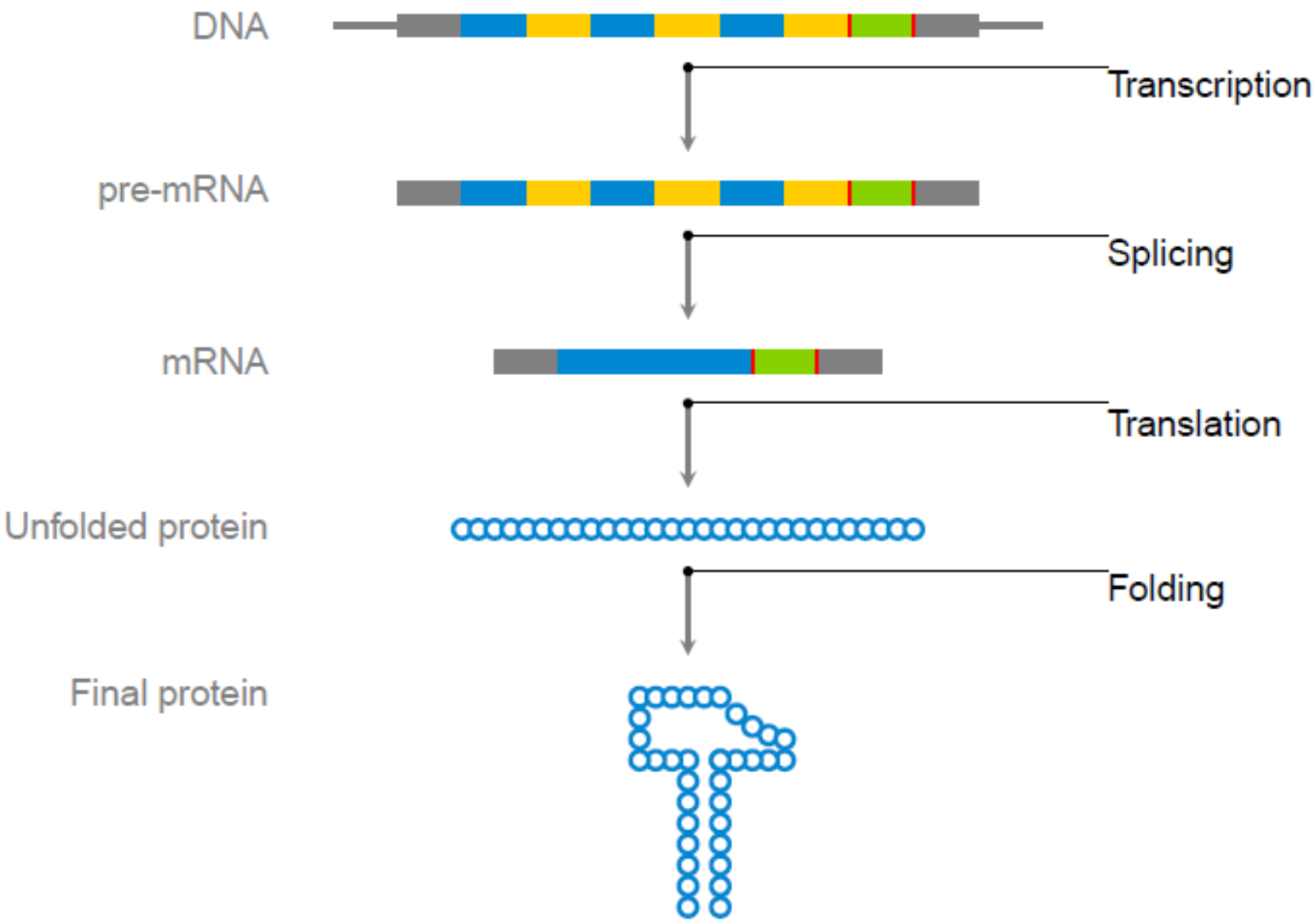
Molecular errors, cryptic sequences, and evolvability

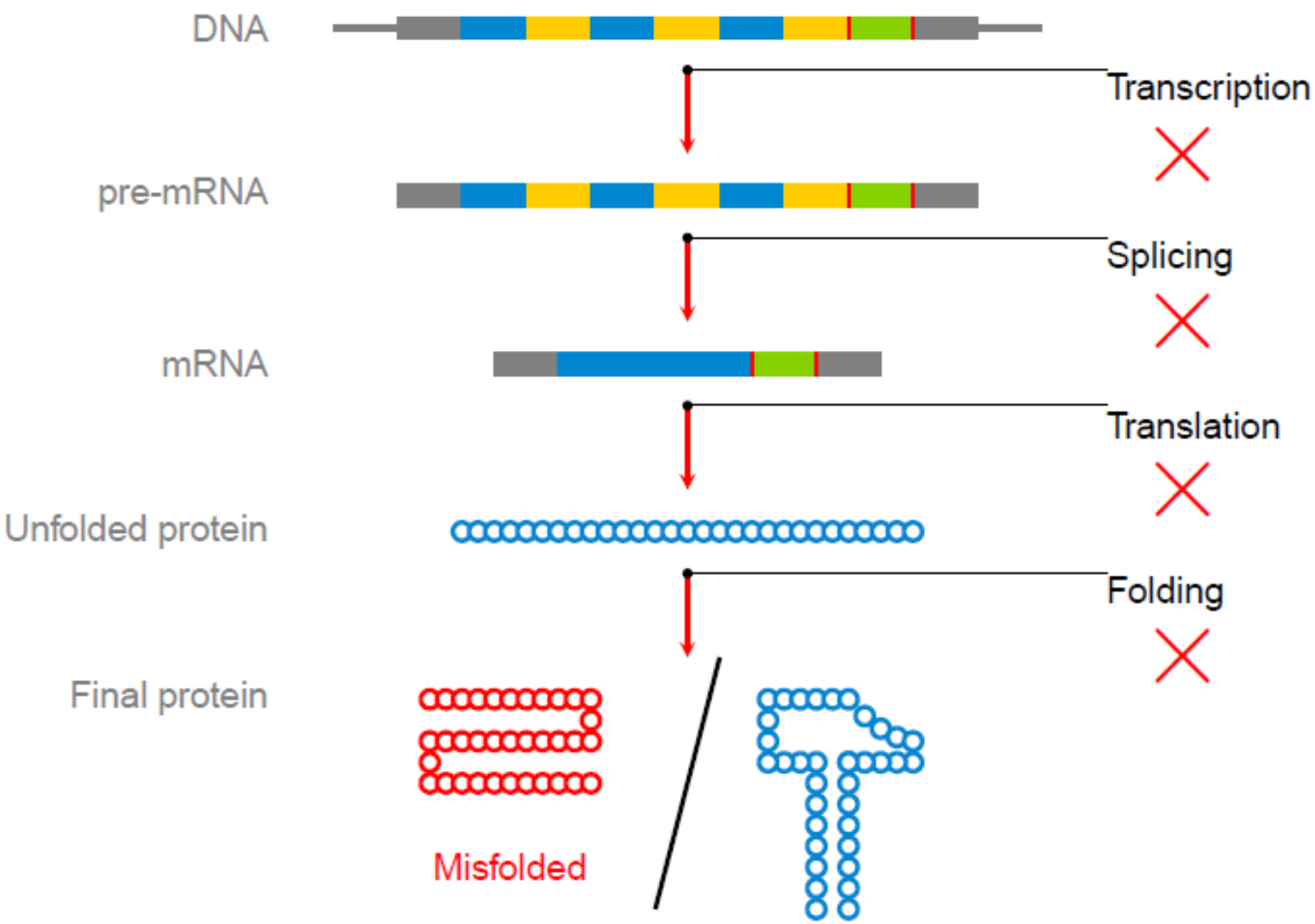
Joanna Masel

Ecology & Evolutionary Biology, University of Arizona

Gene expression



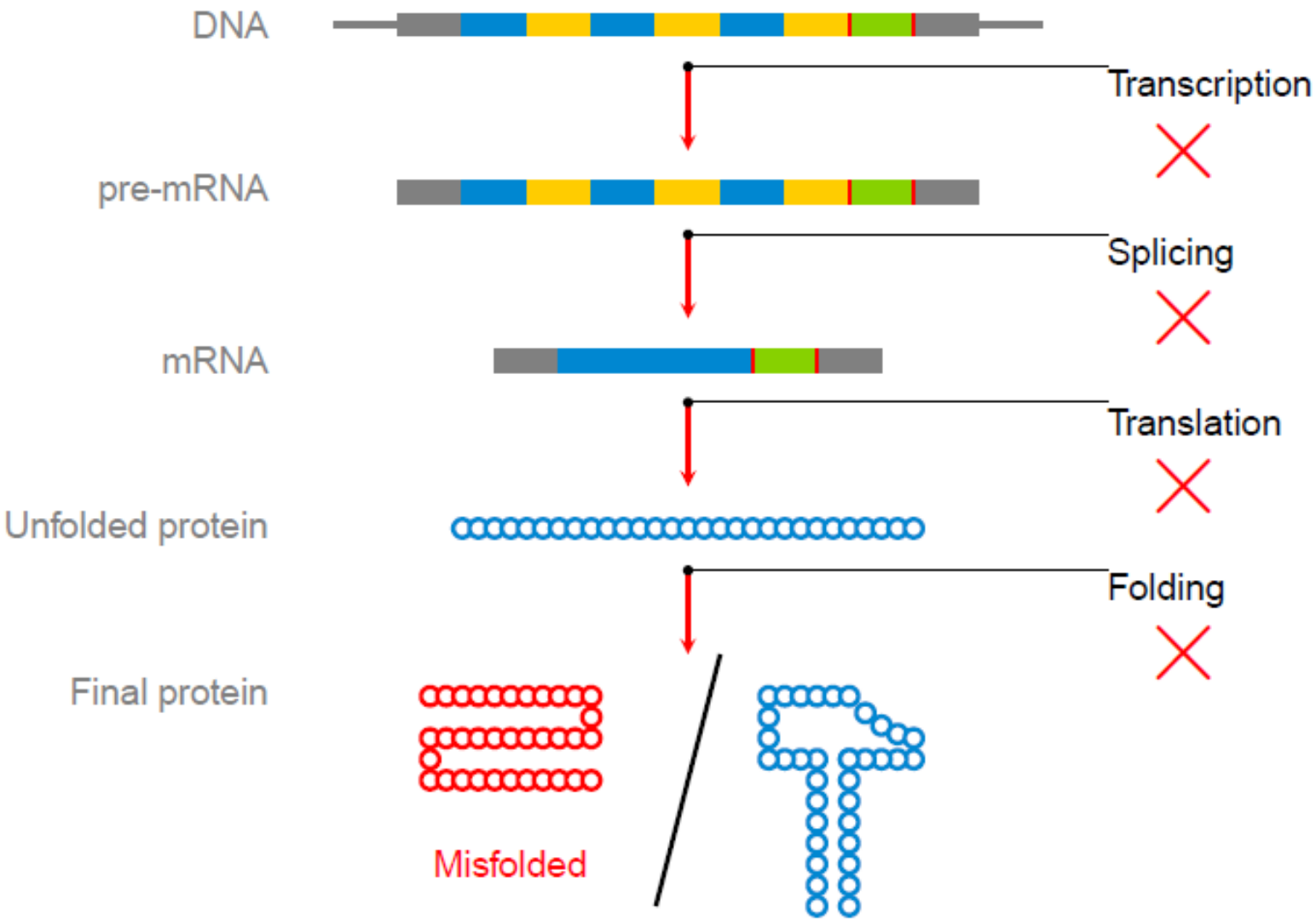
Errors can occur at any stage



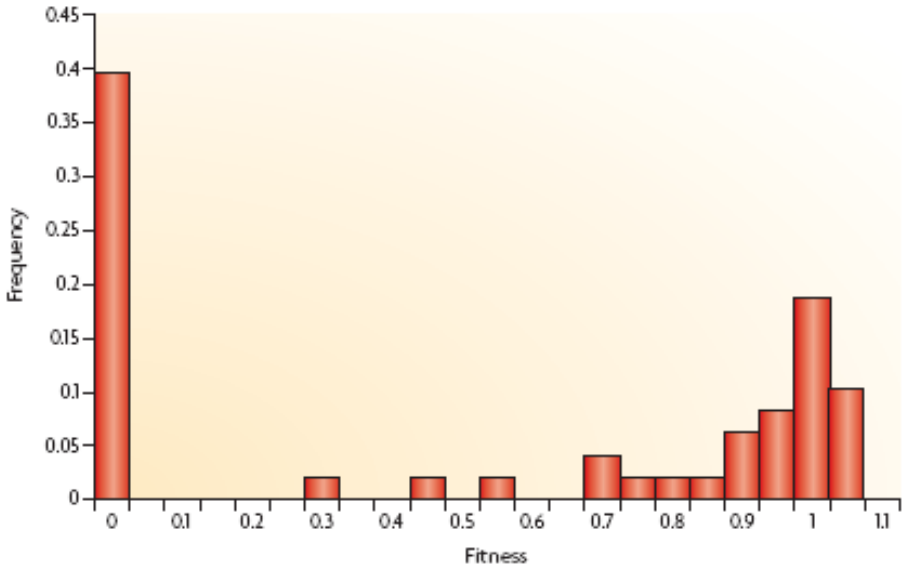
Outline

1. Evolution of error rates under a
   speed vs. accuracy tradeoff
2. Molecular errors pre-screen future variants,
   and so promote evolvability
3. Genetic polymorphism
   is not required for evolvability
4. Protein coding sequences can evolve de novo
   from pre-screened noncoding sequences

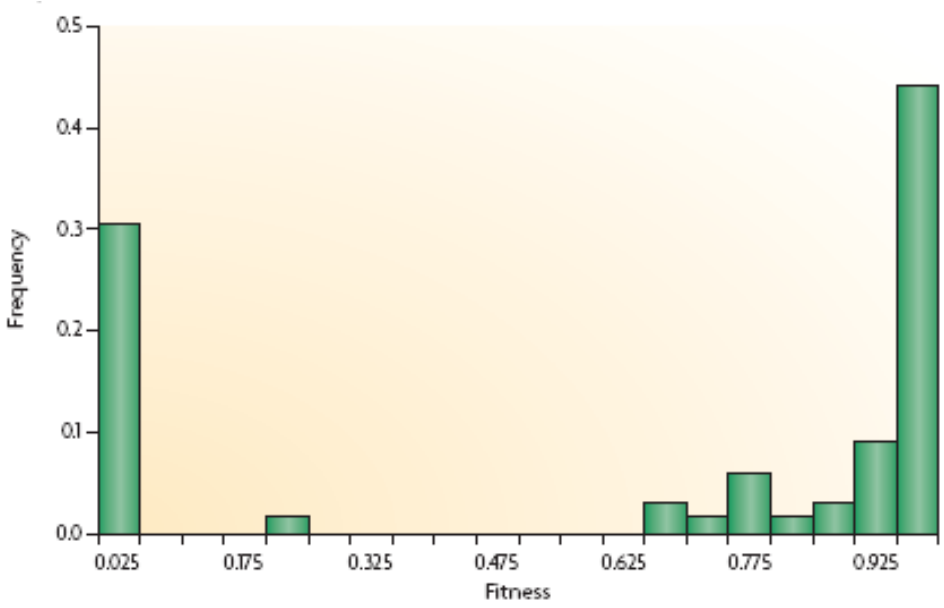
Consequences of errors are either bad or relatively harmless, rarely in between



Distribution of fitness effects of new mutations



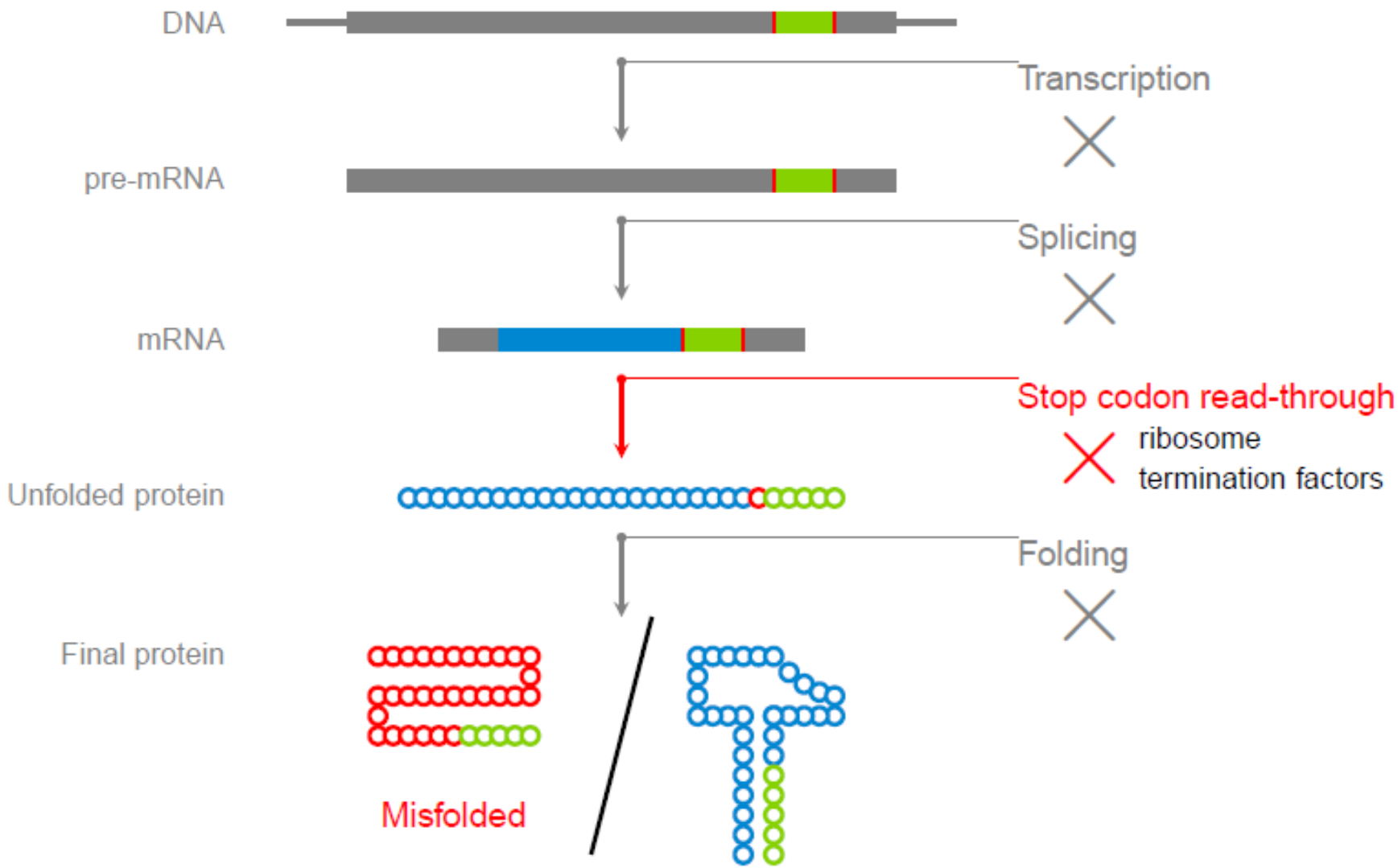
vesicular stomatic virus



yeast

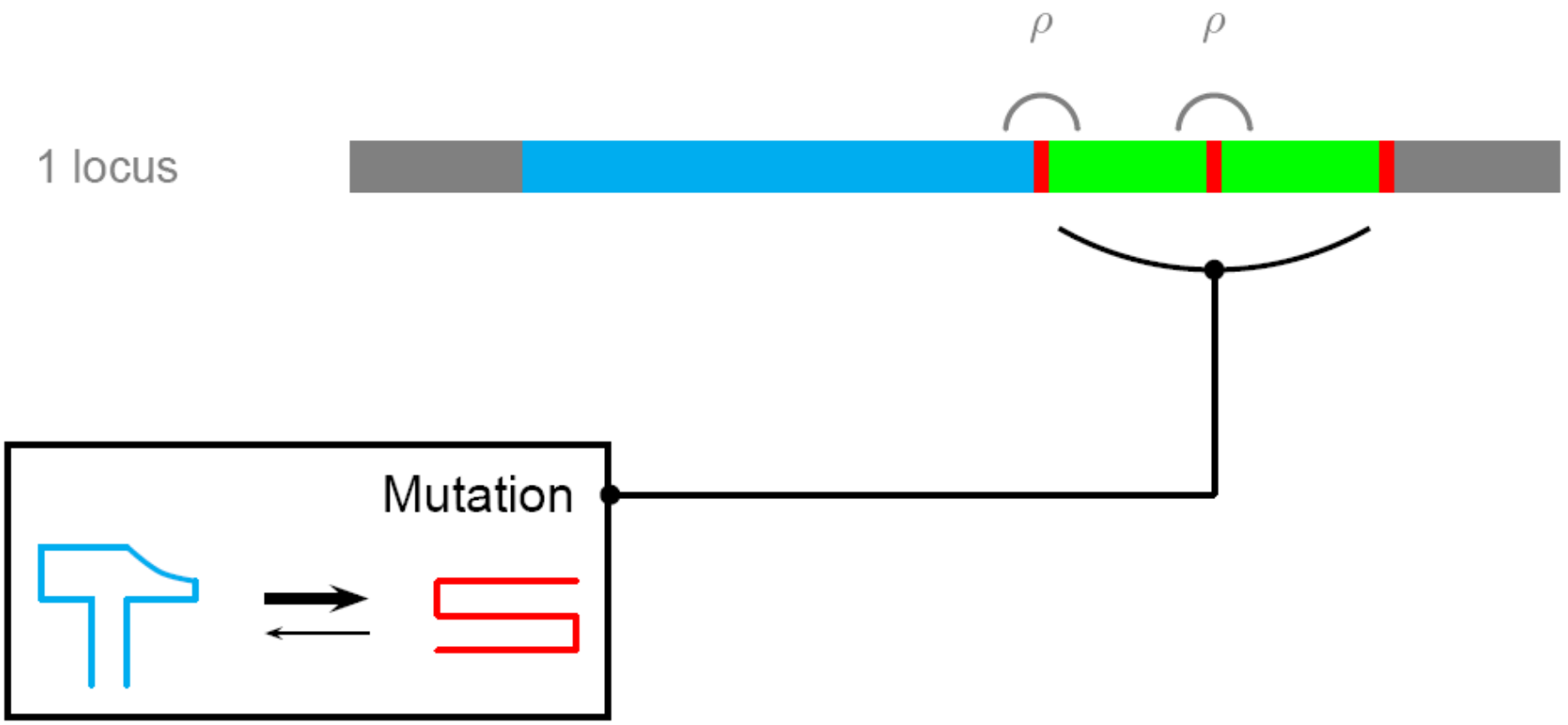
Eyre-Walker & Keightley 2007

Stop codon readthrough: case study of molecular errors

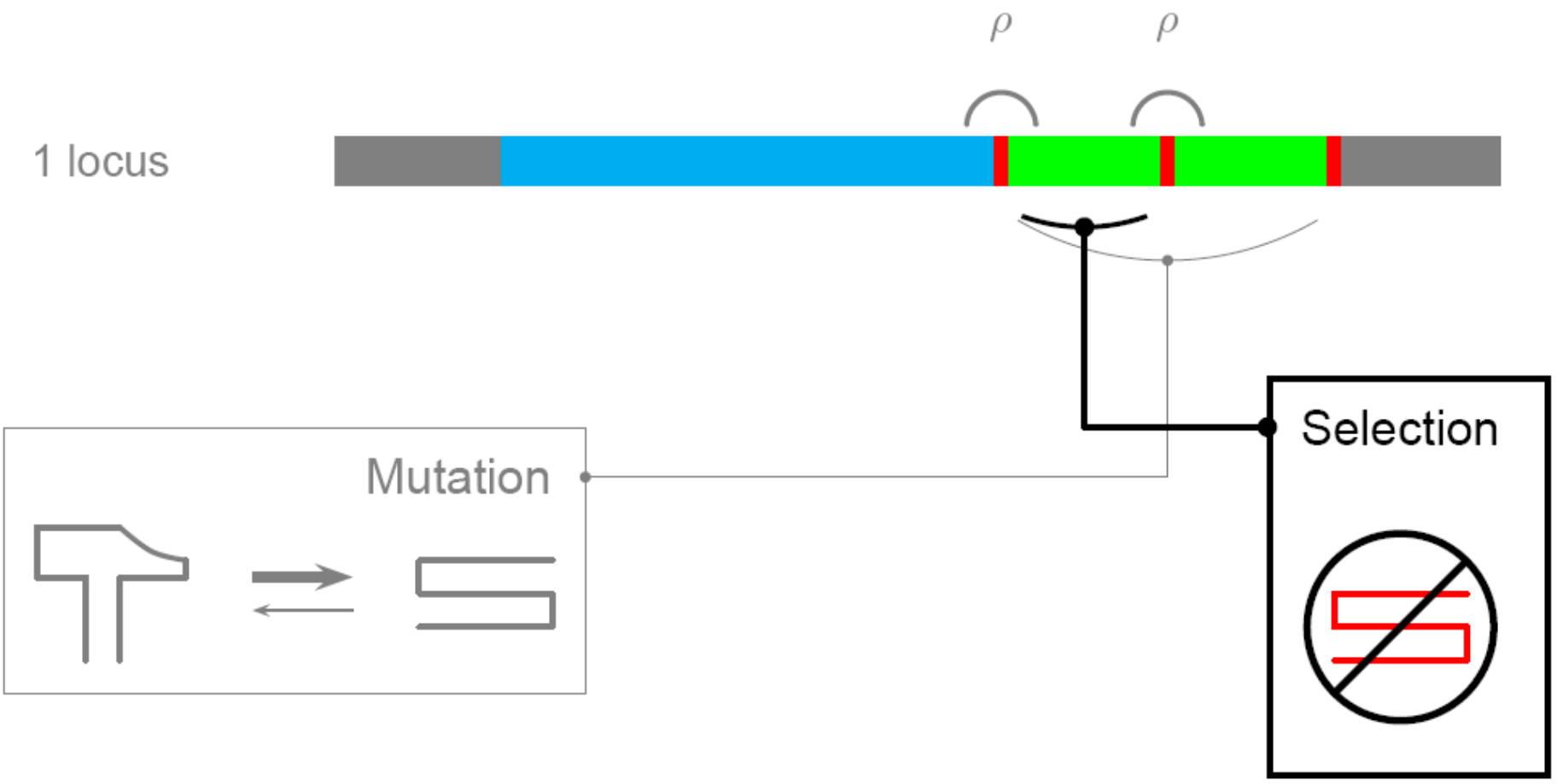


Readthrough at error rate ρ

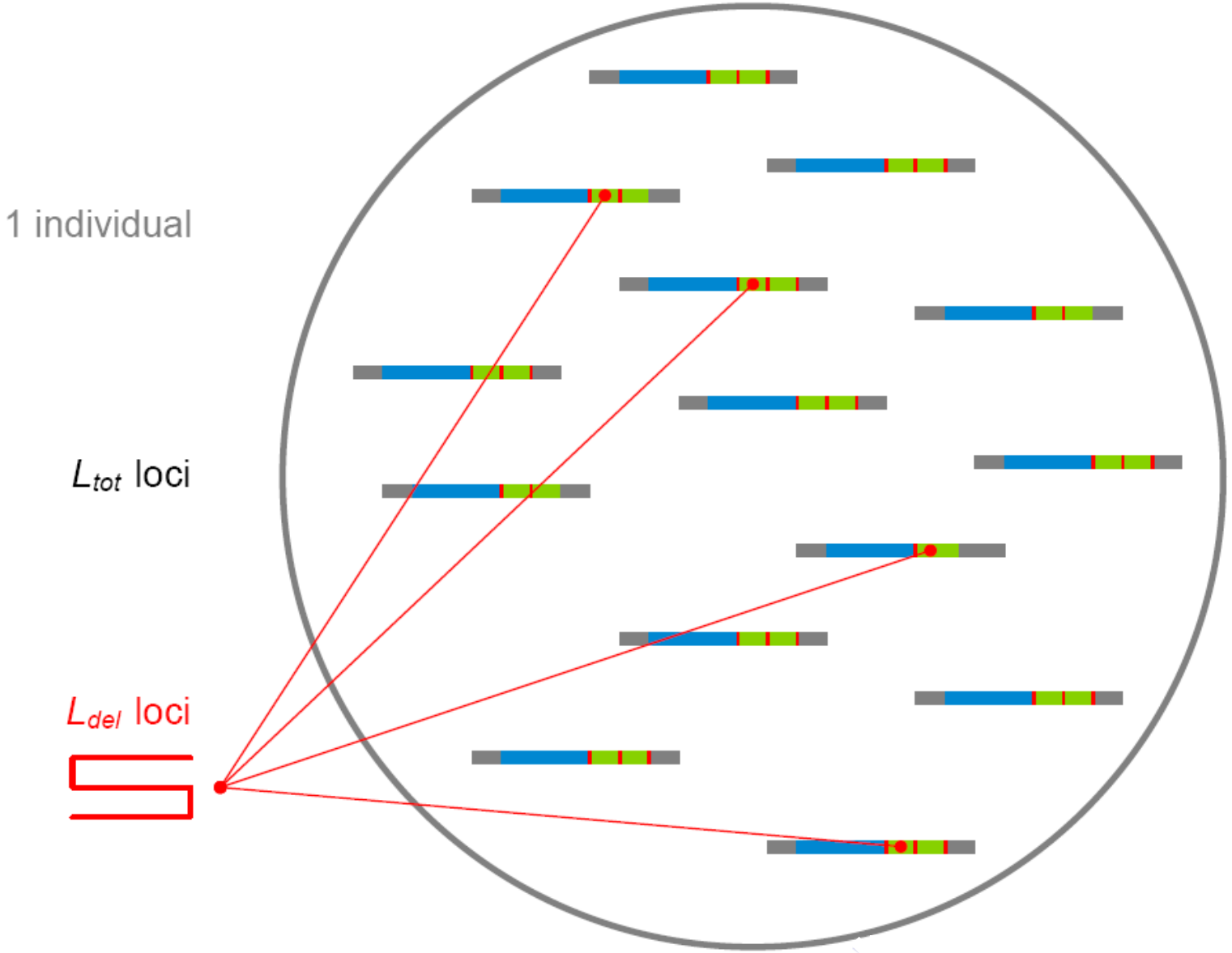
Mutation bias favors misfolding



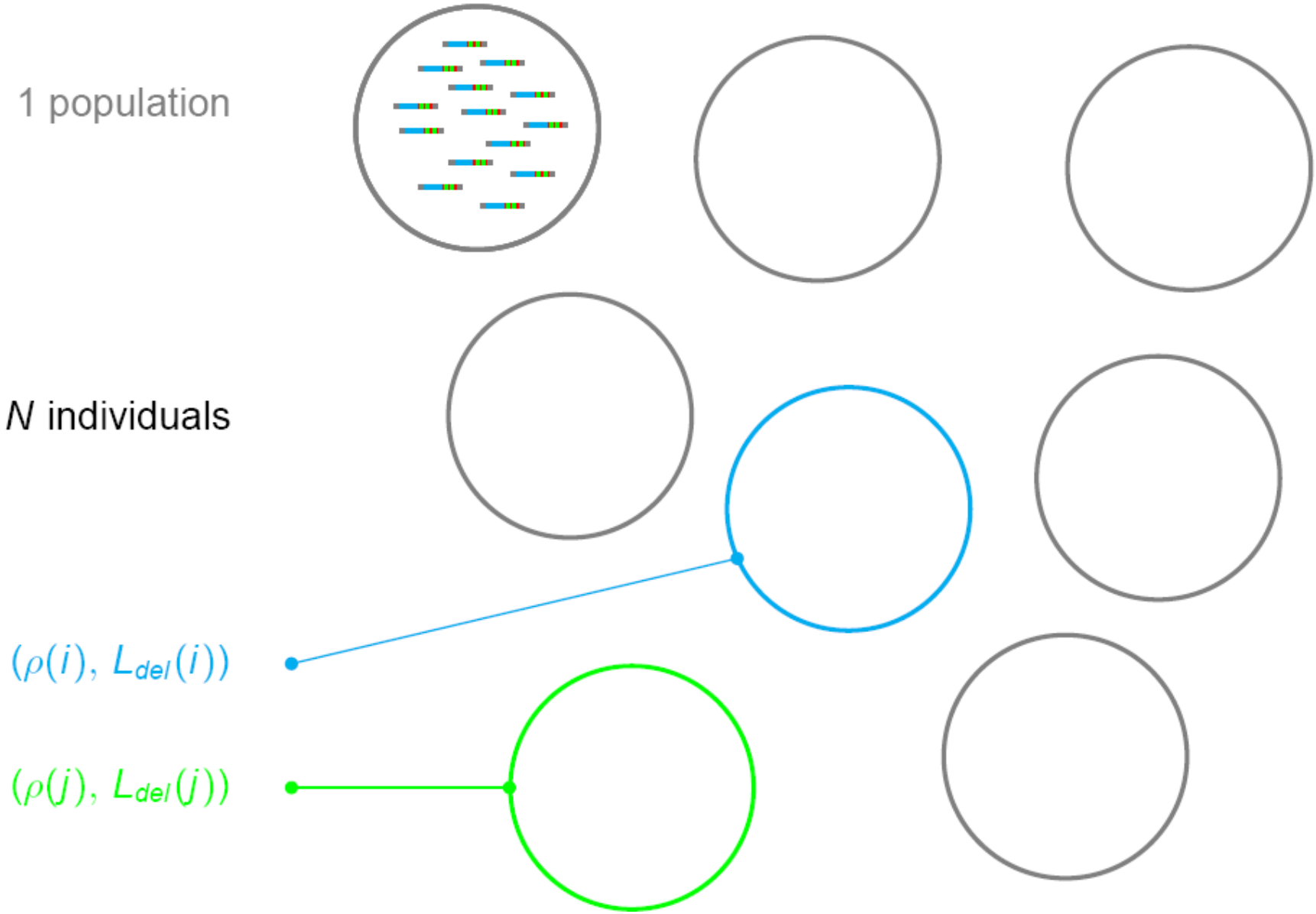
Selection for a stable fold even after a readthrough error



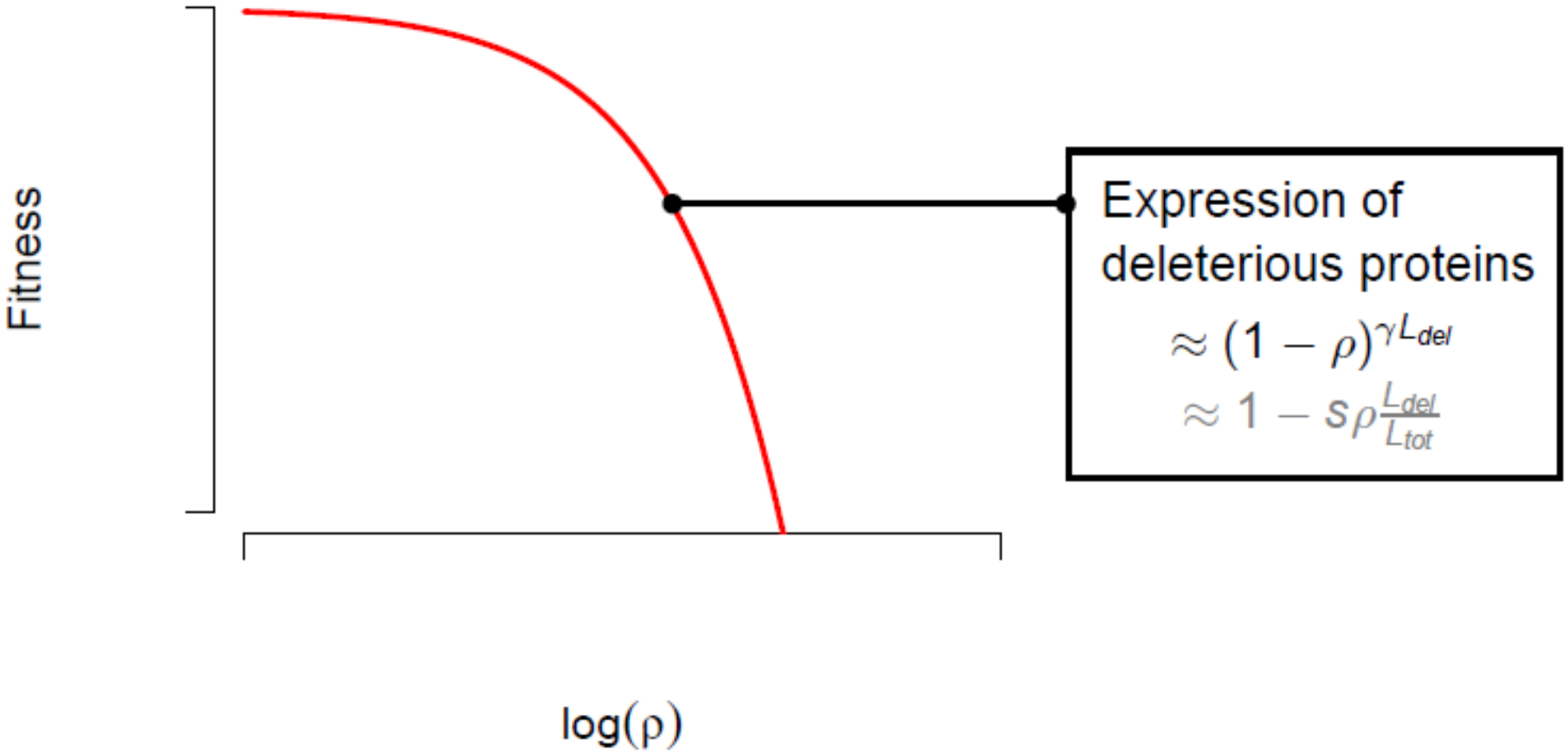
Readthrough errors happen at many loci. Some are sensitive.



Individual genotype
= error rate, #sensitive loci

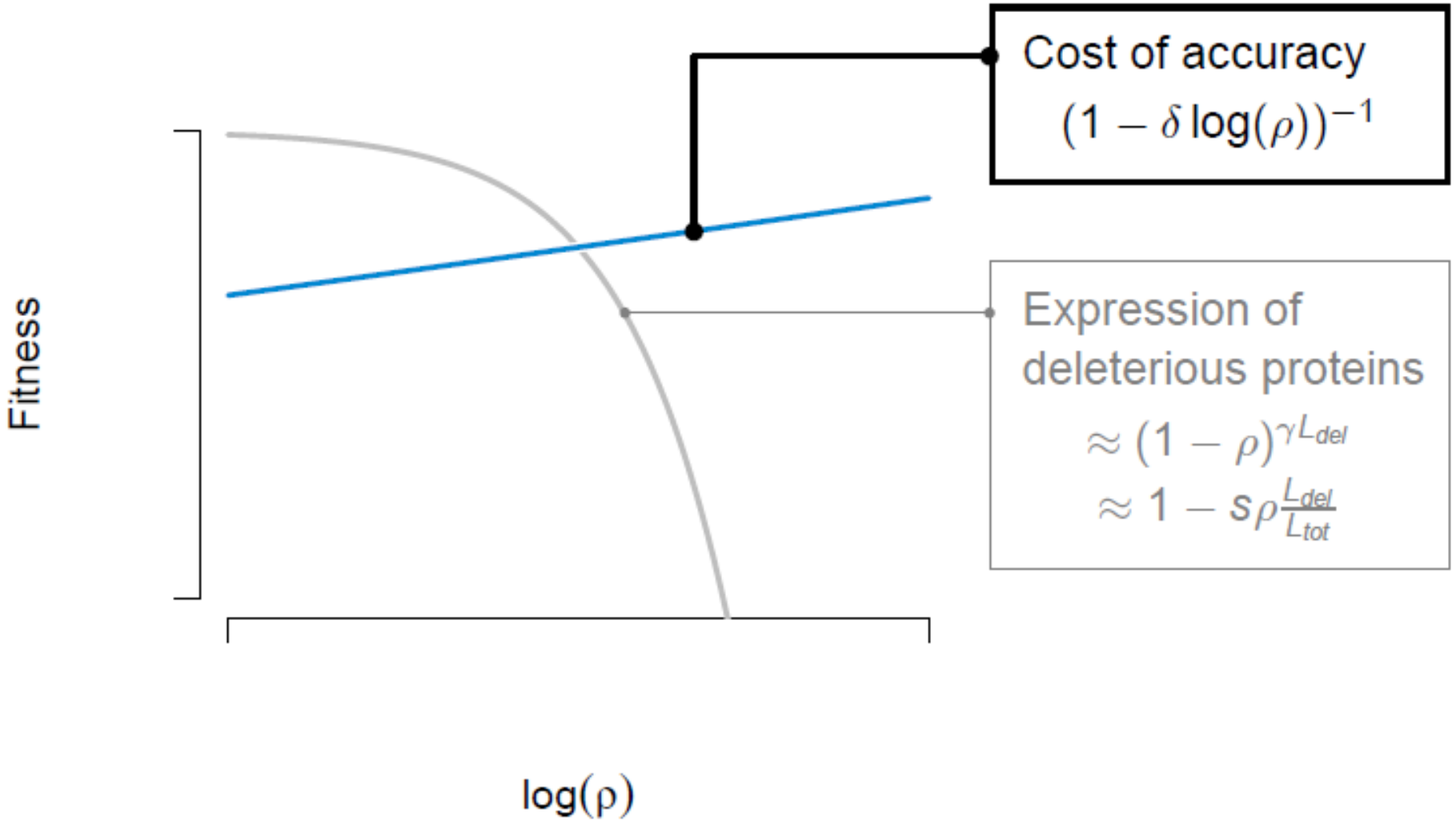


Costs and benefits of proofreading



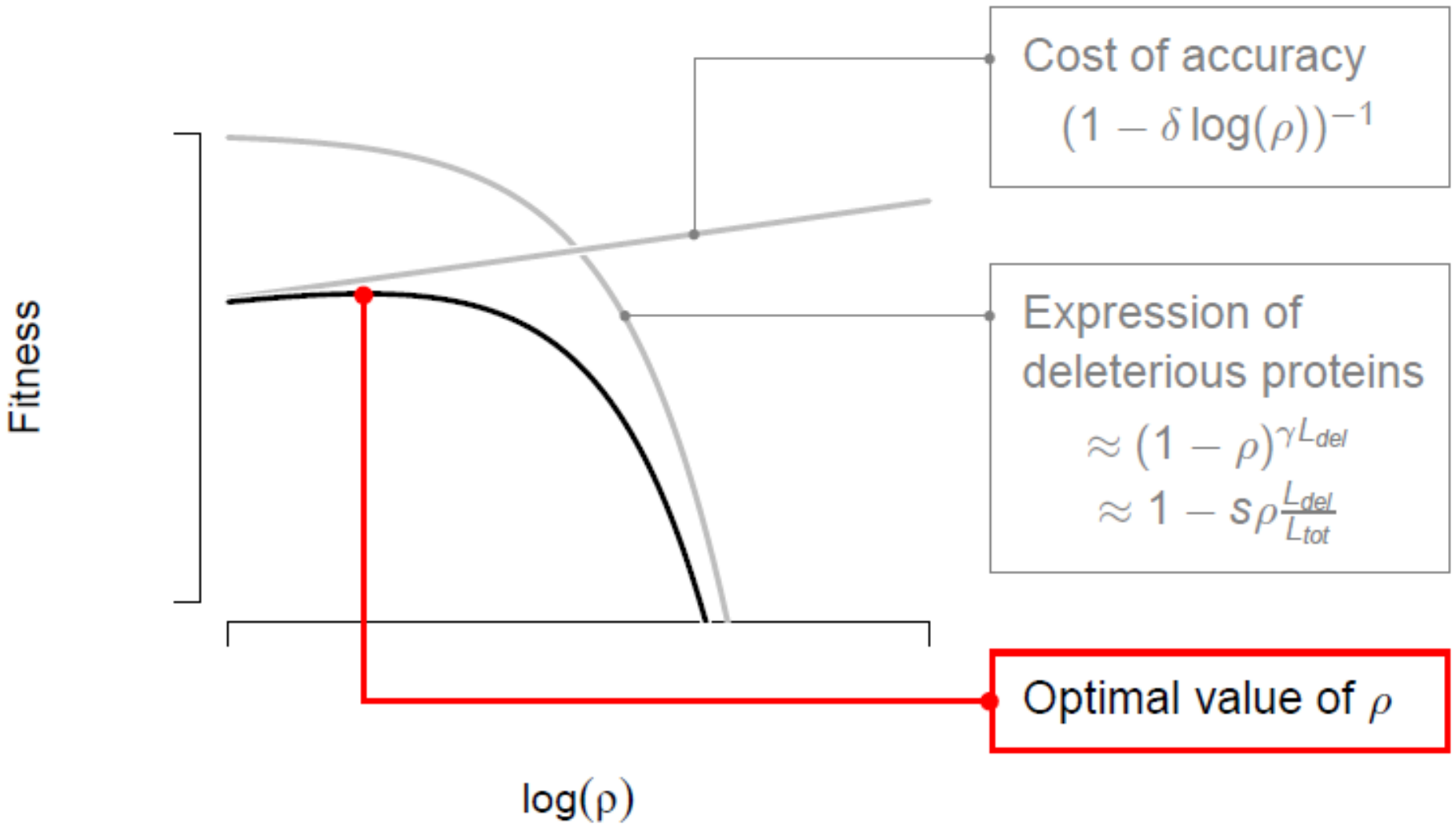
Rajon & Masel (2011)

Costs and benefits of proofreading



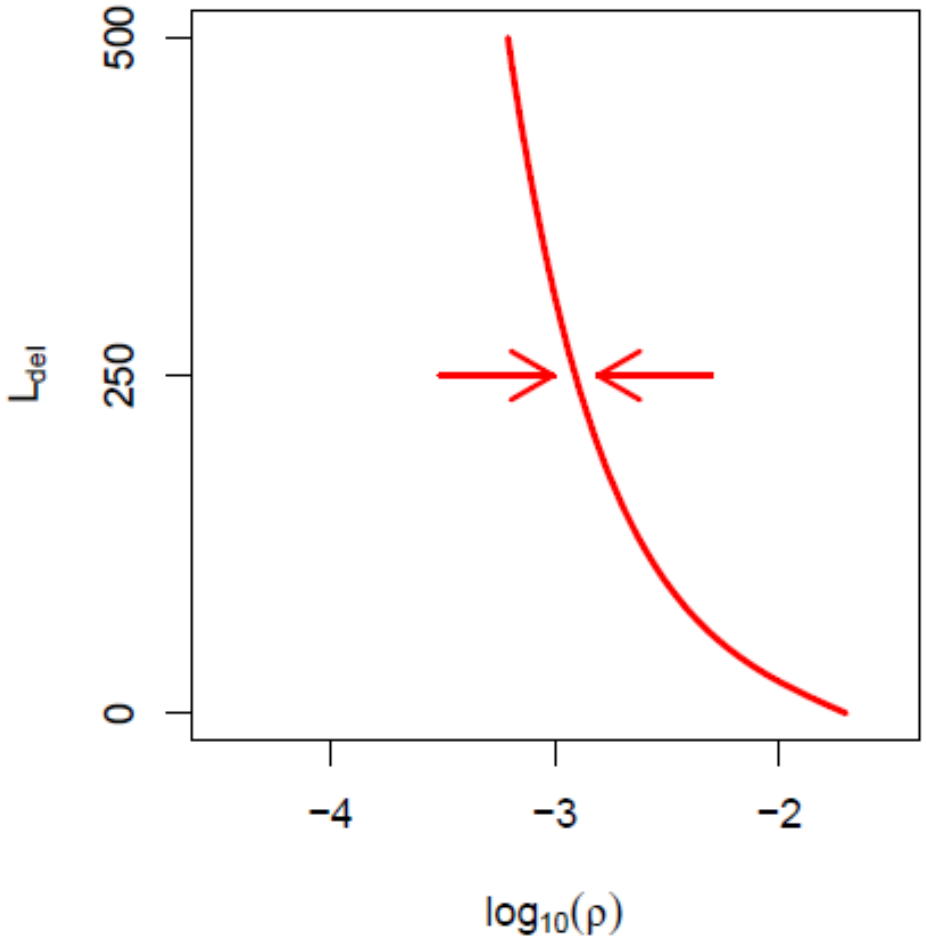
Rajon & Masel (2011)

Costs and benefits of proofreading



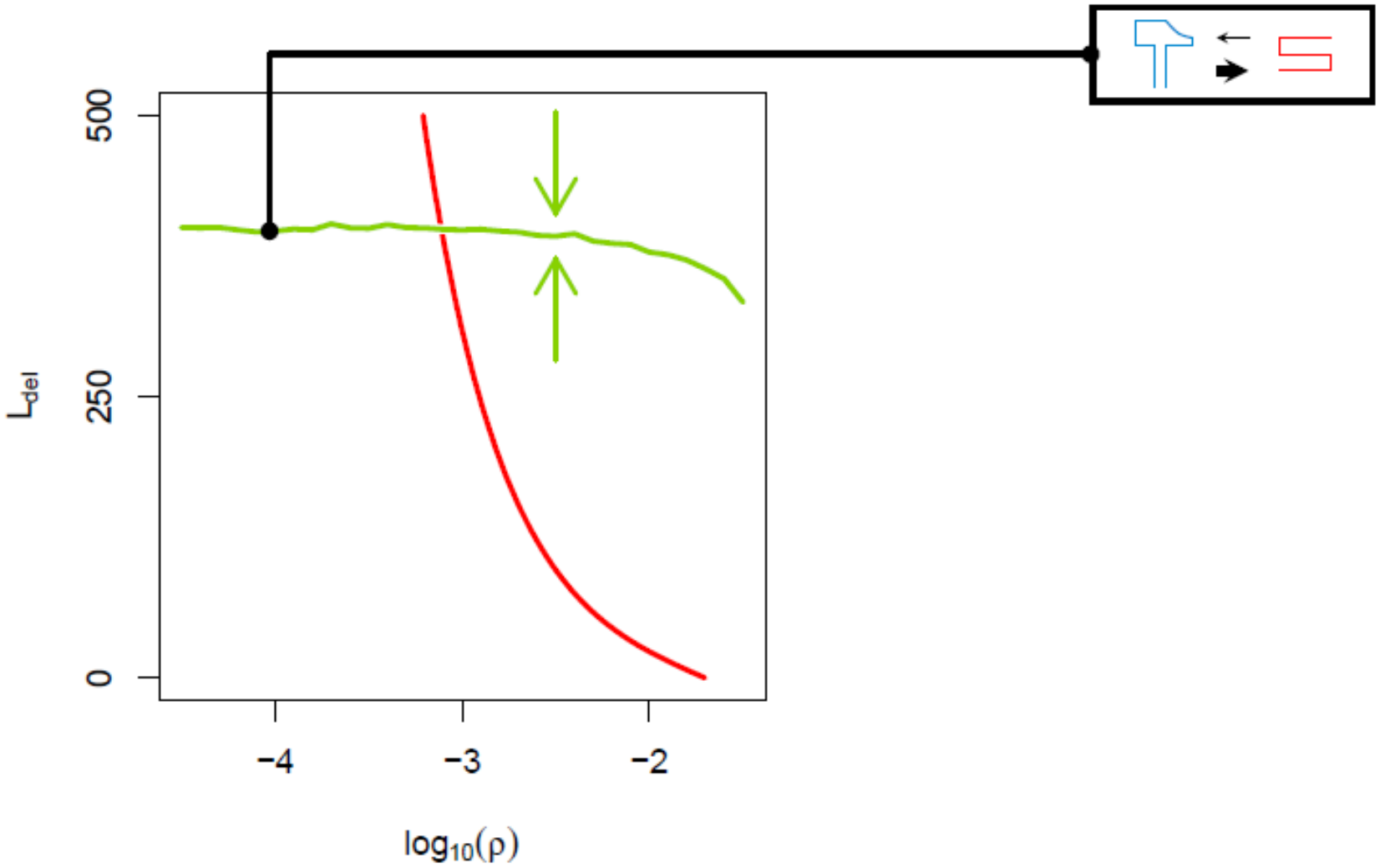
Rajon & Masel (2011)

Coevolution of ρ and L_{del}



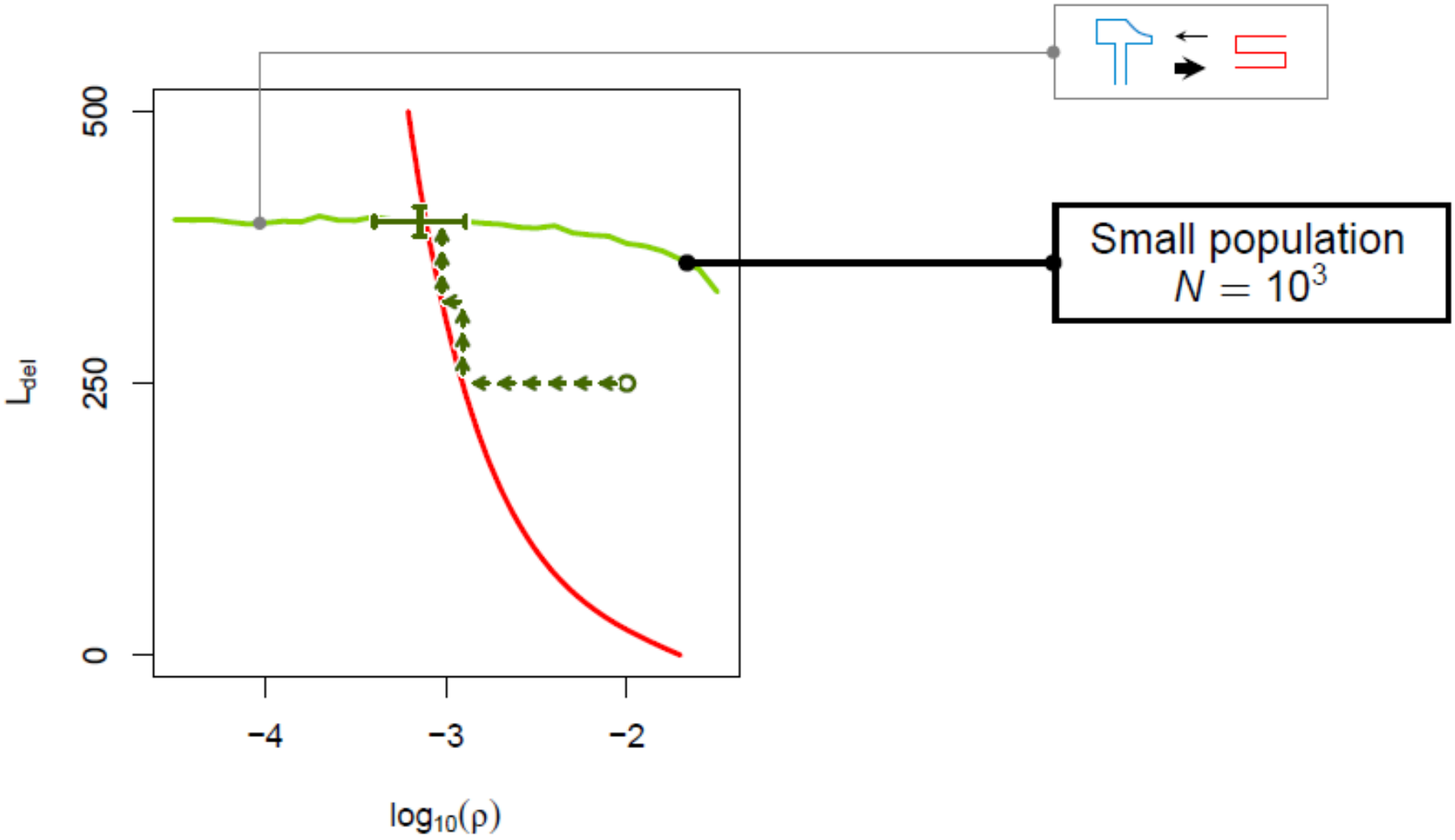
Rajon & Masel (2011)

Coevolution of ρ and L_{del}



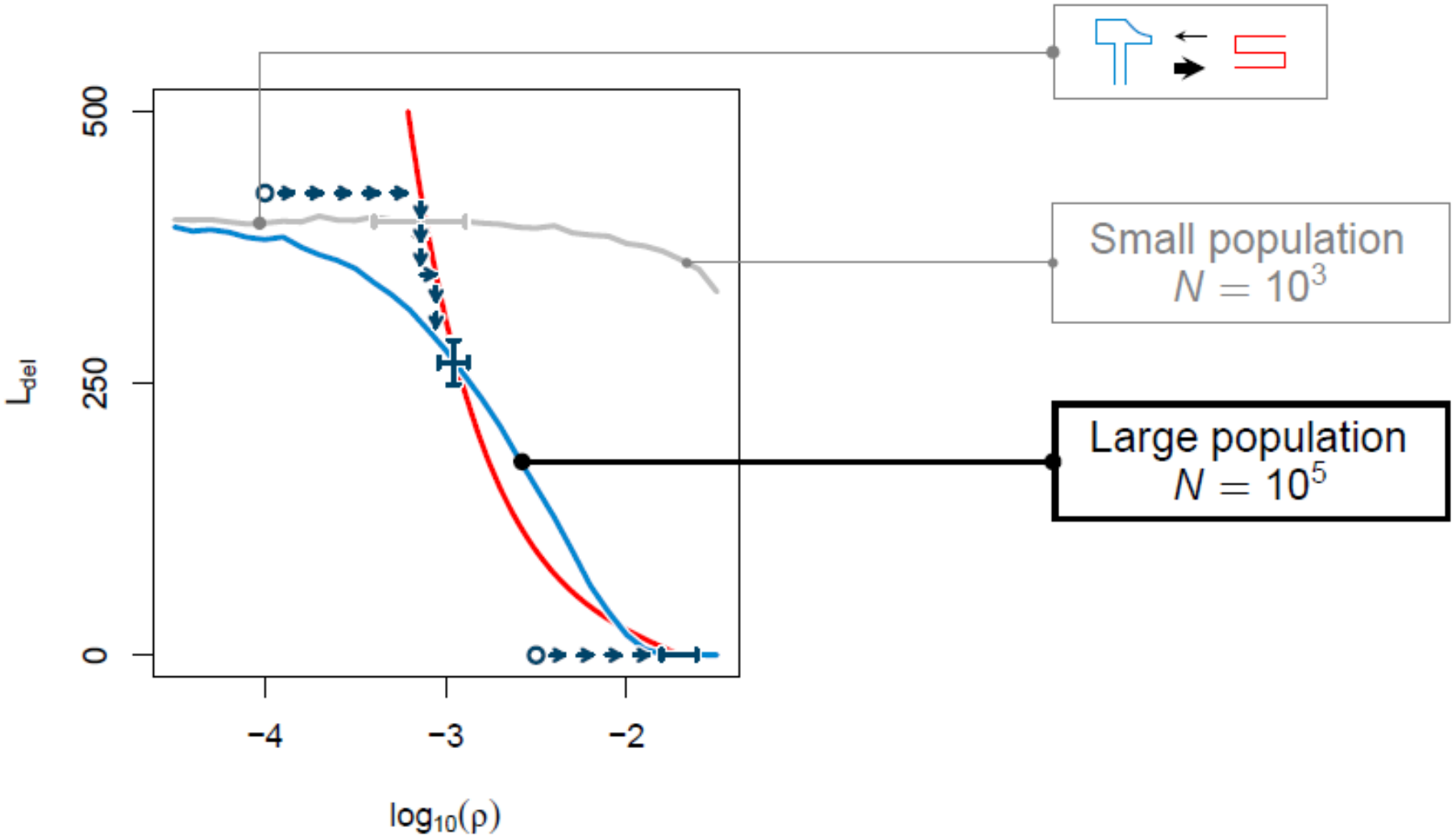
Rajon & Masel (2011)

Coevolution of ρ and L_{del}



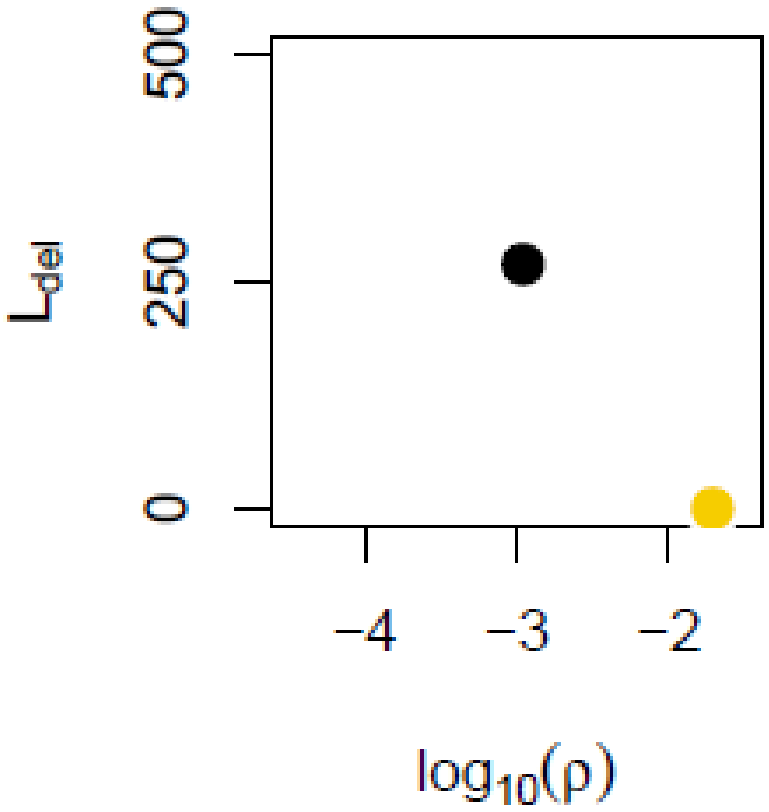
Rajon & Masel (2011)

Two attractors in large populations



Rajon & Masel (2011)

Two strategies are quite different



2 strategies:

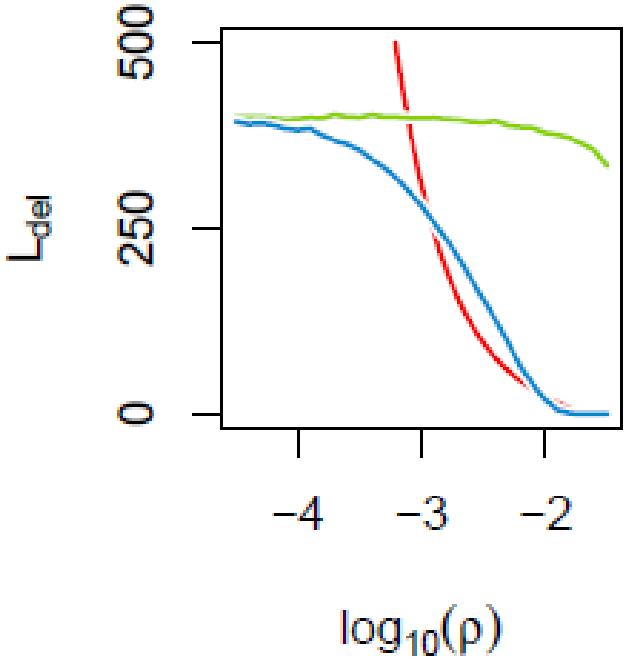
●: allowing deleterious sequences, but hiding them

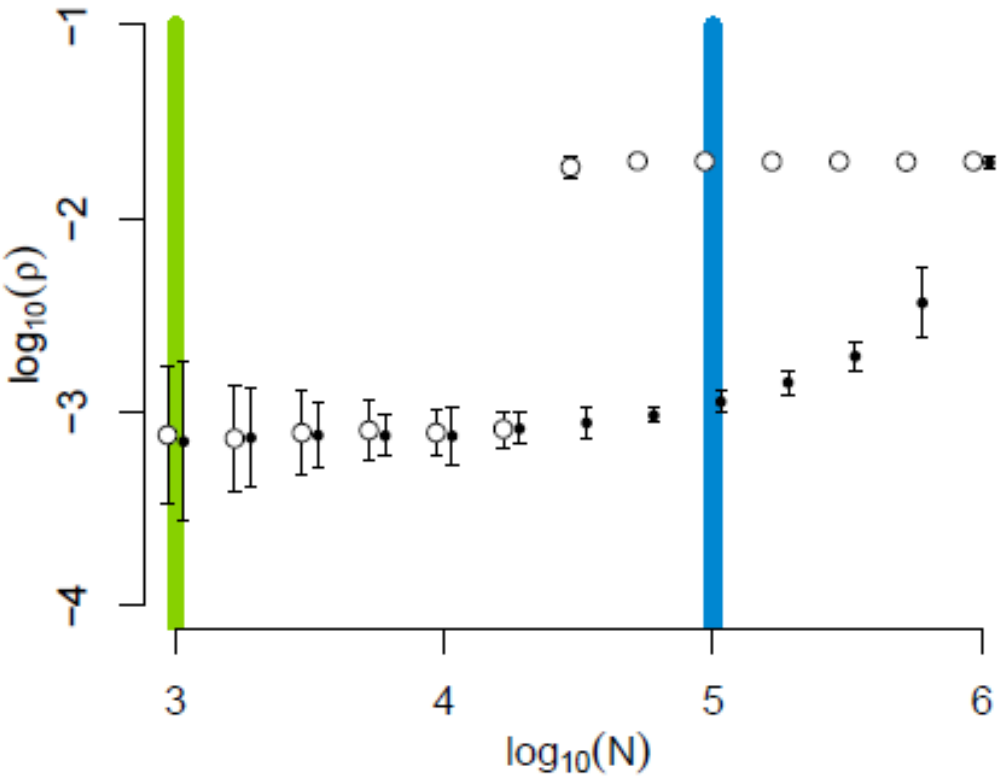
●: eliminating deleterious sequence by expressing them

● or ●?

Rajon & Masel (2011)

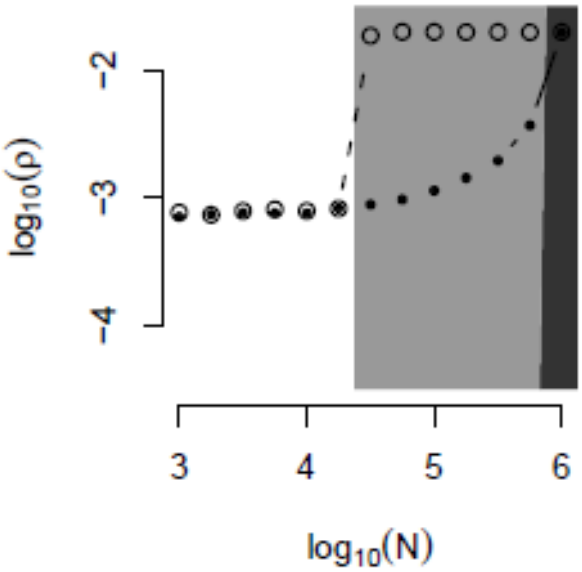
Two attractors for a range of population sizes (i.e. range of limits to weak selection)

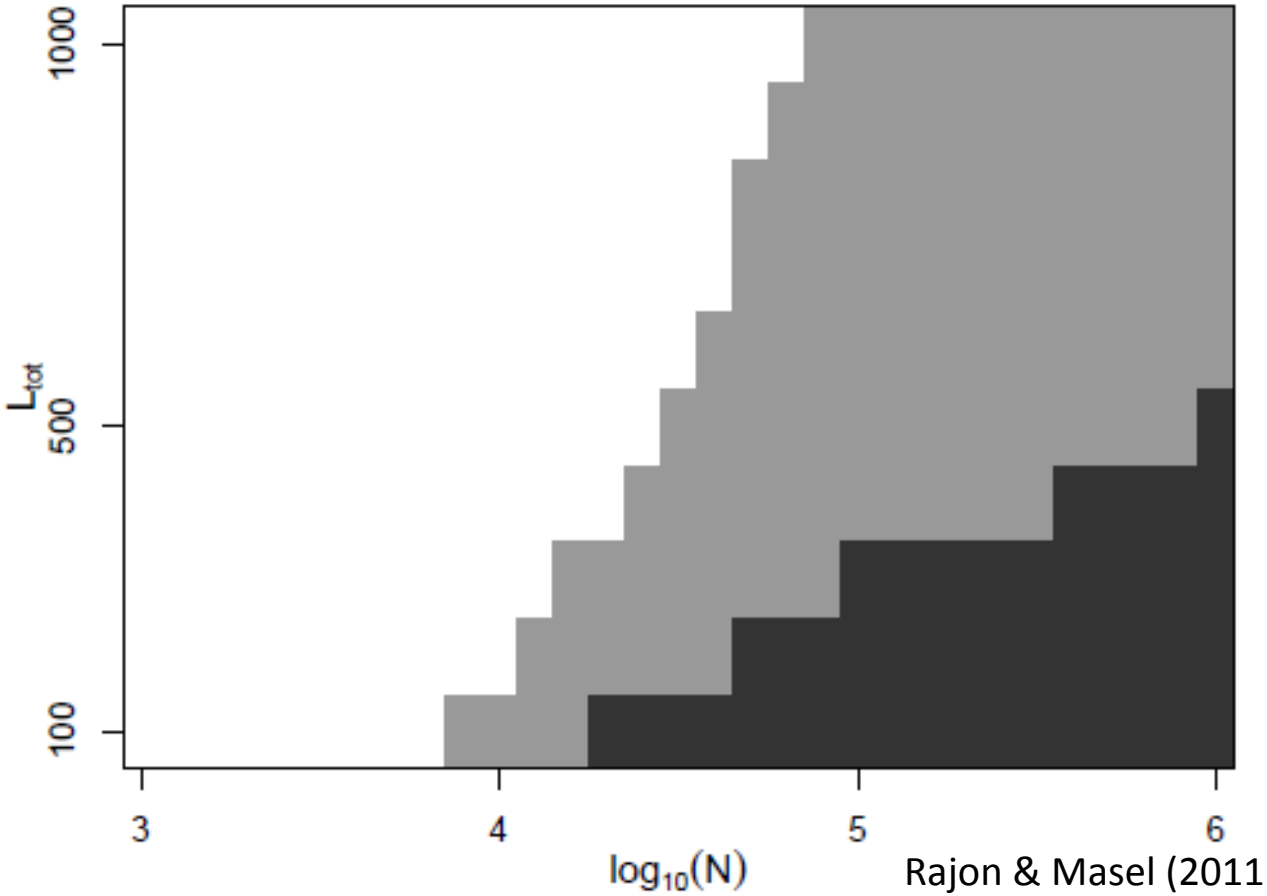




Rajon & Masel (2011)

Larger bistable range with more loci





Rajon & Masel (2011)

Model applies to many kinds of molecular errors

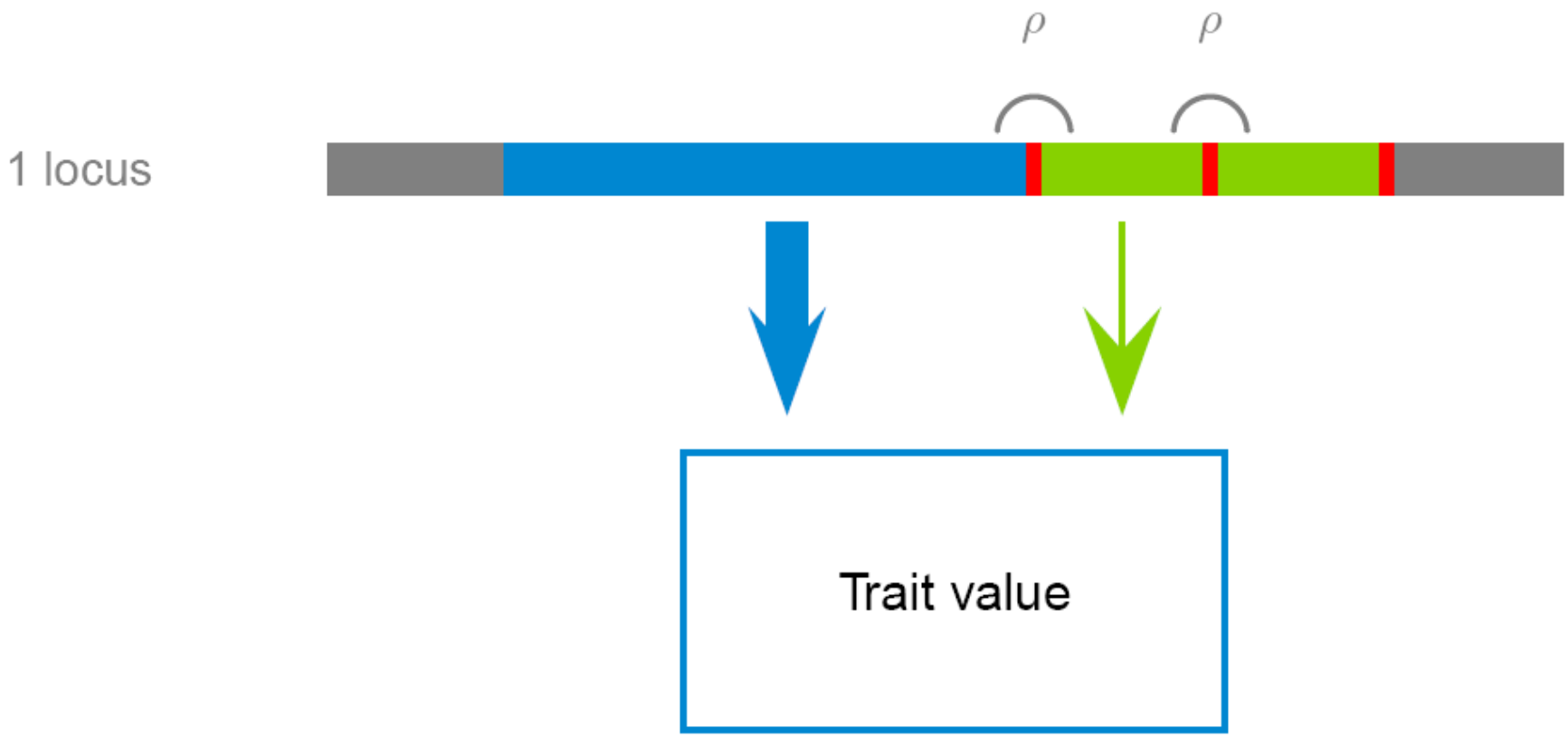
Error	Global solution	Local solution
Stop codon readthrough	Accurate ribosome & release factors	Benign 3'UTR

Rajon & Masel (2011)

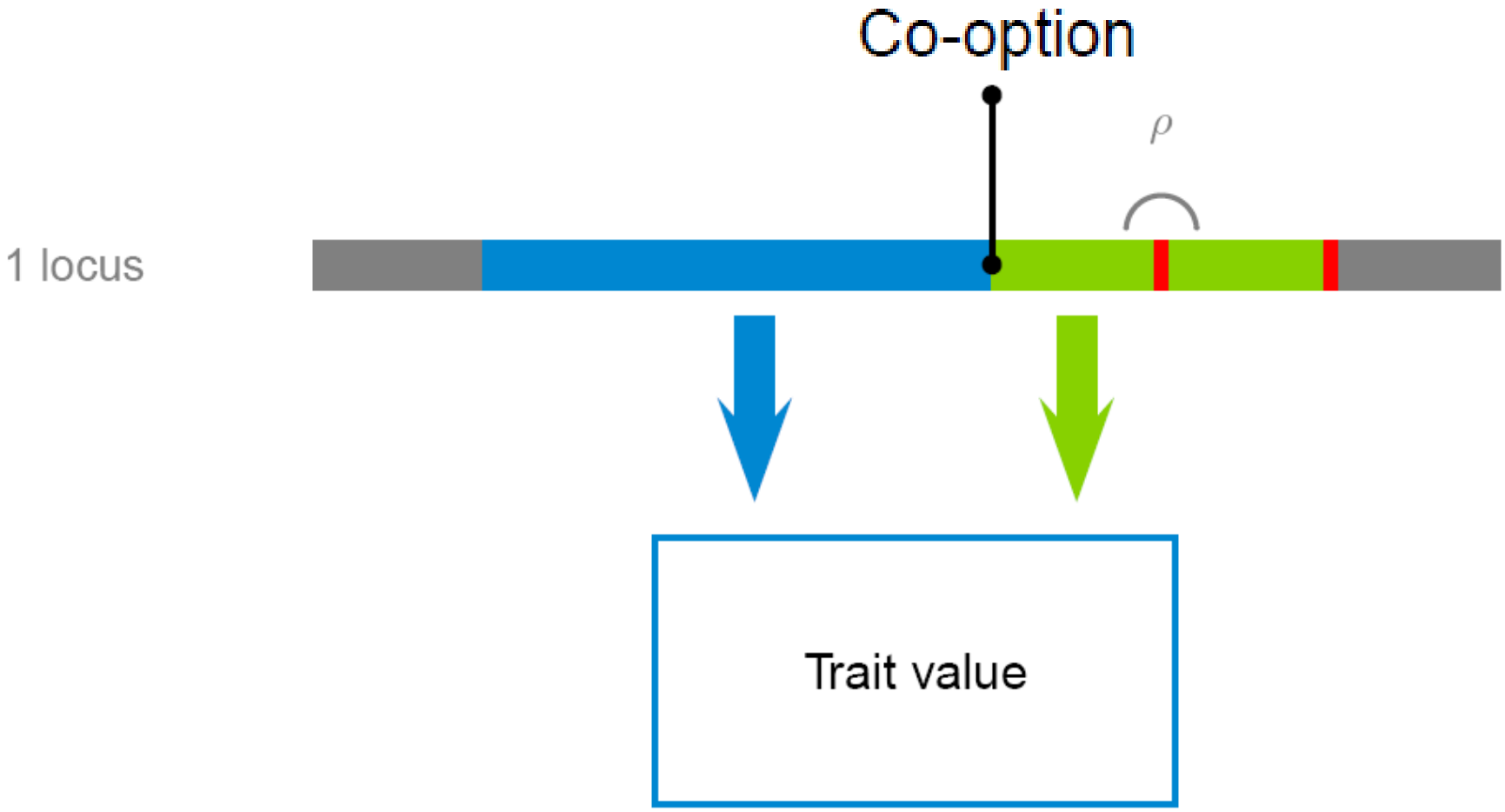
Outline

1. Evolution of error rates under a
   speed vs. accuracy tradeoff
2. Molecular errors pre-screen future variants,
   and so promote evolvability
3. Genetic polymorphism
   is not required for evolvability
4. Protein coding sequences can evolve de novo
   from pre-screened noncoding sequences

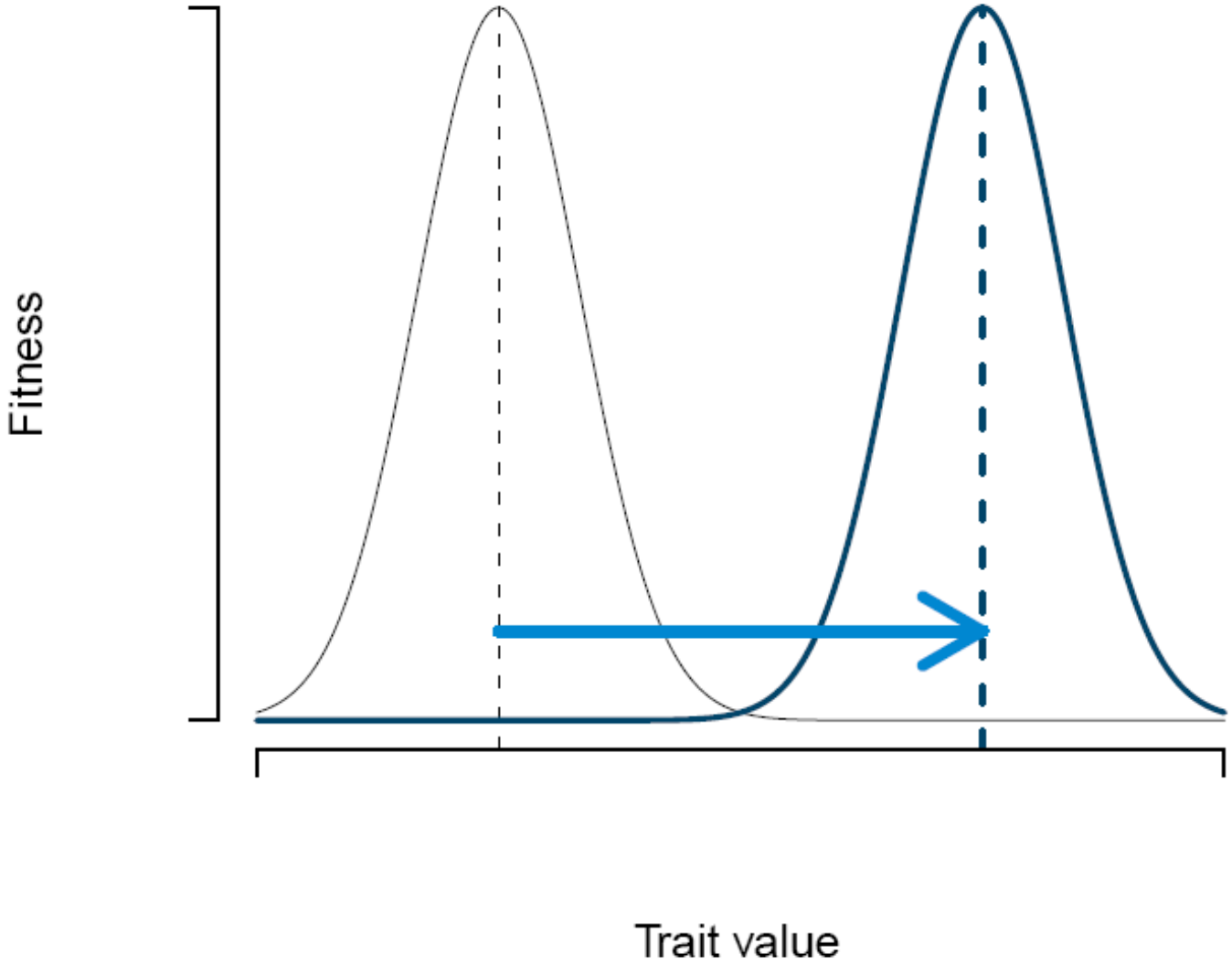
Effect on quantitative trait proportional to expression



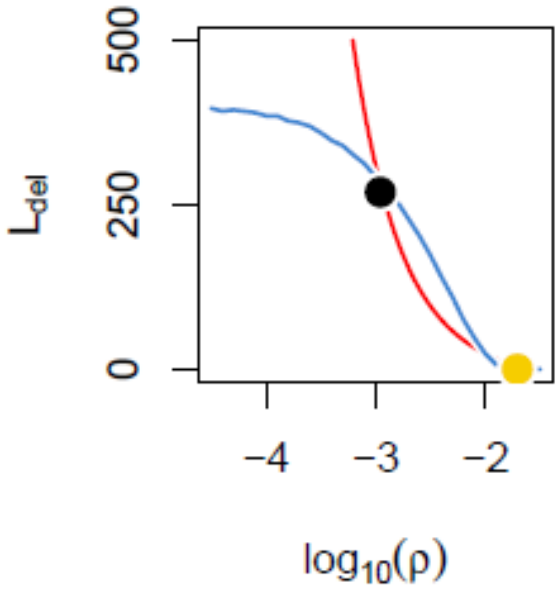
Point mutation in stop codon →
full expression of previously cryptic sequence
(that won't misfold if error rate was high)

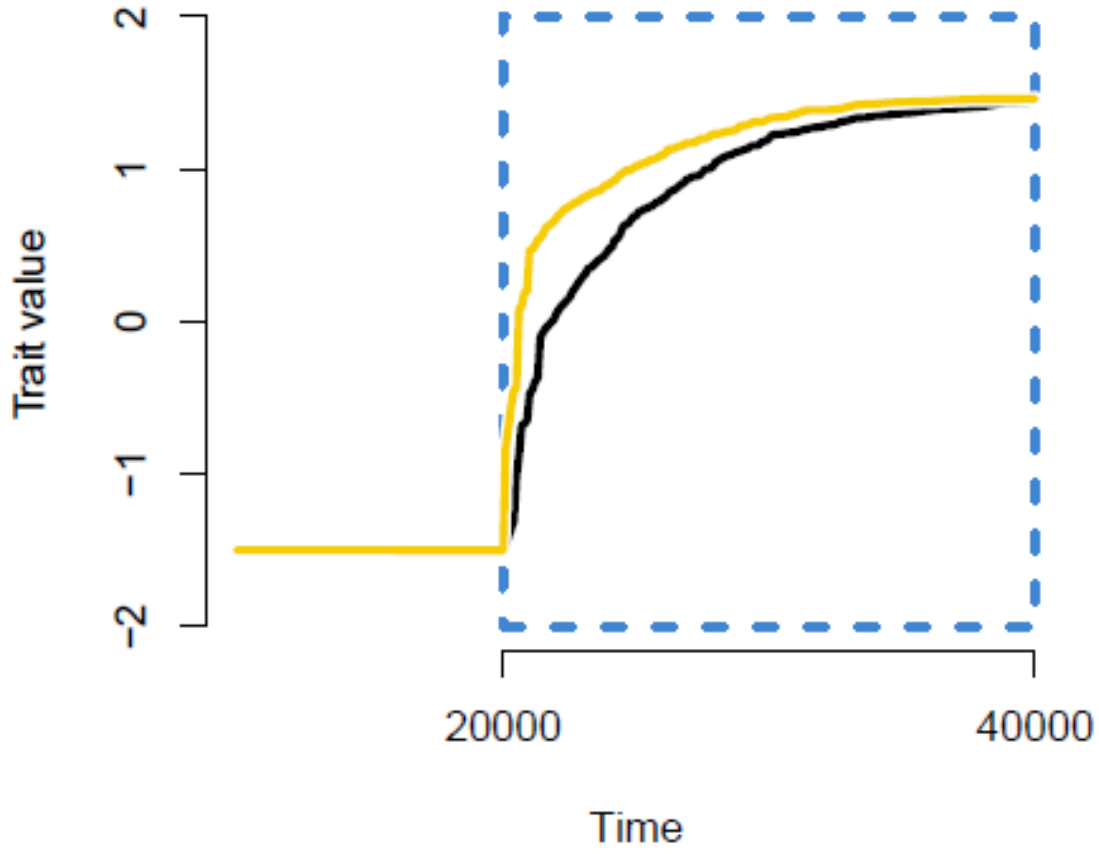


Environmental change in optimal trait value



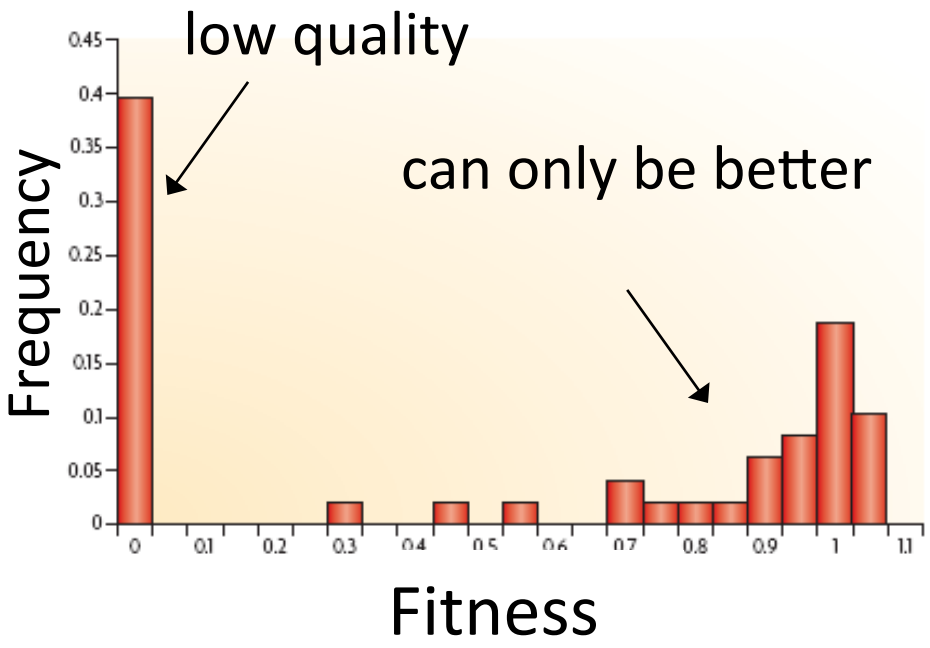
Populations with high error rates evolve faster



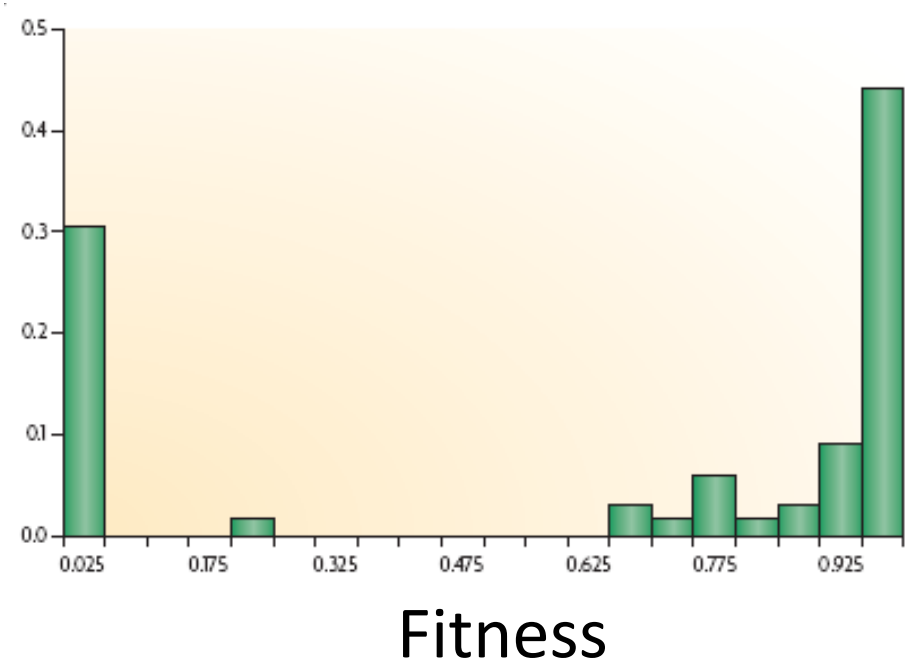


Rajon & Masel (2011)

New mutations



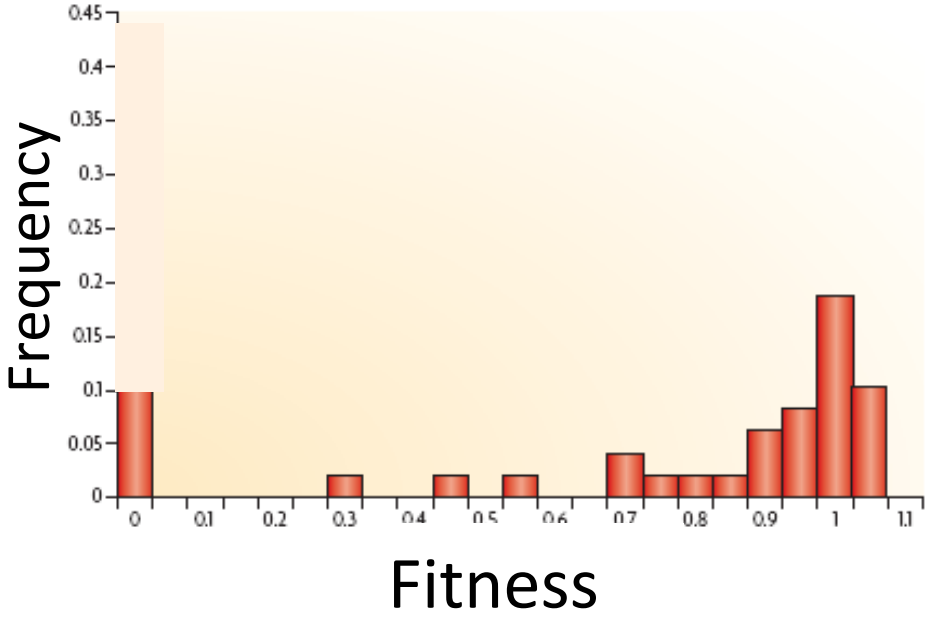
vesicular stomatic virus



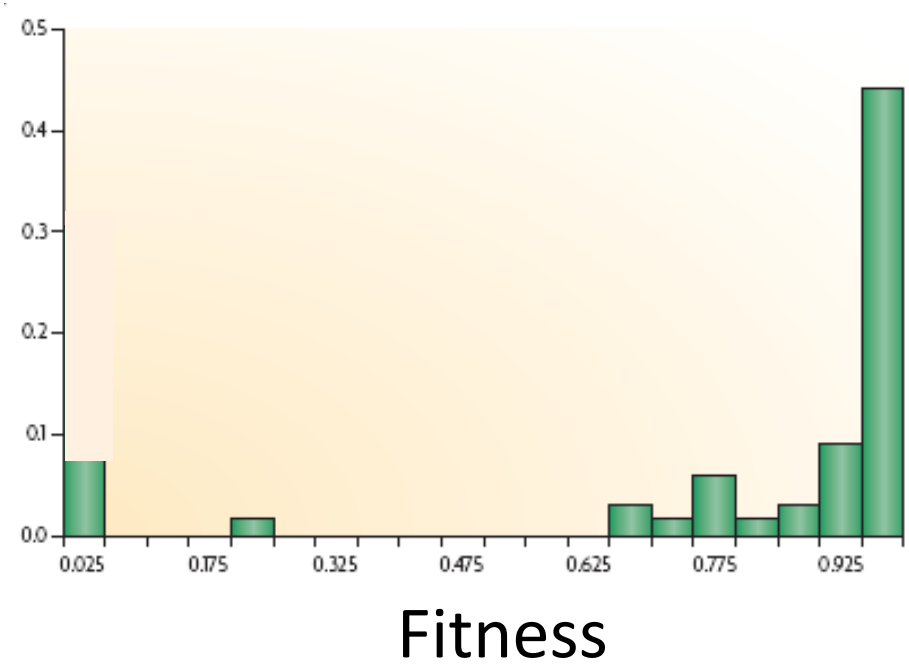
yeast

Eyre-Walker & Keightley 2007

Cryptic variants



vesicular stomatic virus



yeast

Pre-adapting selection

Masel 2006, Rajon & Masel 2011

Evolvability comes from tapping into cryptic variants

- Molecular errors in the present
  mimic mutations in the future
- Strongly deleterious sequences are
  pre-purged in favor of benign ones
- Benign sequences are co-optable for adaptation

Benefits go to any "high error" locally benign cryptic sequences

More examples

- Promiscuous enzyme activities
- Rare protein-protein interactions (PPIs) that lose crypticity when proteins see each other more often

Aside: "cryptic" PPIs (deliberately bad Y2H data) are biologically meaningful

They predict gene noise and plasticity better than "real" PPIs (best practice affinity capture mass spec)

"Stickiness" trumps "hubness"

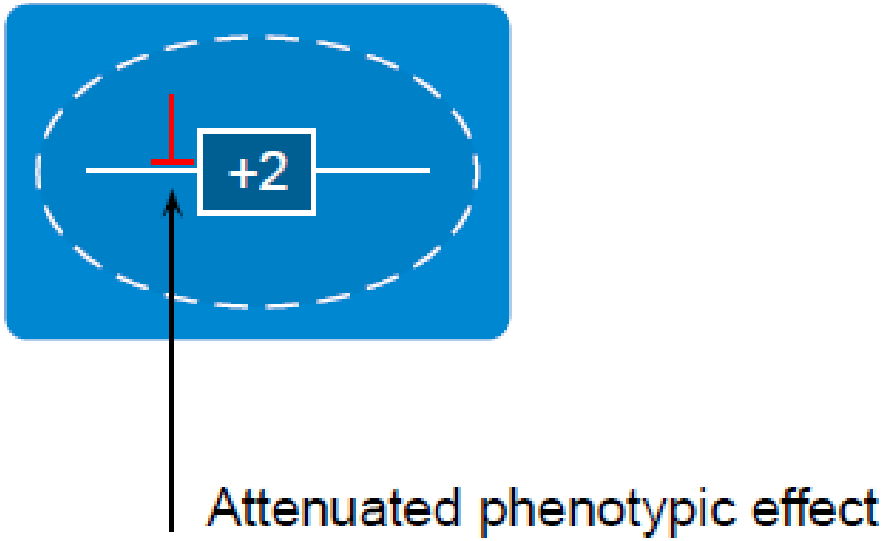
Brettner & Masel (2012)

Outline

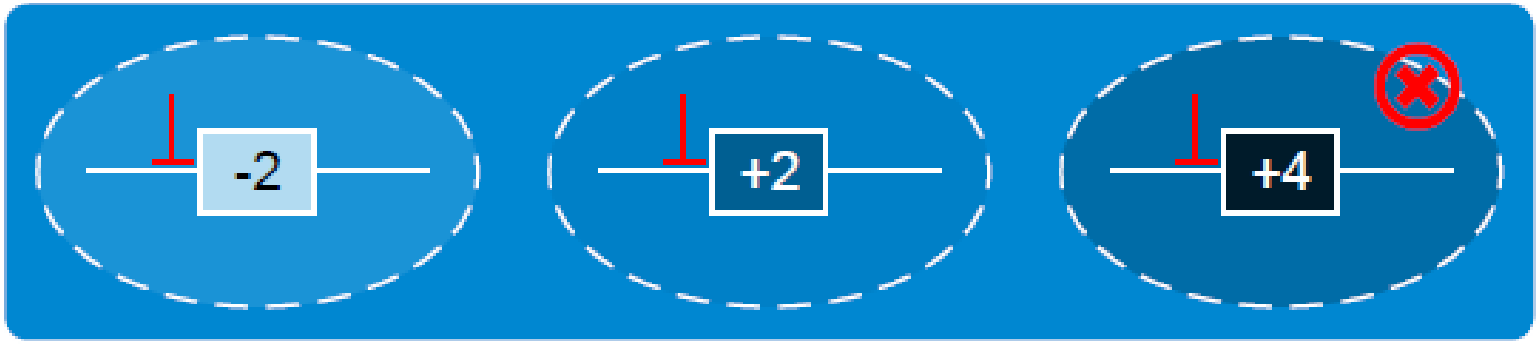
1. Evolution of error rates under a
   speed vs. accuracy tradeoff
2. Molecular errors pre-screen future variants,
   and so promote evolvability
3. Genetic polymorphism
   is not required for evolvability
4. Protein coding sequences can evolve de novo
   from pre-screened noncoding sequences

Let's look at cryptic sequences with and without genetic diversity

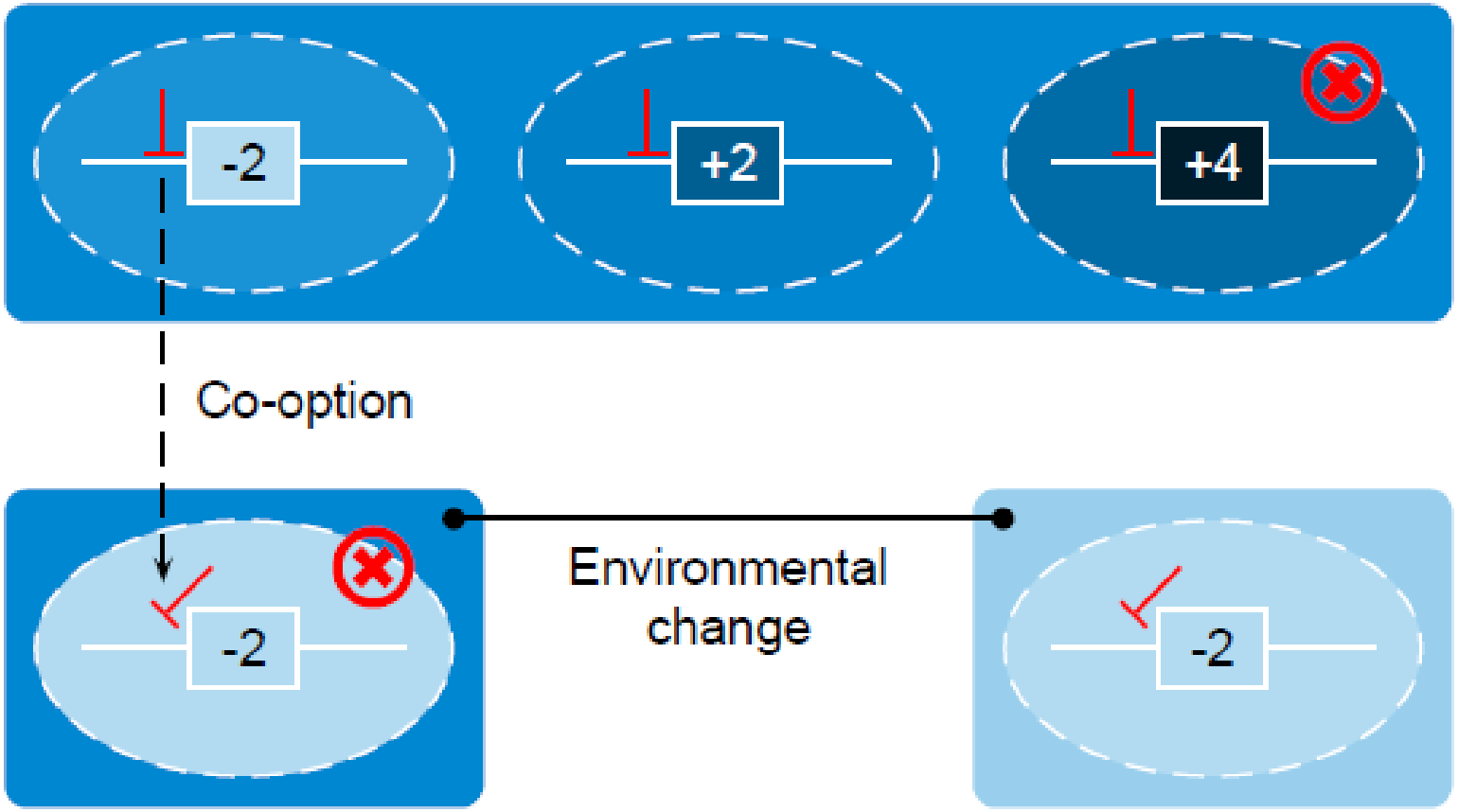
Consider only benign sequences, with different phenotypic effect sizes (i.e. in parameter regime where misfolded cryptic sequences are purged)



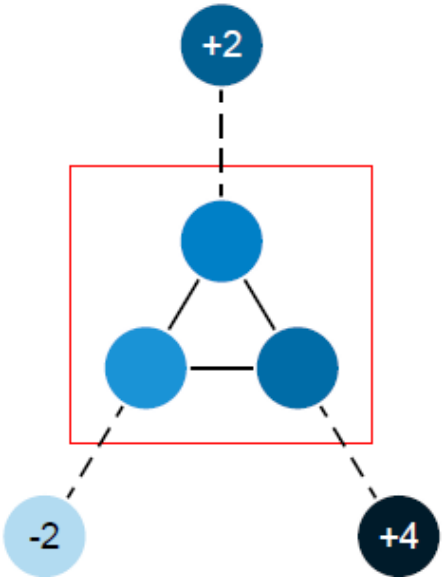
Relaxed selection → cryptic genetic diversity



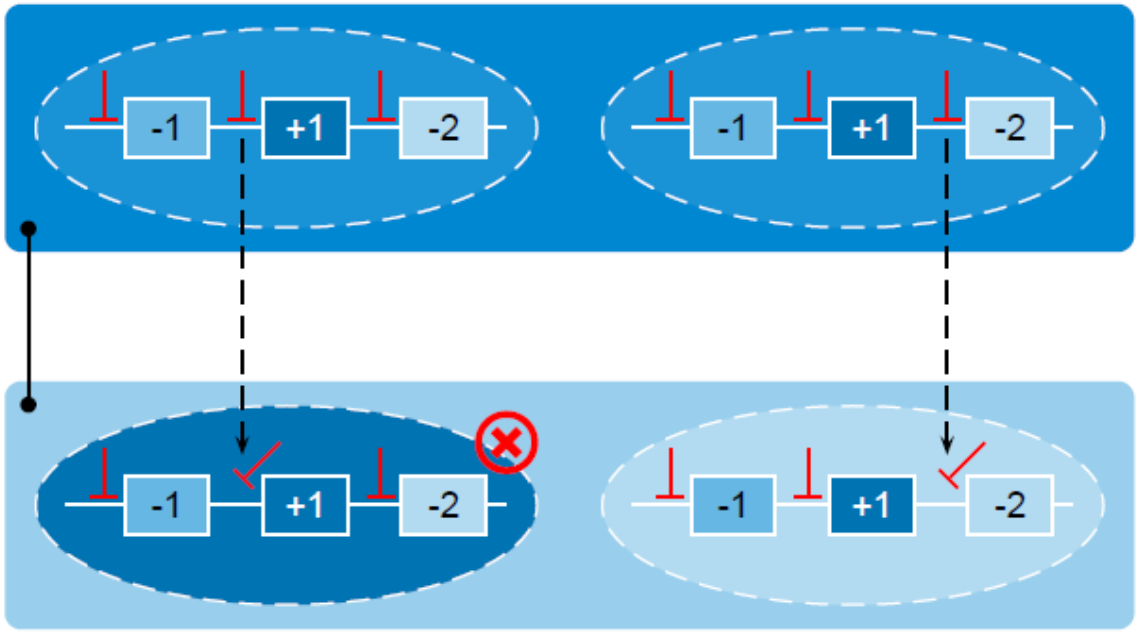
Co-opted variants can be adaptive in a new environment



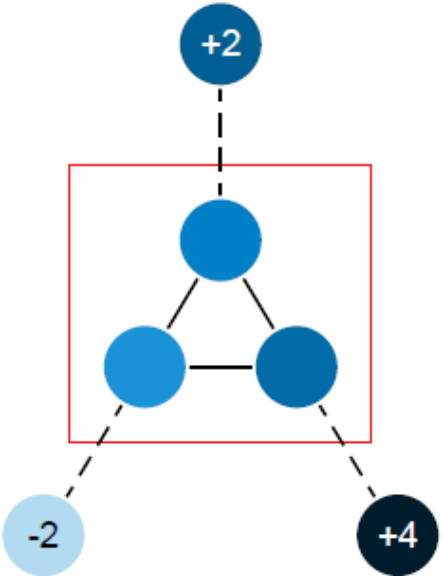
Genotype space / neutral network



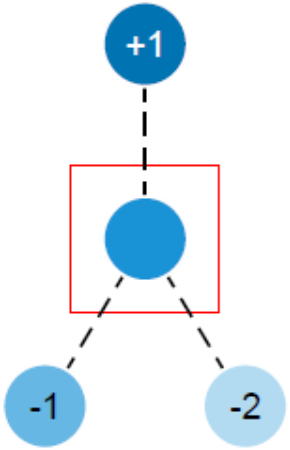
Multiple cryptic loci provide more adaptive options, even in the absence of genetic diversity across population



Two ways to access more novel phenotypes: genetic polymorphism or neighborhood richness

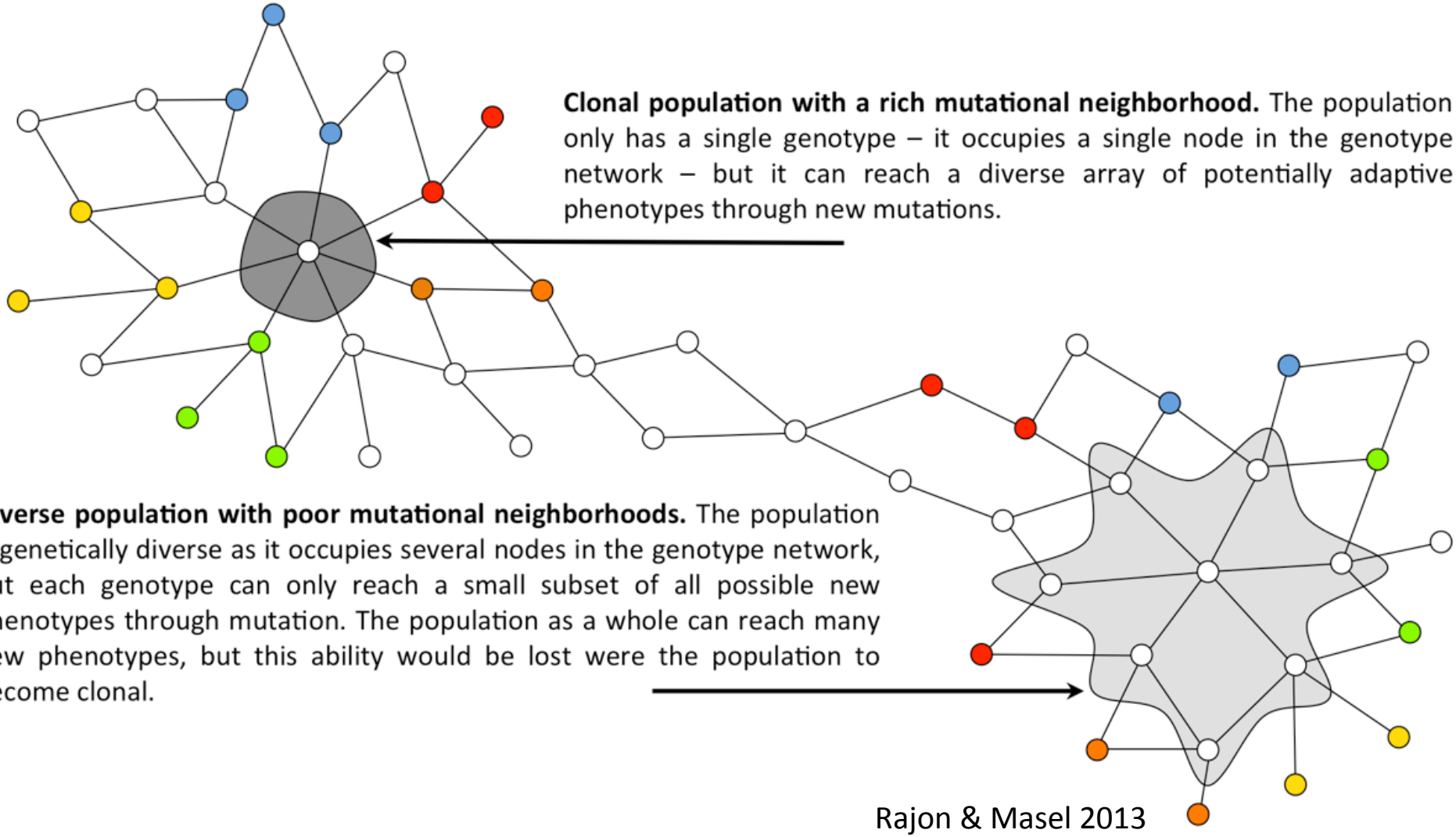


1 locus, 3 genotypes, each accessing one new phenotype



3 loci, 1 genotype can access 3 phenotypes

Two ways to access more novel phenotypes: genetic polymorphism or neighborhood richness

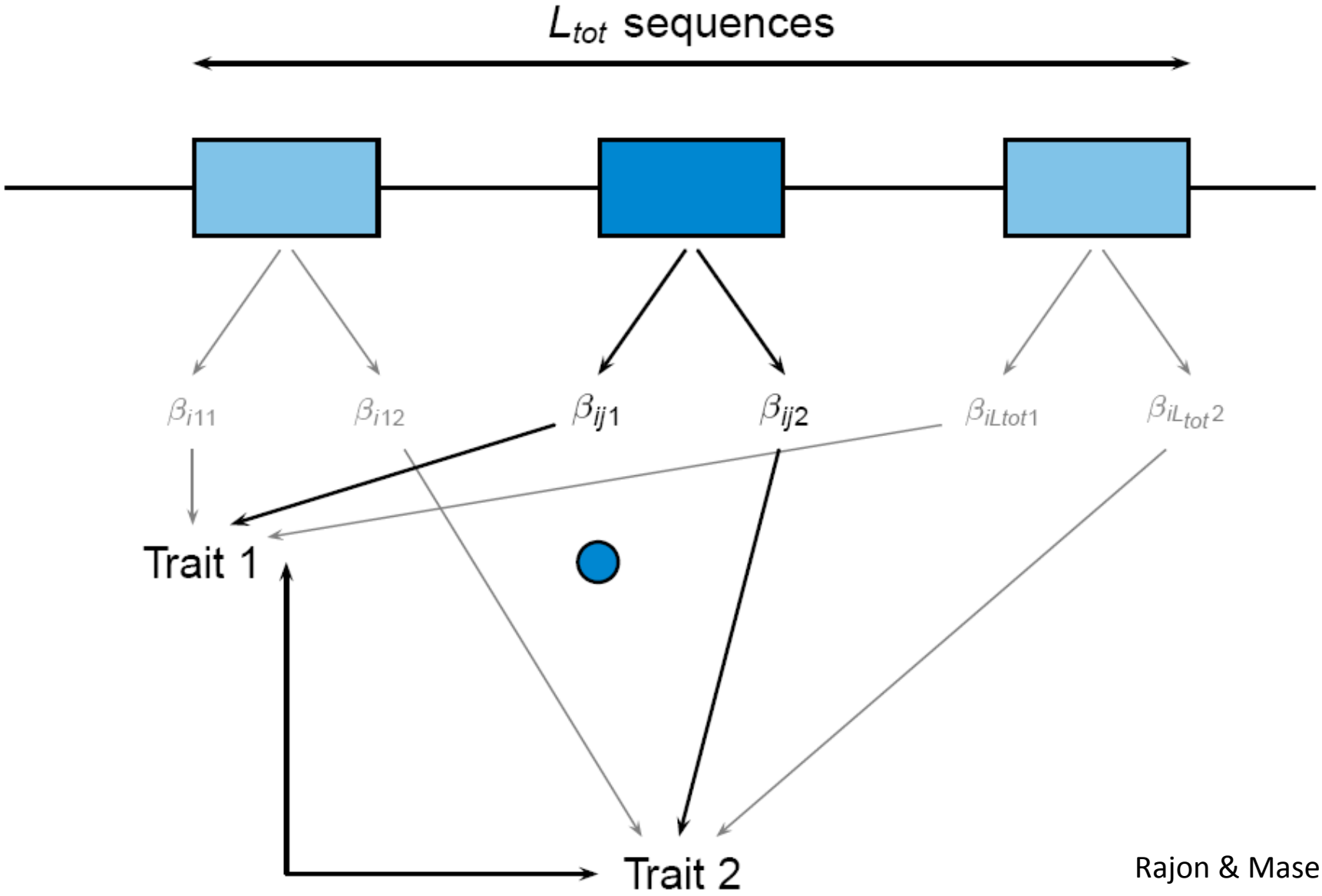


Clonal population with a rich mutational neighborhood. The population only has a single genotype – it occupies a single node in the genotype network – but it can reach a diverse array of potentially adaptive phenotypes through new mutations.

Diverse population with poor mutational neighborhoods. The population is genetically diverse as it occupies several nodes in the genotype network, but each genotype can only reach a small subset of all possible new phenotypes through mutation. The population as a whole can reach many new phenotypes, but this ability would be lost were the population to become clonal.

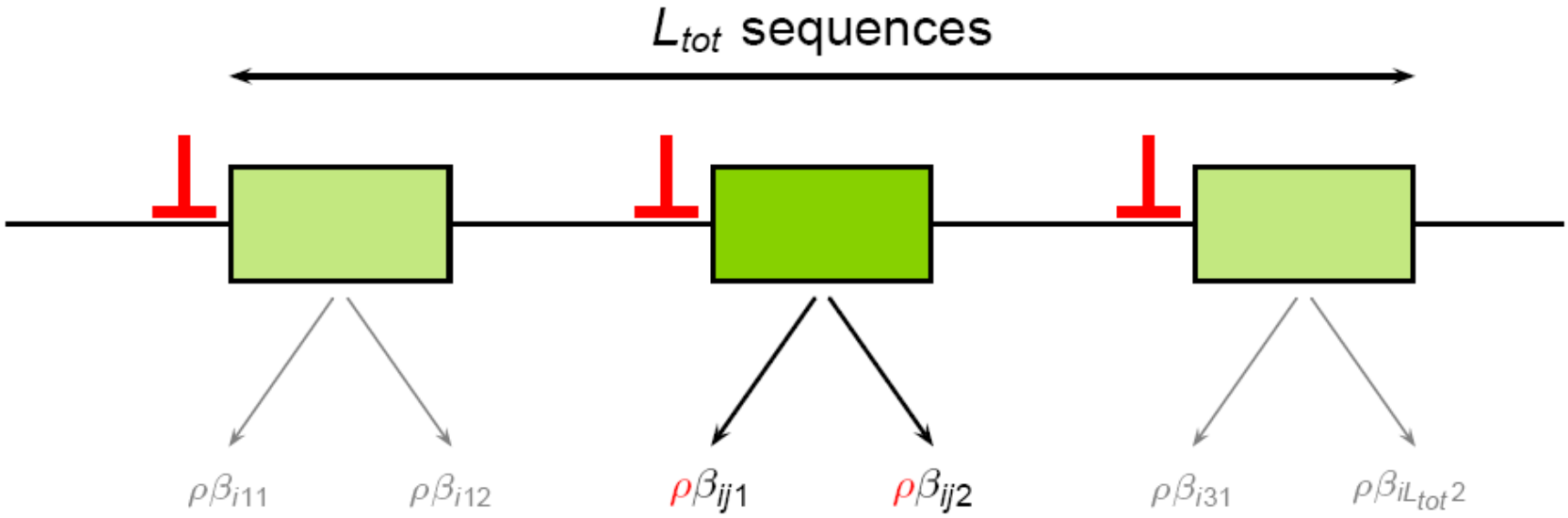
Rajon & Masel 2013

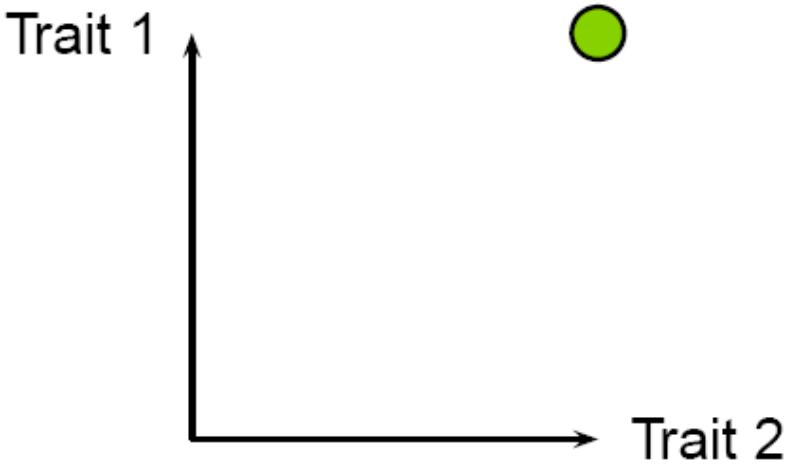
Each cryptic sequence affects multiple traits



Rajon & Masel 2013

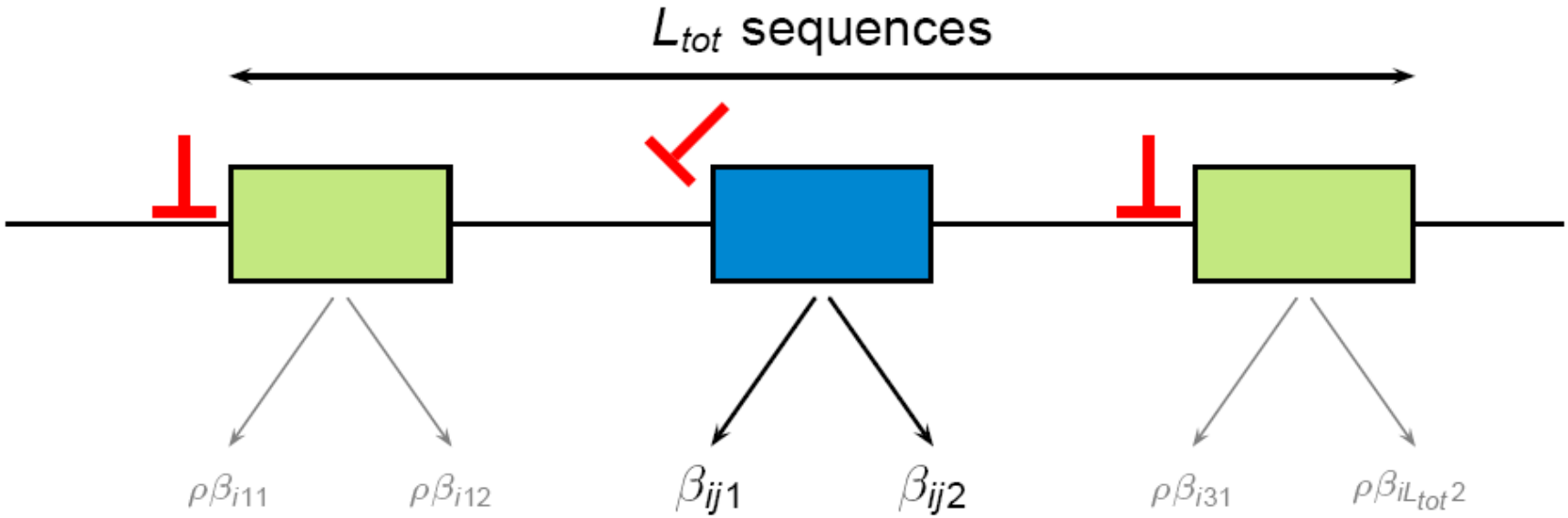
Effects are dampened while cryptic

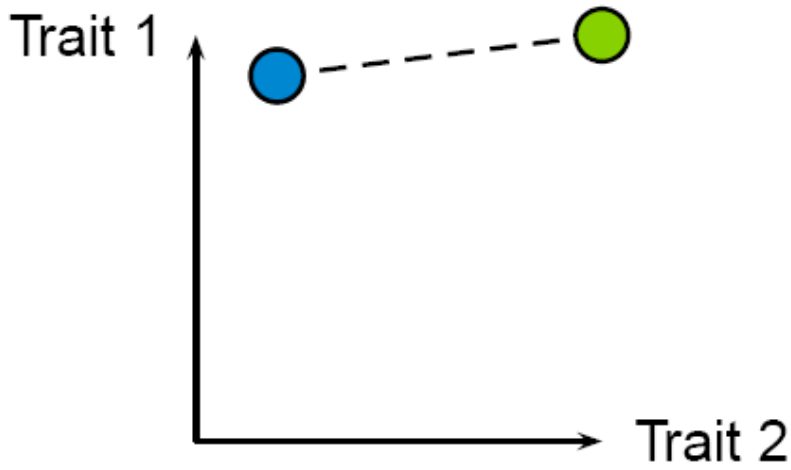




Rajon & Masel 2013

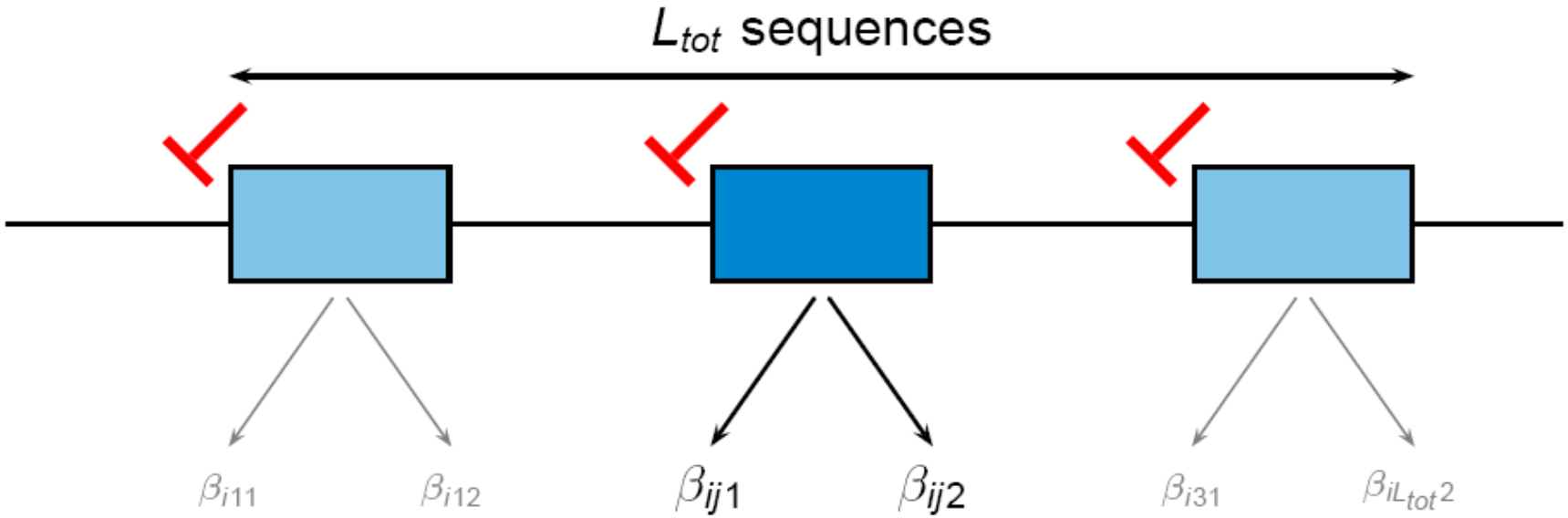
During co-option, crypticity is lost

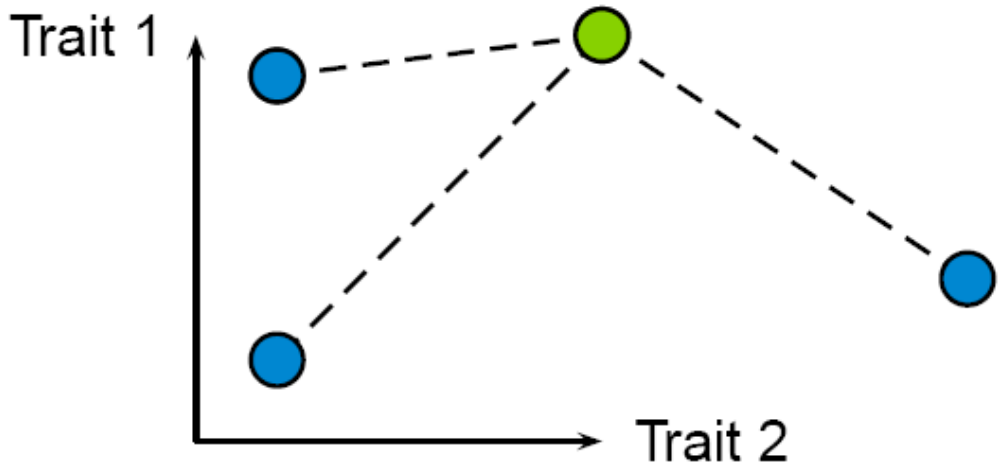




Rajon & Masel 2013

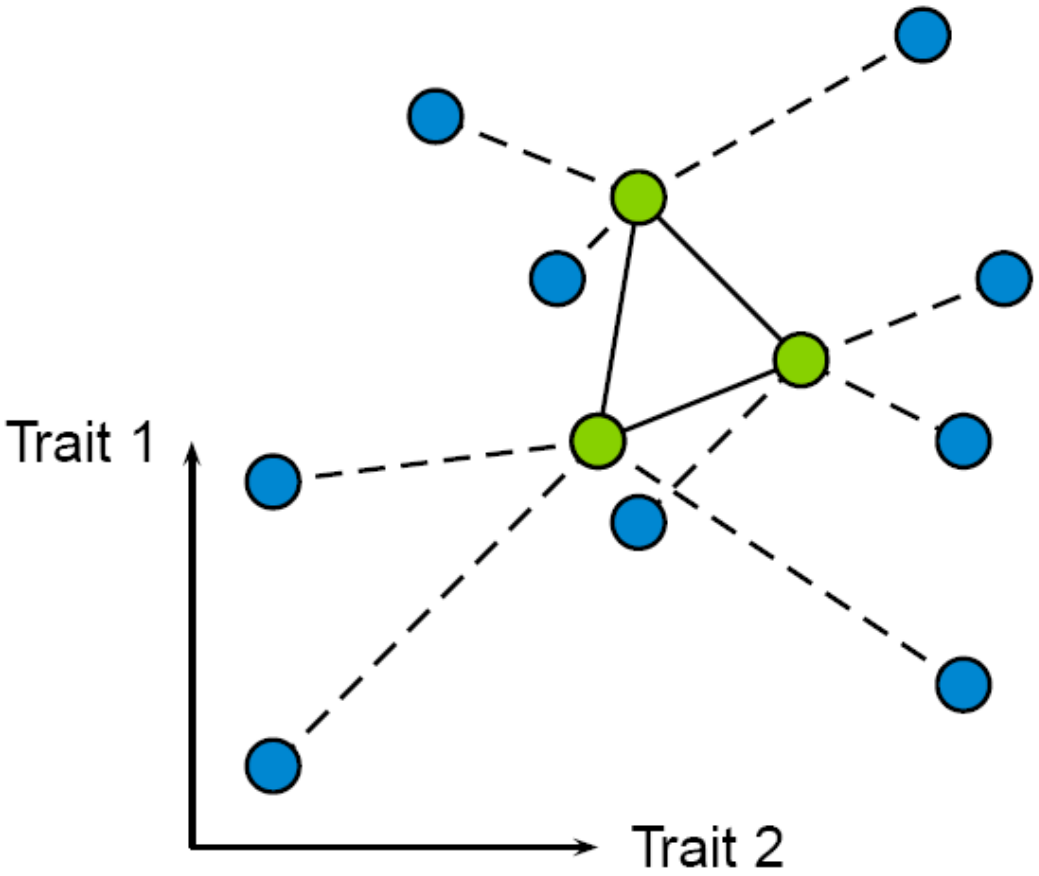
Multiple sequences define neighborhood richness





Rajon & Masel 2013

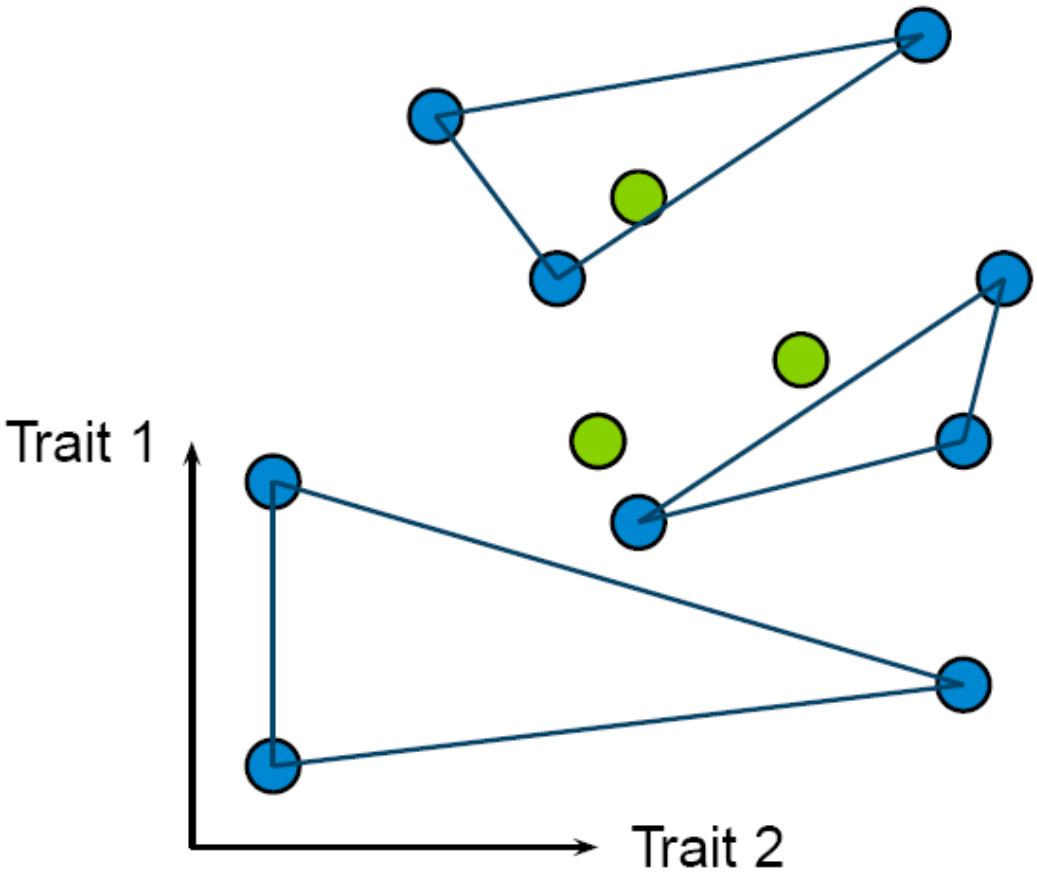
Multiple genotypes increase accessible phenotypes still further



Rajon & Masel 2013

Quantify phenotypic diversity due to neighborhood richness

d_G: mean distance between individuals with the same initial genotype

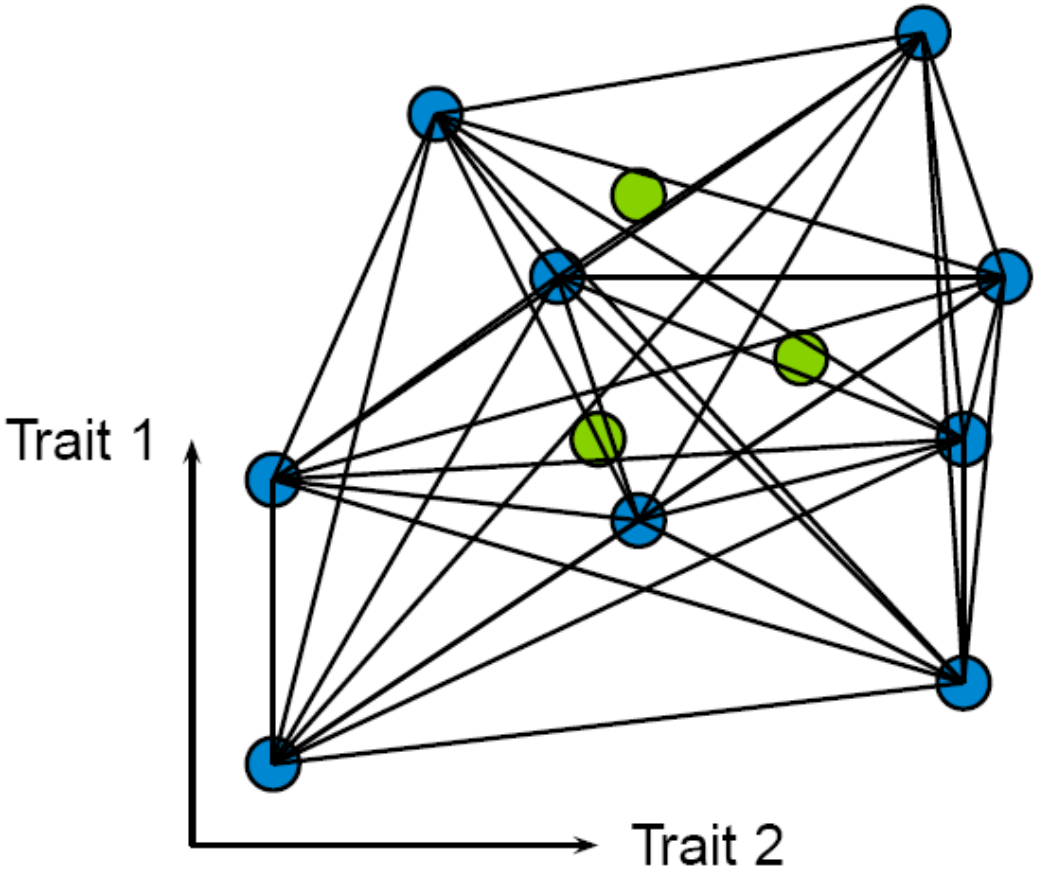


Rajon & Masel 2013

Compare to total phenotypic diversity

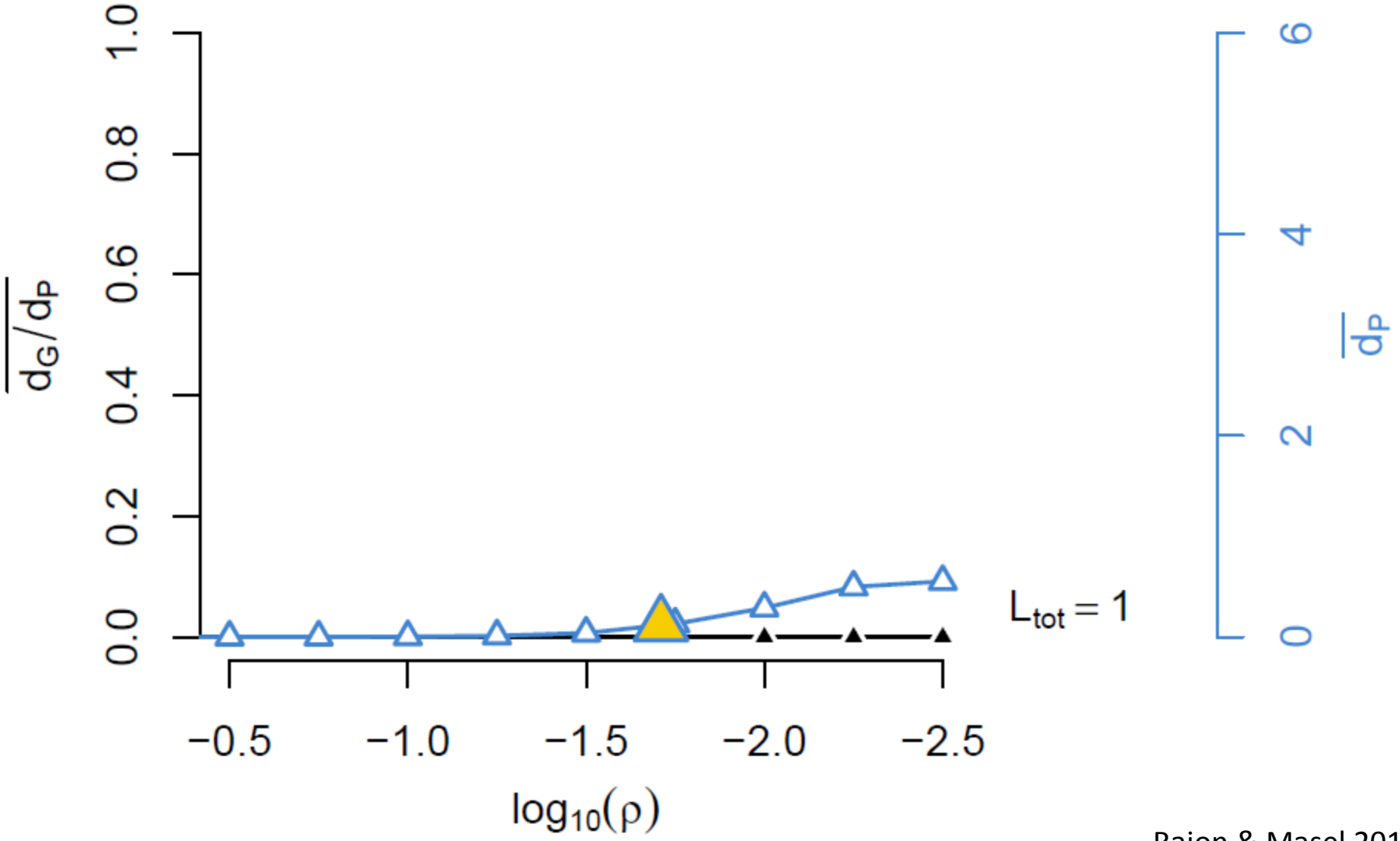
d_G: mean distance between individuals with the same initial genotype

d_P: mean distance between two individuals in the population



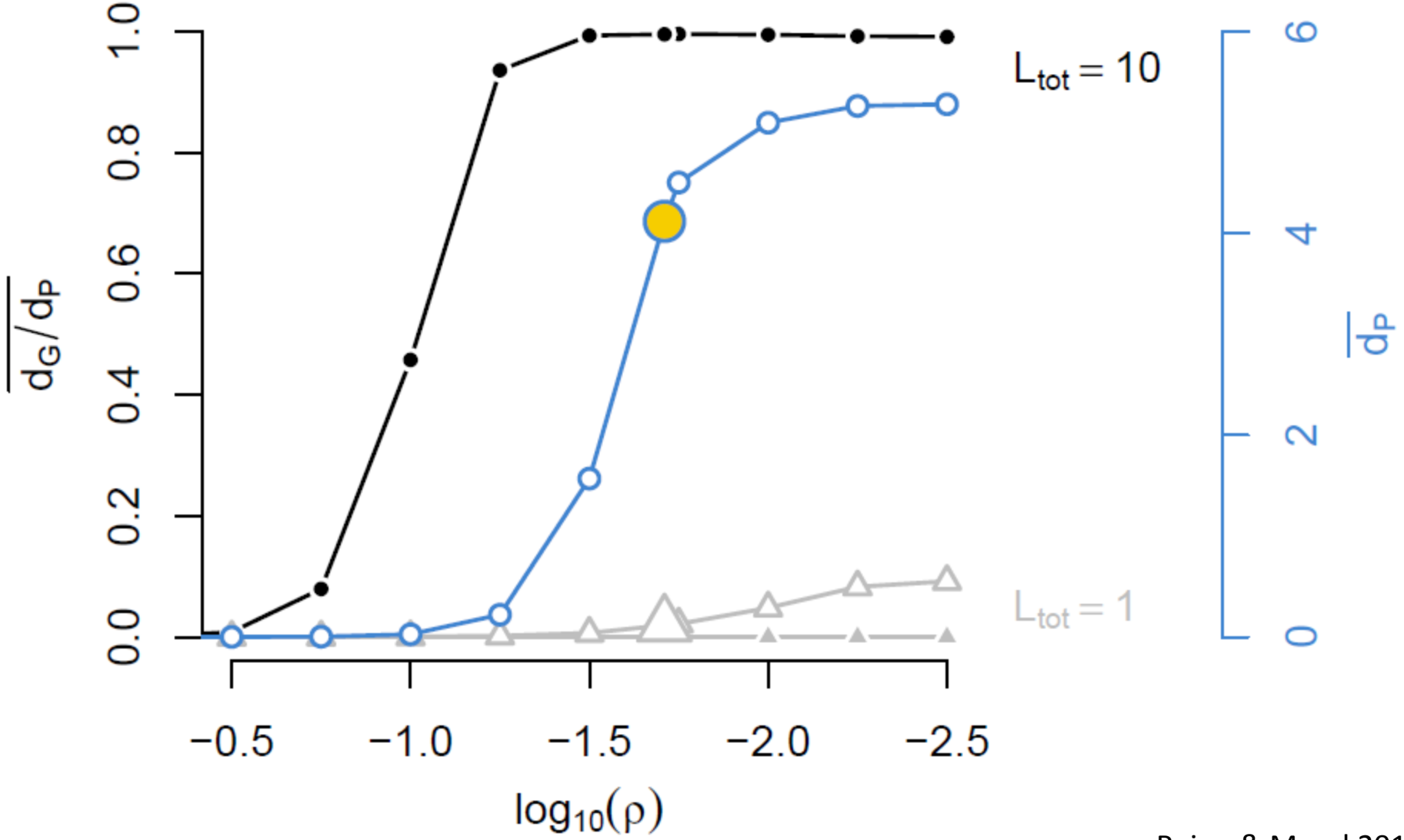
Rajon & Masel 2013

With one locus, all genetic diversity, no neighborhood richness



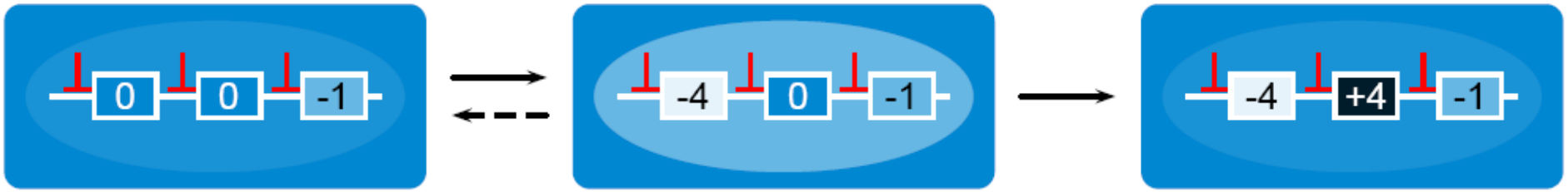
Rajon & Masel 2013

With 10 loci, more phenotypic diversity, dominated by neighborhood richness



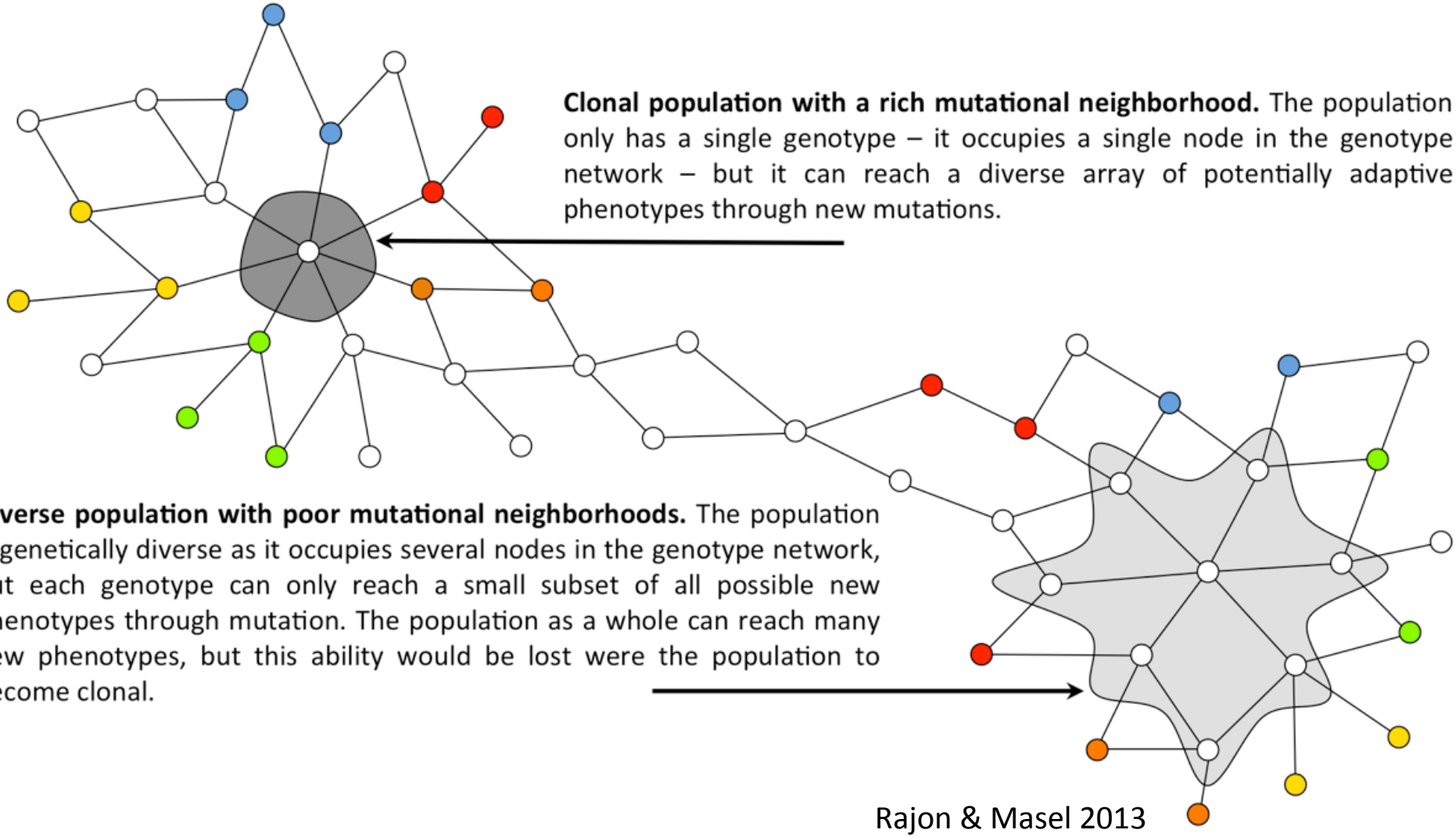
Rajon & Masel 2013

Compensatory evolution drives high neighborhood richness



Rajon & Masel 2013

“Spread” across a genotype space is not required for the high evolvability of polygenic traits in asexuals



Clonal population with a rich mutational neighborhood. The population only has a single genotype – it occupies a single node in the genotype network – but it can reach a diverse array of potentially adaptive phenotypes through new mutations.

Diverse population with poor mutational neighborhoods. The population is genetically diverse as it occupies several nodes in the genotype network, but each genotype can only reach a small subset of all possible new phenotypes through mutation. The population as a whole can reach many new phenotypes, but this ability would be lost were the population to become clonal.

Rajon & Masel 2013

What do we need for evolvability?

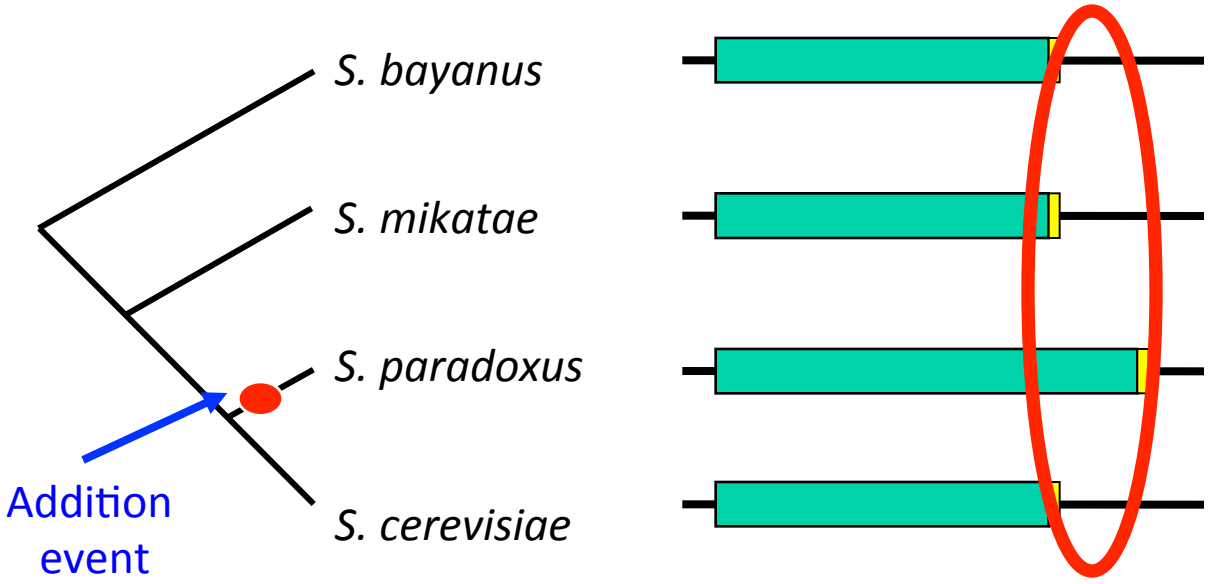
- A minimum level of selection on cryptic sequences, to purge the misfolded options
- Selection as weak as possible above that minimum, to allow maximum compensatory evolution

- This balance is exactly what we get in one attractor of our speed vs. accuracy model!

Outline

1. Evolution of error rates under a
   speed vs. accuracy tradeoff
2. Molecular errors pre-screen future variants,
   and so promote evolvability
3. Genetic polymorphism
   is not required for evolvability
4. Protein coding sequences can evolve de novo
   from pre-screened noncoding sequences

Stop codon readthrough can be coopted for de novo C-terminal pieces of genes



- Conversion of non-coding to coding confirmed by homologous phylogenetic comparisons
  - 75 events in Saccharomyces
  - 67 events in mouse/rat

Giacomelli, Hancock & Masel (2007)

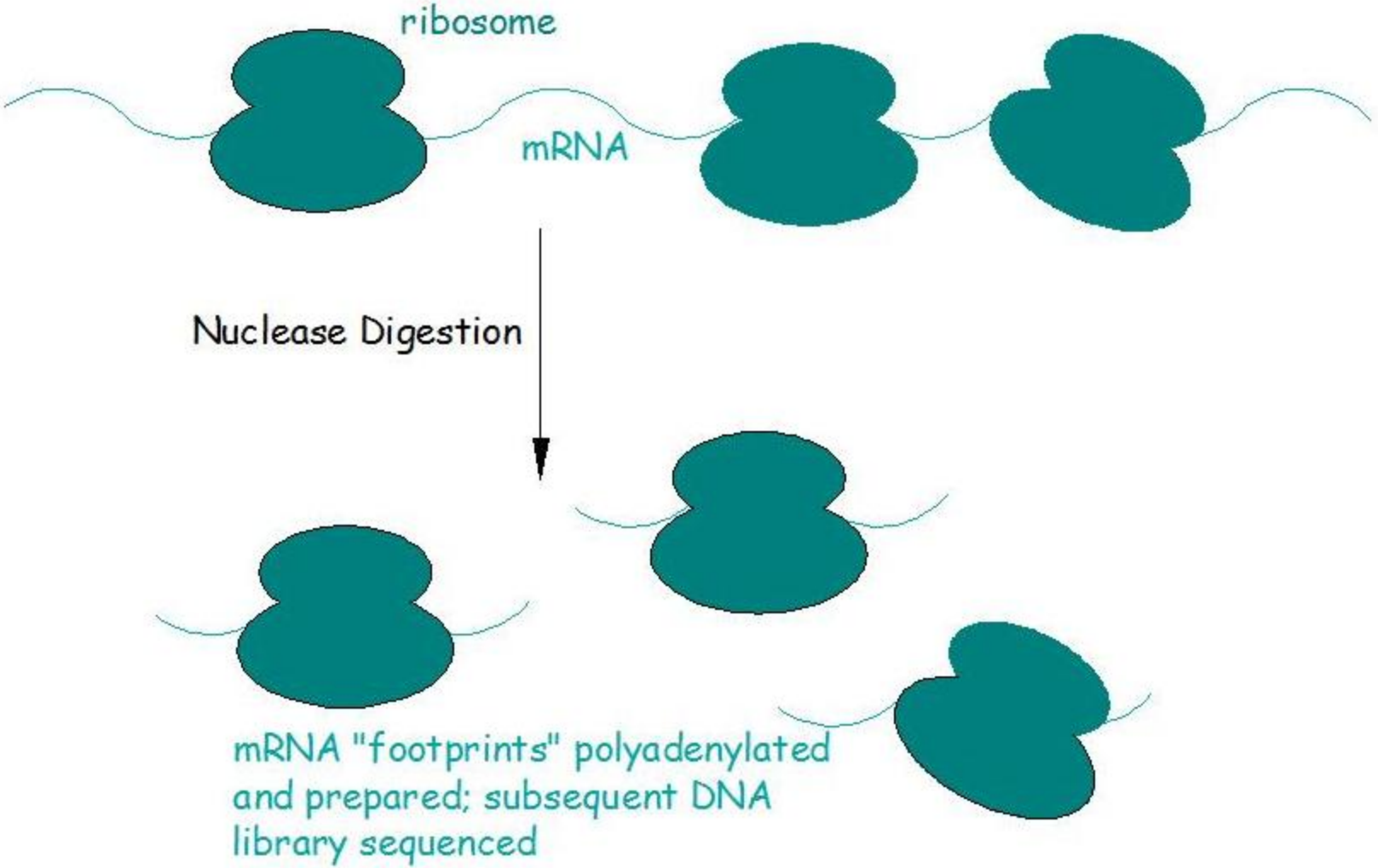
Complete genes evolve *de novo* too. How is this possible?

1. Accidental, low level transcription, transcript rapidly degraded
2. Transcript escapes degradation
3. Transcript occasionally exported to cytoplasm, where it associates with ribosomes and "accidental" ORFs may be translated at low levels
4. New, functional coding gene

Errors at each stage give a "preview" of the next one, allowing pre-adaptation to occur

We tested whether penultimate stage 3 is common

Ribosome Profiling



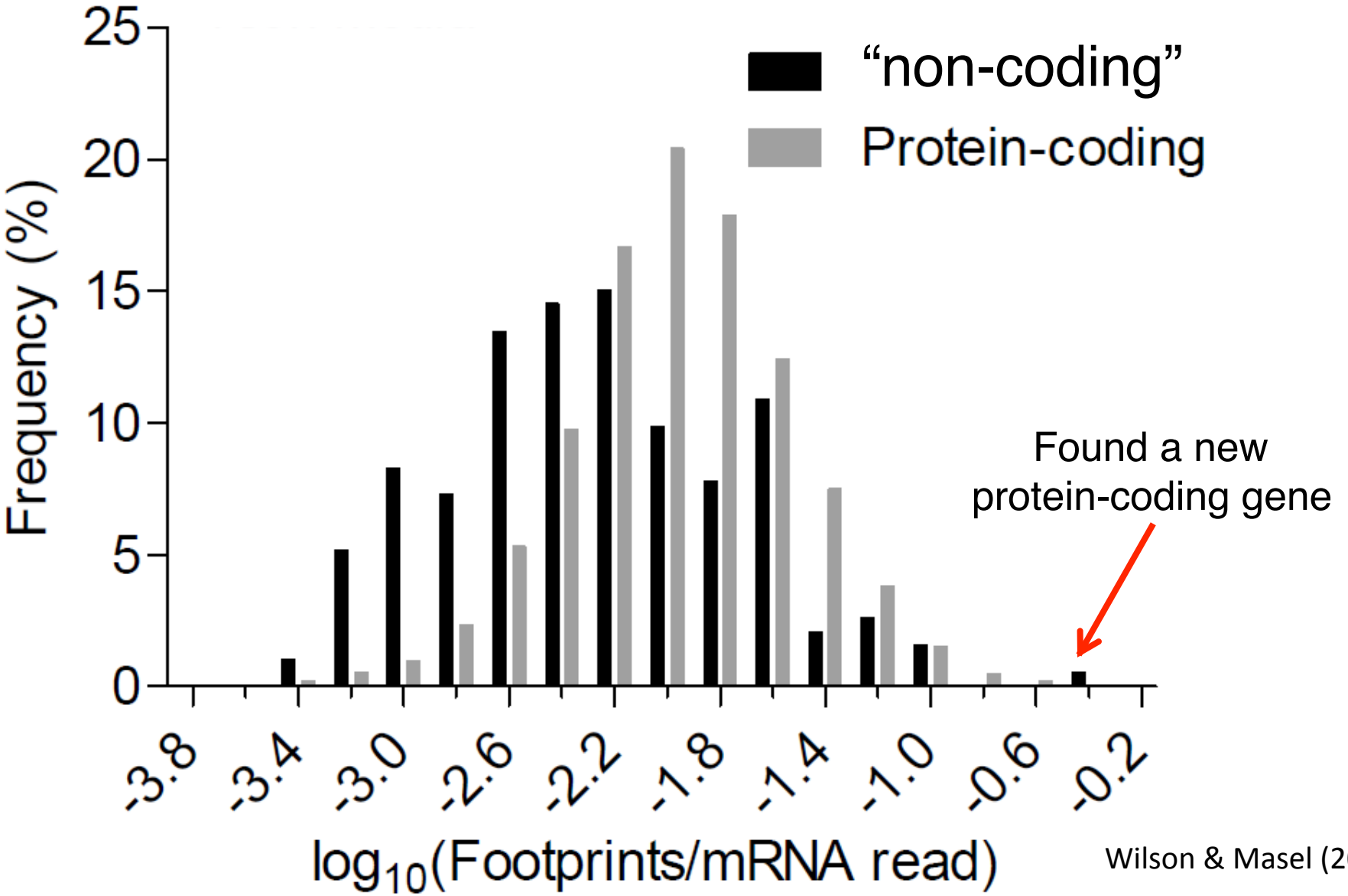
Ingolia et al. 2009

Are “non-coding” transcripts associated with ribosomes?

- Used ribosomal footprints that exactly mapped to unique genome site Ingolia et al. 2009
- 217/404 “non-coding” transcripts showed ribosomal association

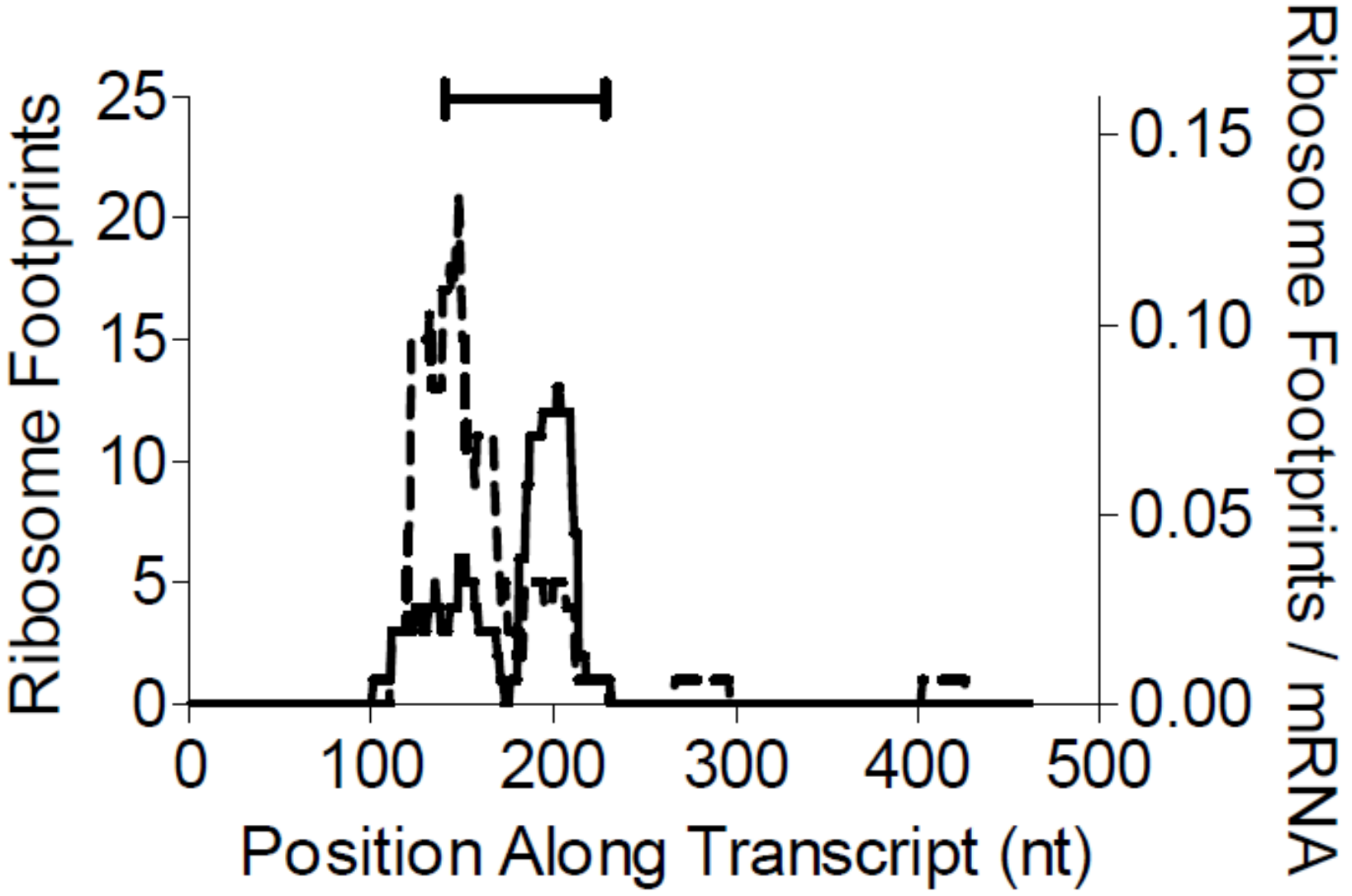
Wilson & Masel 2011

Many individual “non-coding” transcripts have ORF-like ribosome densities



Wilson & Masel (2011)

Ribosomal footprint locations match a 28aa ORF



Wilson & Masel (2011)

Summary of ribosome profiling results

- Looks like a new coding sequence, but we don't know if polypeptide is functional
- Looks like de novo evolution
- Proof of principle of powerful method to annotate short de novo proteins
- Penultimate stage of gene birth is widespread

Wilson & Masel (2011)

Conclusions

- Molecular errors are common and important (eg PPIs)
- 2 solutions to many molecular errors
  - low error rate via a proofreading mechanism for all sites
  - high error rate, but robustness to each separate error
- High error rates pre-screen future variants,
  and so promote evolvability
- With multiple loci,
  genetic diversity is not required for evolvability
- De novo genes may have been prescreened by
  widespread ribosomal association to “non-coding”
  sequences

Broader picture

- Waste and mess and errors are not just a typical biological nuisance
- Without waste and mess, creative evolutionary innovations may not be possible
- Looking for a clean molecular machine can miss the essence of biology

Thanks!

NIH
Pew Charitable Trusts
John Templeton Foundation

Etienne Rajon

Ben Wilson

Mike Giacomelli

Leandra Brettner

Now hiring postdocs