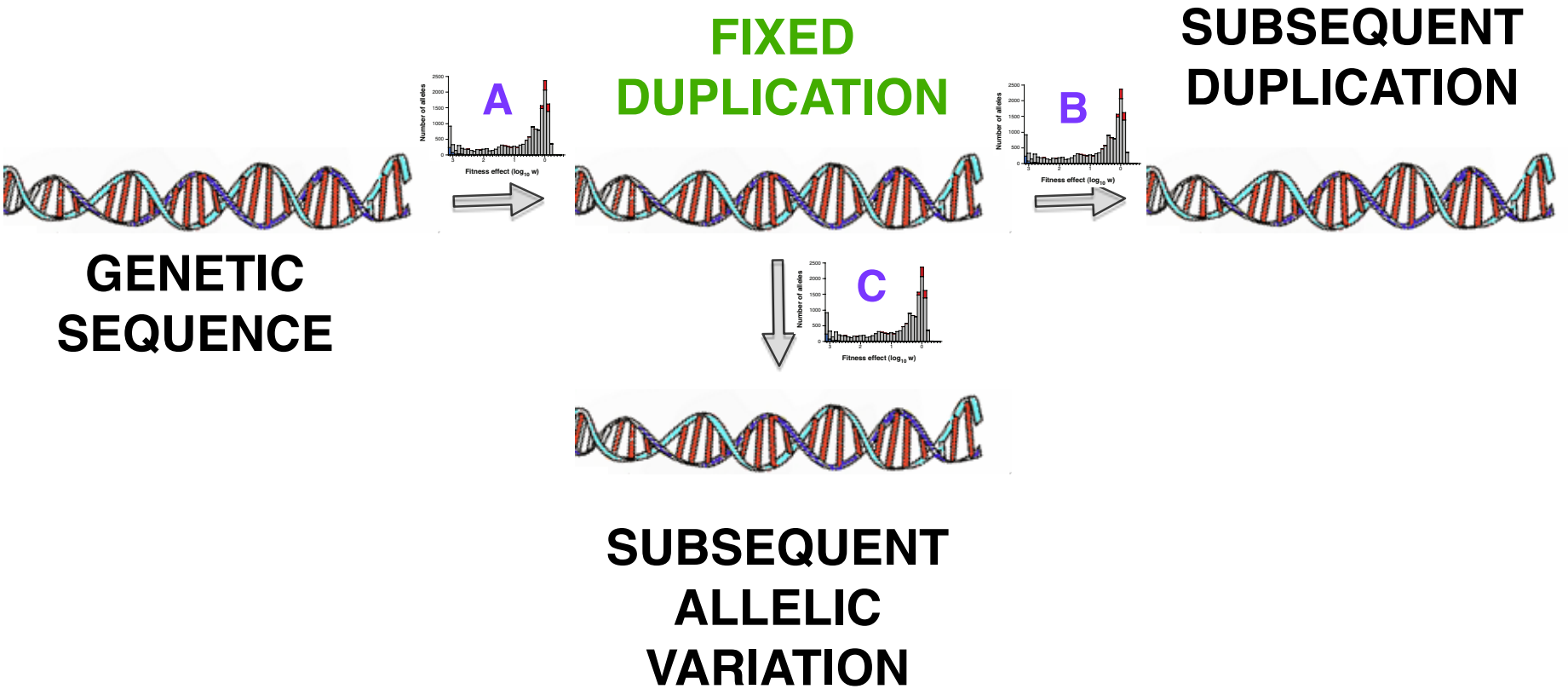
Unique Effects of Gene Duplication:

When a gene duplication is retained by the genome, two things are changed:

1. The probability of allelic variation in that gene family is increased
2. The probability of subsequent gene duplications of that gene family is increased.

'Constructional selection', in brief:

1. The probability that a gene duplication goes to fixation in the population is a strongly increasing function of the fitness effect of its creation;
2. Different distributions of fitness effects among all the possible gene duplications in the genome means that new genes are enriched with those with fat upper fitness tails;
3. Creation of a new gene not only produces its current phenotypic effect, but carries with it a new "neighborhood" in "sequence space" — the kinds of variants that it can in turn give rise to — both allelic and subsequent duplicates.


FIXED
DUPLICATION
SUBSEQUENT
DUPLICATION
A
B
C
Number of alleles
Fitness effect ($\log_{10}$ w)
GENETIC
SEQUENCE
SUBSEQUENT
ALLELIC
VARIATION

'Constructional selection', cont'd:

A Because a gene's creation, its subsequent duplication, and allelic diversification, are most likely to be acting on the same set of organismal functions, the distribution of fitness effects should be related between these three events:

1. a gene's creation
2. subsequent allelic variation of the gene
3. subsequent duplications of the gene.

B Thus, the very fact of a gene's existence is a condition that biases the distribution of fitness effects of a gene's allelic variation and subsequent duplication.

C Item (A) is claimed on first principles, but the magnitude of the relationship is the principal determinant of this mechanism, and requires empirical quantification.

Prevalence of Gene Duplication

Table 1. Prevalence of gene duplication in all three domains of life[a]

	Total number of genes	Number of duplicate genes (% of duplicate genes)	Refs
Bacteria			
Mycoplasma pneumoniae	677	298 (44)	[65]
Helicobacter pylori	1590	266 (17)	[66]
Haemophilus influenzae	1709	284 (17)	[67]
Archaea			
Archaeoglobus fulgidus	2436	719 (30)	[68]
Eukarya			
Saccharomyces cerevisiae	6241	1858 (30)	[67]
Caenorhabditis elegans	18 424	8971 (49)	[67]
Drosophila melanogaster	13 601	5536 (41)	[67]
Arabidopsis thaliana	25 498	16 574 (65)	[69]
Homo sapiens	40 580[b]	15 343 (38)	[11]

[a]Use of different computational methods or criteria results in slightly different estimates of the number of duplicated genes [12].
[b]The most recent estimate is ~30 000 [61].

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in ecology & evolution*, *18*(6), 292-298.

A model with perfect inheritance of β_i between duplications

- Consider for illumination what happens in the extremum of the relationship between the distribution of fitness effects of a gene's creation and its subsequent duplications.
- Suppose they are identical, and inherited perfectly between gene duplications.

A model with perfect inheritance of β_i, cont'd

We have the following model:

$$\frac{\mathrm{d}n_i(t)}{\mathrm{d}t} = \Pr[i \text{ duplicates}] \Pr[i \text{ fixes}] \Pr[i \text{ maintained}]\ n_i(t)$$

where

$$\Pr[\text{i fixes}] = \int_0^\infty \phi(w) f_i(w) \mathrm{d}w,$$

w is the fitness of the organism with the duplicated gene,

$f_i(w)$ is the distribution of fitness effects from duplication of i,

$\phi(w)$ is the fixation probability as a function of fitness (dependent on population size, etc.)

A model with perfect inheritance of β_i, cont'd

- Let

$$\alpha := \Pr[i \text{ duplicates}],$$
$$\beta_i := \Pr[i \text{ fixes}] \Pr[i \text{ maintained}].$$

- The dynamics of genome expansion are then:

$$\frac{\mathrm{d}n_i(t)}{\mathrm{d}t} = \alpha\beta_i \; n_i(t)$$

hence

$$n_i(t) = e^{\alpha\beta_i t} n_i(0).$$

- This is exponential growth with parameter

$$\beta_i = \Pr[i \text{ fixes}] \Pr[i \text{ maintained}]$$

A model with perfect inheritance of β_i, cont'd

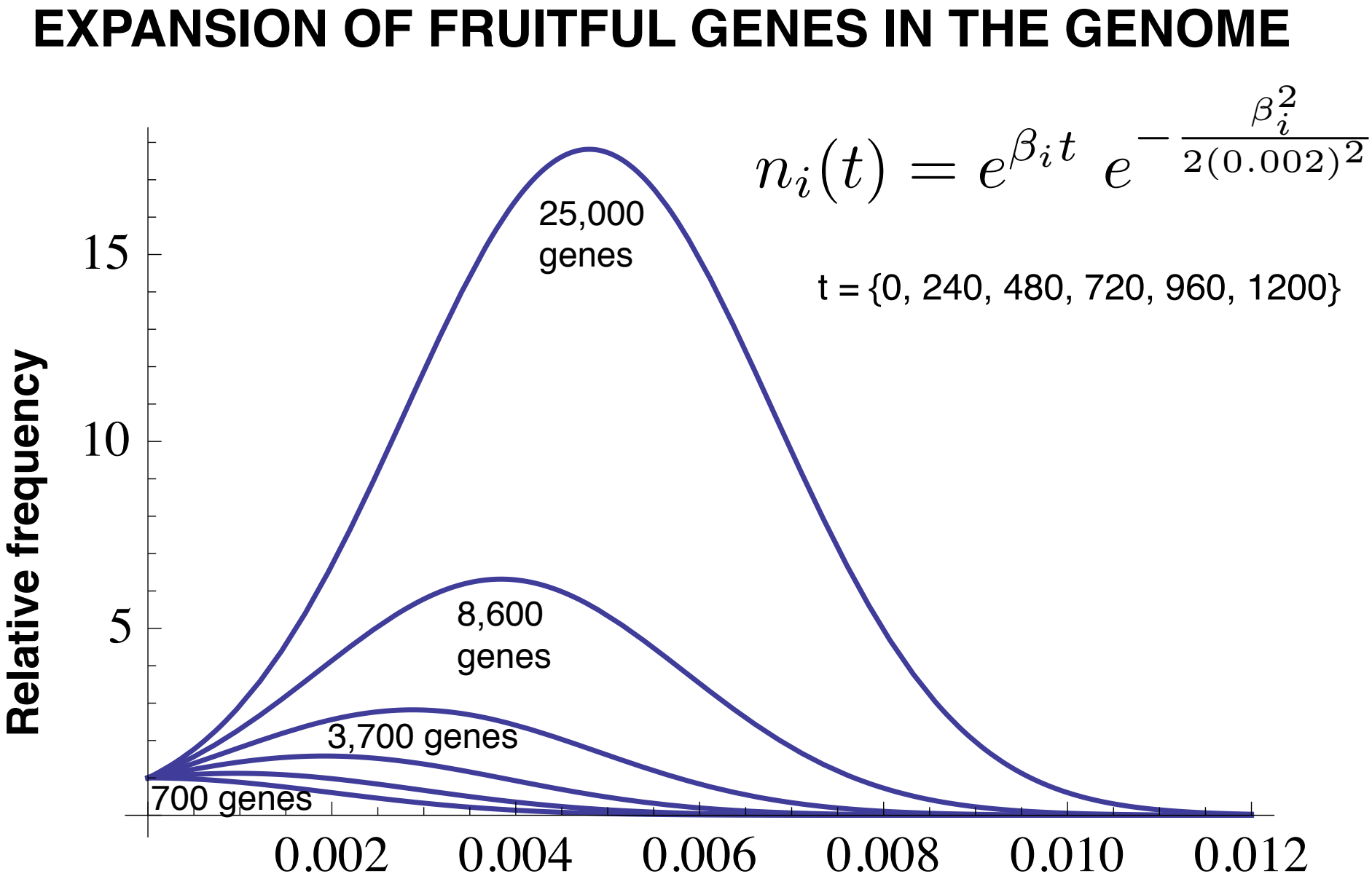
For concreteness, let $n_i(0) \sim N(0, \sigma)$:

$$n_i(0) = e^{-\frac{\beta_i^2}{2\sigma^2}}, \qquad \beta_i \in [0,1], \sigma << 1$$

Then

$$\begin{aligned} n_i(t) &= e^{\alpha\beta_i t}\ e^{-\frac{\beta_i^2}{2\sigma^2}} \\ &= e^{\beta_i(\alpha t - \frac{\beta_i}{2\sigma^2})}. \end{aligned}$$

A model with perfect inheritance of β_i, cont'd



A model with perfect inheritance of β_i, cont'd

Result (Fisher's Theorem applied to genome growth)

Assuming that β_i is perfectly transmitted between gene duplications, the fraction of new duplicated genes that go to fixation and are maintained, $\overline{\beta}(t)$,

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

increases at rate

$$\frac{d}{dt}\overline{\beta}(t) = \alpha \, \mathrm{Var}(\beta_i) \geq 0.$$

A model with perfect inheritance of β_i, concluded

Proof.

$$\frac{d}{dt}\overline{\beta}(t) = \sum_{i\in\mathcal{G}} \beta_i \frac{d}{dt}\left(\frac{n_i(t)}{N(t)}\right)$$

$$= \sum_{i\in\mathcal{G}} \beta_i [\frac{d}{dt} n_i(t) \,/\, N(t) - n_i(t)\frac{d}{dt} N(t) \,/\, N(t)^2]$$

$$= \frac{\alpha}{N(t)} \sum_{i\in\mathcal{G}} \beta_i^2 \, n_i(t) - \frac{\alpha}{N(t)^2} \left(\sum_{i\in\mathcal{G}} \beta_i \, n_i(t)\right)^2$$

$$= \alpha \left[\sum_{i\in\mathcal{G}} \beta_i^2 \frac{n_i(t)}{N(t)} - \overline{\beta}(t)^2\right] = \alpha \, \mathrm{Var}(x) \geq 0. \qquad \square$$

Less-than-perfect heritability of β_i

Define a transmission function, $T(i \leftarrow j)$, which is the probability that a gene of type j gives rise to a copy of type i. It satisfies conditions

$$\sum_{i \in \mathcal{G}} T(i \leftarrow j) = 1 \text{ for all } j \in \mathcal{G}, \text{ and } T(i \leftarrow j) \geq 0 \text{ for all } i, j \in \mathcal{G}.$$

Here, the fraction of the new genes that are of type i is

$$p_i(t) = \sum_{i,j \in \mathcal{G}} T(i \leftarrow j)\, n_j(t) \,/\, N(t).$$

Less-than-perfect heritability of β_i, cont'd

- The genome expansion dynamics now become:

$$\frac{d}{dt} n_i(t) = \alpha \, \beta_i \sum_{j \in \mathcal{G}} T(i \leftarrow j) \, n_j(t).$$

- Price's Covariance and Selection equation (Price, 1970, 1972) emerges when we consider selection in the presence of arbitrary transmission:

Less-than-perfect heritability of β_i, cont'd

Result (Price's Equation in genome expansion)

For a gene of type j, let

- β_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:* $\xi_j = \sum_{i \in \mathcal{G}} \beta_i \, T(i \leftarrow j)$.

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

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

where

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

Selection FOR and OF evolvability

Therefore:

1. selection for individual fitness gives
2. selection of the upper tail of the fitness distribution,
3. when there is covariance between them.

A Novel Darwinian Process: Genome-as-Population

- We see that the requirements for Darwinian evolution:
  1. Heritable
  2. Variation in
  3. Fitness (viability and fecundity)
- actually emerges in the genome as population of genes.

Viability in the 'genome-as-population':

- The viability of a genetic sequence is simply its survival in the genome.
- This will depend on whether selection:
  1. establishes it in the population (fixation or stable polymorphism), and
  2. maintains it against mutational degradation or replacement by other genes.

Viability, cont'd:

- This in turn depends on:
  1. there being adaptive opportunity for properties of the sequence;
  2. the sequence having functional properties which are not disrupted by new functional contexts; and
  3. the sequence having properties that allow its duplication without disrupting existing functions of genes with which it interacts.

Fecundity in the 'genome-as-population':

- The fecundity of a genetic sequence is the rate at which copies of it appear in the genome.
- This depends on:
  1. the rate of 'illegitimate' recombination events involving that sequence; and
  2. whether the sequence codes for its own duplicative transposition.

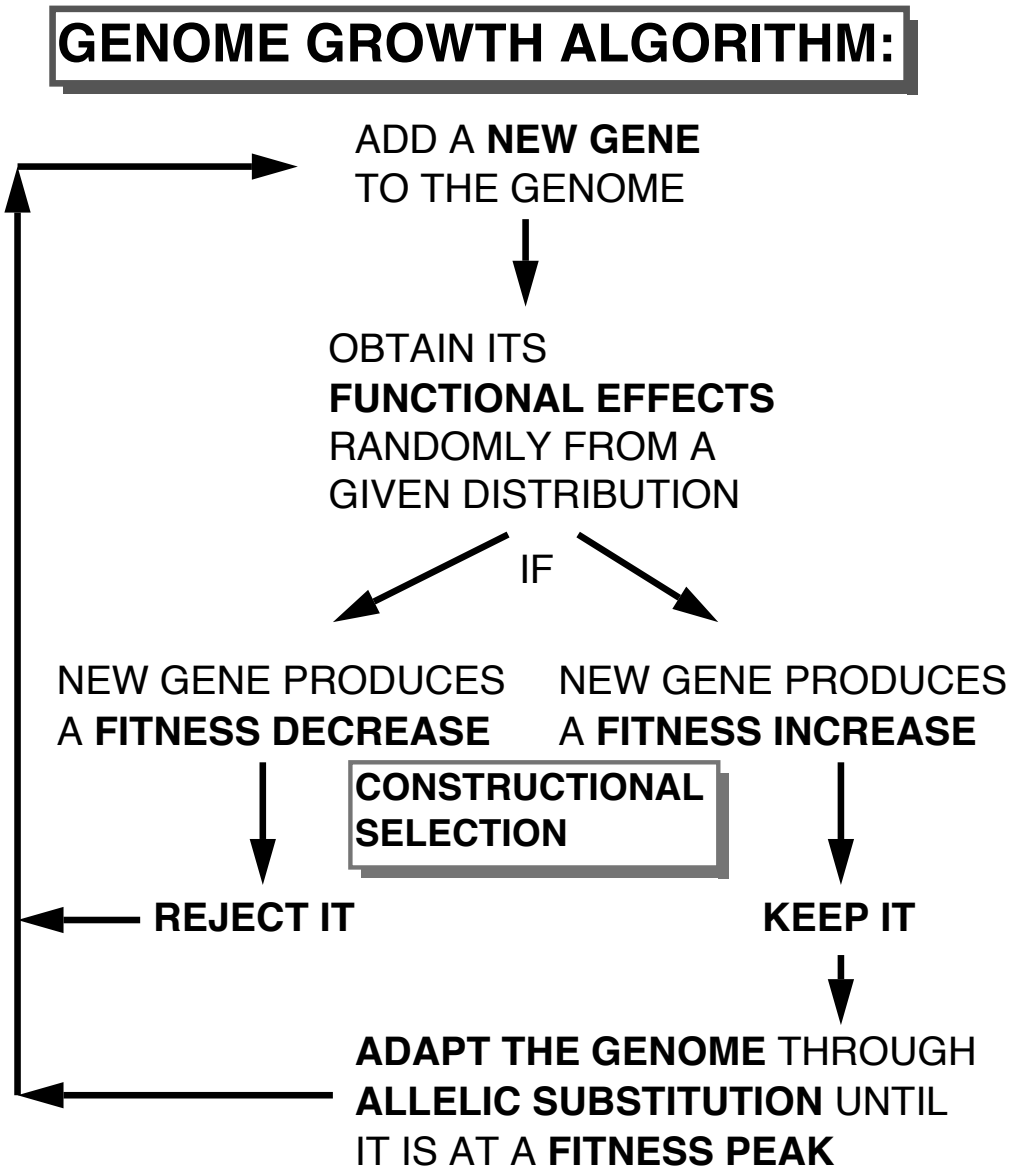
Heritability in the 'genome-as-population':

- Heritability in the genome-as-population refers to ancestral and offspring genes having correlated properties.
- This depends on:
  1. Conservation of the property of a gene over the time scale on which gene duplications occur; and
  2. Carry-over of the property from ancestral to offspring genes.

Simplified Model of Genome Expansion

- Strong selection.
- Rare allelic mutation.
- Rarer gene duplication.
- House-of-cards sampling of phenotypic effects of new genes.

So selective sweeps and allelic evolution occur in instants between time increments.



What changes when new genetic material is added to the genome? In addition to any effects on fitness, there are two changes:

- There are new degrees of freedom for genetic variation
- The probability distribution of point mutations is changed
- The probability distribution of gene duplications is altered.

Ohno (1970, p. 59)

- *"Only by the accumulation of forbidden mutations at the active sites can the gene locus change its basic character and become a new gene locus. An escape from the ruthless pressure of natural selection is provided by the mechanism of gene duplication.*
- *By duplication, a redundant copy of a locus is created.*
- *Natural selection often ignores such a redundant copy, and, while being ignored, it accumulates formerly forbidden mutations and is reborn as a new gene locus with a hitherto non-existent function."*

Q. Are duplicate genes necessarily redundant and ignored by natural selection, or could this be an evolved feature of genes in the genome?

HUMAN DISEASES CAUSED BY GENE DUPLICATION:

Table 1 Duplication phenotypes

Species	Gene	Gene category	Genomic alteration	Disease	Phenotype	Mechanism	References
Similar under- and overexpression phenotypes							
Hs/Mm	*MECP2*	Methylated DNA binding	Duplication *MECP2*	Progressive neurodevelopmental syndrome in males	Mental retardation, epilepsy	Loss of gene function due to under- and overexpression	(52, 98)
Hs	*SOX3*	Developmental regulator (TF)	Duplication *SOX3*	X-linked hypopituitarism (XLHP)	X-linked hypopituitarism and infundibular hypoplasia	Loss of gene function due to under- and overexpression	(104)
Hs	ND (*TBX1*)	ND	Duplication 22q11.2	Velocardiofacial syndrome (VCFS)	Variable: normal to developmental delay and malformations	Loss of gene function due to under- and overexpression	(106)
Hs	ND (*ELN*)	ND	Duplication 7q11.23	Williams Beuren syndrome (WBS)	Delay expressive language	Loss of gene function due to under- and overexpression	(92)
Hs	ND (*JAGGED1*)	Developmental regulator (TF)	Duplication 20p11	Alagille syndrome (AS)	Cardiovascular-, ocular-, bile duct-, and skeletal anomalies	Loss of gene function due to under- and overexpression	(63)
Hs	*RAI1*	Developmental regulator (TF)	Duplication 17p11.2	Smith-Magenis syndrome (SMS)	Mild mental retardation and dental abnormalities	Loss of gene function due to under- and overexpression	(77)
Hs	*PLP1*	Proteolipid protein	Duplication *PLP1*	Pelizaeus-Merzbacher (PM)	Demyelination disorder CNS	Loss of gene function due to under- and overexpression	(26)
Dissimilar under- and overexpression phenotypes							
Hs	*PMP22*	Myelin protein	Duplication *PMP22*	Charcot Marie Tooth 1A (CMT1A)	Peripheral myelin neuropathy	Loss of gene function due to under- and overexpression	(54, 64)
Hs	*NSD1*	Histone methyltransferase	Duplication *NSD1*	Growth retardation syndrome	Growth retardation	Loss of gene function due to under- and overexpression	(63)
Hs	ND (*MMP23*)	ND	Duplication 1p36	Premature closure cranial sutures	Craniosynostosis	Incremental gene function	(31)
Complex, yet unresolved expression phenotypes							
Hs	*LMB1*	Laminar nuclear envelope protein	Duplication *LMB1*	Autosomal dominant leukodystrophy (ADLD)	Demyelination disorder CNS	ND	(73)
Hs	ND	ND	Duplication 10q24	Split hand/split foot malformation 3 (SHFM3)	Split hand/split foot	ND	(21)
Hs	ND	ND	Duplication 2q13	Orofacial clefting/cleft palate only	Mental retardation and orofacial clefting	ND	(72)

Table 1 (*Continued*)

Species	Gene	Gene category	Genomic alteration	Disease	Phenotype	Mechanism	References
Hs	ND	ND	Duplication 16p13	ATR-X-like	X-linked α-thalassemia/mental retardation	ND	(2)
Protein aggregation due to overexpression							
Hs	*SNCA*	Molecular chaperone	Duplication *SNCA*	Parkinson disease	Nigrostriatal neuron degeneration	Protein aggregation and underlying gene mutations	(91)
Hs	*APP*	Amyloid precursor protein	Duplication *APP*	Alzheimer disease	Parenchymal/vascular amyloid deposition	Protein aggregation and underlying gene mutations	(37, 79)
Phenotypes of CNVs related to environment and immunity							
Hs	*CYP2D6*	P-450 isoenzyme	*P-450* CNV	Altered drug metabolism	Adverse drug effects	Incremental/linear gene function model	(12, 40, 75)
Hs	*CCL3L1*	Chemokine receptor	*CCL3L1* CNV	Altered HIV susceptibility	Enhanced HIV/AIDS susceptibility	Incremental/linear gene function model	(33)
Common complex phenotypes of CNVs in defense-related genes							
Hs/Rn	*FCGR3B/ Fcgr3-rs*	Fc receptor for IgG	*FCGR3* CNV	Glomerulonephritis/ systemic lupus erymathosus	Susceptibility to glomerulonephritis	Pathogenic underexpression phenotype	(1)
Mm	*TLR7*	Toll-like receptor	*TLR7* CNV	Systemic lupus erymathosus-like disease	Autoantibody-elicted autoimmunity	Pathogenic overexpression phenotype	(76)
Hs	*bBD2*	Antimicrobial peptides	*bBD2* CNV	Crohn's disease of the colon	Inflammatory bowel disease	Pathogenic underexpression phenotype	(28)

Conrad, B., & Antonarakis, S. E. (2007). Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu. Rev. Genomics Hum. Genet.*, *8*, 17-35.

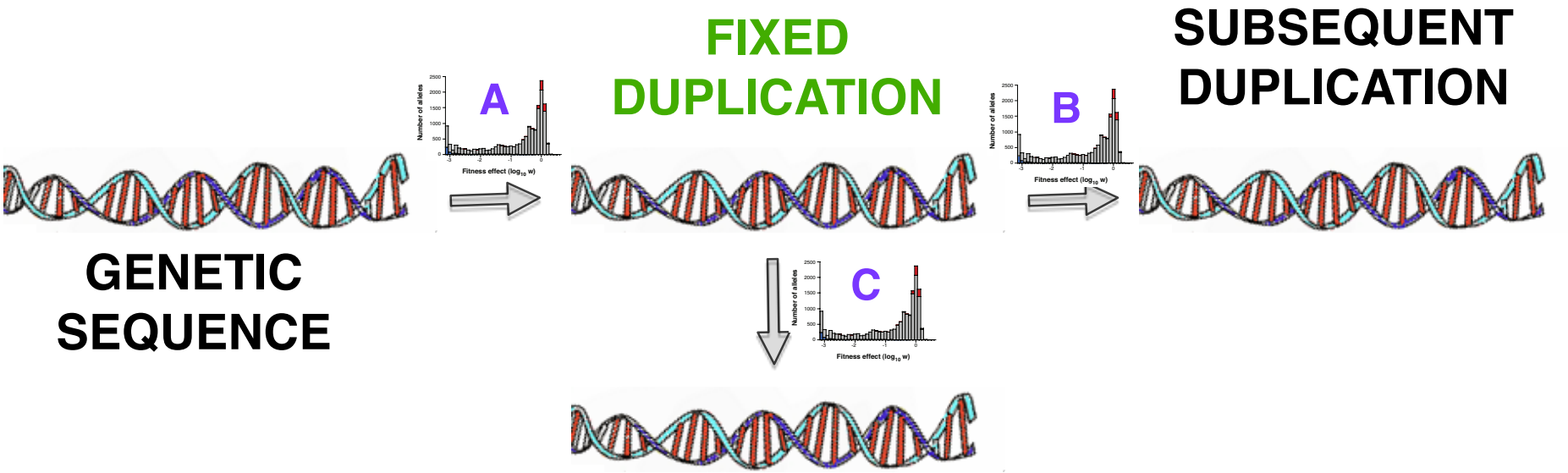
Heritability of the Distribution of Fitness Effects

Prediction

If the distribution of fitness effects of duplication is generally heritable between duplications, then genes that produce deleterious duplications (e.g. diseases of gene duplication) ought to have either

1. *higher than average rates of duplication; or*
2. *fewer and older paralogs than average over the genome.*

A ↔ C Relationship



The Correlated Allelic Variation Effect

- How selection on the creation of new genes can cause subsequent allelic variation of the genes to be more likely to be adaptive.
- Look at:
  1. the fitness distributions of alleles from all new genes, and
  2. from only those genes that selection stably incorporates into the genome.

The Correlated Allelic Variation Effect

The relationship between gene origin and subsequent allelic variation

- Suppose that a newly created gene of type i gives rise to allelic variants.
- Let the allelic fitnesses, w', be distributed with probability density $f_i(w')$.
- No assumptions need to be made about this density, so it would certainly include the biologically plausible case in which most of the alleles are deleterious.
- This effect applies to *de novo* generated new genes (Joanna Masel's talk) as well as duplicates.

The Correlated Allelic Variation Effect, cont'd

For a gene or type i, we see that the proportion

$$F_i(w) = \int_w^\infty f_i(y)\,\mathrm{d}y,$$

of its alleles are fitter than w.

The Correlated Allelic Variation Effect, cont'd

Result (Correlated allelic variation)

Let

$\overline{F}(w)$ *be the proportion of new alleles of randomly created genes that are fitter than* w*, and*

$F^*(w)$ *be the proportion of new alleles of stably incorporated genes that are fitter than* w*.*

Then

$$F^*(w) = \overline{F}(w) + \mathrm{Cov}[F_i(w),\ \beta_i/\overline{\beta}]. \tag{1}$$

Proof.

Let $p_i(t) = n_i(t)/N(t)$ be the proportion of genes of type i in the genome. The proportion of alleles that are fitter than w, among randomly created gene, is

$$\overline{F}(w) = \sum_{i \in \mathcal{G}} F_i(w)\, p_i,$$

while among genes stably incorporated in the genome, it is

$$\begin{aligned} F^*(w) &= \Pr[w' > w \mid \text{gene was incorporated}] \\ &= \frac{\Pr[w' > w \ \& \ \text{gene was incorporated}]}{\Pr[\text{gene was incorporated}]} \\ &= \sum_{i \in \mathcal{G}} F_i(w)\, \beta_i\, p_i \,/\, \overline{\beta} = \overline{F}(w) + \mathrm{Cov}[F(w_i),\ \beta_i/\overline{\beta}]. \square \end{aligned}$$

If there is a positive correlation between the fixation probability

$$\beta_i = \int_0^\infty \phi(w)\, g_i(w)\, \mathrm{d}w$$

of a new gene, and the fitness distribution

$$F_i(w) = \int_w^\infty f_i(y)\, \mathrm{d}y$$

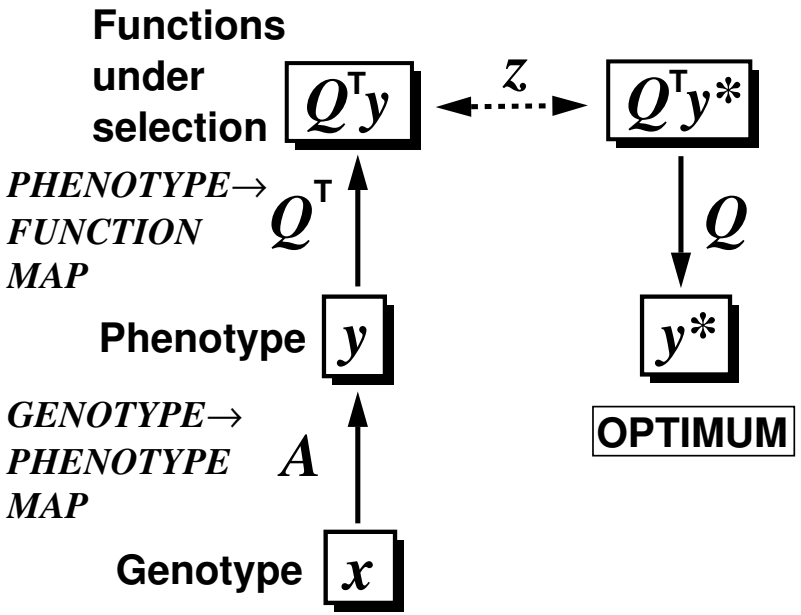
of its alleles, then $F^*(w)$ is greater than $\overline{F}(w)$. Similarity between the functions $g_i(w)$ and $f_i(w)$ would produce a positive covariance.

The biological foundation for a positive covariance would include:

1. there continuing to be adaptive opportunity for variation in the phenotype controlled by the gene, and
2. the same suite of phenotypic characters being affected by the alleles of the gene as were affected during the gene's origin.

With these plausible and general provisions, we see how selection on new genes can also select on the fitness distributions of the alleles that these genes generate.

B-Matrix (Wagner, 1989) Model with Genome Growth



$$w(\mathbf{x}) = \mathrm{Exp}[-(\mathbf{Ax} - \mathbf{y}^*)^\top \mathbf{Q\Lambda Q}^\top (\mathbf{Ax} - \mathbf{y}^*)]$$

$\mathbf{x}$ is the genotype vector

$\mathbf{A}$ maps the genotype to the phenotype

$\mathbf{y}^*$ is the optimal phenotype vector

$\mathbf{Q}$ maps $(\mathbf{Ax} - \mathbf{y}^*)$ to fitness components $\mathbf{\Lambda}$.

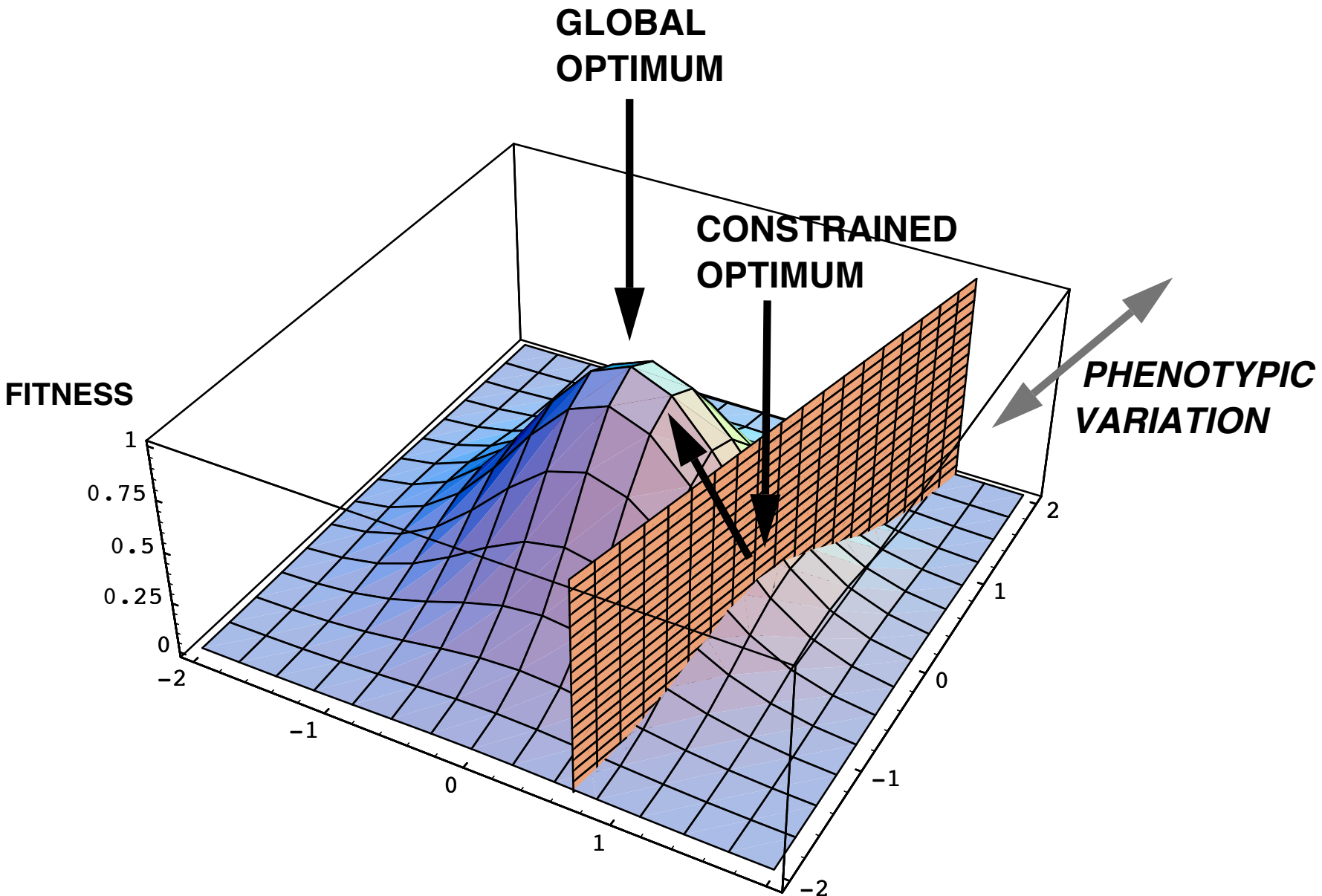
B-Matrix (Wagner, 1989) Model with Genome Growth

- Now, add a new column, $\mathbf{a}_{n+1}$, to $\mathbf{A}$, to yield $\mathbf{A}'$,
- along with a new allelic value x_{n+1} to $\mathbf{x}$, to yield $\mathbf{x}'$.
- $\mathbf{a}_{n+1}$ and x_{n+1} are sampled from constant distributions.
- The new fitness with the added gene is:

$$
\begin{aligned}
& w(\mathbf{x}') \\
& \quad = \mathrm{Exp}[-(\mathbf{A}'\mathbf{x}' - \mathbf{y}^*)^\top \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top(\mathbf{A}'\mathbf{x}' - \mathbf{y}^*)] \\
& \quad = \mathrm{Exp}[-(\mathbf{A}\hat{\mathbf{x}} + x_{n+1}\mathbf{a}_{n+1} - \mathbf{y}^*)^\top \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top(\mathbf{A}\hat{\mathbf{x}} + x_{n+1}\mathbf{a}_{n+1} - \mathbf{y}^*)].
\end{aligned}
$$

- Gene $n+1$ with vector $\mathbf{a}_{n+1}$ goes to fixation if $w(\mathbf{x}') > w(\mathbf{x})$.
- If $w(\mathbf{x}') \leq w(\mathbf{x})$, $\mathbf{a}_{n+1}$ goes extinct, and a new gene $n+1$ is sampled.

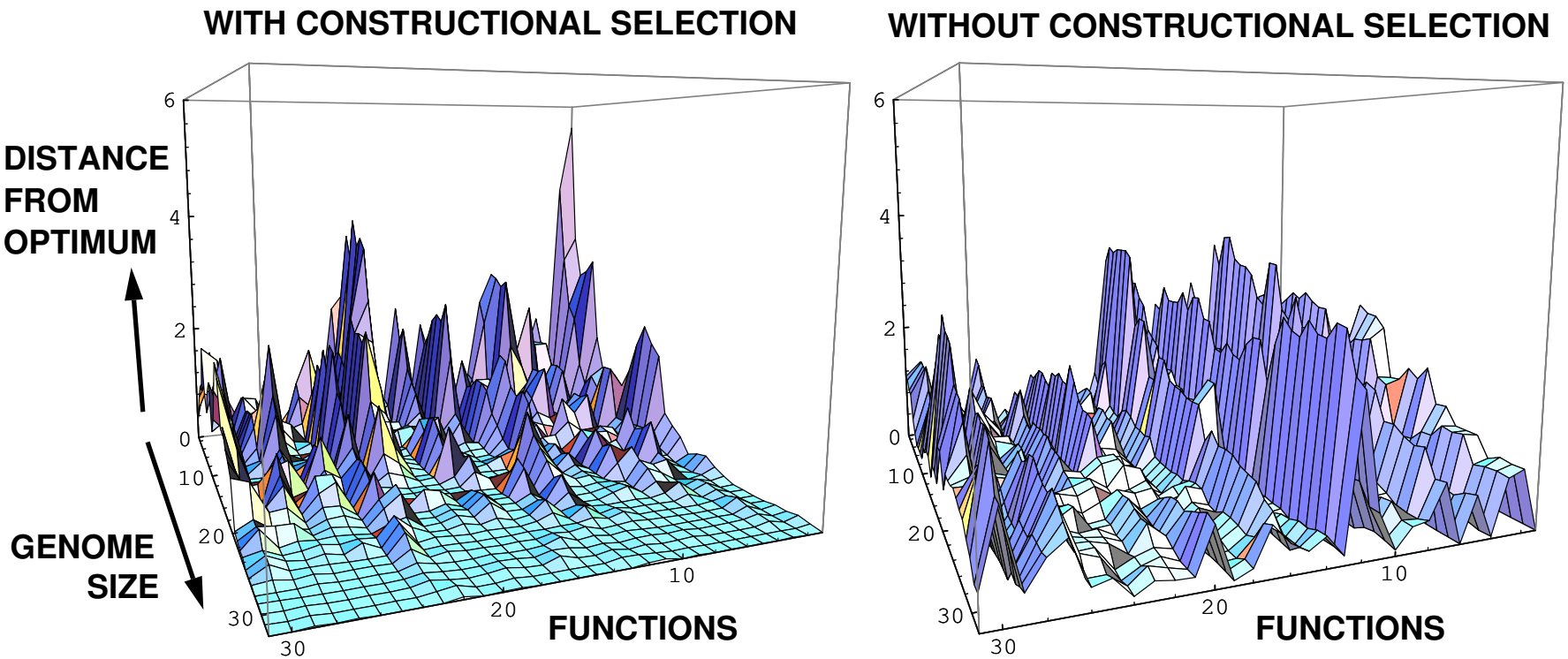
B-Matrix (Wagner, 1989) Model with Genome Growth



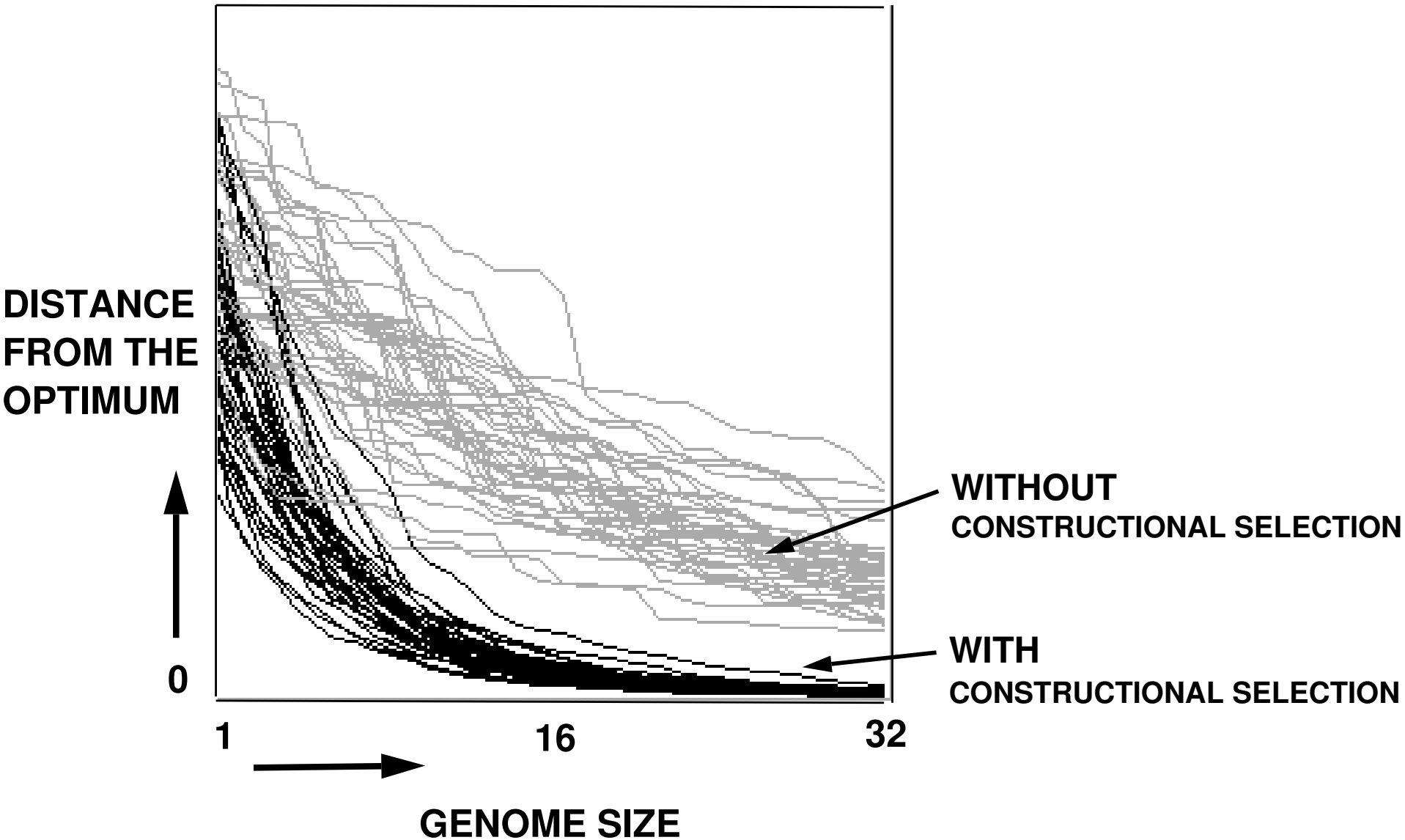
B-Matrix (Wagner, 1989) Model with Genome Growth

Plotted are fitness components $\lambda_i z_i^2$, where $\mathbf{q}_i := [\mathbf{Q}]_i$,

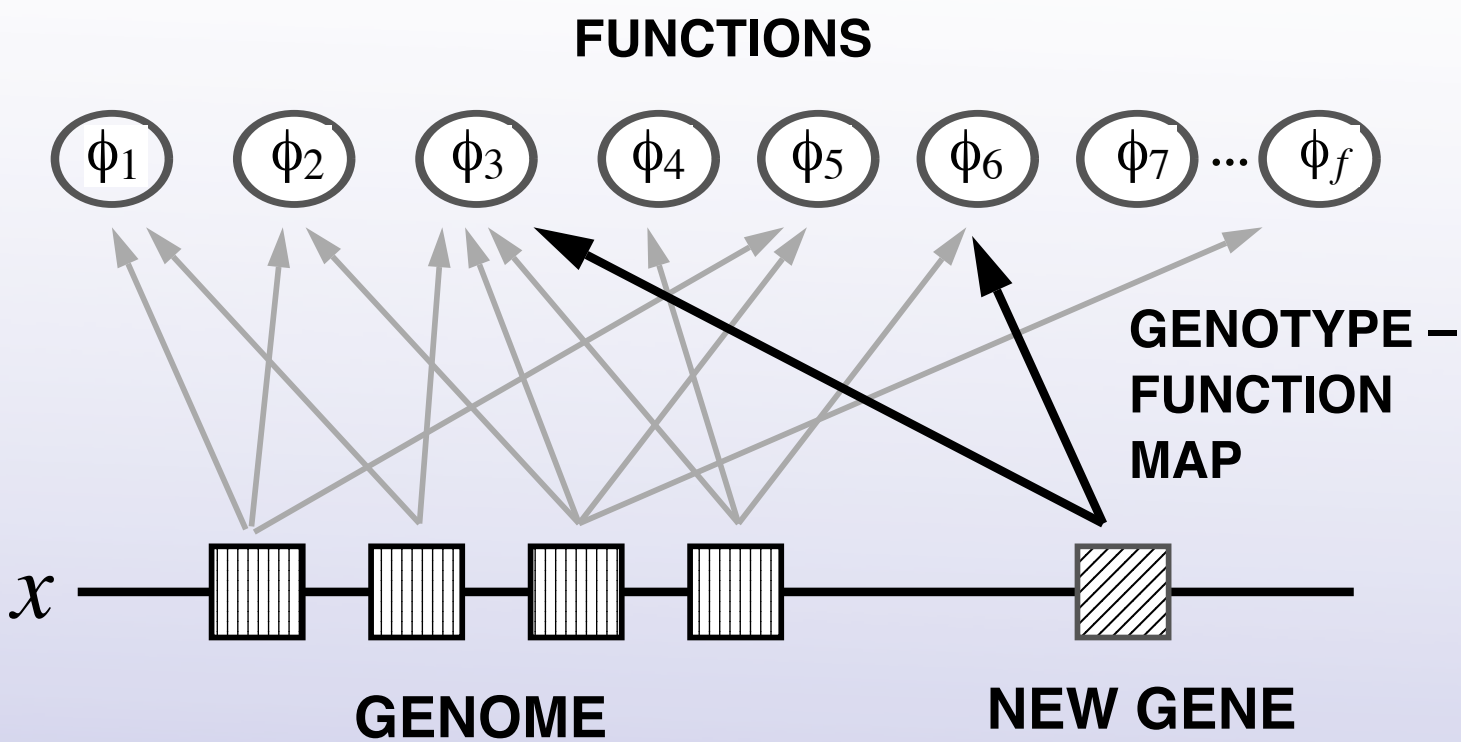
$$z_i = \mathbf{q}_i^\top (\mathbf{A}\mathbf{x} - \mathbf{y}^*) \qquad \text{and} \quad w(\mathbf{x}) = \sum_i \lambda_i z_i^2.$$



B-Matrix (Wagner, 1989) Model with Genome Growth

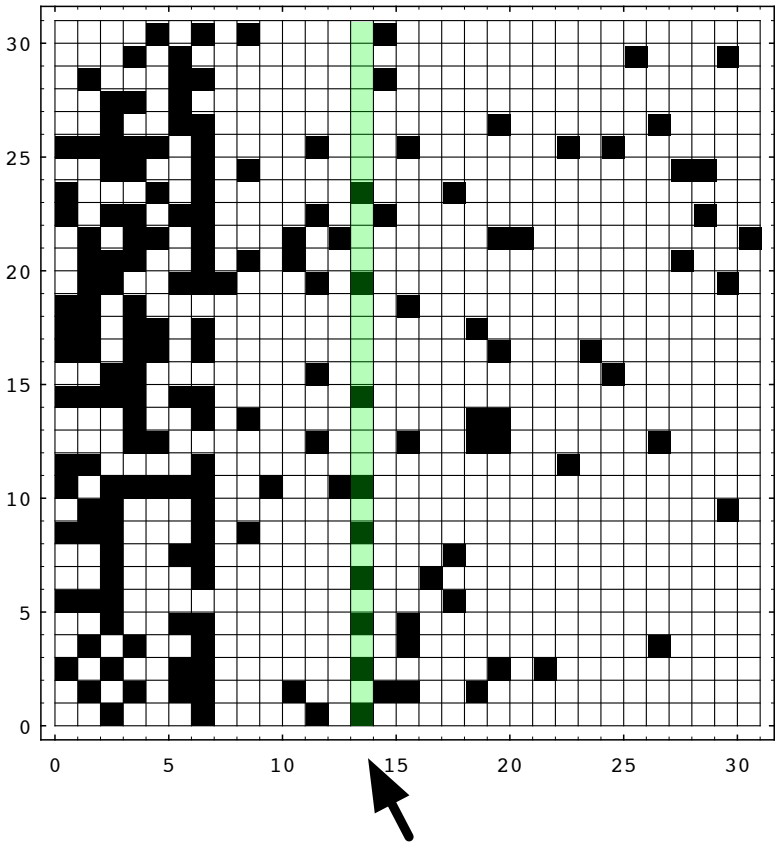


NK (Kauffman and Levin, 1987) Model with Genome Growth

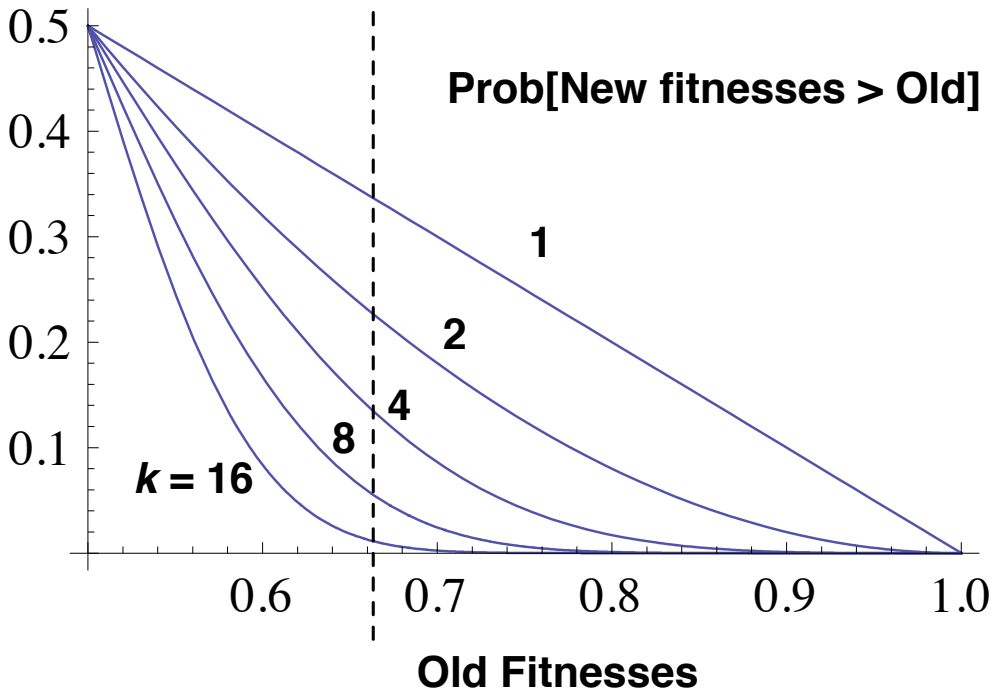


$$w(x) = \frac{1}{f}\sum_{i=1}^{f}\phi_i(x)$$

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

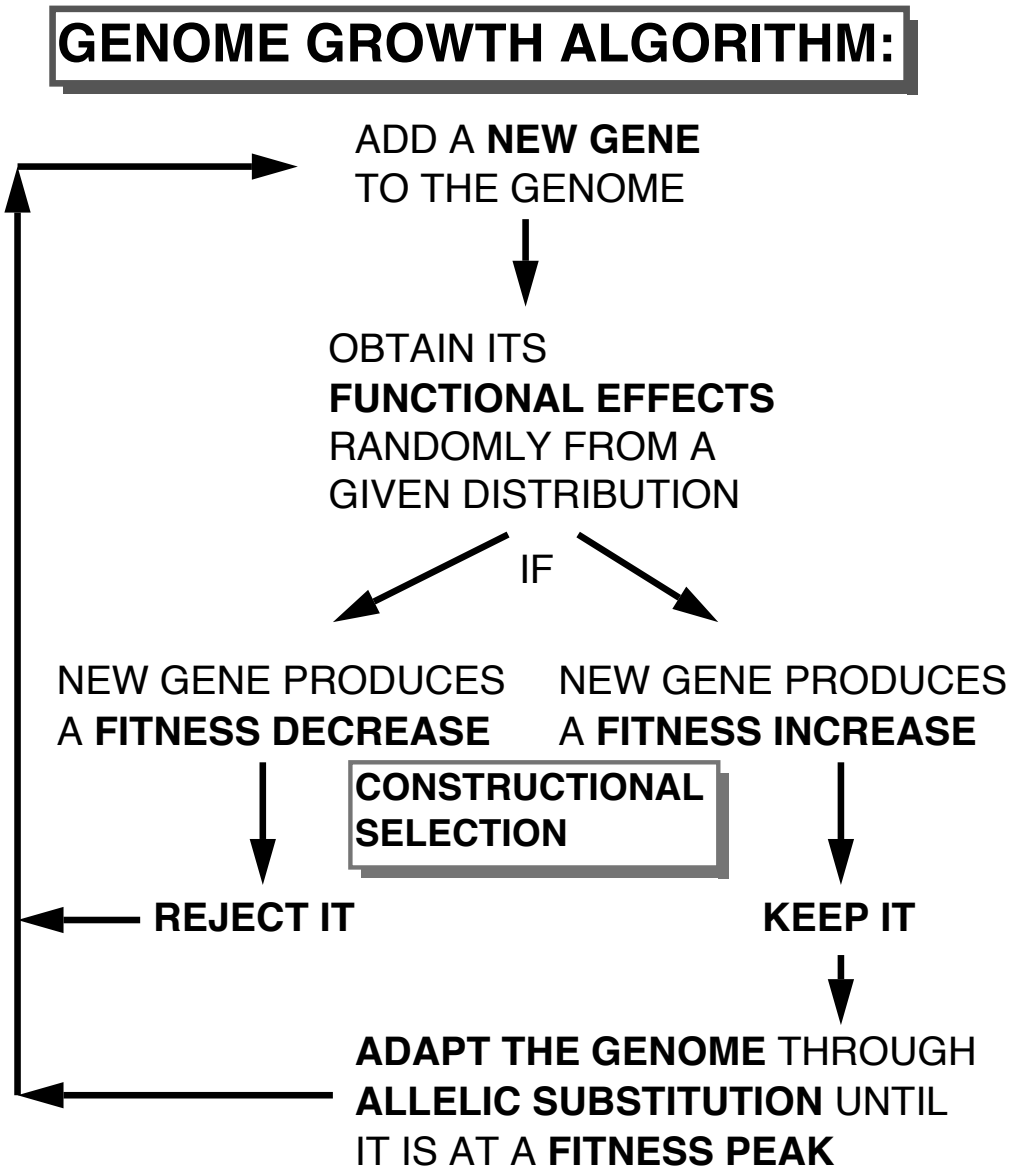


NEW GENE ALTERS *k* FITNESS COMPONENTS

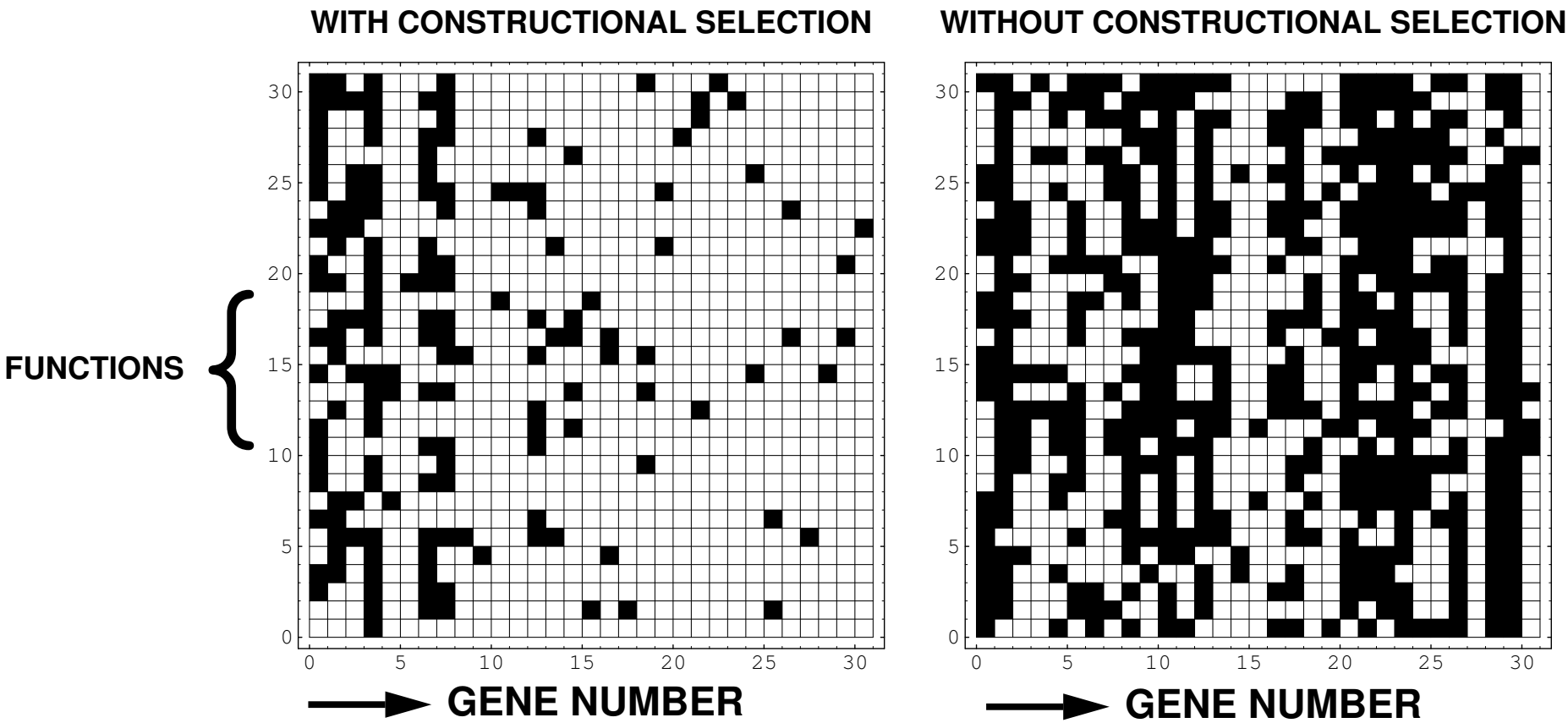


Simplified Model of Genome Expansion, reprise

- Strong selection
- Rare allelic mutation
- Rarer gene duplication
- Hence: Selective sweeps and allelic evolution occur between time increments

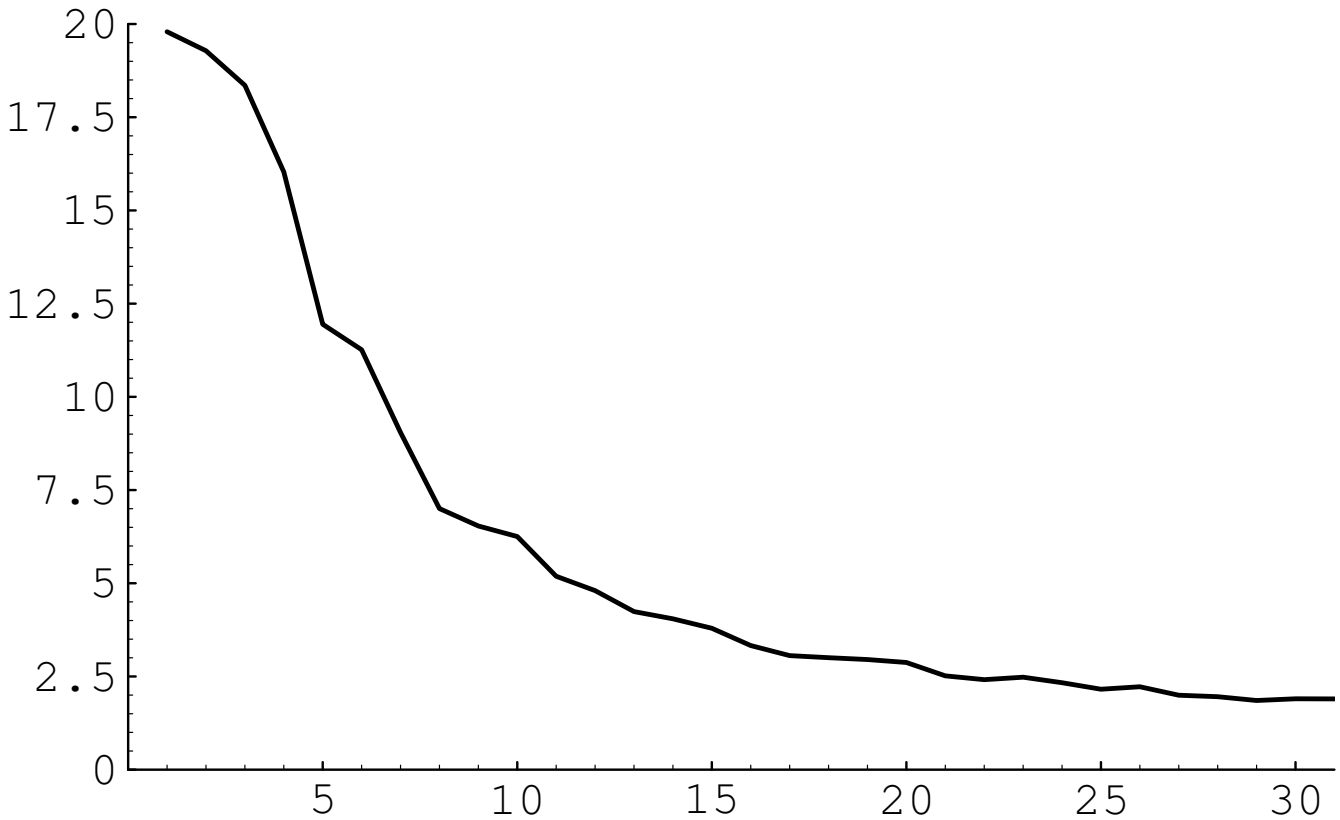


NK (Kauffman and Levin, 1987) Model with Genome Growth



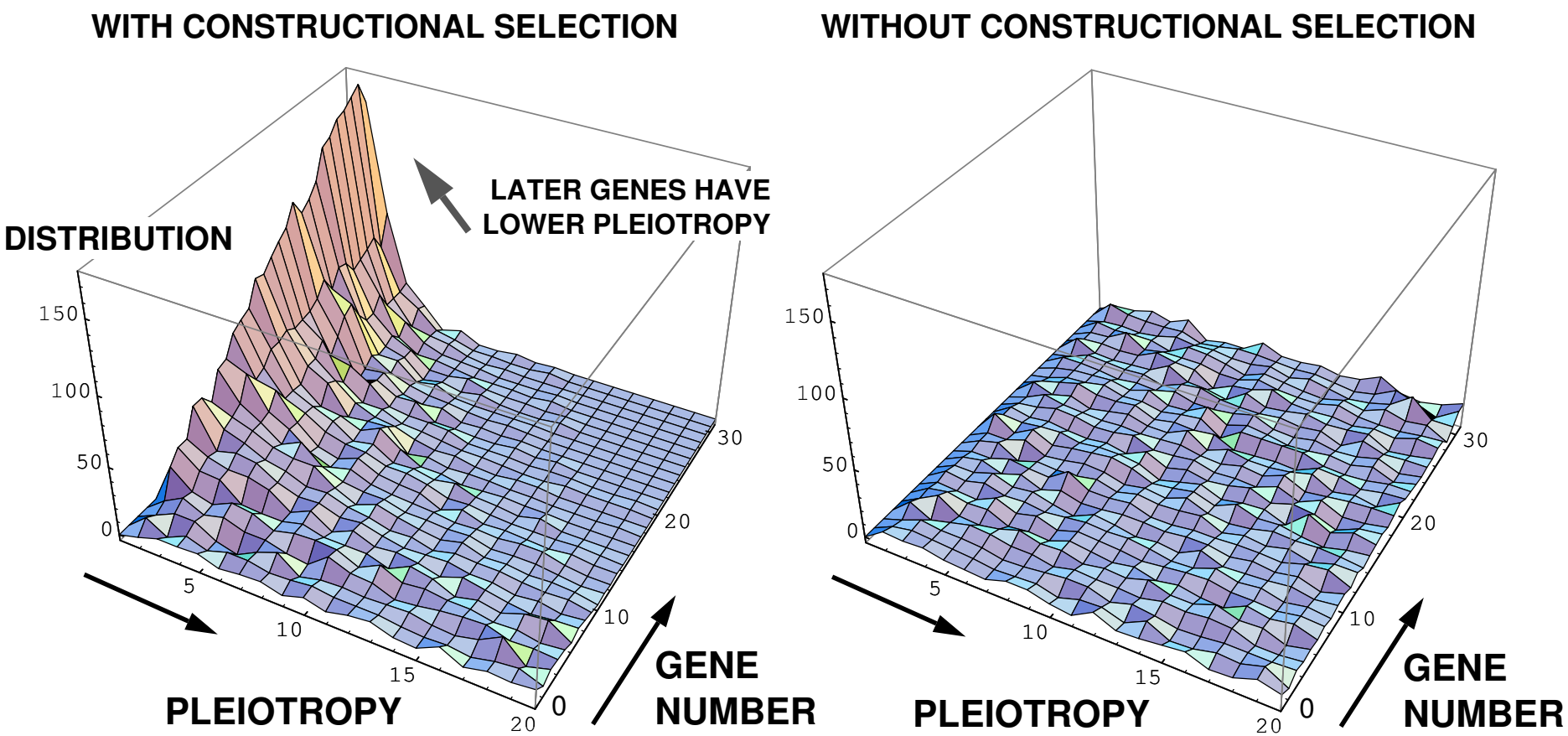
NK (Kauffman and Levin, 1987) Model with Genome Growth

AVERAGE PLEIOTROPY, k_n

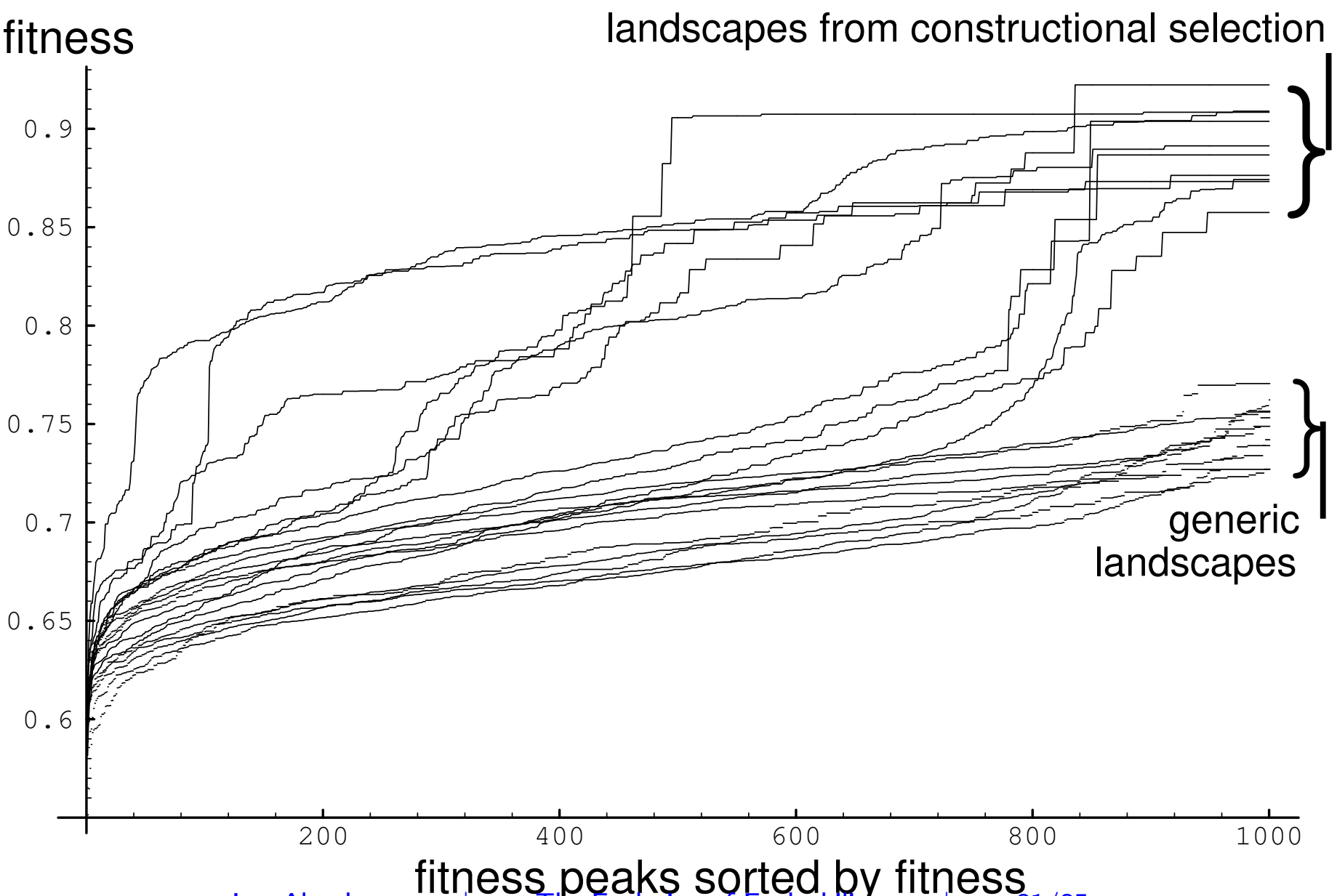


GENE NUMBER, n

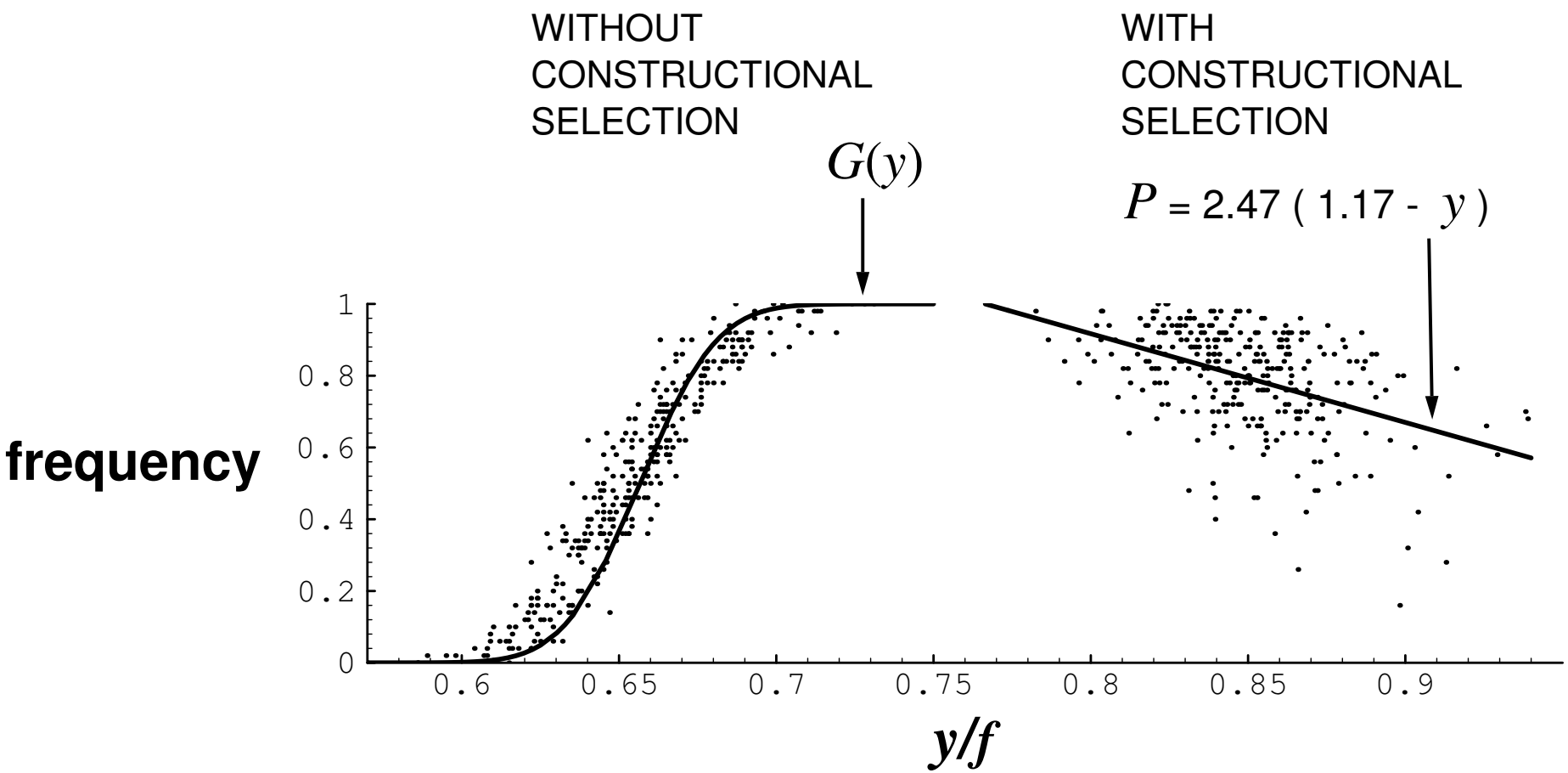
NK (Kauffman and Levin, 1987) Model with Genome Growth



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



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



Conclusion

- So, we've seen some of the early history and continuing controversy about the evolution of evolvability.
- I've described a general mechanism by which evolvability can evolve through individual selection.
- The mechanisms occurs through the relationships between distributions of fitness effects of
  1. gene origin
  2. subsequent allelic variation, and
  3. subsequent gene duplication.

Conclusion, cont'd

- This mechanism is illustrated with
  1. A simple but general mathematical model with a Price equation
  2. Wagner's B-Matrix model of real vector genotype/phenotype maps
  3. Kauffman's NK landscape model of discrete genotype/phenotype maps.

Conclusion, cont'd

I have attempted to show how gene origins can shape the distribution of fitness effects of both gene duplication and allelic variation.

> *"Why such selection 'is a means by which selection can come to act indirectly on evolutionary potential' is beyond my understanding."* – Reviewer for *Evolution*

- I hope my presentation has been clearer to you that it was for that reviewer.
- Details can be found in A. (1995), Genome growth and the evolution of the genotype-phenotype map. In *Evolution and Biocomputation: Computational Models of Evolution*, *LNCS* 899: 205–259.

Conclusion, concluded

- If these ideas have piqued your interest, I welcome theoretical and empirical collaborations.

Thank you for your attention!

A. 1993. Constructional selection and the evolution of the genotype-phenotype map. Technical report, Institute for Statistics and Decision Sciences, Duke University.

A. 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.

Alberch, P. 1991. From genes to phenotype: dynamical systems and evolvability. *Genetica*, 84:5–11.

Conrad, M. 1977. Evolutionary adaptability of biological macromolecules. *Journal of molecular evolution*, 10(1):87–91.

Conrad, M. 1979. Bootstrapping on the adaptive landscape. *BioSystems*, 11:167–182.

Conrad, M. 1990. The geometry of evolution. *BioSystems*, 24(1):61–81.

Conrad, M. and Volkenstein, M. 1981. Replaceability of amino acids and the self-facilitation of evolution. *Journal of Theoretical Biology*, 92(3):293–299.

Dawkins, R. 1988. The evolution of evolvability. In Langton, C. G., editor, *Artificial Life, The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, volume VI of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 201–220, Redwood City, CA. Addison-Wesley.

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

Graves, C. J., Ros, V. I., Stevenson, B., Sniegowski, P. D., and Brisson, D. 2013. Natural selection promotes antigenic evolvability. *PLoS Pathogens*, 9(11):e1003766.

Kauffman, S. A. 1990. Requirements for evolvability in complex systems: Orderly dynamics and frozen components. *Physica D*, 42:135–152.

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.

Lynch, M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences U.S.A.*, 104(Suppl 1):8597–8604.

Ohno, S. 1970. *Evolution by Gene Duplication*. Springer Verlag, Berlin.

Price, G. R. 1970. Selection and covariance. *Nature*, 227:520–521.

Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics*, 35:485–489.

Riedl, R. J. 1975. *Die Ordnung des Lebendigen: Systembedingungen der Evolution*. Parey, Hamburg and Berlin.

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

Sniegowski, P. D. and Murphy, H. A. 2006. Evolvability. *Current Biology*, 16(19):R831–R834.

Wagner, G. P. 1989. Multivariate mutation-selection balance with constrained pleiotropic effects. *Genetics*, 122:223–234.

Zhong, W. and Priest, N. K. 2011. Stress-induced recombination and the mechanism of evolvability. *Behavioral ecology and sociobiology*, 65(3):493–502.