

www.gephebase.org

GepheBase

The Database of Evolutionary Genotype-Phenotype Relationships



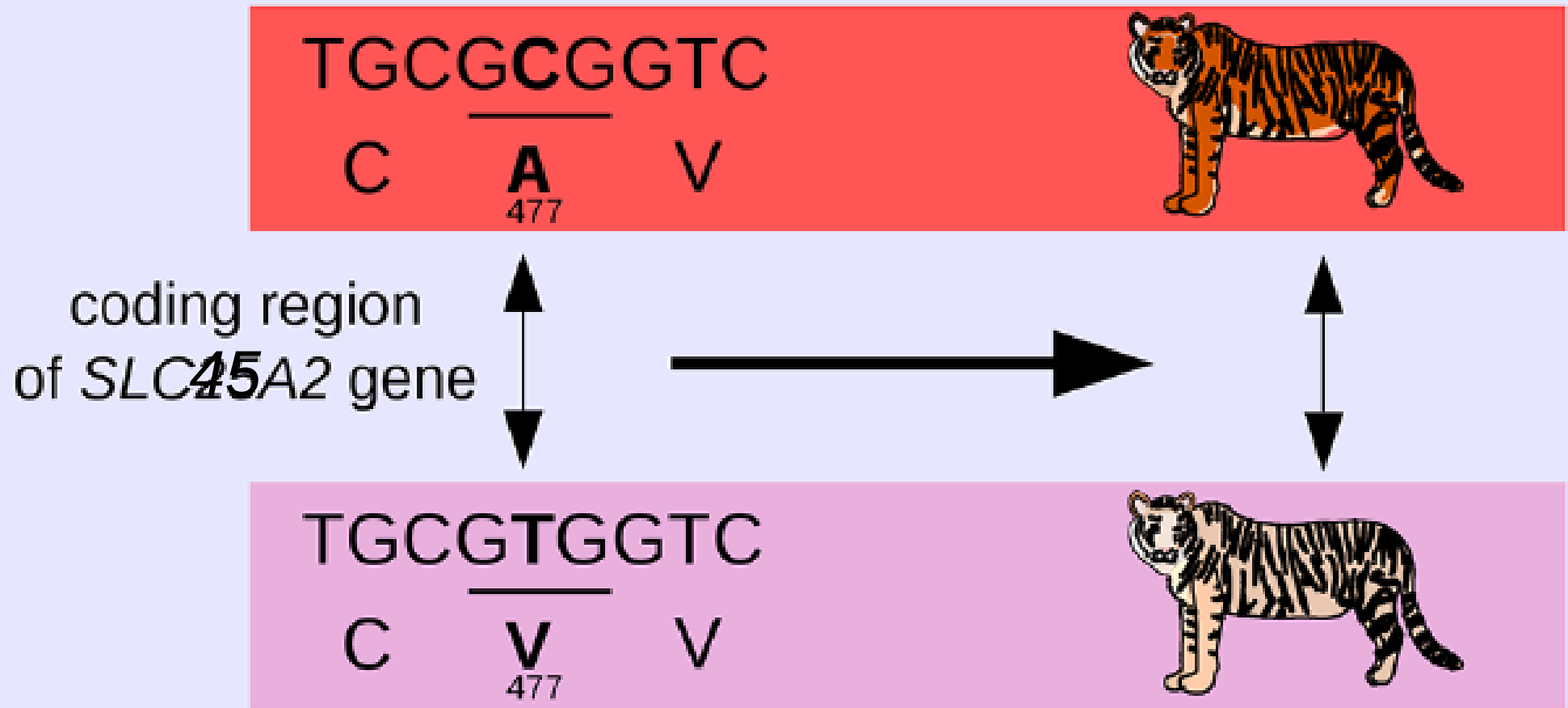
Includes Natural, Domesticated and Experimental Variation
but **NO LAB MUTANTS** and **NO CLINICAL TRAITS**

>2000 genes and mutations
associated with
natural phenotypic changes
in animals and plants

- What is Gephebase?
- Main findings so far using Gephebase
- Your work with Gephebase


Ge-phe

a Genetic **VARIATION** *causing*
a Phenotypic **VARIATION**



CHKov1

GePheBase



Home

Power

keyword

Search

Zoom

Help

Suggest an Article

The Database of Genotype-Phenotype Relationships

Search Gephebase for genes, phenotypes, taxa, mutations, articles

Most Visited

Les plus visités

Getting Started

Latest Headlines

PubMed

BiblioVIE

GoogScholar

FlyBase

DicoCNRS

BLAST

IJM-WebMail

GEPHE SUMMARY

Gephebase Gene	CHKov1	GepheID	GP00000182
Entry Status	-	Main curator	Martin

PHENOTYPIC CHANGE

Show All Details

Trait #1	
Trait Category	Physiology
Trait	Pathogen resistance
Trait State in Taxon A	Drosophila melanogaster - susceptible
Trait State in Taxon B	Drosophila melanogaster - resistant

Trait #2	
Trait Category	Physiology
Trait	Xenobiotic resistance (insecticide)
Trait State in Taxon A	-
Trait State in Taxon B	-

GO - Cellular Component

UniProtKB

AY047531

Main-Reference -

To understand the molecular basis of how hosts evolve resistance to their parasites, we have investigated the genes that cause variation in the susceptibility of *Drosophila melanogaster* to viral infection. Using a host-specific pathogen of *D. melanogaster* called the sigma virus (Rhabdoviridae), we mapped a major-effect polymorphism to a region containing two paralogous genes called *CHKov1* and *CHKov2*. In a panel of inbred fly lines, we found that a transposable element insertion in the protein coding sequence of *CHKov1* is associated with increased resistance to infection. Previous research has shown that this insertion results in a

RELATED GEPHE

Related Genes

15 (18-wheeler, Dipteracin, GGBP1, GGBP2, Immune deficiency, pastrel, PGRP-LC, SR-CII, Tehao, Acetylcholinesterase (Ace-2), alcohol dehydrogenase (Adh), Cyp12d1, cyp6g1, GSTE1-E10 cluster, resistance to dieldrin)

Related Haplotypes

No matches found.

EXTERNAL LINKS

FLYBASE

A database of drosophila genes and genomes.

[FBal0190391](#)

DFAM

Open database of collection of DNA Transposable Element sequence alignments, hidden Markov Models (HMMs), consensus sequences, and genome annotations.

[DF0001587](#)

COMMENTS

@TE

ADVANCED SEARCH

Field

Trait Category

Molecular Type

Comments

Aberration Type

Term

Morphology

Cis-regulatory

@TE

Deletion



+ Add search criteria

- ☐ Split Mutations
- ☐ Group Haplotypes
- ☐ Group Genes

Submit

Your research retrieved 3 results.

Show All Additional References

Show 100 entries

Select All	Gene	Trait	Mutation	Taxon B	Taxonomic Status	Evidence	Main Reference
<input type="checkbox"/>	Kit (type III receptor protein-tyrosine kinase)	Coloration (coat)	Cis-regulatory, Deletion	<i>Felis catus</i> domestic cat - (species) D	Domesticated	Association Mapping	David VA; Menotti-Raymond M; Wallace AC ; et al. (2014) Endogenous retrovirus insertion in the KIT oncogene determines white and white spotting in domestic [...]
<input type="checkbox"/>	VvMYBA1	Coloration (fruit)	Cis-regulatory, Deletion	<i>Vitis vinifera</i> wine grape - (species) D	Domesticated	Candidate Gene	Kobayashi S; Goto-Yamamoto N; Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. 1 Additional References
<input type="checkbox"/>	VvMYBA1	Coloration (fruit)	Cis-regulatory, Deletion	<i>Vitis vinifera</i> wine grape - (species) D	Domesticated	Candidate Gene	Ferreira V; Matus JT; Pinto-Carnide O ; et al. (2019) Genetic analysis of a white-to-red berry skin color reversion and its transcriptomic and metabolic [...]

Presumptive Null	No
Molecular Type	Cis-regulatory
Aberration Type	Deletion
Deletion Size	1-10 kb
Molecular Details of the Mutation	Excision of the full-length FERV1 element leaving the two LTR residues
Experimental Evidence	Association Mapping
Main Reference	Endogenous retrovirus insertion in the KIT oncogene determines white and white spotting in domestic cats. (2014)
Authors	David VA; Menotti-Raymond M; Wallace AC; Roelke M; Kehler J; Leighty R; Eizirik E; Hannah SS; et al. ... show more
Abstract	The Dominant White locus (W) in the domestic cat demonstrates pleiotropic effects exhibiting complete penetrance for absence of coat pigmentation and incomplete penetrance for deafness and iris hypopigmentation. We performed linkage analysis using a pedigree segregating White to identify KIT (Chr. B1) as the feline W locus. Segregation and sequence analysis of the KIT gene in two pedigrees (P1 and P2) revealed the remarkable retrotransposition and evolution of a feline endogenous retrovirus (FERV1) as responsible for two distinct phenotypes of the W locus, Dominant White, and white spotting. A full-length (7125 bp) FERV1 element is associated with white spotting, whereas a FERV1 long terminal repeat (LTR) is associated with all Dominant White individuals. For purposes of statistical analysis, the alternatives of wild-type sequence, FERV1 element, and LTR-only define a triallelic marker. Taking into account pedigree relationships, deafness is genetically linked and associated with this marker; estimated P values for association are in the range of 0.007 to 0.10. The retrotransposition interrupts a DNAase I hypersensitive site in KIT intron 1 that is highly conserved across mammals and was previously demonstrated to regulate temporal and tissue-specific expression of KIT in murine hematopoietic and melanocytic cells. A large-population genetic survey of cats (n = 270), representing 30 cat breeds, supports our findings and demonstrates statistical significance of the FERV1 LTR and full-length element with Dominant White/blue iris (P < 0.0001) and white spotting (P < 0.0001), respectively.

RELATED GEPHE

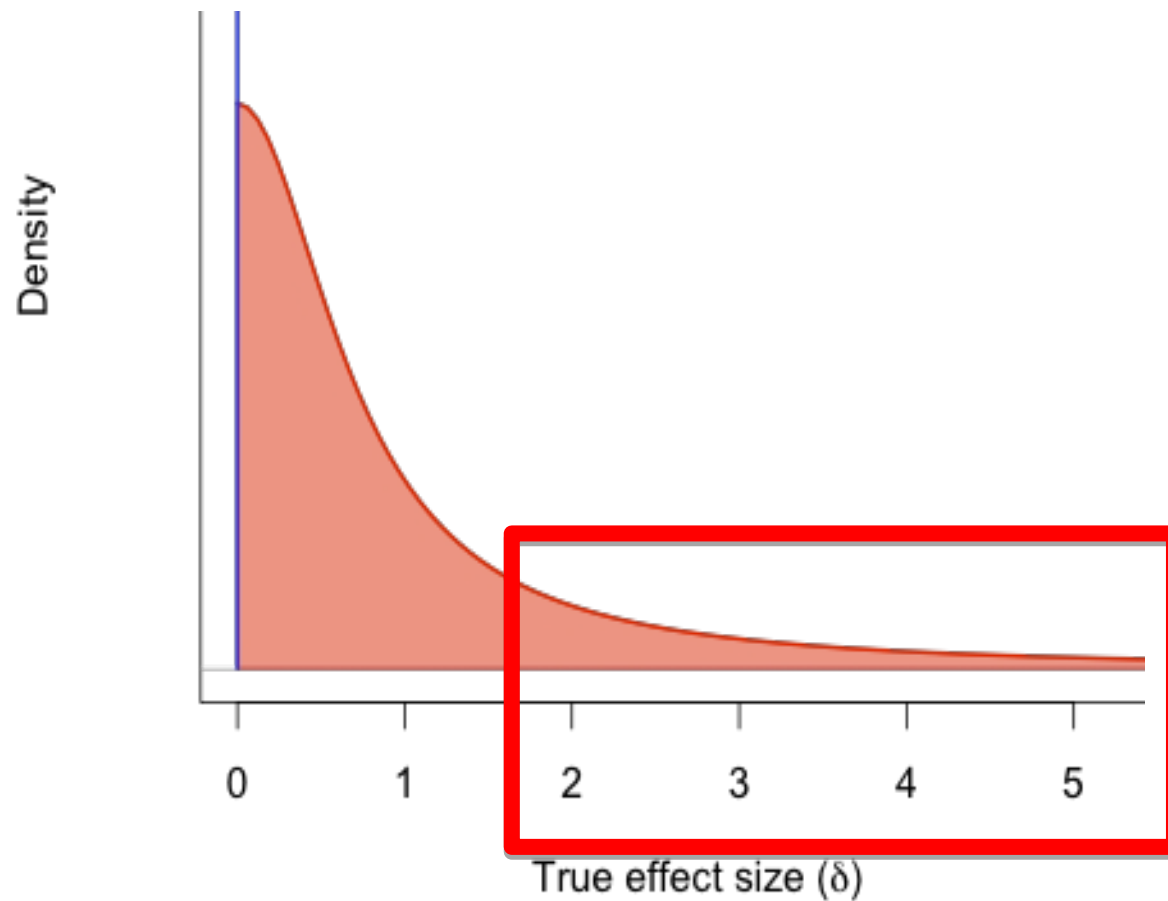
Related Genes	6 (Agouti, MC1R, Melanophilin (MLPH), Taqpep, tyrosinase (TYR), tyrosinase-related protein 1 (TYRP1))	Related Haplotypes	2
---------------	---	--------------------	---

COMMENTS

@Pleiotropy @TE; Affects a conserved melanocyte enhancer

THE QTN PROGRAM AND THE ALLELES THAT MATTER FOR EVOLUTION: ALL THAT'S GOLD DOES NOT GLITTER

Matthew V. Rockman^{1,2}



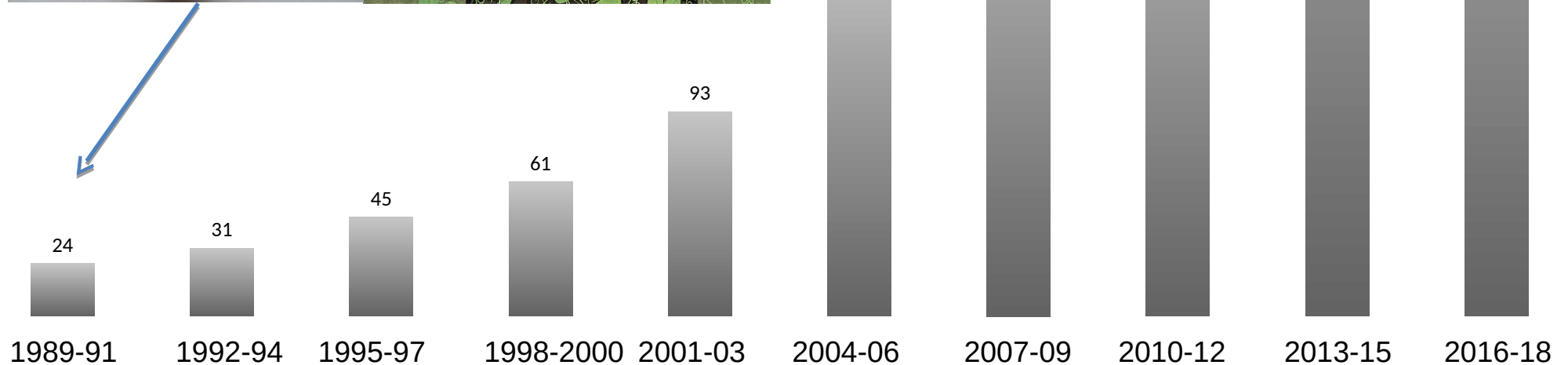
The shiny
“**large effect**” loci we can
document experimentally

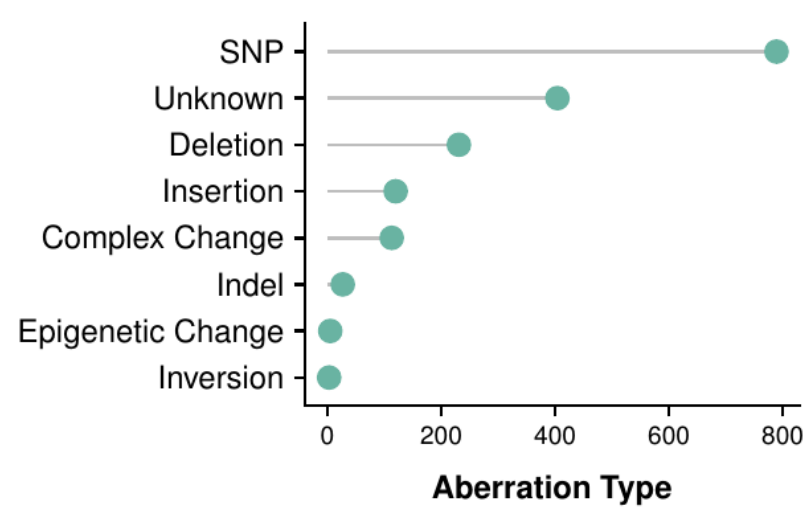
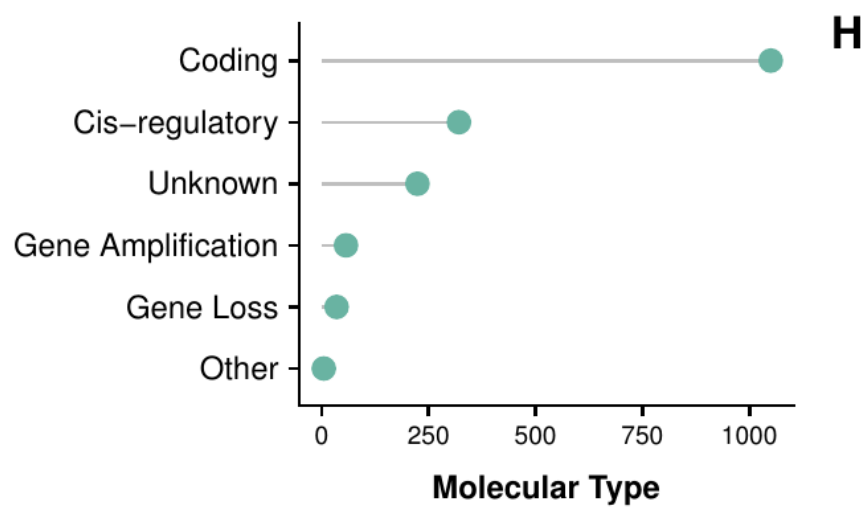
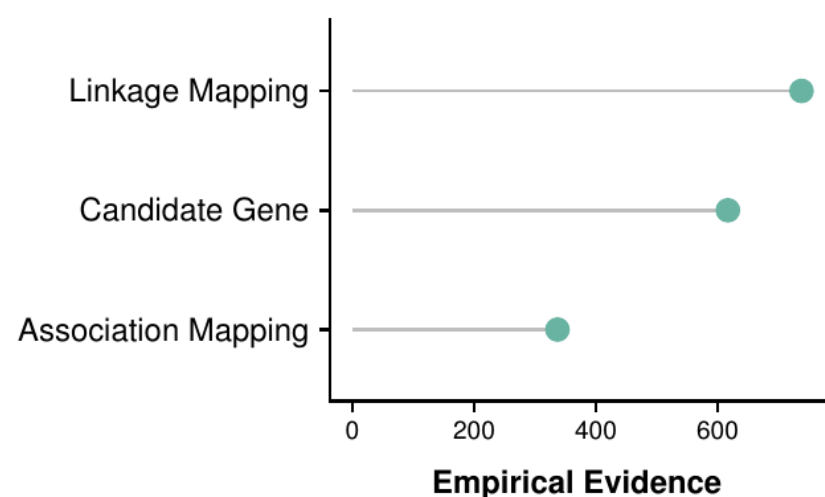
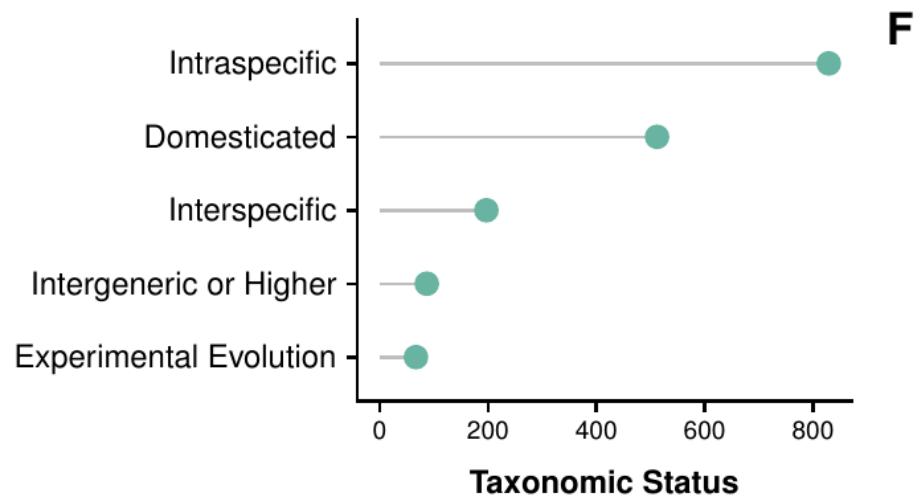
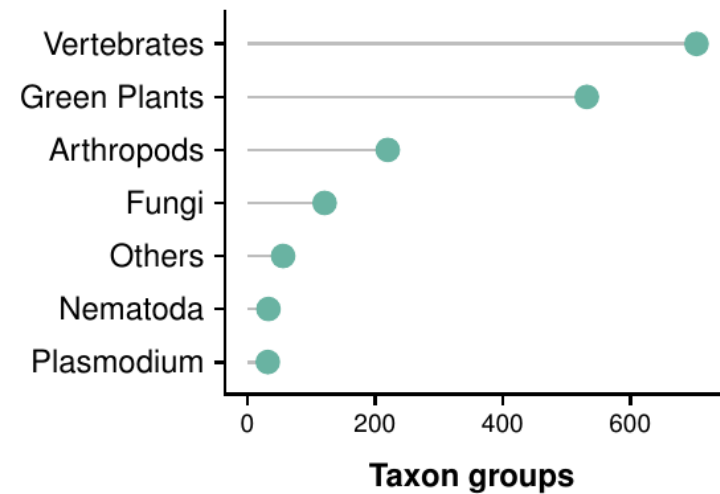
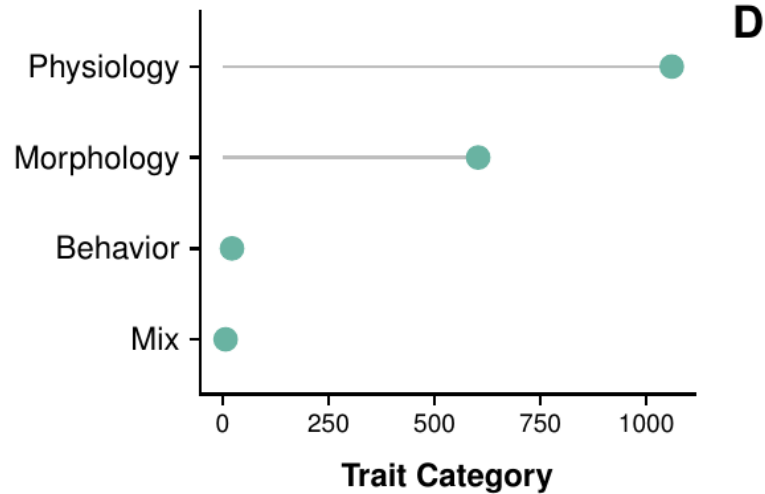




Since 2015:
hard to keep up
with all the
publications...

The Wrinkled-Seed Character of Pea Described by Mendel Is Caused by a Transposon-like Insertion in a Gene Encoding Starch-Branching Enzyme





Experimental Evidence

3 categories, each with biases



Candidate Gene

Reverse Genetics:

looking for sequence differences and trait effects based on previous studies of a given gene



Linkage Mapping

Forward Genetics:

trait mapping in hybrids obtained from laboratory crosses, using recombination over a few generations



Association Mapping

Forward Genetics:

statistical SNP/character state association in large cohorts, using recombination over many generations

Experimental Principle

Example

66 cases of color variation associated to *MC1R* coding mutations in vertebrates

F2 crosses between melanic and amelanic phenotypes in cavefish : identification of *MC1R* and *Oca2* alleles in distinct cave populations

GWAS of human pigmentation (skin, hair, eyes): identification and confirmation of causal variants at >15 genes including *Oca2* p.His615Arg in Eastern Asia

Ascertainment Bias on Locus Identification

High

Low to Intermediate
(depending on resolution / cross size)

Low

Molecular Type Bias

Favors identification of **coding mutations**

Little molecular bias

Can miss structural variants (short read genotyping)

Trait Type Bias

Favors traits with small molecular targets, large-effect size

Amenable to dissection of complex traits with small-effect size (large crosses, multiparental families)

Most common approach for complex traits with small-effect size

Taxonomic Breadth

Large

Narrow, limited to interfertile lineages (populations or sister species)

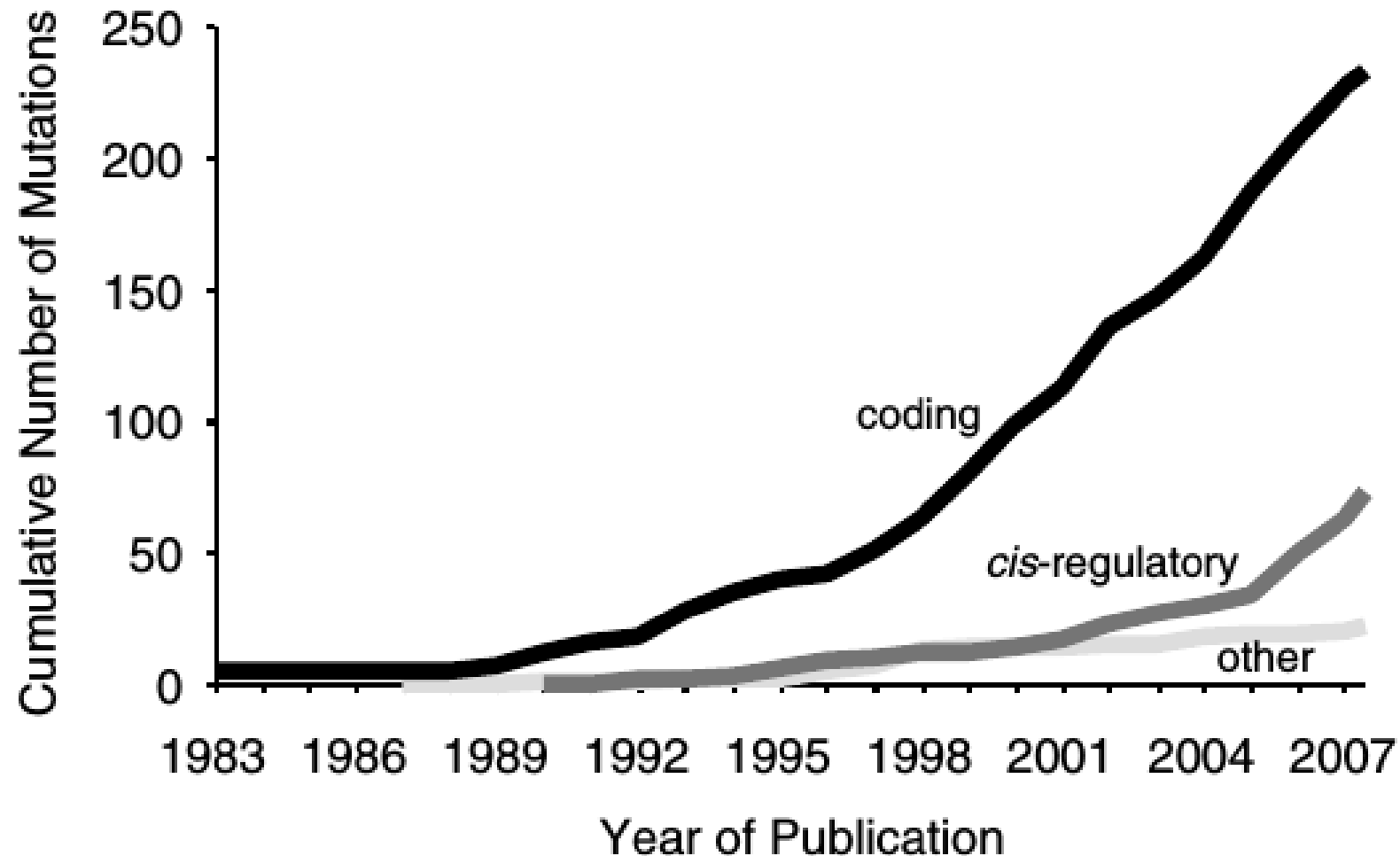
Very narrow, limited to polymorphic or intermixing populations

Gephebase can be used in various ways

- as a powerful bibliographic tool
- as a place to formulate hypotheses
- as a list of potential targets for breeders interested in transferring traits of interest to new species
- as an extensive compilation for broad meta-analyses on the genetic loci of evolution
- as a resource for epistemologists interested in biases and sociological aspects in the field of genetic evolution

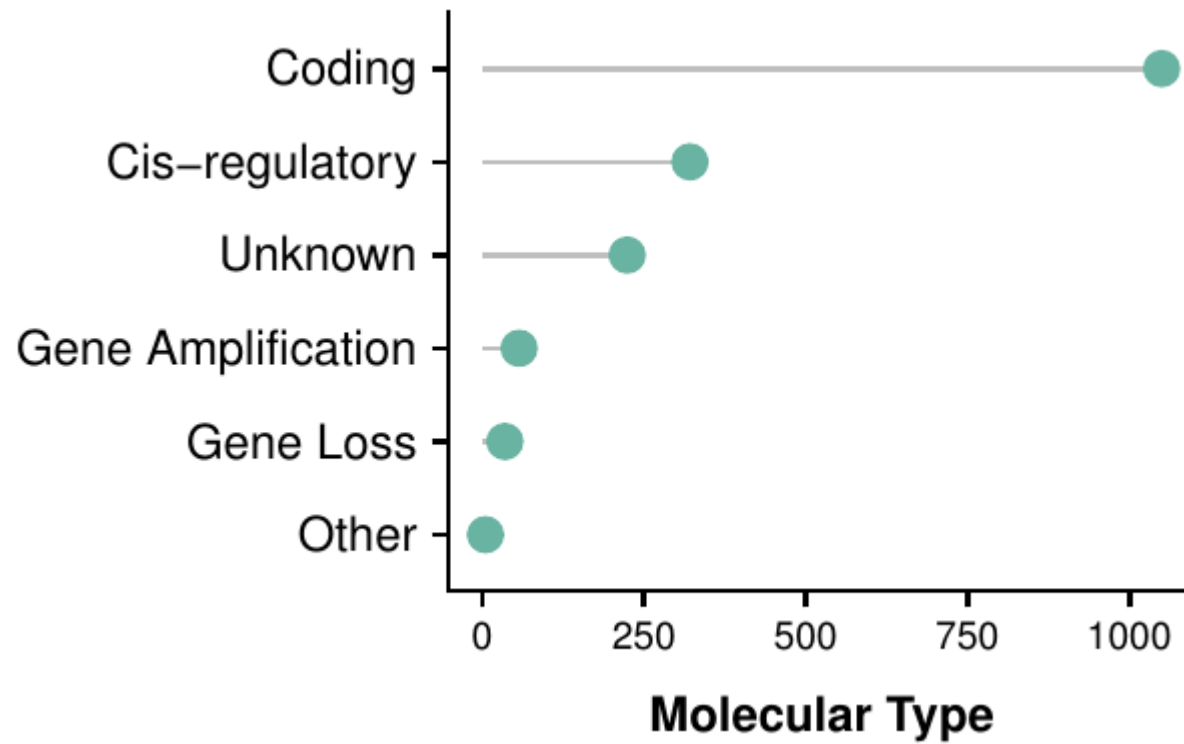
- What is Gephebase?
- Main findings so far using Gephebase
- Your work with Gephebase

More known cases of coding than cis-regulatory mutations

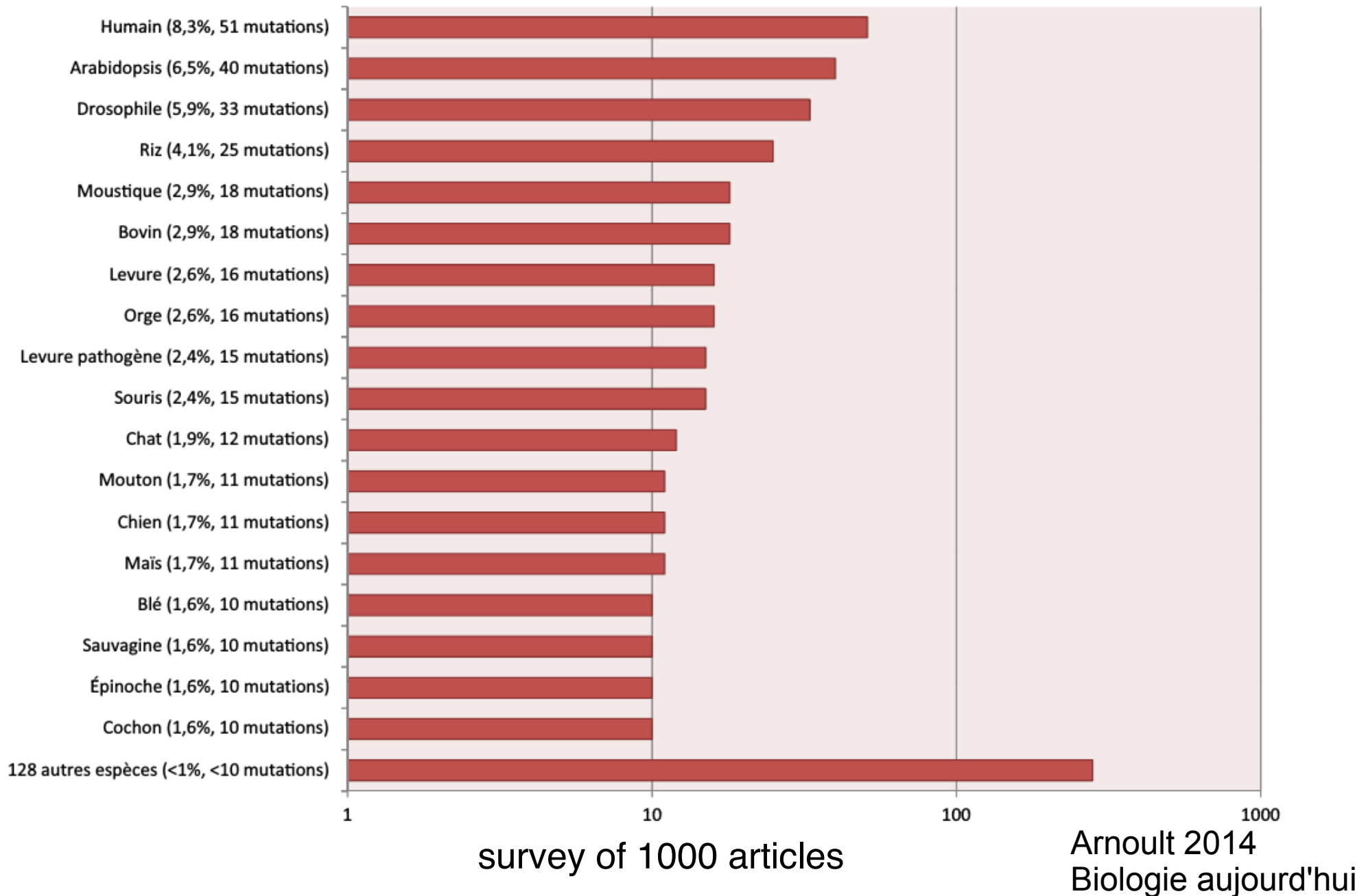


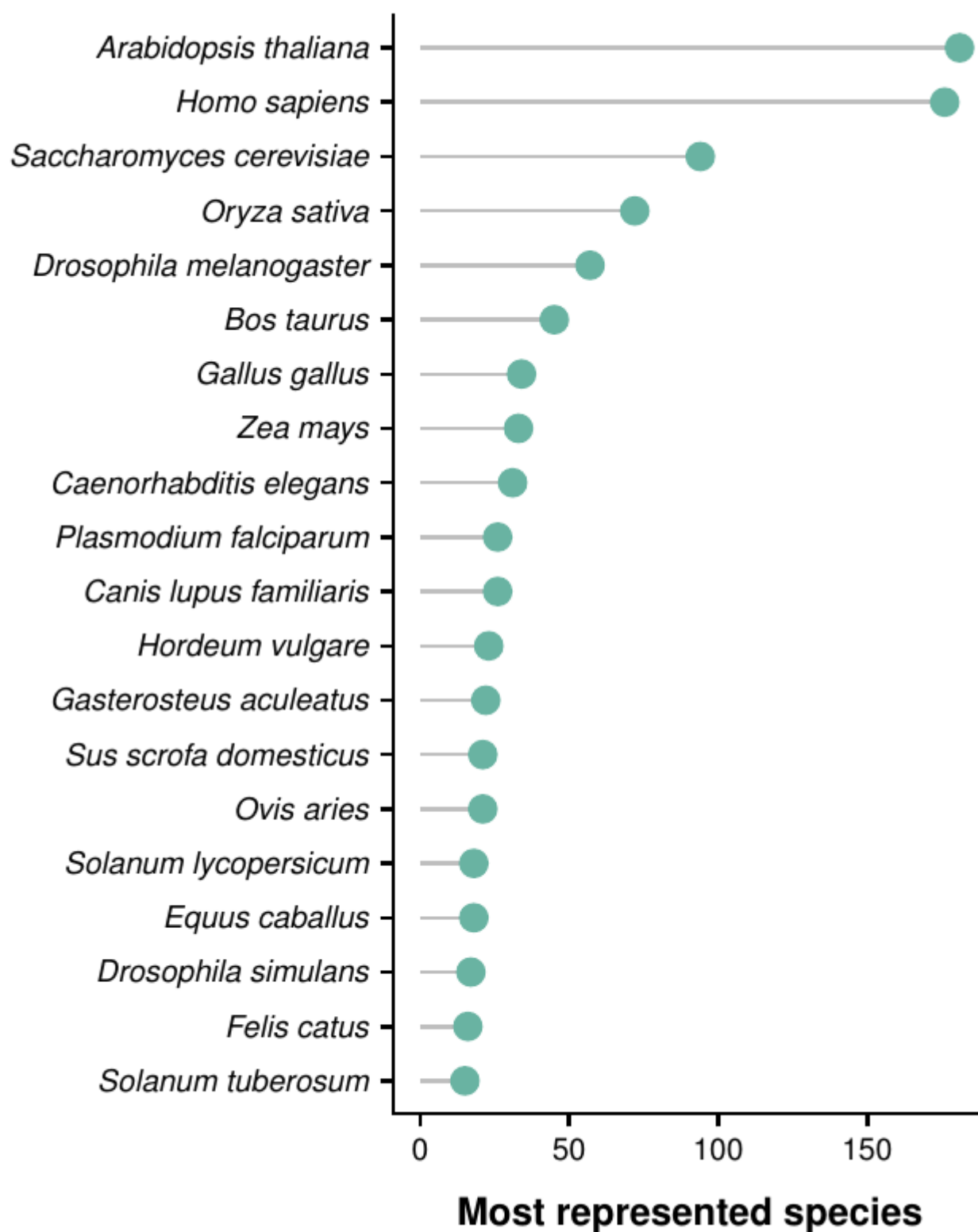
2008: survey of ~300 articles

Stern and Orgogozo 2008 Evolution



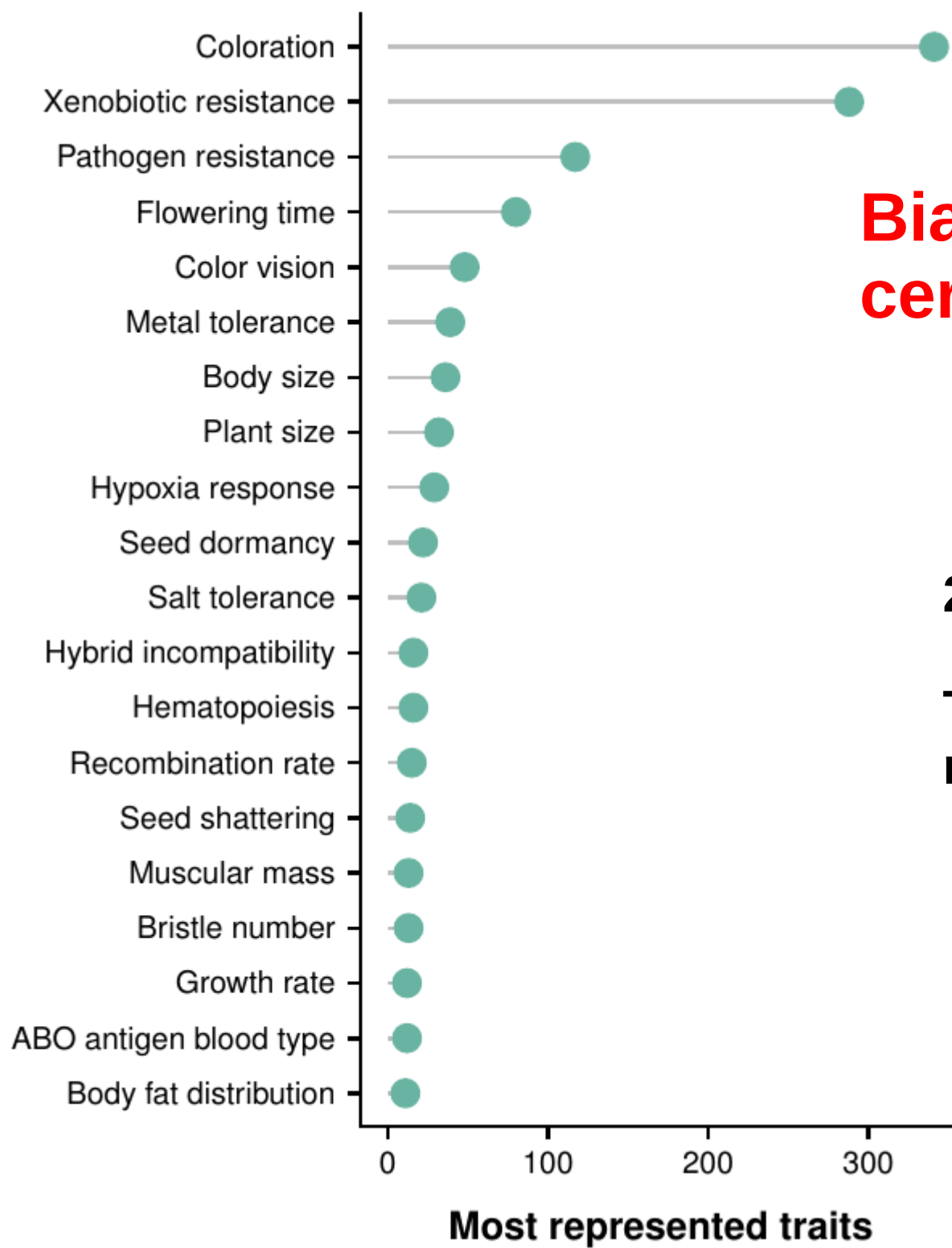
Bias towards certain species





327 species in total

**The other species
represent 752 entries**

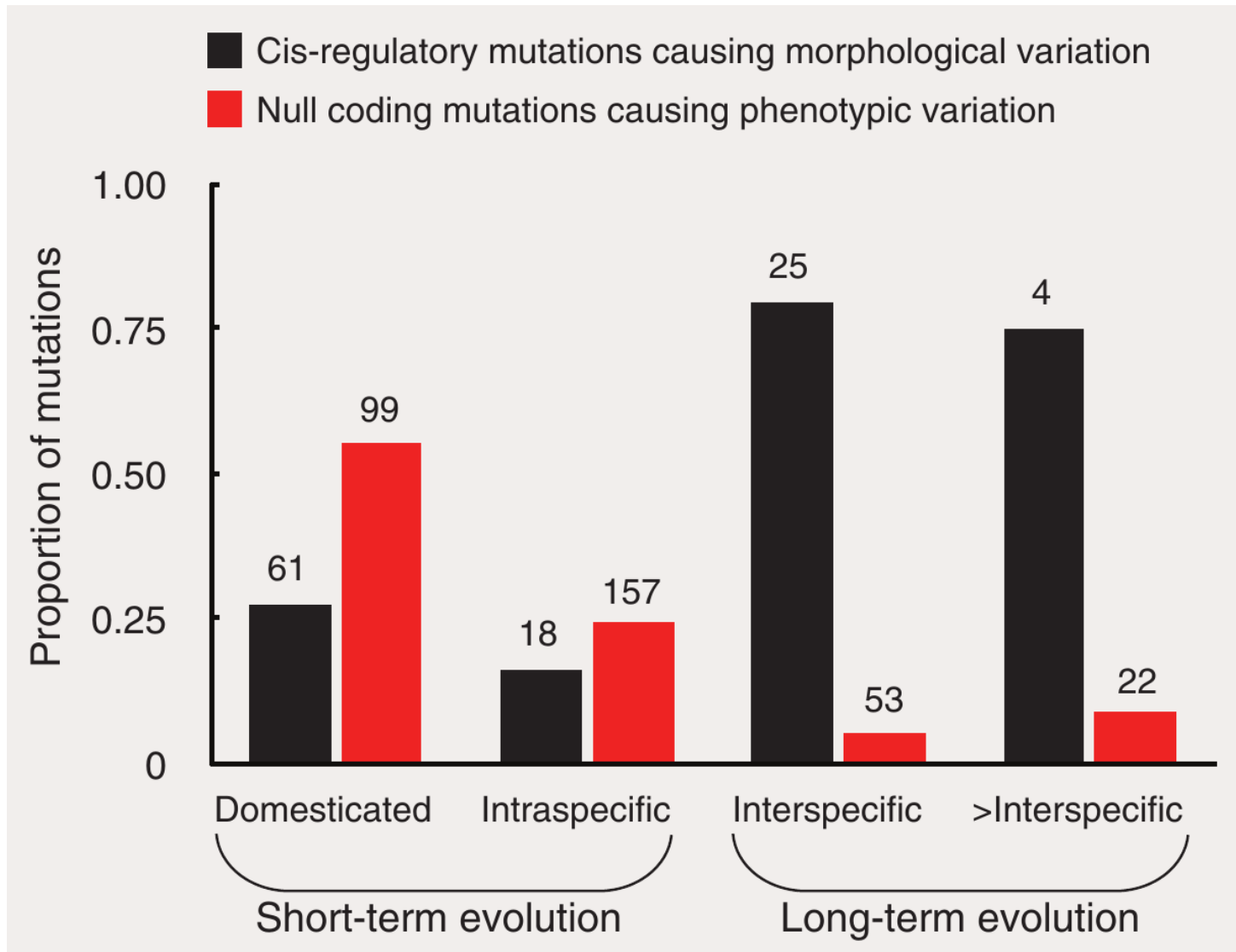


**Bias towards
certain traits**

261 traits in total

**The other traits
represent 517 entries**

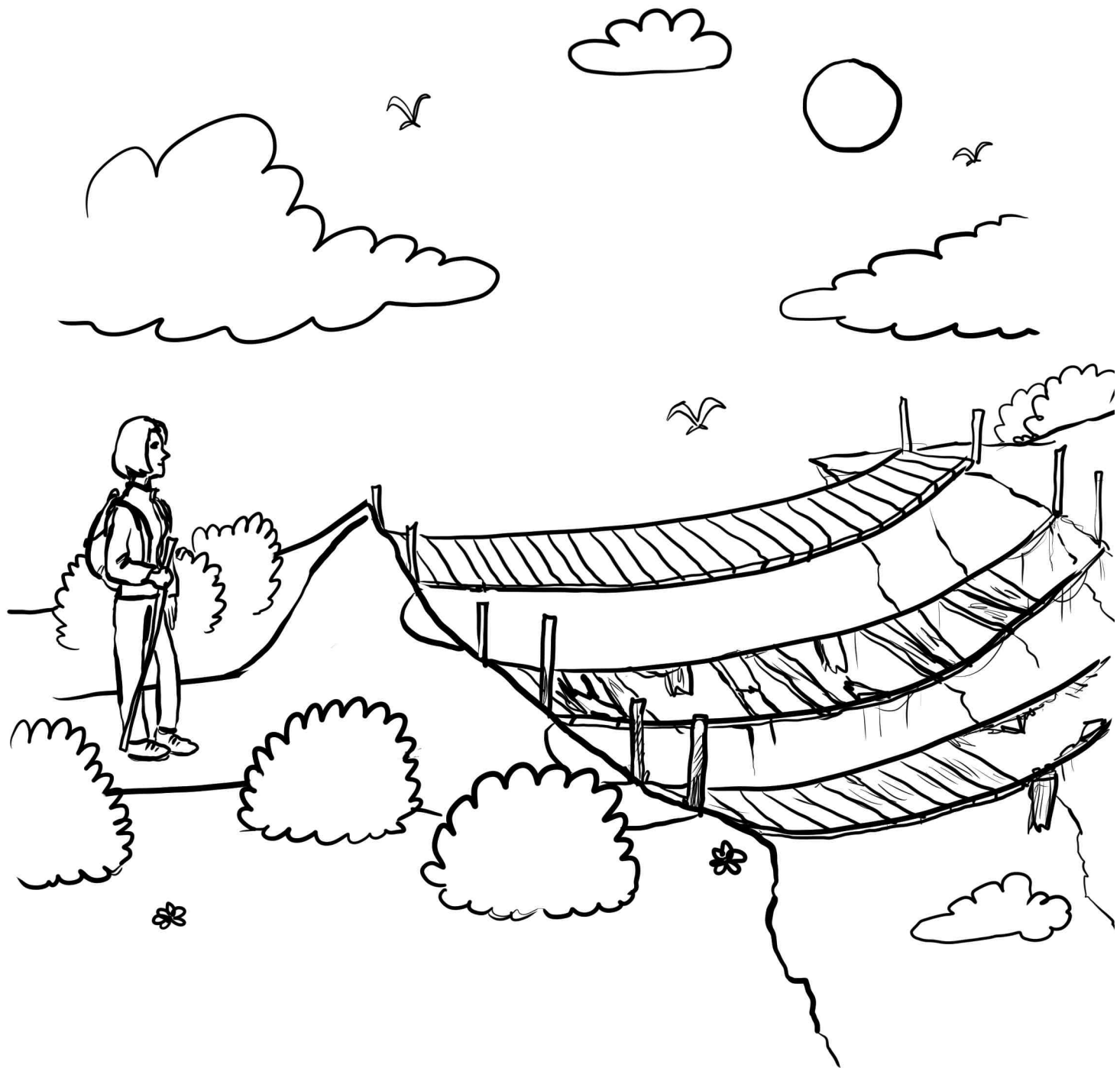
Short-term and long-term evolution involve different types of mutations

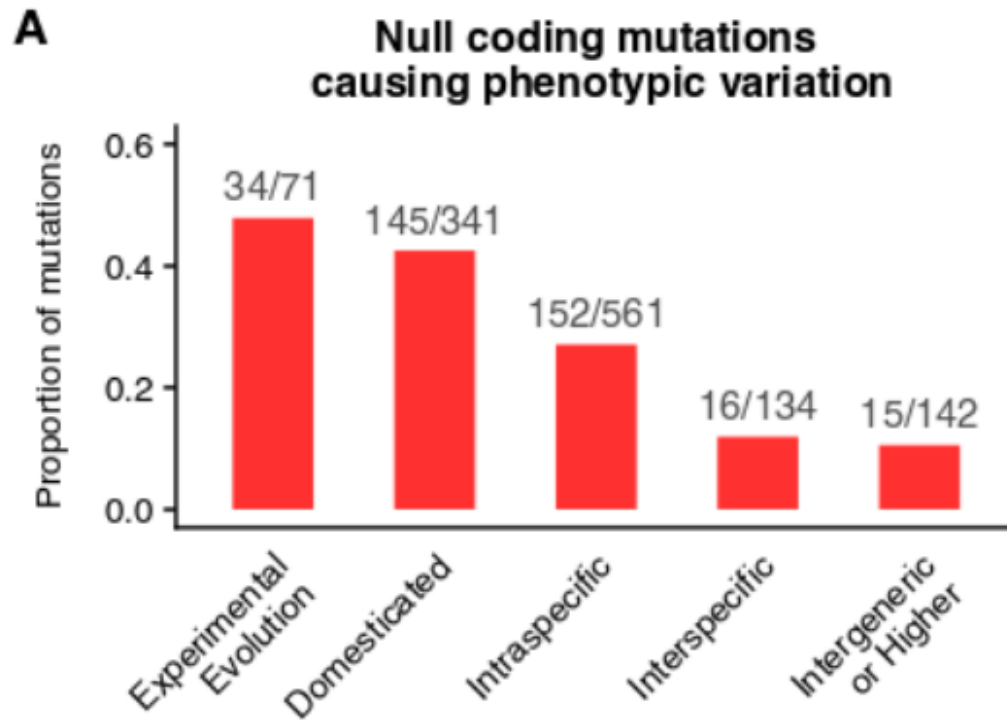


2008: survey of ~300 articles

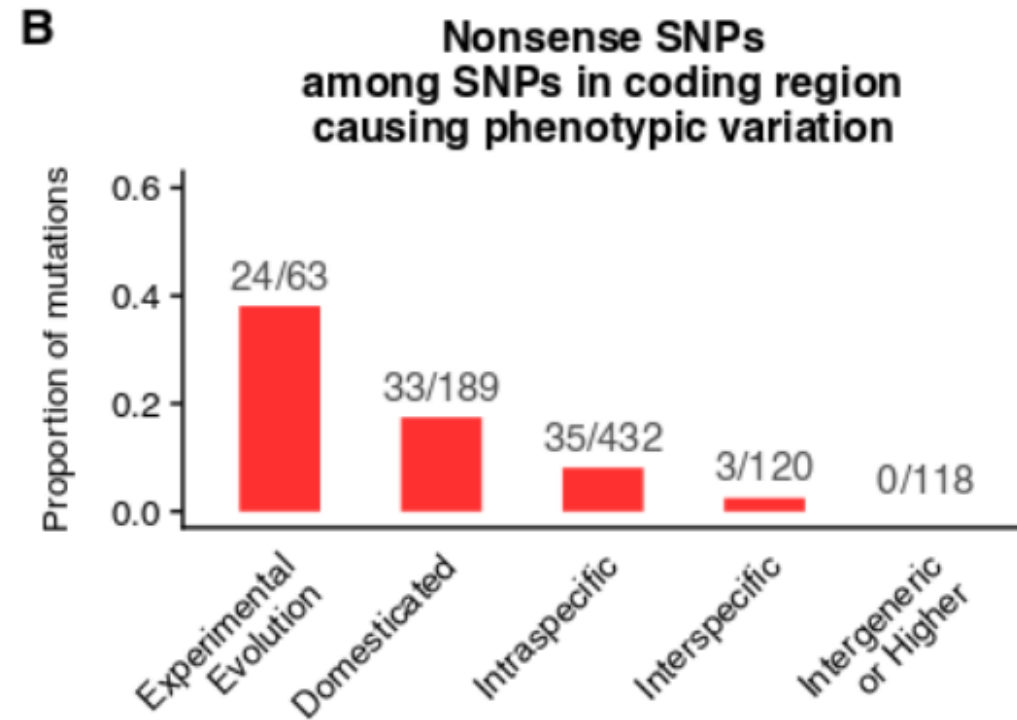
Stern and Orgogozo 2008 Evolution

Stern and Orgogozo 2009 Science



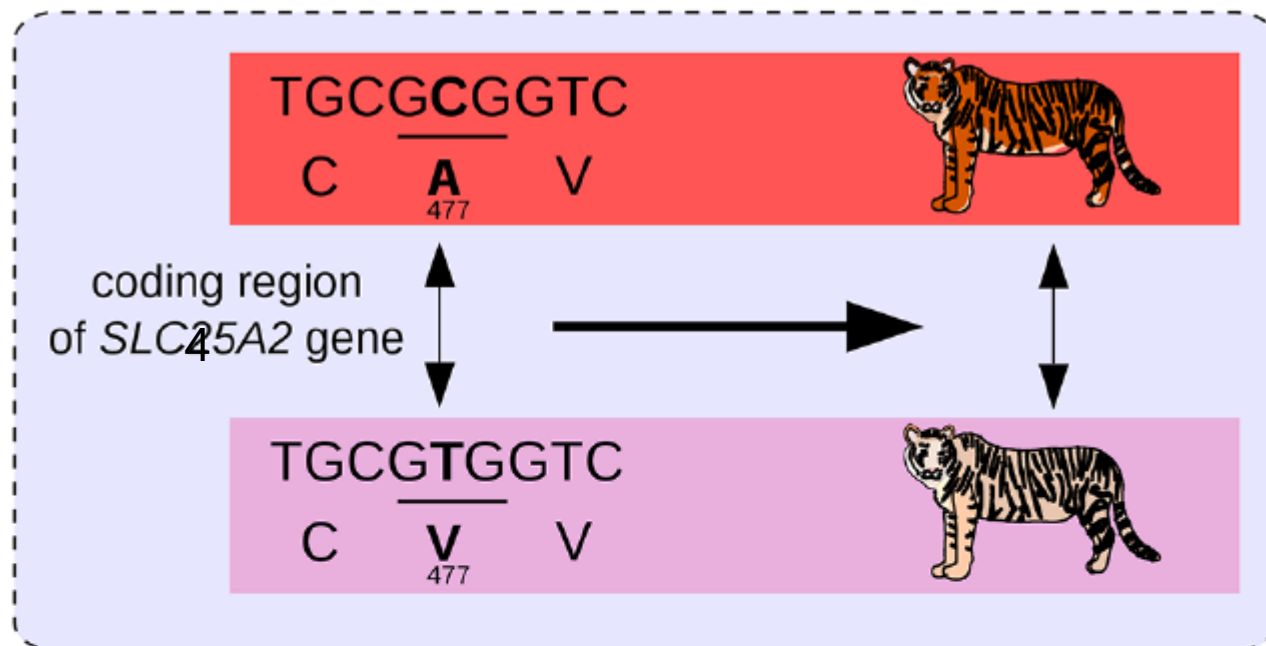


All 1281 coding mutations

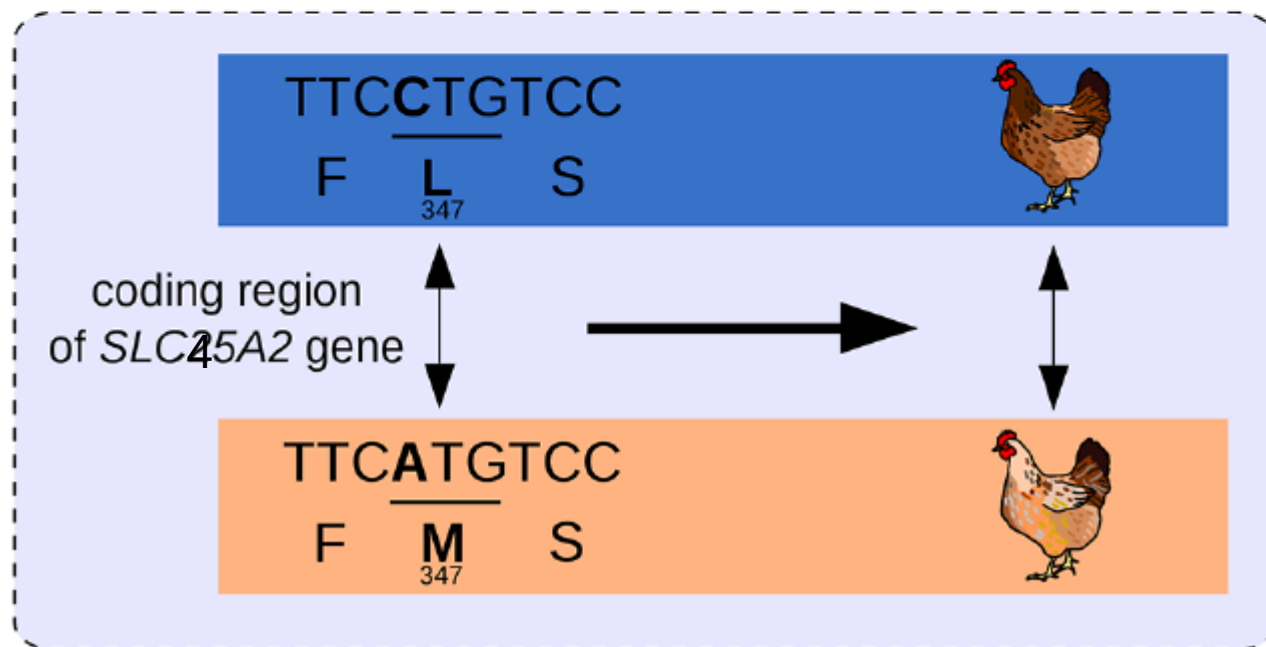


Courtier-Orgogozo et al NAR 2019

Hotspot genes: evolution repeats itself



Also in:
Humans
Horses
Quails
Chickens
Mice
Pigeons



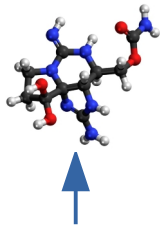
Repeated evolution via the same amino acid change

clam



Nav1.4
X E945D

Saxitoxin

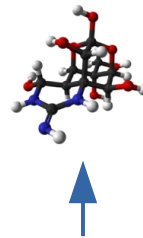


toxic plancton

garter snake



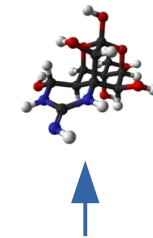
Nav1.4 sodium channel
X E945D



Tetrodotoxin



toxic newt



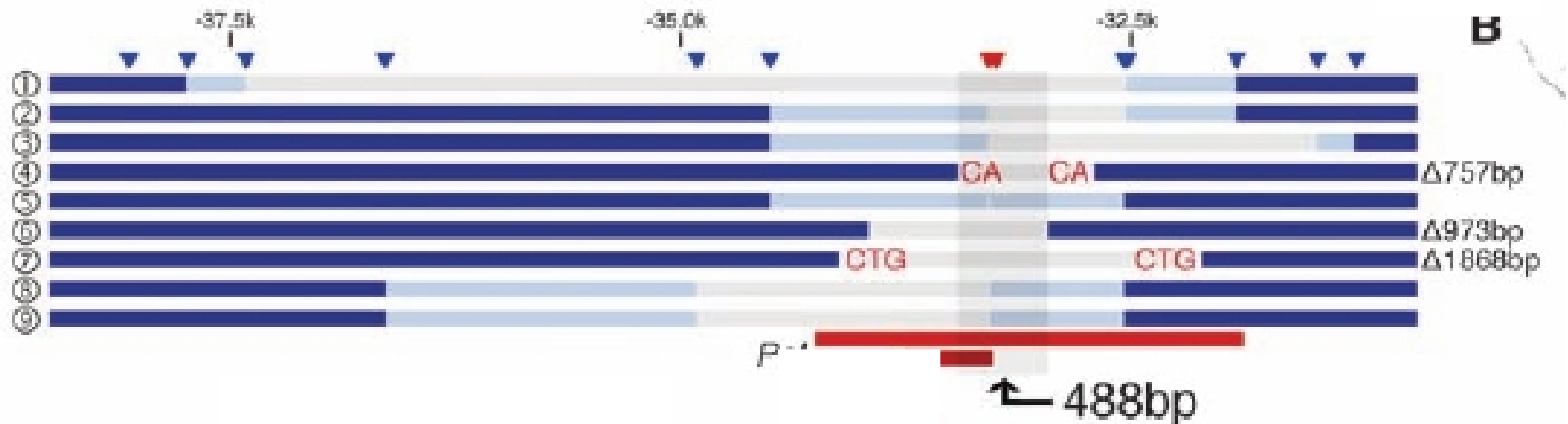
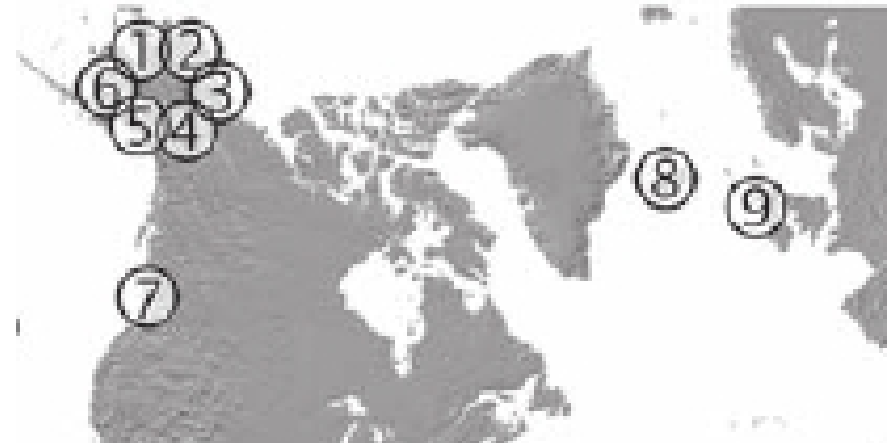
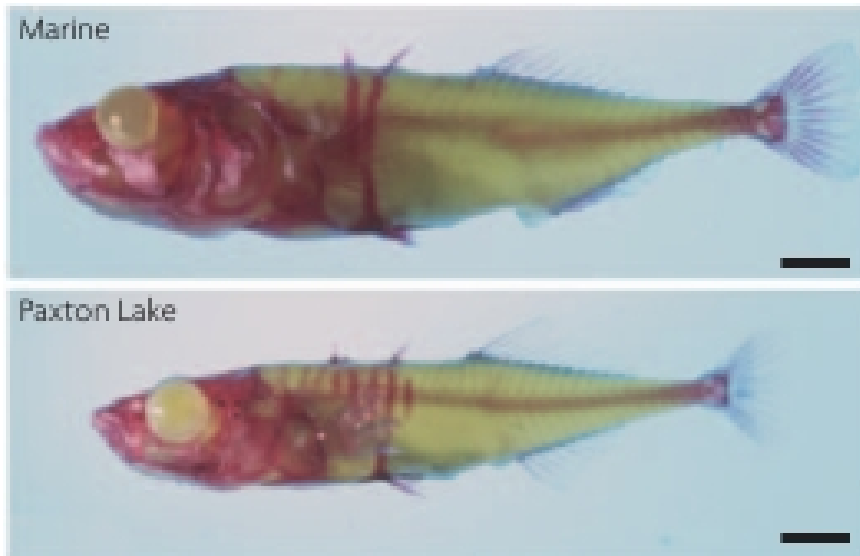
Nav1.4
E945D

X

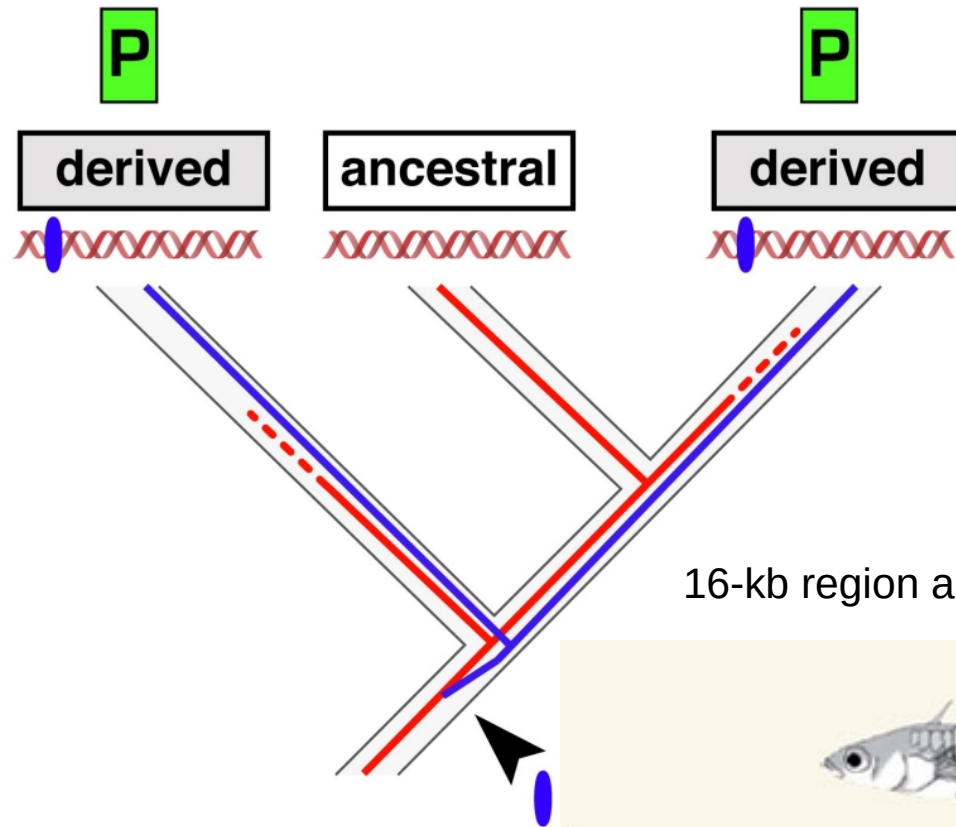


pufferfish (fugu)

9 independent deletions in the cis-regulatory region of *Pitx1*

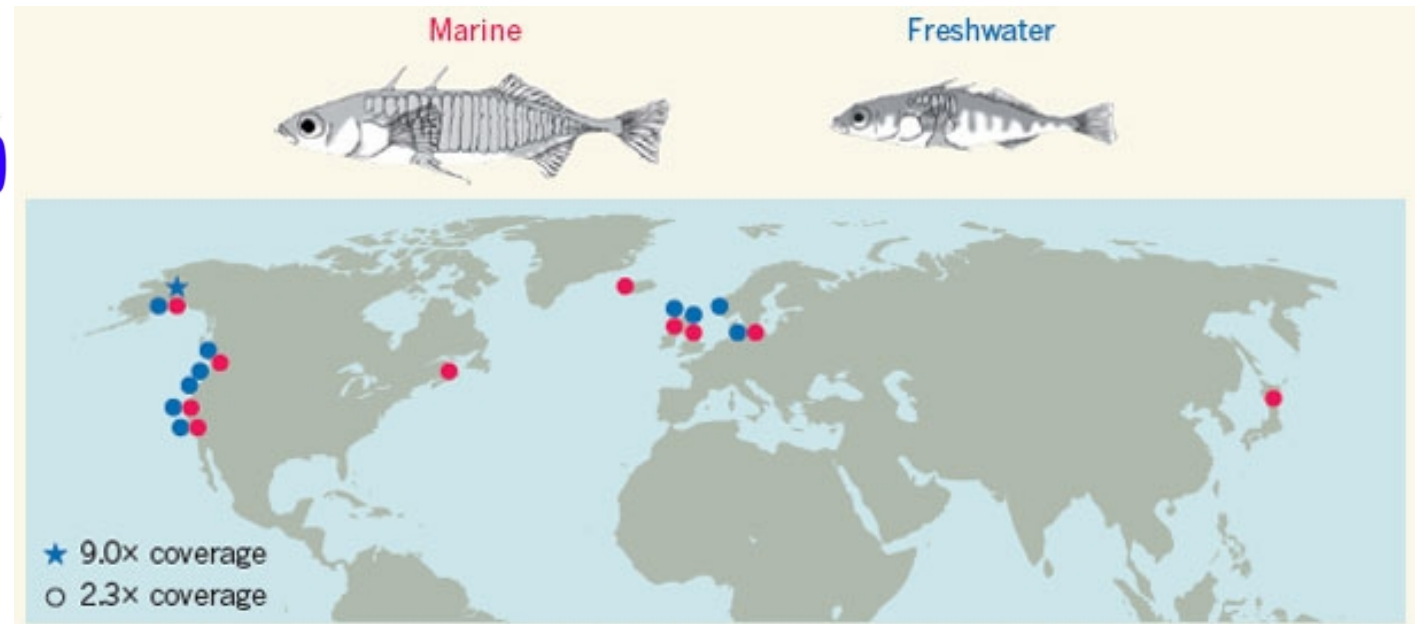


Repeated evolution via ancestral polymorphisms

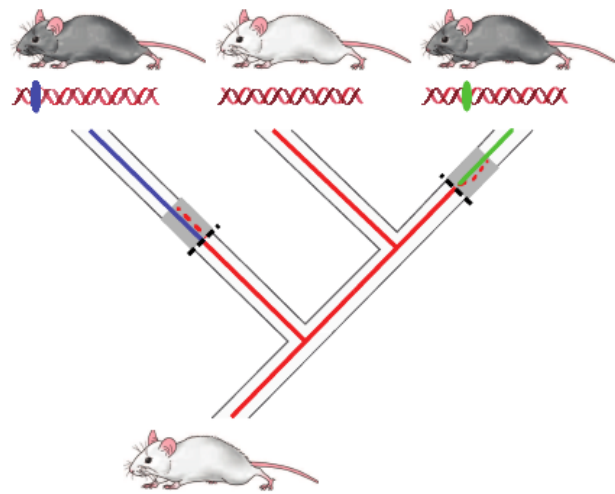


Incomplete lineage sorting
Standing genetic variation

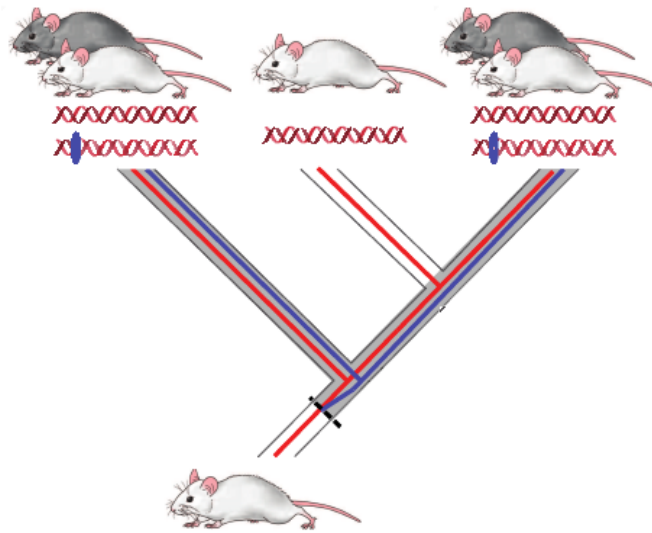
16-kb region around *Eda* shared by all freshwater fishes



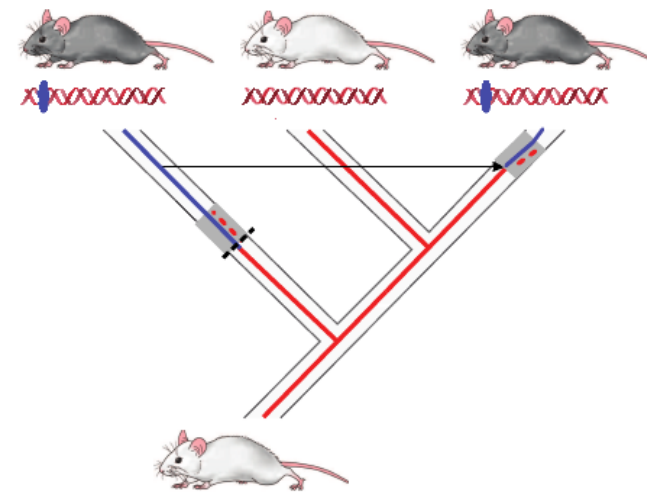
Armor plate
(Colosimo 2005
Jones 2012)



Genetic convergence



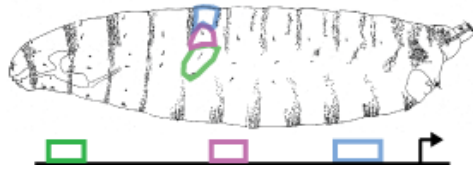
Ancestral polymorphism



Lateral transfer

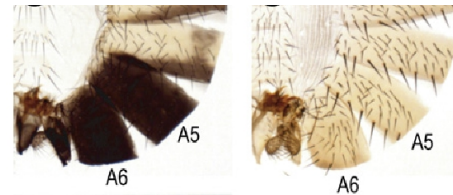
Accumulation of evolutionary-relevant mutations at the same locus

6 mutations in *svb*



(McGregor, Orgogozo et al. 2007 Nature)

3 mutations in *tan*



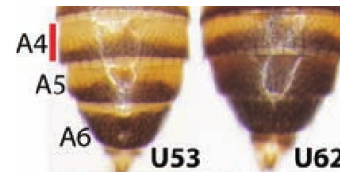
(Jeong et al., 2008 Cell)

2-4 mutations in *nvd*



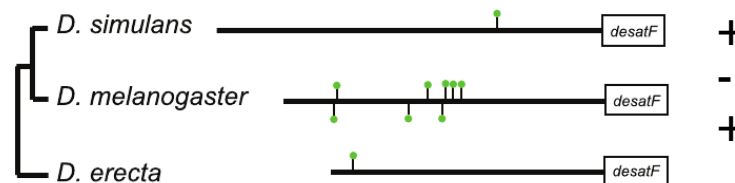
(Lang et al. 2012 Science)

5 mutations in *ebony*



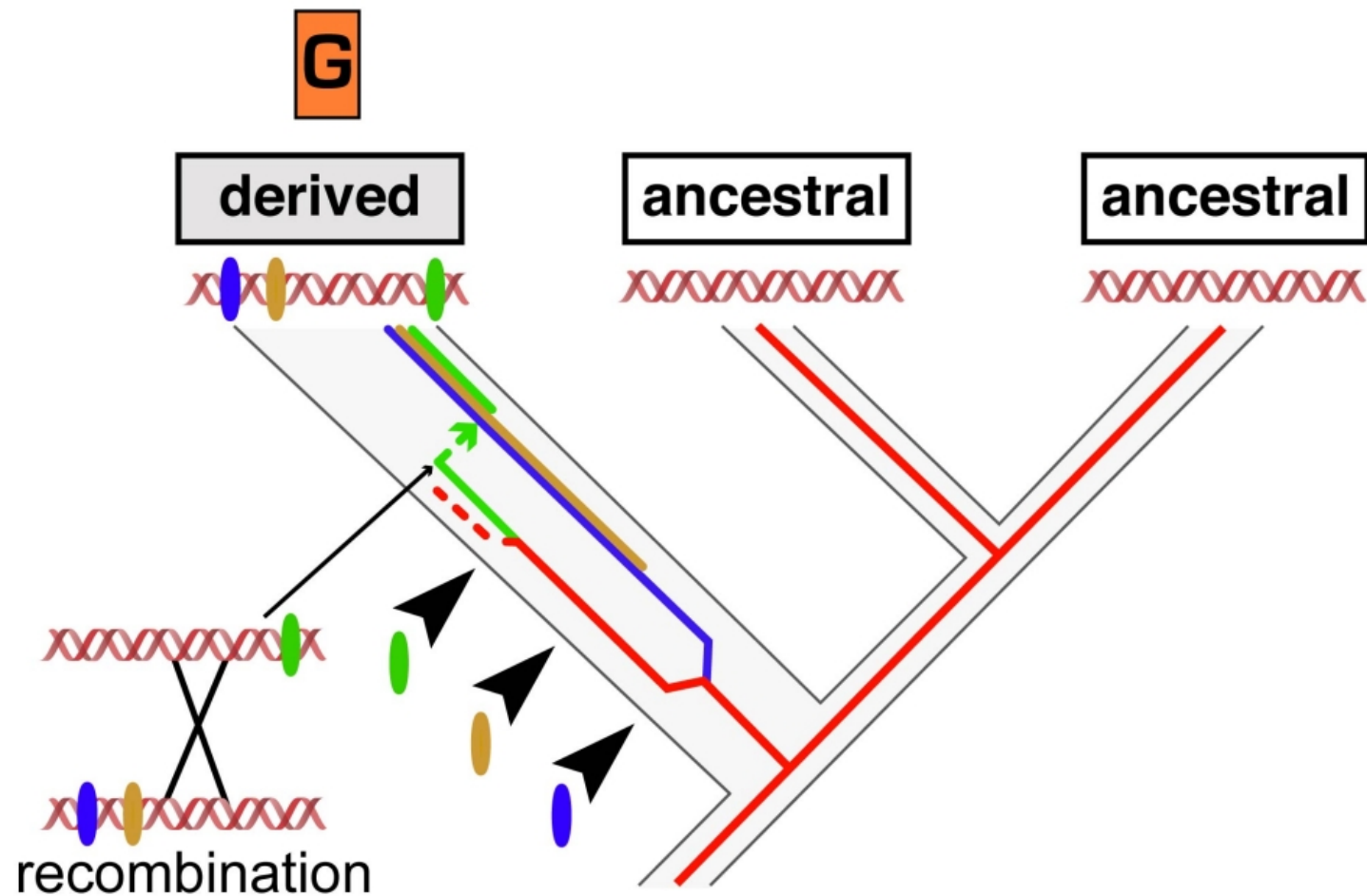
(Rebeiz et al., 2009 Cell)

3 deletions in *desatF*



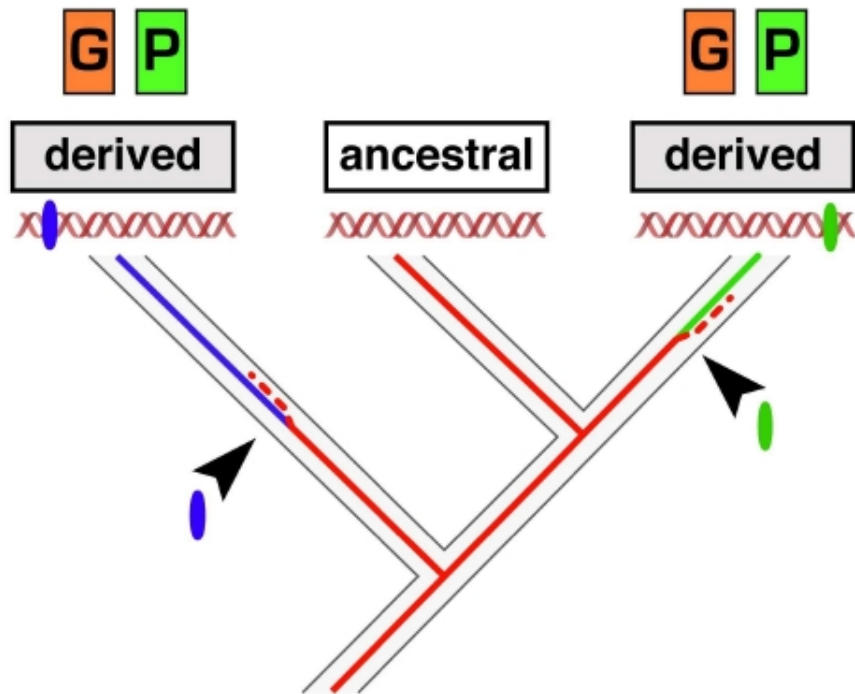
(Shirangi et al., 2009 PloS Biol)

Intralineage hotspot

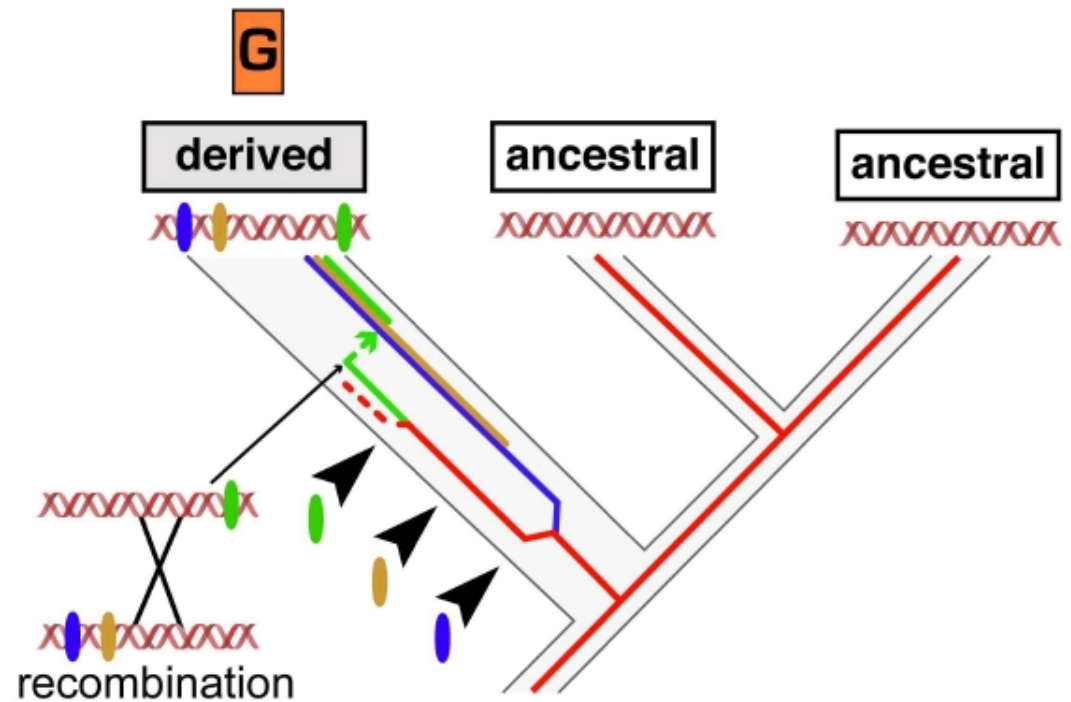


Hotspot genes: preferred targets of evolution

Interlineage hotspot



Intralineage hotspot



Repeats in..



.. the genes responsible for natural evolution

Ex : *hemoglobin* in dogs and humans in Tibet
(Wang et al 2014 GBE)



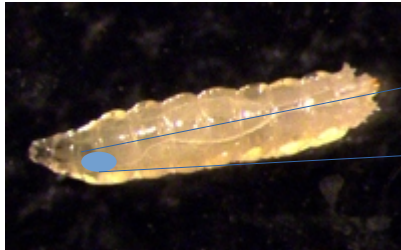
.. the genes responsible for experimental evolution

Ex : *sulfate transporter SUL1* in yeasts in low sulfate
(Gresham et al 2008 PloS Genetics)

**Why is the set of genetic paths
limited?**

There are specialized genes in a genome

Steroid hormone biosynthesis

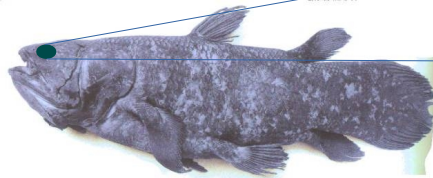


a specialized tissue
specialized enzymes

2-4 mutations in *nvd*



Color vision



a specialized tissue
specialized molecules

mutations in *opsin*
genes

Hypoxia resistance



a specialized tissue
specialized molecules

mutations in
haemoglobin genes



McCracken
2009

Specialized genes are usually genes that interact with external parameters

Why is the set of genetic paths limited?

genes with specialized functions

But what about phenotypes involving multifunctional genes?

Evolution appears to use a restricted set of all possible paths

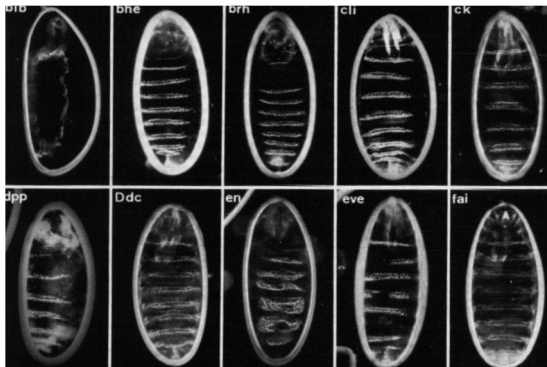
Changes in trichome pattern

EVOLUTION



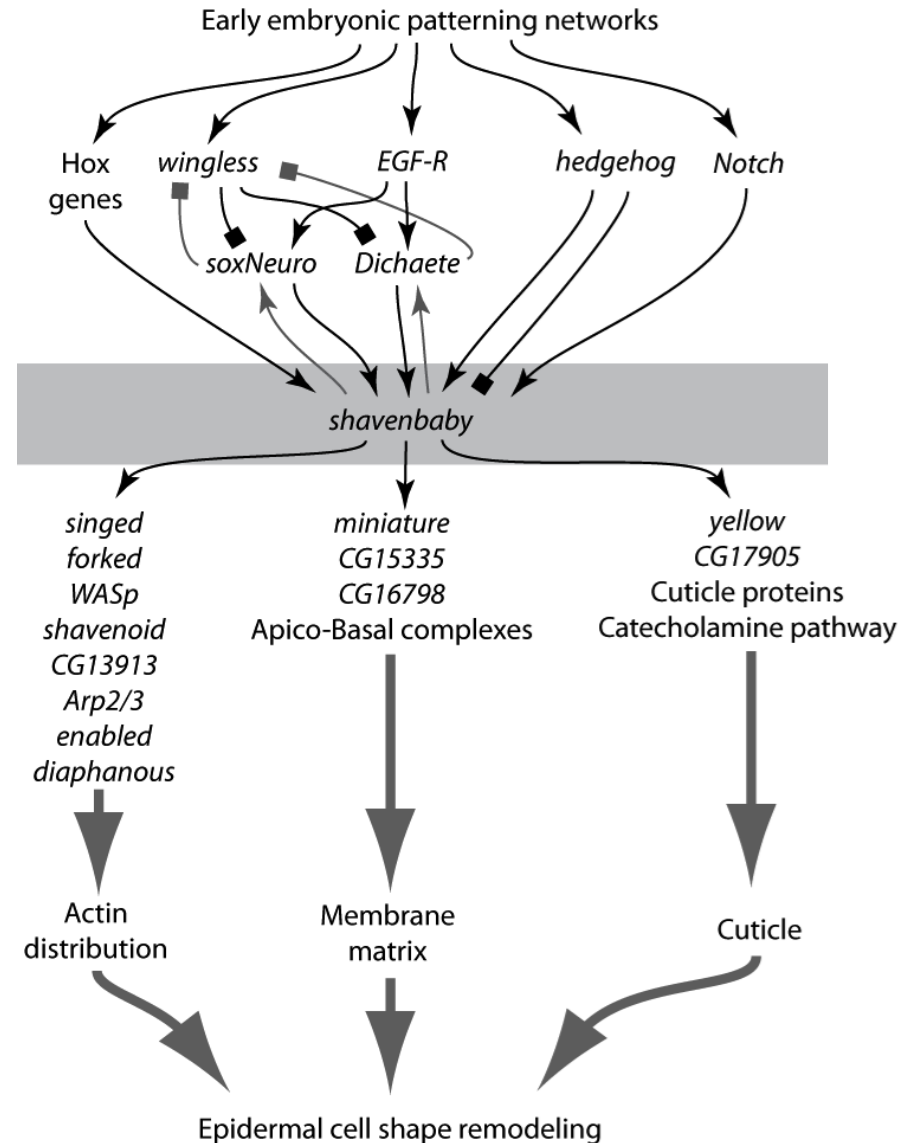
A single
gene

MUTAGENESIS

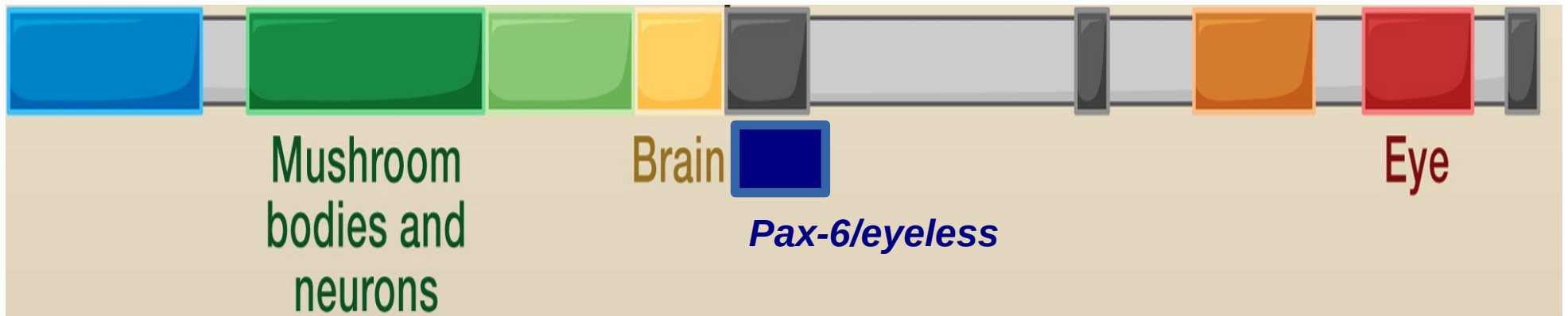


(Nüsslein-Volhard et Weischaus)

~100 genes

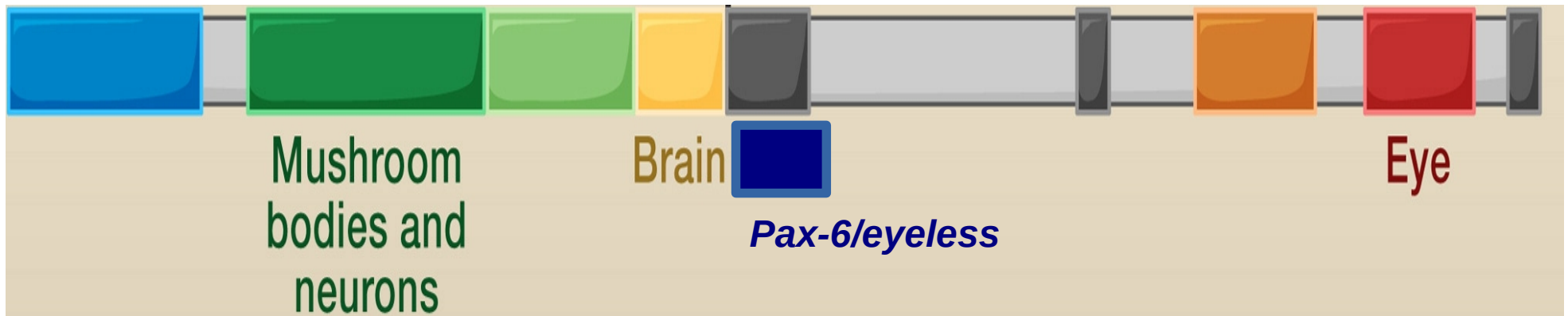


Specialized loci within multifunctional genes



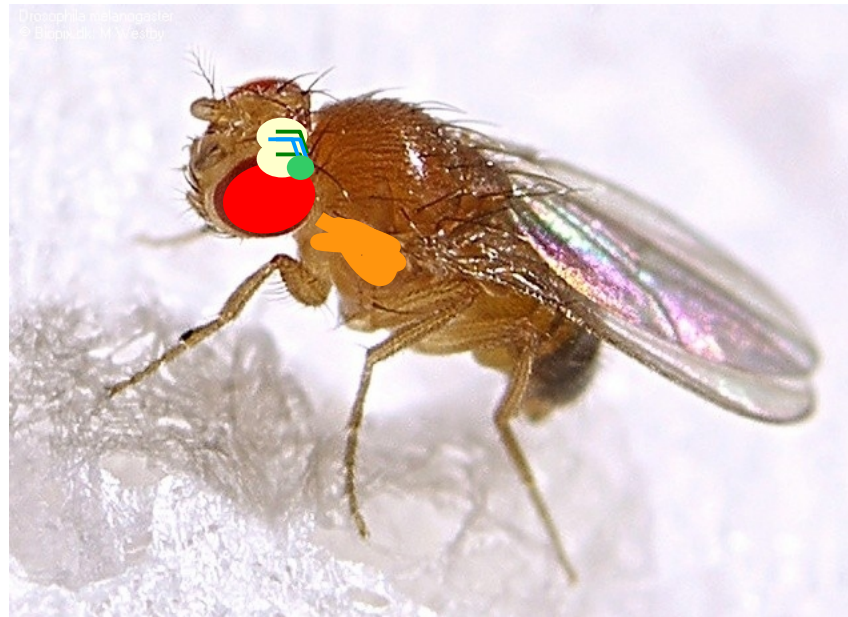
Carroll 2008

Specialized loci within multifunctional genes



Carroll 2008

Modularity of cis-regulatory elements is reflected in modularity of body parts



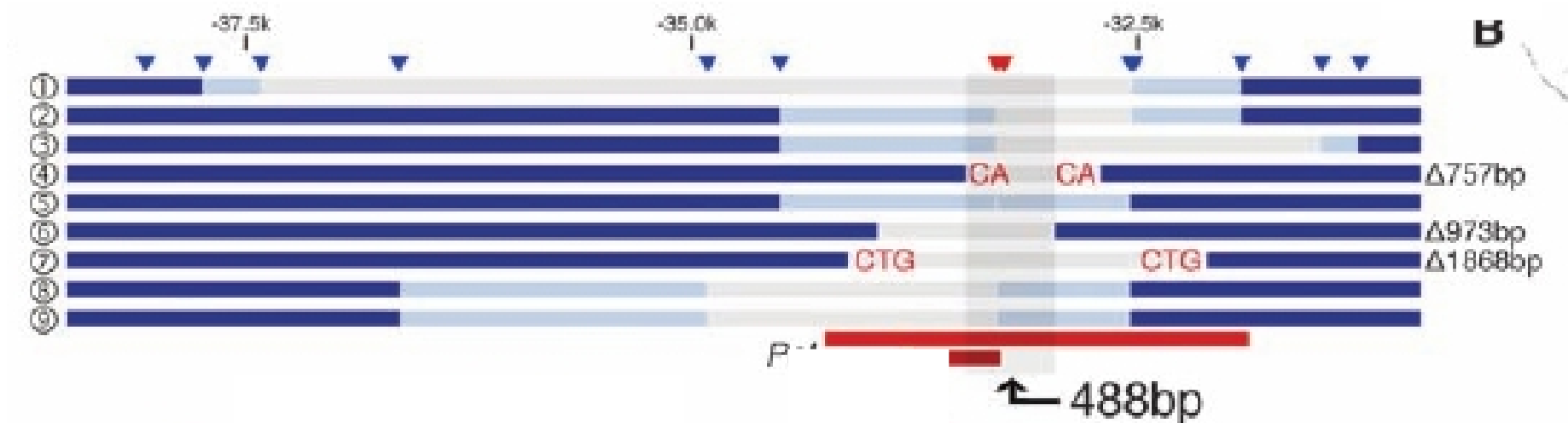
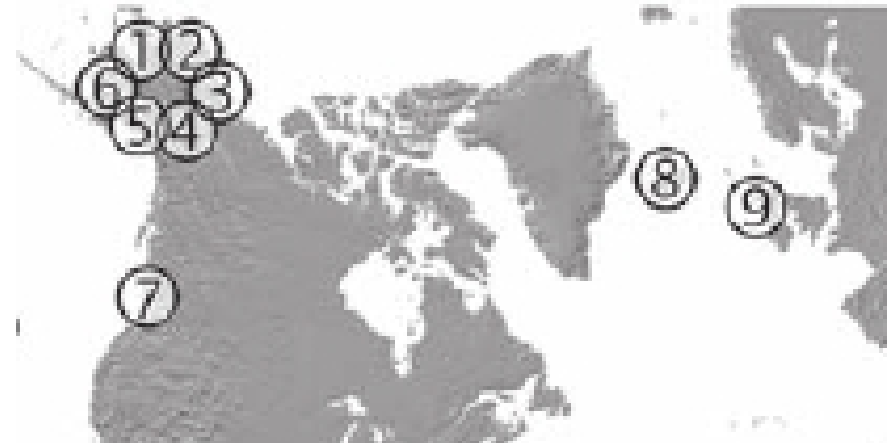
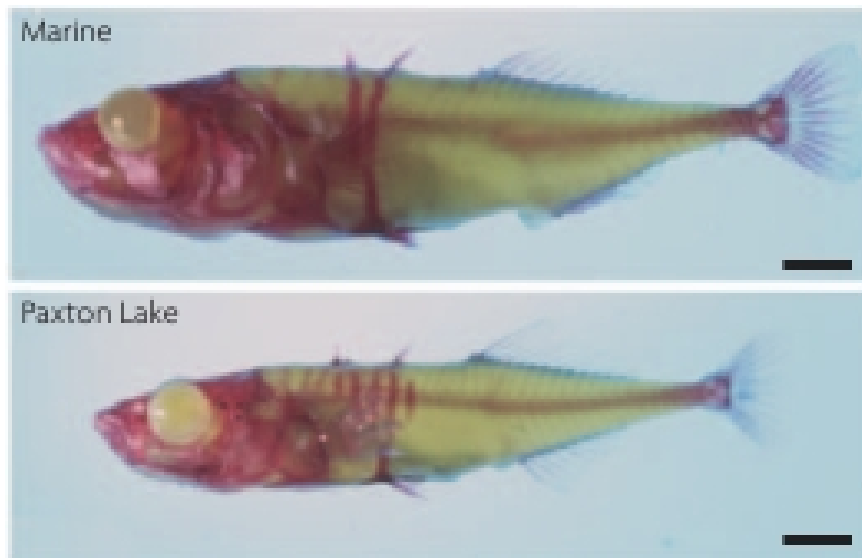
Why is the set of genetic paths limited?

genes with specialized functions

multifunctional genes with specialized regions

mutational bias

9 deletions in the cis-regulatory region of *Pitx1* due to region sensitive to chromosome breaks



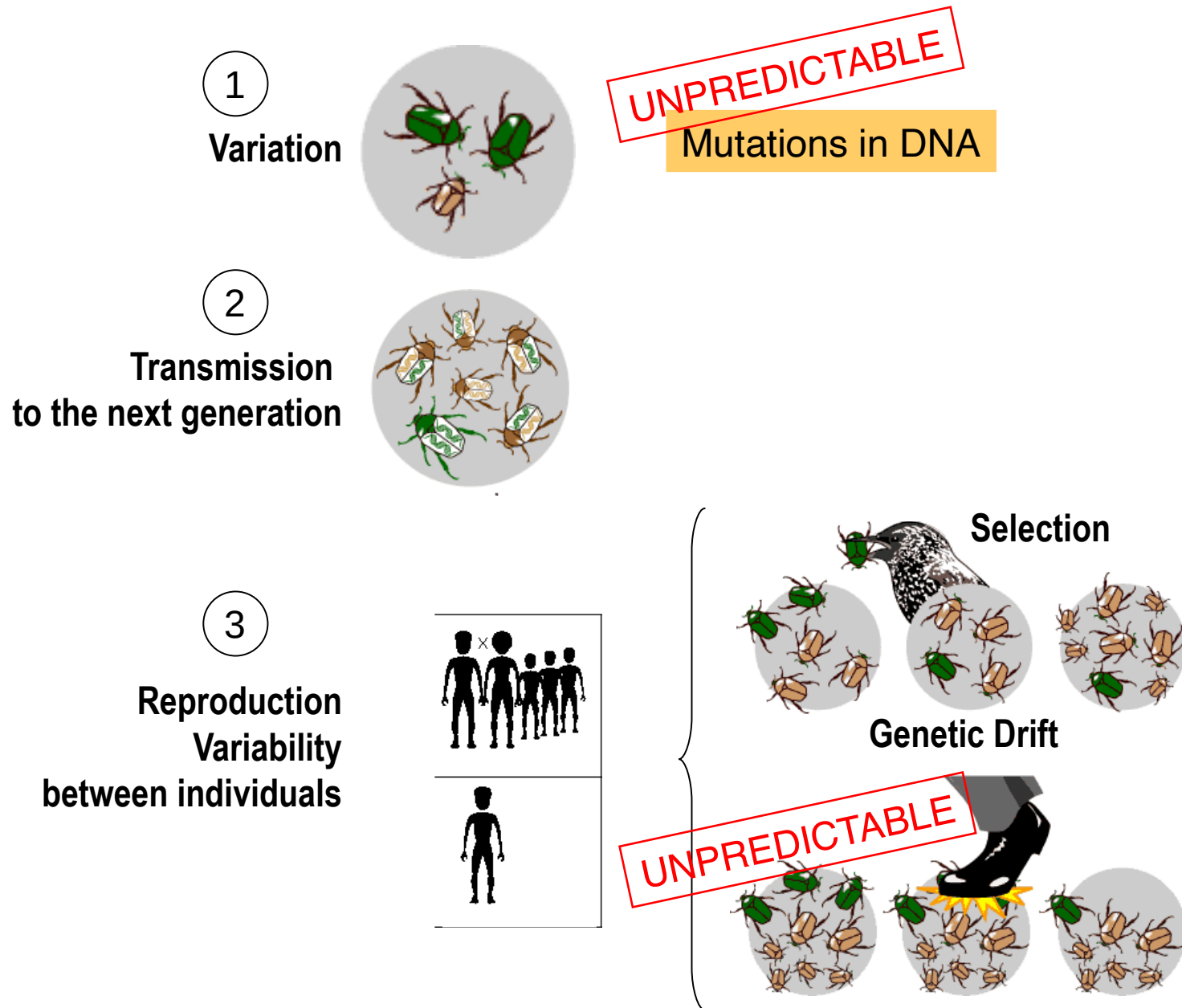
Why is the set of genetic paths limited?

genes with specialized functions

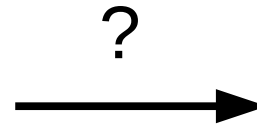
multifunctional genes with specialized regions

mutational bias

Classical Darwinian Evolution



Many unpredictable processes
at a low level



Predictable Evolution
at the genetic level

Mutations in DNA
Chromosome segregation during meiosis
Assortative mating
Gamete competition during fecondation
Life history traits
Genetic linkage
Environmental changes (meteorite, etc.)
...



From random processes can emerge predictability

Many unpredictable processes
at a low level

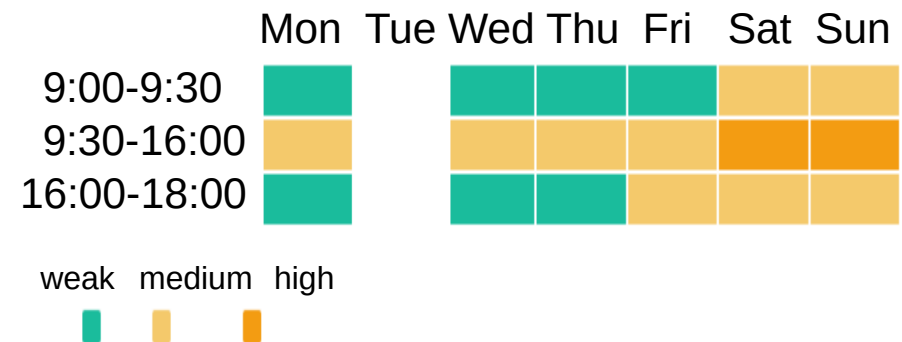


Predictable
at a higher level

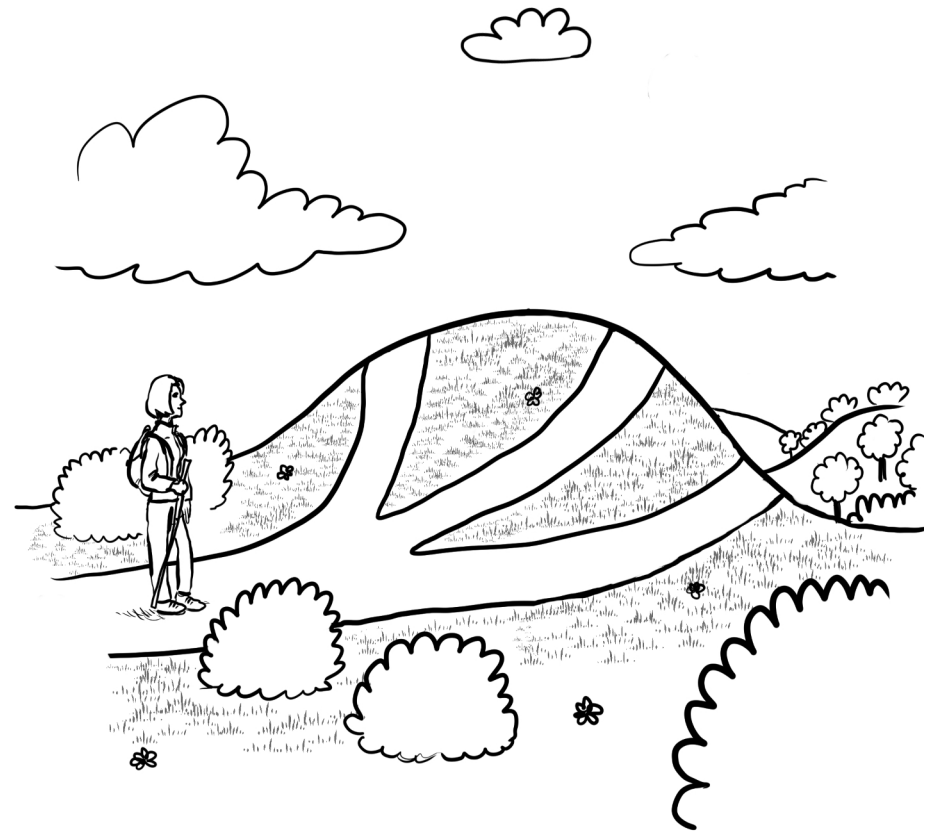
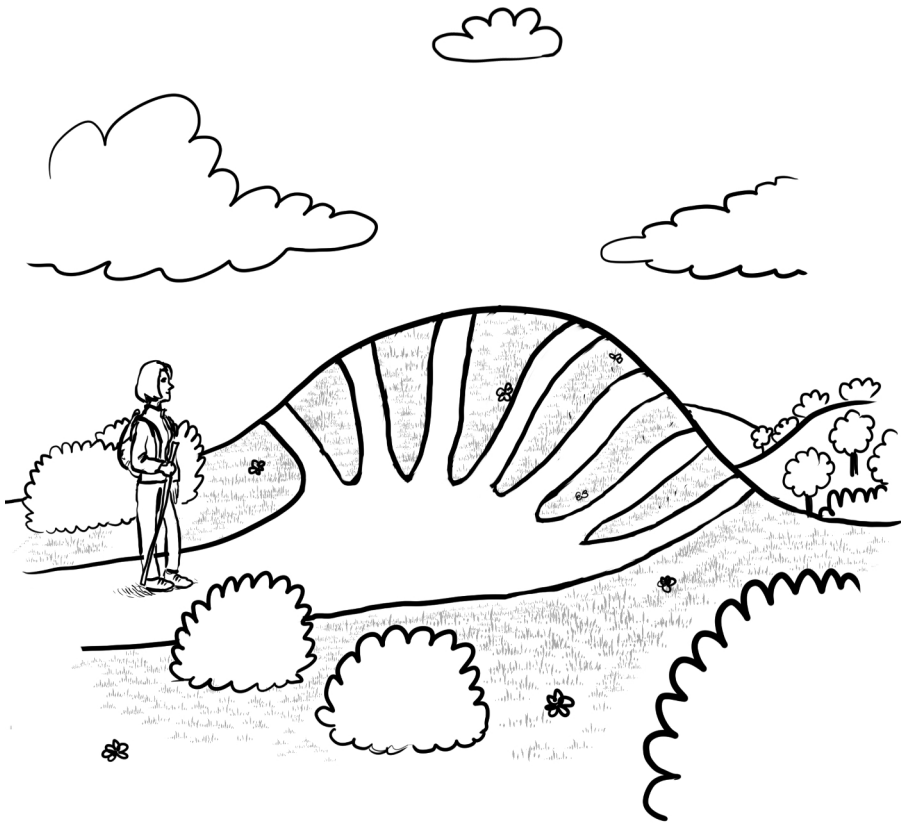
Trajectory of single individuals



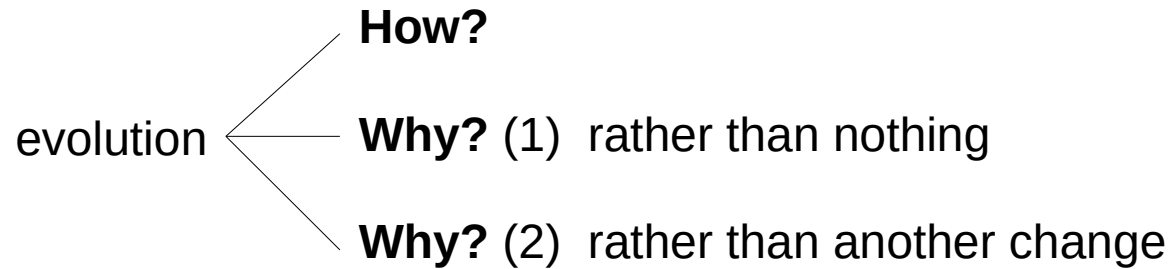
Estimated waiting time for Louvre museum



The number of possible paths for evolution is smaller than previously thought



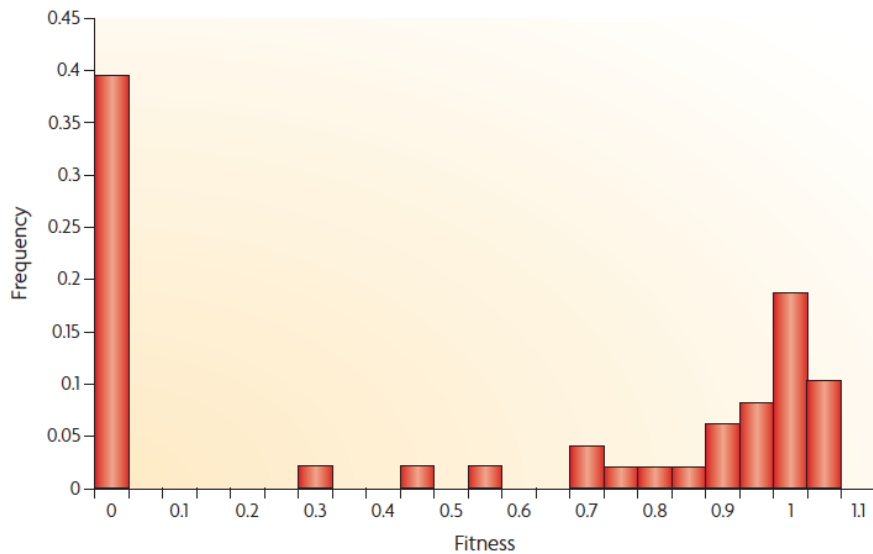
What are all the possible evolutionary paths?



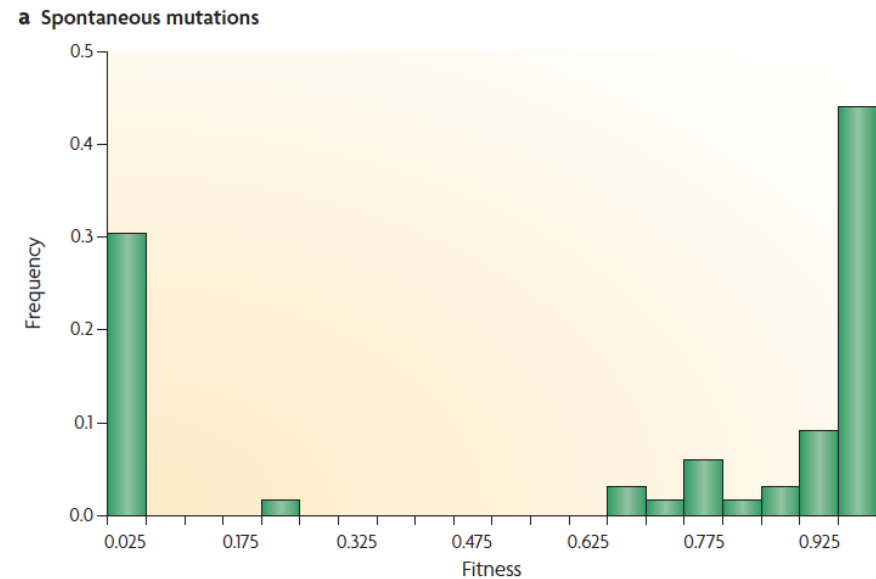
Test all possible paths:

Examine all possible mutations in a protein sequence,
in a cis-regulatory sequence, etc.

Mostly deleterious and neutral mutations



Vesicular Stomatitis Virus



Saccharomyces cerevisiae
diploids

Evolution: unconstrained and unpredictable?

[past and present organisms are] a subset of workable, but basically fortuitous, survivals among a much larger set that could have functioned just as well, but either never arose, or lost their opportunities, by historical happenstance.

Stephen Jay Gould, 2002

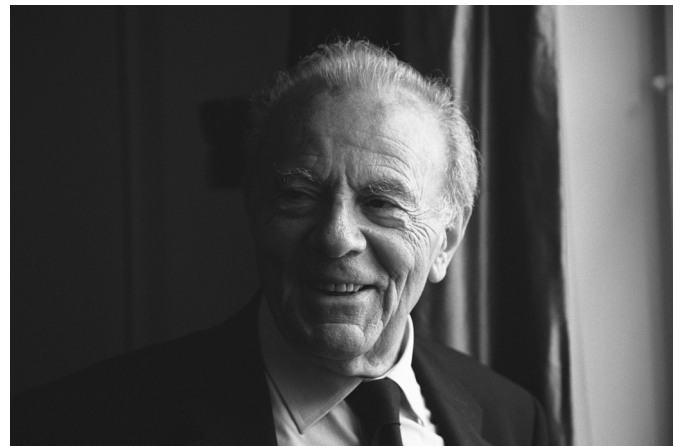


It is hard to realize that the living world as we know it is just one among many possibilities; that its actual structure results from the history of the earth.

1977

Evolution and Tinkering

François Jacob



Would life evolve again, would it produce similar living beings?



evolution { **How?**
Why? (1) rather than nothing
Why? (2) rather than another change



COLUMBA LIVIA or ROCK-PIGEON.

THE MILLION COPY INTERNATIONAL BESTSELLER

RICHARD DAWKINS THE SELFISH GENE

WITH A NEW INTRODUCTION
BY THE AUTHOR
30th
Anniversary
edition

WHY EVOLUTION IS TRUE

'Compelling . . . masterful . . .
outstandingly good.'
Richard Dawkins, *TLS*

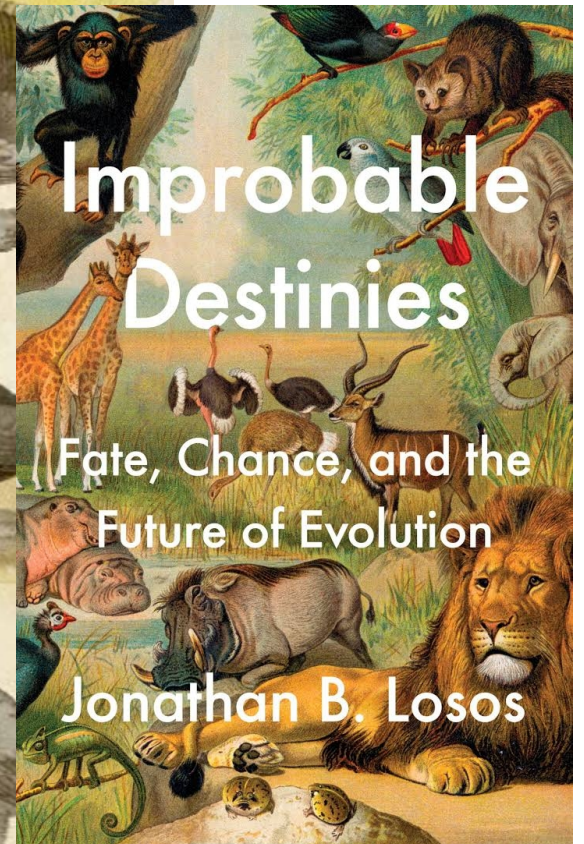


JERRY A. COYNE

Improbable Destinies

Fate, Chance, and the
Future of Evolution

Jonathan B. Losos



Diverse conclusions from Gephebase

More known cases of coding than cis-regulatory mutations

Bias towards certain species

Bias towards certain traits

Long-term versus short-term evolution: fewer null mutations, more cis-regulatory mutations

Hotspot genes

Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes

Virginie Courtier-Orgogozo ^{1,*}, Laurent Arnoult¹, Stéphane R. Prigent¹, Séverine Wiltgen² and Arnaud Martin^{3,*}

¹Institut Jacques Monod, CNRS, UMR 7592, Université de Paris, Paris, France, ²Version Ops, Montpellier, France and ³Department of Biological Sciences, The George Washington University, Washington, DC, USA

Received June 19, 2019; Revised August 21, 2019; Editorial Decision August 30, 2019; Accepted September 06, 2019

