The genotype-phenotype relationship

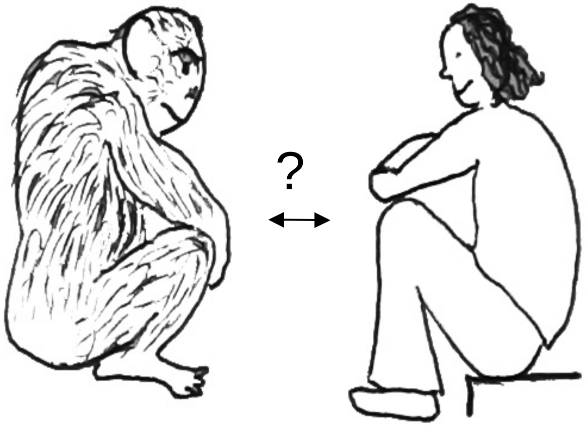
homologies, convergences, environnemental factors

Virginie Orgogozo
Institut Jacques Monod
Paris

What makes us different?

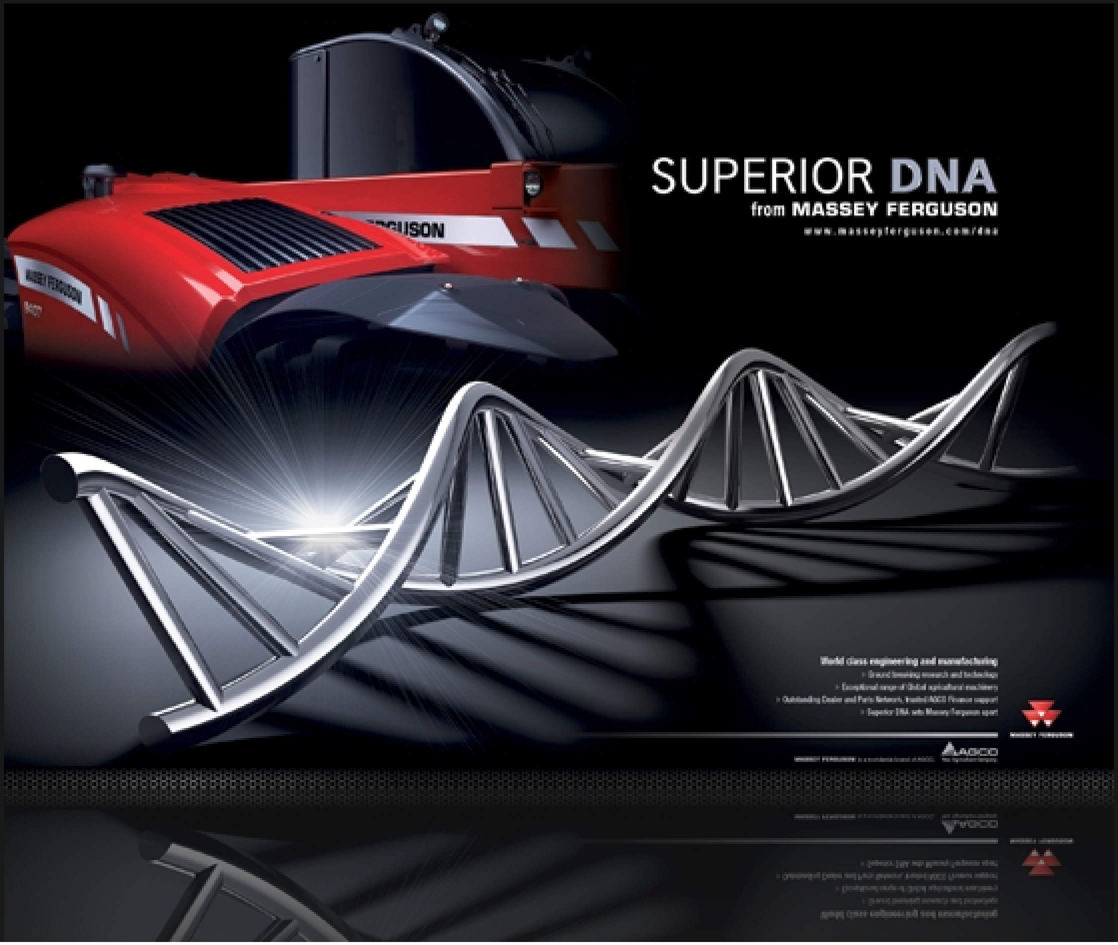
between individuals

between species



Where do we come from?

The importance of DNA



SUPERIOR **DNA**
from **MASSEY FERGUSON**
www.masseyferguson.com/dna

World class engineering and manufacturing
> Ground breaking research and technology
> Exceptional range of Global agricultural machinery
> Outstanding Dealer and Parts Network, trusted AGCO Finance support
> Superior DNA sets Massey Ferguson apart

AGCO

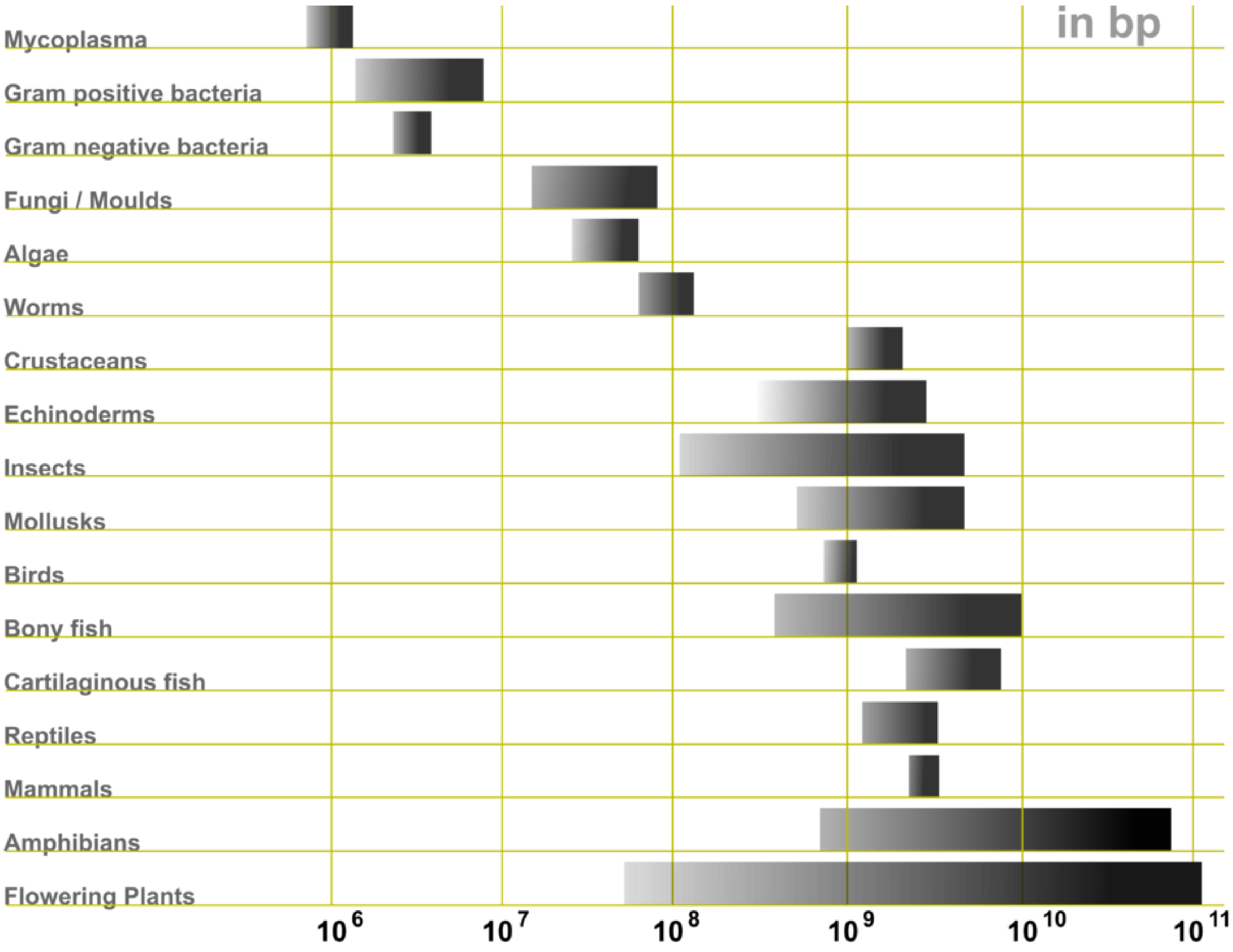
OUR LESS IS MORE
DESIGN PHILOSOPHY IS
BEST APPRECIATED ON
15% INCLINES.

GENETICALLY
ENGINEERED
FIT

P.R.O. ROAD II

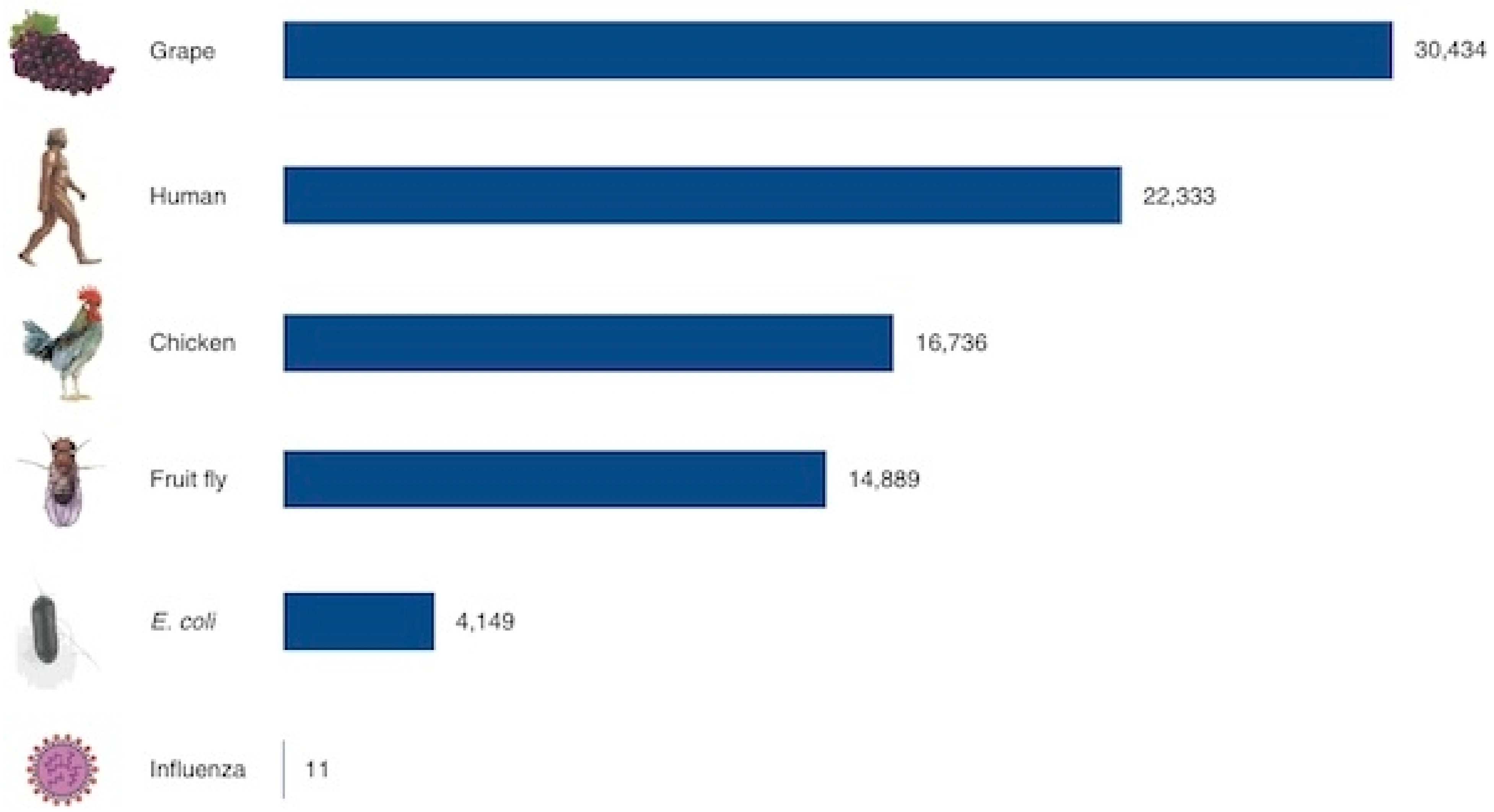
P.R.O. MTN

Genome size is uncorrelated to phenotypic complexity



Wikipedia

Gene number is uncorrelated to phenotypic complexity



M. Lynch 2007, p46

How do we explain genome size ?

Genome size correlates with various features:

- inverse of mutation rate
- cell size
- cell division rate
- body size
- metabolic rate/developmental rate
- inverse of population size

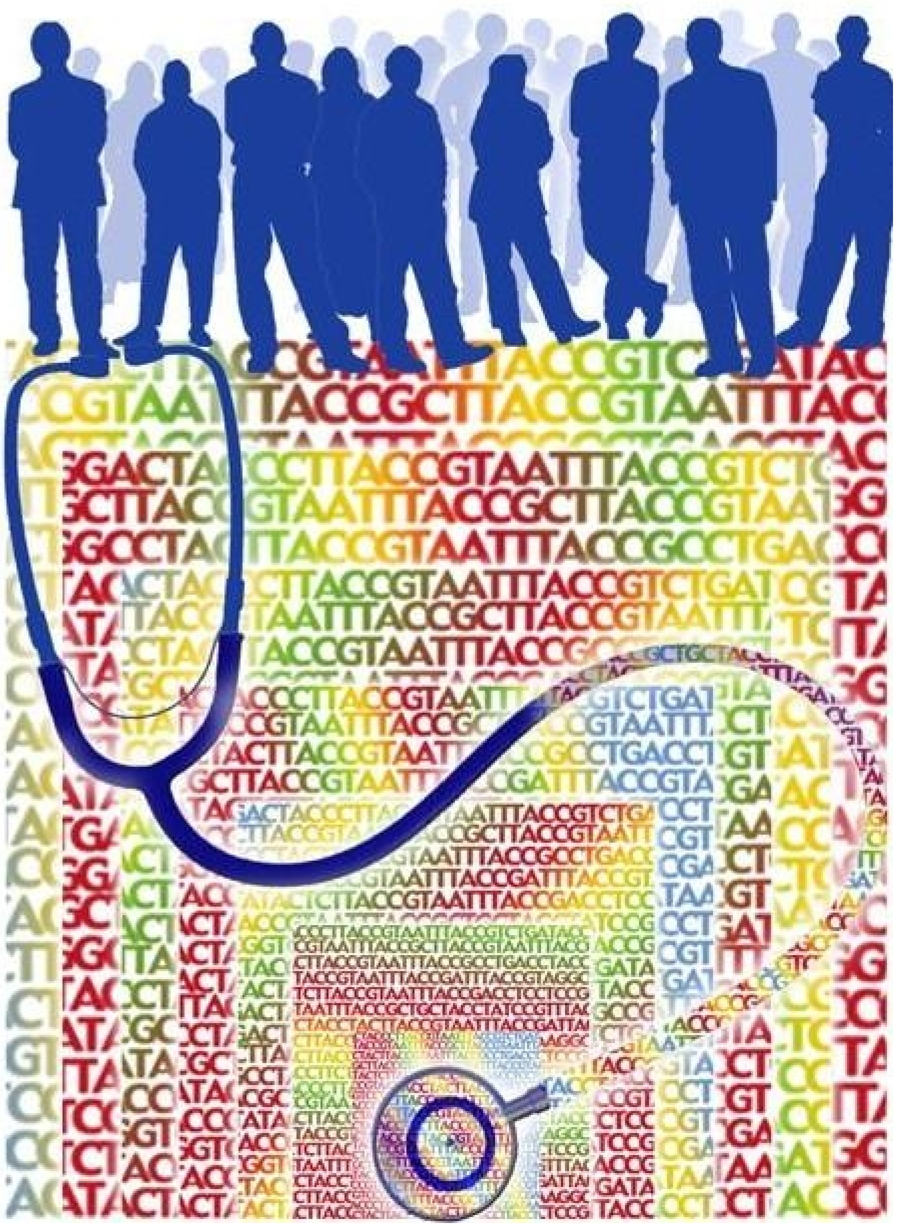
How do we explain genome size ?

Genome size correlates with various features:

- inverse of mutation rate
- cell size
- cell division rate
- body size
- metabolic rate/developmental rate
- inverse of population size

Larger individuals ↔ smaller population sizes ↔reduced ability to eradicate excess DNA

Human genetic diversity



Genome size: 2.9 Gb
Gene number: 25 000
(1% of coding sequences)

In one individual:
~70 new mutations compared to his parents
~20 lethal mutations (heterozygous)

Genetic difference between two humans?

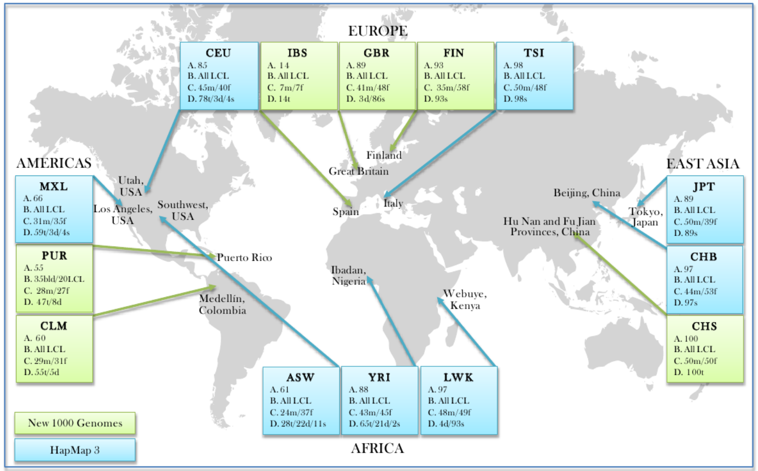
Genetic differences between humans and chimps?

Genetic differences between two humans

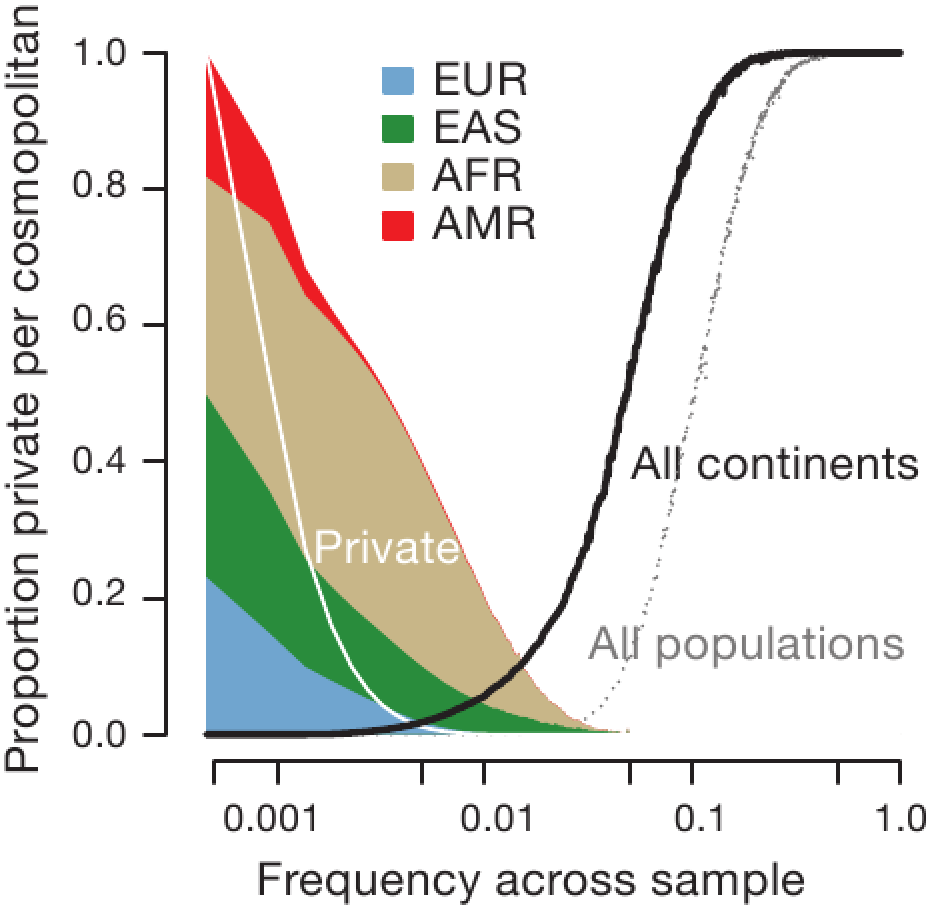
~0.1% (SNP only, accessible genome < 90%)

1000 Genome Project, Nature 2012

SNP
Indels (here < 50 bp)
CNV



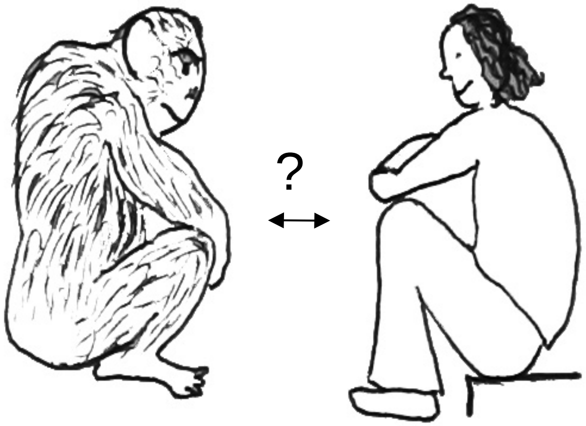
Rare SNP are population-specific



Genetic differences between humans and chimps

1.2% (SNP only, accessible genome < 80%)

up to 5% if taking indels and repeats into account



Pbs:
Fraction of unassembled primate genome sequences
Large inversions and translocations are refractory to detection
Higher sequence polymorphism in primates than in humans

O'Bleness 2012 Nature Reviews Genetics

What makes humans special?

~~Gene number~~

Other aspects of our genome?
cis-regulatory evolution, junk DNA, alternative splicing, post-translational modifications of proteins, new genes, small RNA, transposable elements, active pseudogenes...

Non genetic factors?
environment, culture, bacterial community, GxE

Small-scale genetic changes with large effects?
epistasis, GxE

Humans not so special?


99.4% human?
Banners by www.zephyr-tvc.com

What makes humans special?

~~Gene number~~

Other aspects of our genome?
cis-regulatory evolution, junk DNA, alternative splicing, post-translational modifications of proteins, new genes, small RNA, transposable elements, active pseudogenes...

Non genetic factors?
environment, culture, bacterial community, GxE

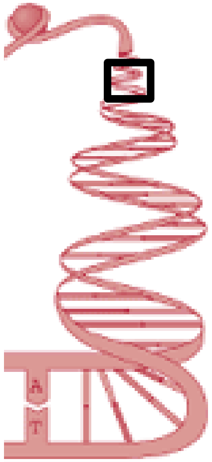
Small-scale genetic changes with large effects?
epistasis, GxE

Humans not so special?

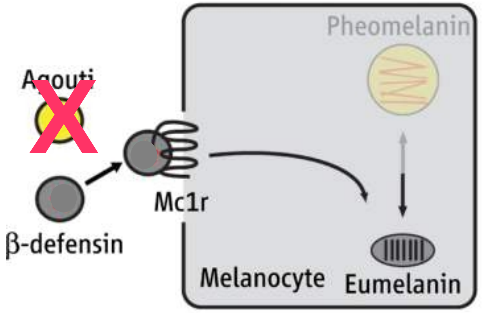
The black panther

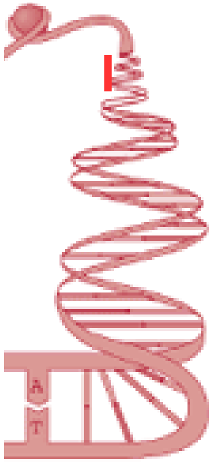
Schneider 2012

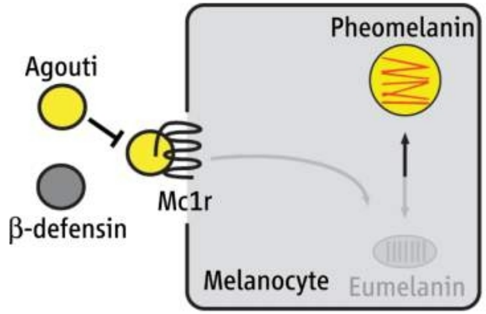
Agouti → Body color



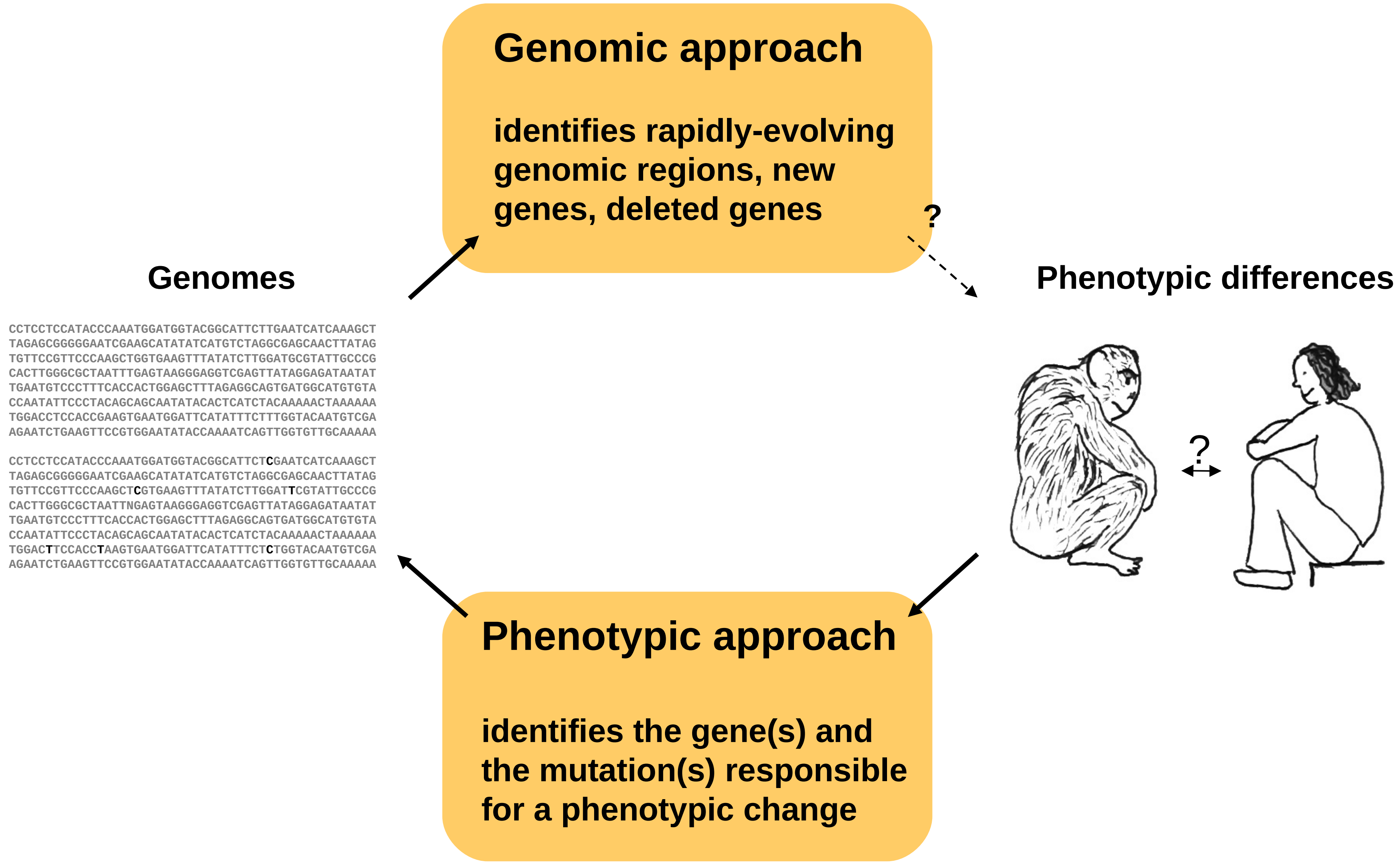
48-bp deletion in the *Agouti* gene







What are the mutations responsible for phenotypic differences?



Genomic approach

identifies rapidly-evolving genomic regions, new genes, deleted genes

Genomes

```
CCTCCTCCATACCCAAATGGATGGTACGGCATTCTTGAATCATCAAAGCT
TAGAGCGGGGGAATCGAAGCATATATCATGTCTAGGCGAGCAACTTATAG
TGTTCCGTTCCCAAGCTGGTGAAGTTTATATCTTGGATGCGTATTGCCCG
CACTTGGGCGCTAATTTGAGTAAGGGAGGTCGAGTTATAGGAGATAATAT
TGAATGTCCCTTTCACCACTGGAGCTTTAGAGGCAGTGATGGCATGTGTA
CCAATATTCCCTACAGCAGCAATATACACTCATCTACAAAAACTAAAAAA
TGGACCTCCACCGAAGTGAATGGATTCATATTTCTTTGGTACAATGTCGA
AGAATCTGAAGTTCCGTGGAATATACCAAAATCAGTTGGTGTTGCAAAAA

CCTCCTCCATACCCAAATGGATGGTACGGCATTCTCGAATCATCAAAGCT
TAGAGCGGGGGAATCGAAGCATATATCATGTCTAGGCGAGCAACTTATAG
TGTTCCGTTCCCAAGCTCGTGAAGTTTATATCTTGGATTCGTATTGCCCG
CACTTGGGCGCTAATTNGAGTAAGGGAGGTCGAGTTATAGGAGATAATAT
TGAATGTCCCTTTCACCACTGGAGCTTTAGAGGCAGTGATGGCATGTGTA
CCAATATTCCCTACAGCAGCAATATACACTCATCTACAAAAACTAAAAAA
TGGACTTCCACCTAAGTGAATGGATTCATATTTCTCTGGTACAATGTCGA
AGAATCTGAAGTTCCGTGGAATATACCAAAATCAGTTGGTGTTGCAAAAA
```

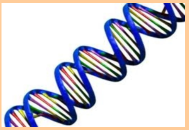
Phenotypic differences

Phenotypic approach

identifies the gene(s) and the mutation(s) responsible for a phenotypic change

What makes us different?

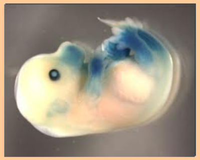
Genetics
DNA structure, universal genetic code
Mutations = changes in DNA sequence (+ epimutations)



Genomics
total gene number unrelated to phenotypic complexity
conserved genes in distinct species

```
CCTCCTCCATACCCAAA
TGGATGGTACGGCATTC
TTGAATCATCAAAGCTT
AGAGCGGGGGAATCGAC
CTCCTCCATACCCAAAT
```

Developmental biology
1990s: toolkit genes (set of genes with similar functions in multiple species)
Importance of cis-regulatory mutations for morphological evolution



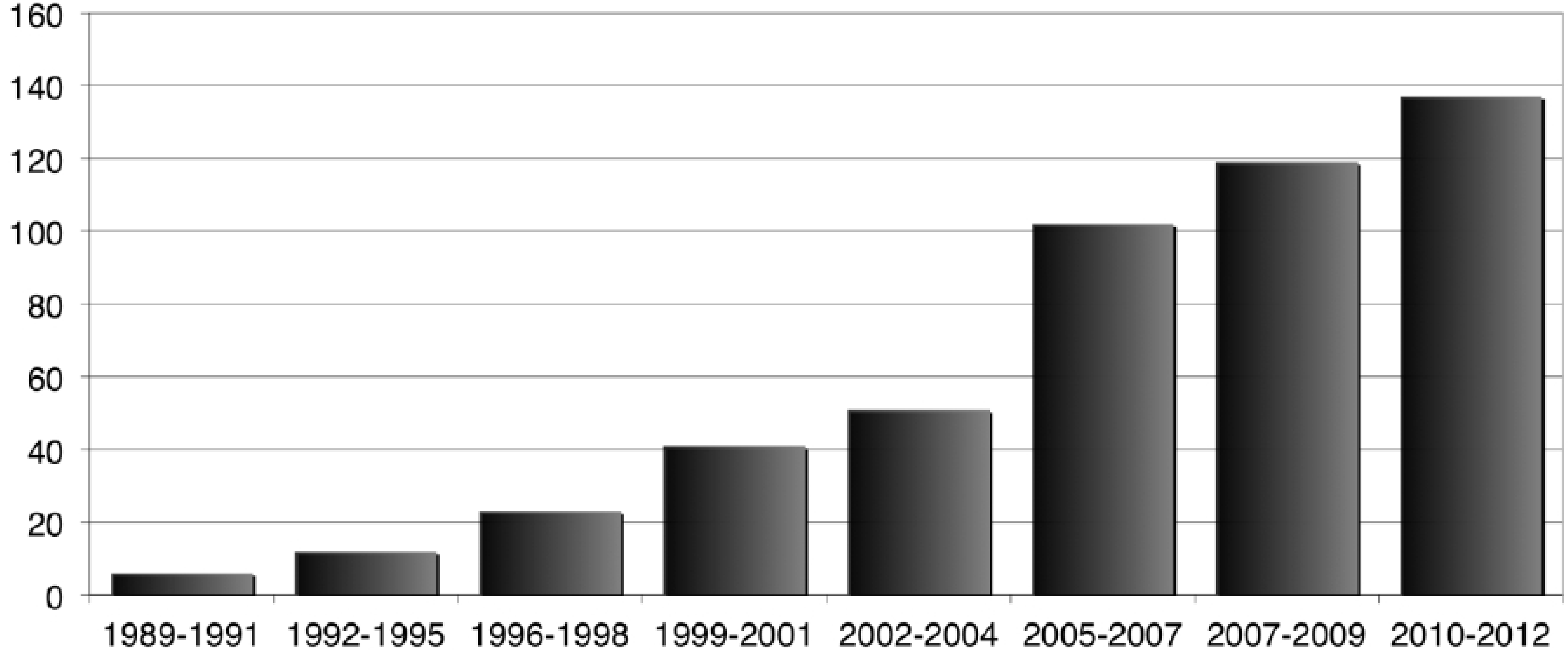
Evolutionary genetics
What are the genes and mutations responsible for phenotypic differences?

SIMILARITIES

Where do we come from?

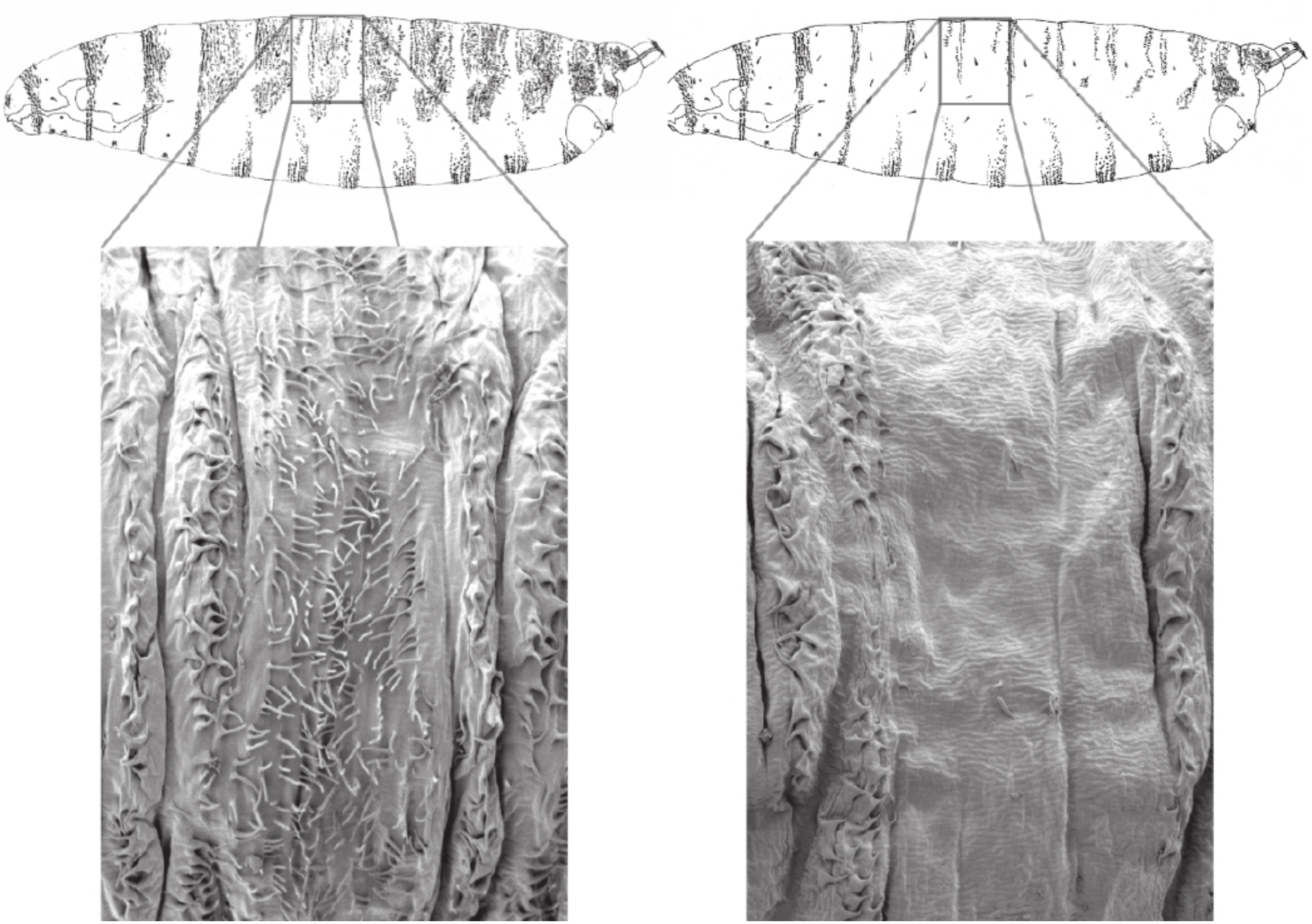
1859: Darwin and Wallace. Theory of evolution.
1920-40s: population genetics (evolution = change in allele frequencies)
Evolutionary genetics: what is the origin of phenotypic changes?

Number of alleles identified as responsible for a phenotypic difference in animals and plants (including domestication)

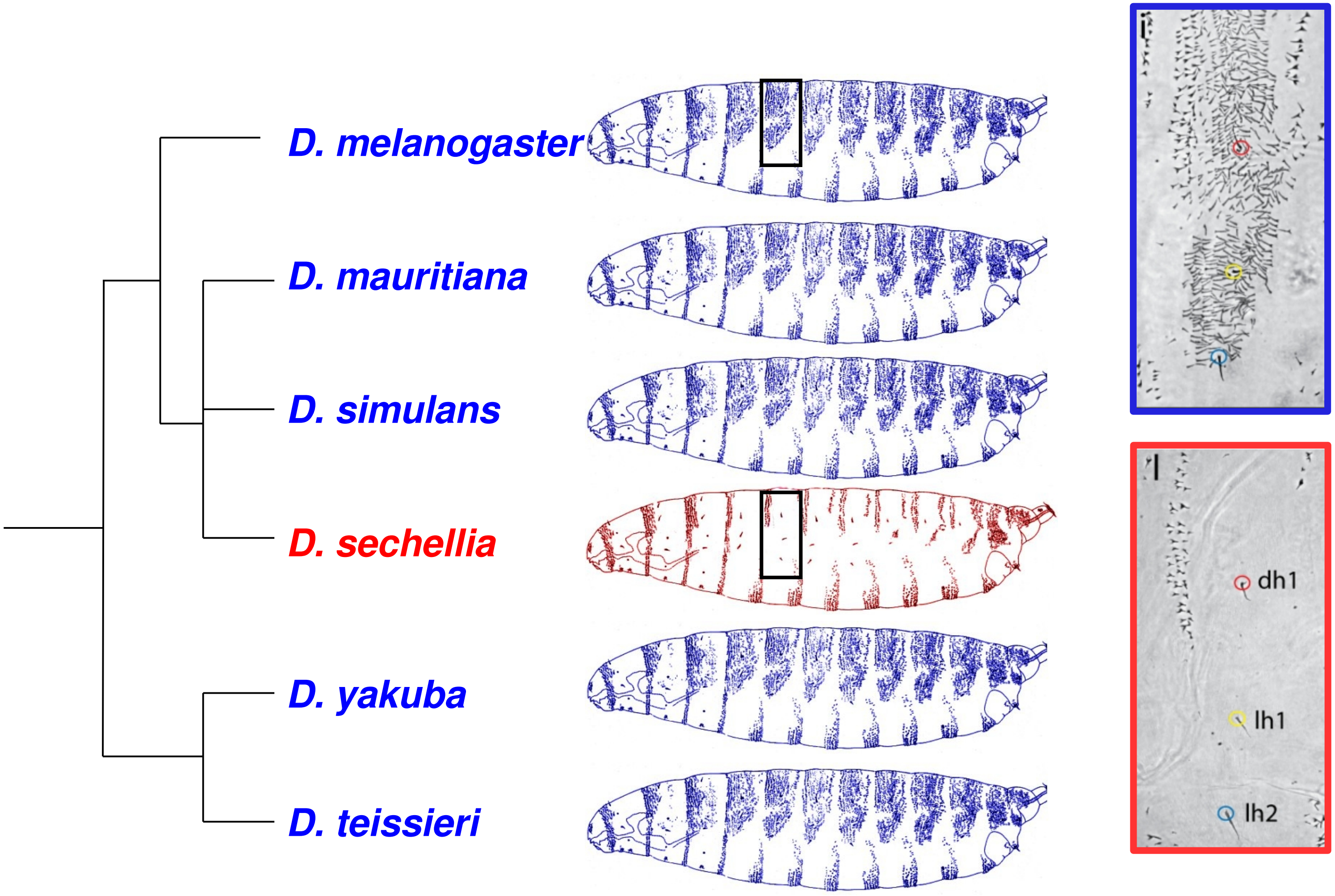


Martin and Orgogozo 2013 Evolution

Hair pattern in flies



Succena and Stern 2000 PNAS
McGregor, Orgogozo et al. 2007 Nature
Frankel, Wangl and Stern 2012 PNAS



D. melanogaster

D. mauritiana

D. simulans

D. sechellia

D. yakuba

D. teissieri

The causing mutation is on the X chromosome

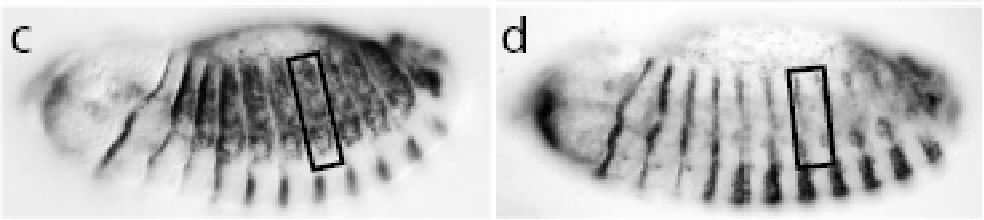
Sucena and Stern, 2000

The causing mutation is on the X chromosome

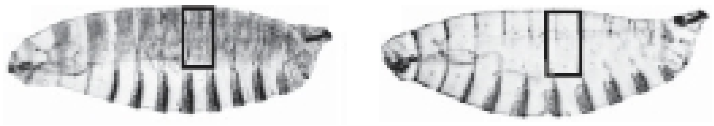
Evolution caused by a change in the *svb* gene

1 **Transcription factor that promotes trichome formation**

2 **Correlation between expression pattern and phenotype**



D. melanogaster ***D. sechellia***



D. melanogaster ***D. sechellia***

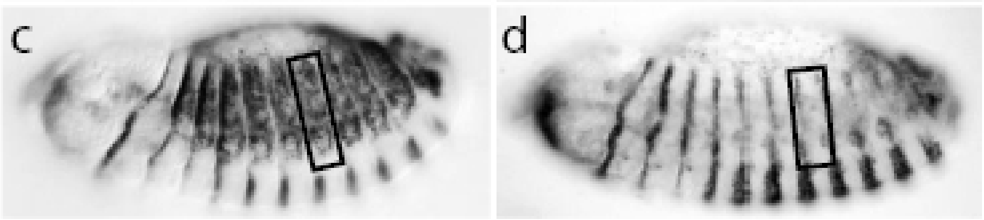
Sucena and Stern, 2000

The causing mutation is on the X chromosome

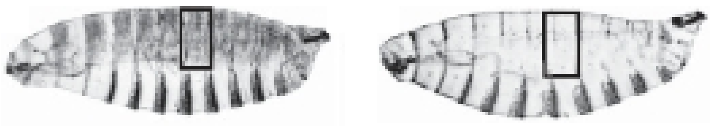
Evolution caused by a change in the *svb* gene

1 **Transcription factor that promotes trichome formation**

2 **Correlation between expression pattern and phenotype**



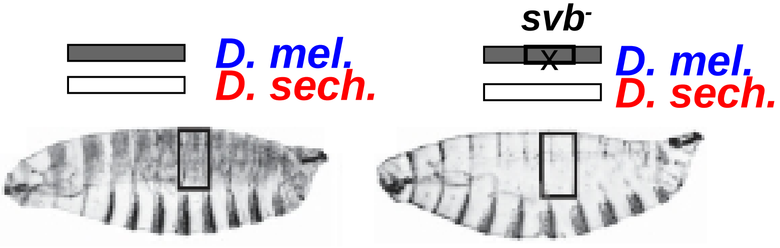
D. melanogaster ***D. sechellia***



D. melanogaster ***D. sechellia***

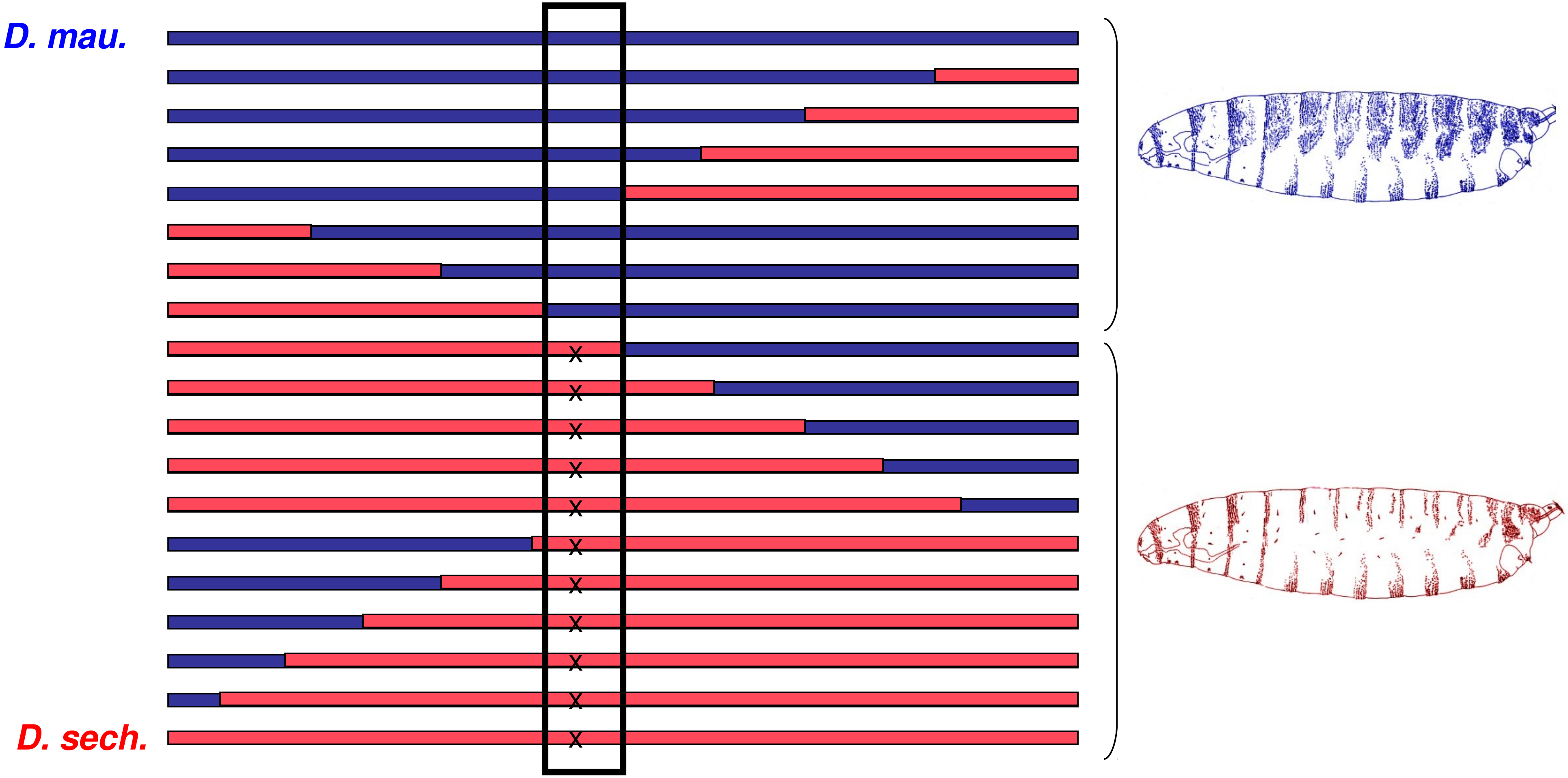
3 **Complementation assay with *D. melanogaster* svb^{WT} or svb^{-}**

hybrids *D. mel./D.sech.*

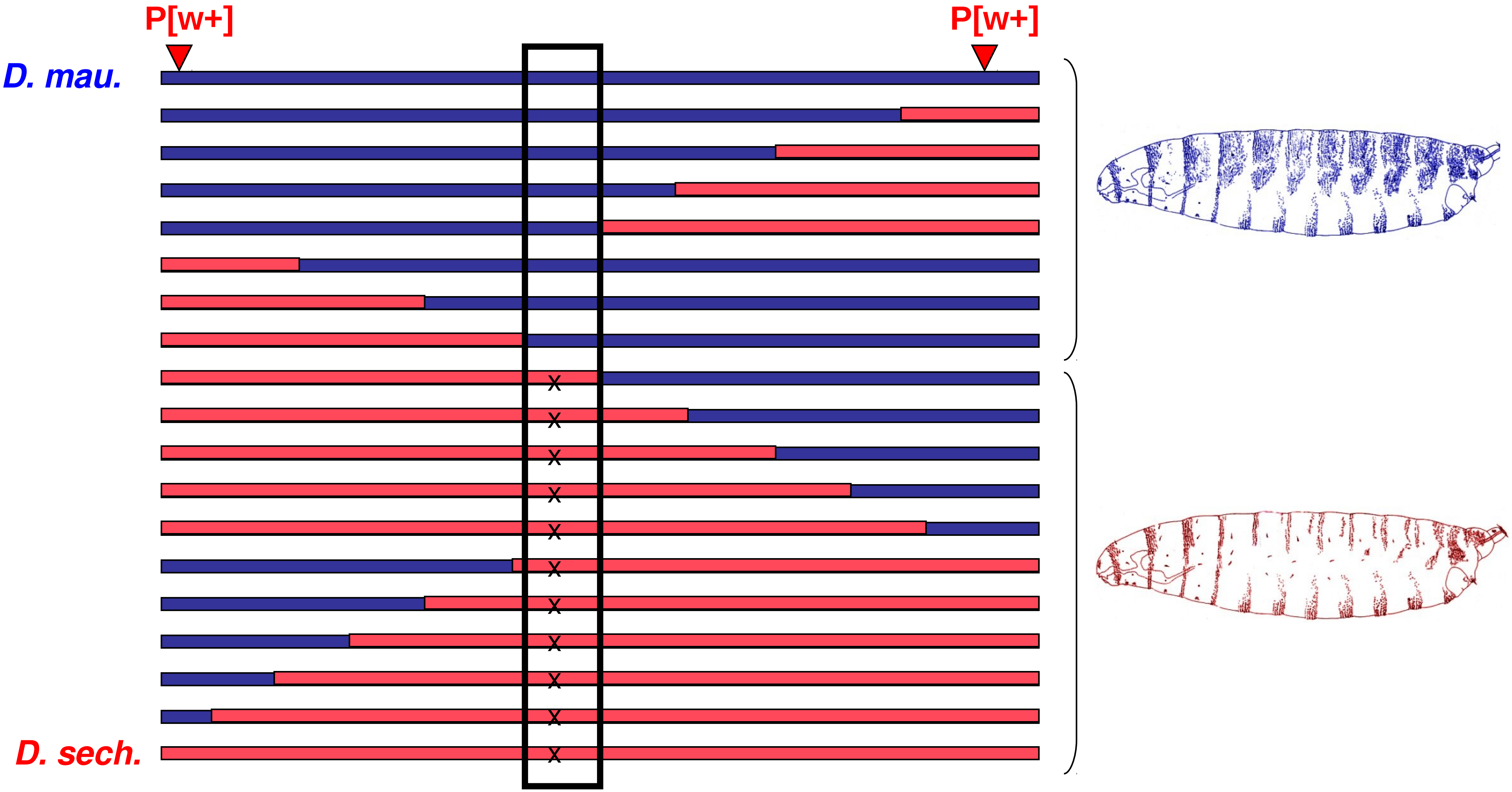


Sucena and Stern, 2000

Where is the mutation ?

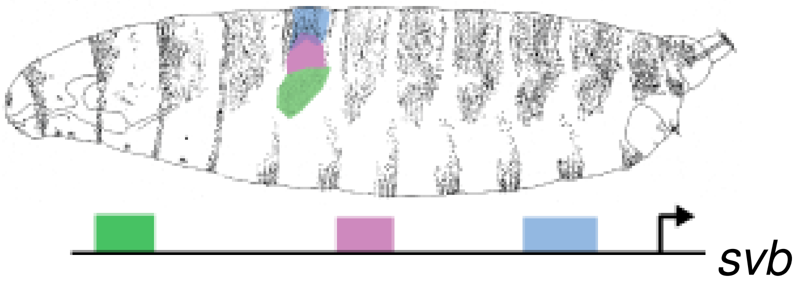


Where is the mutation ?

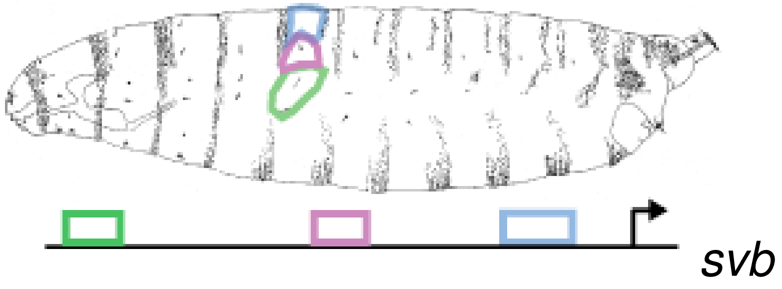


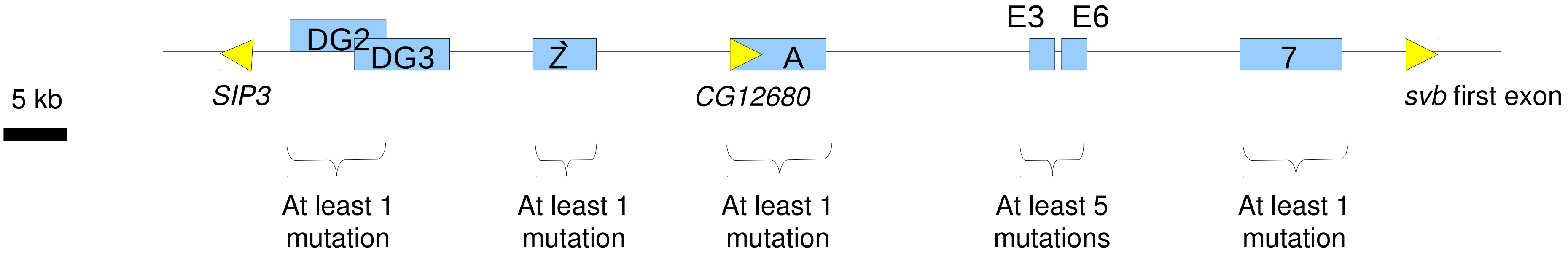
Several mutations in a single gene

D. mauritiana



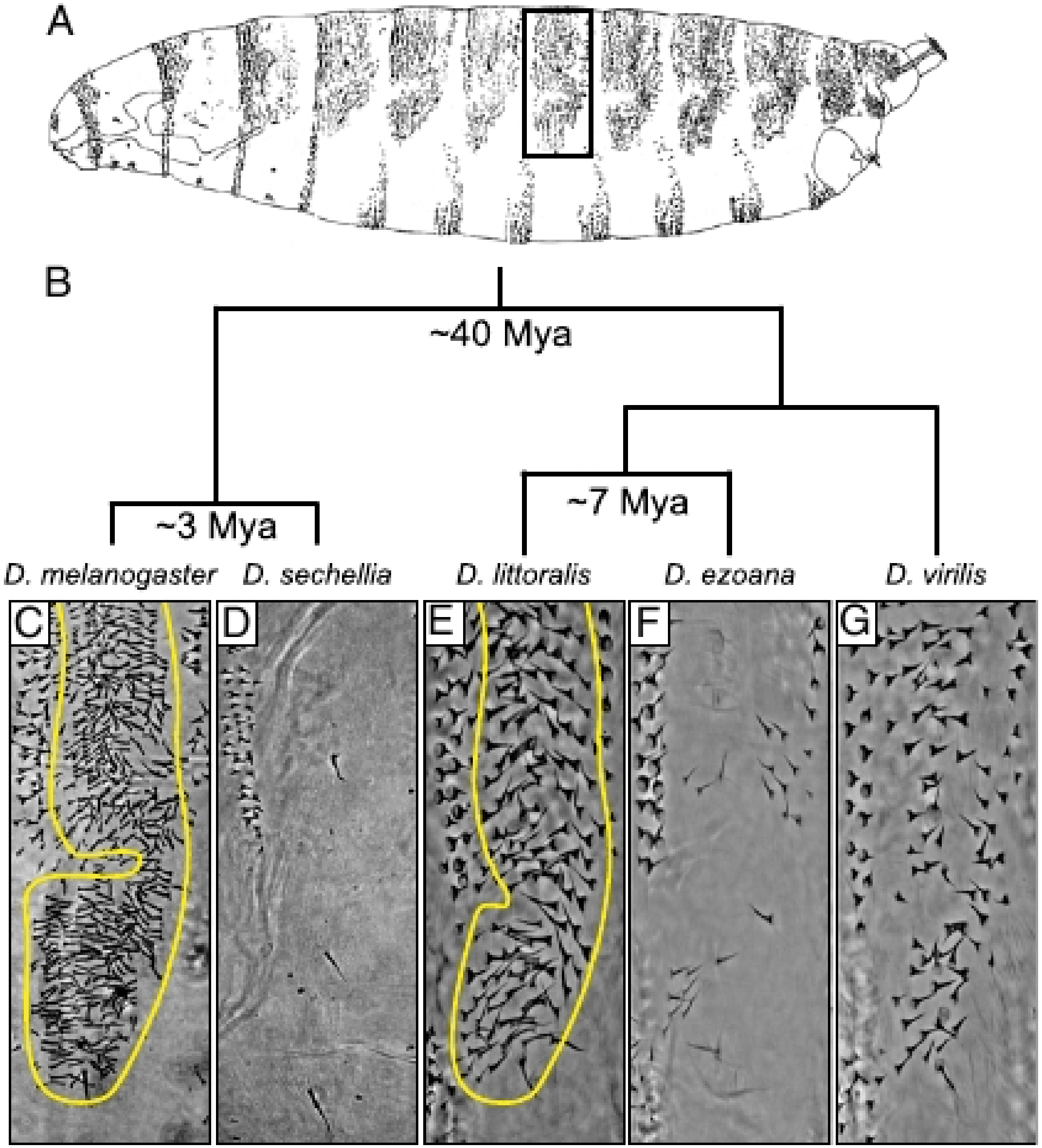
D. sechellia



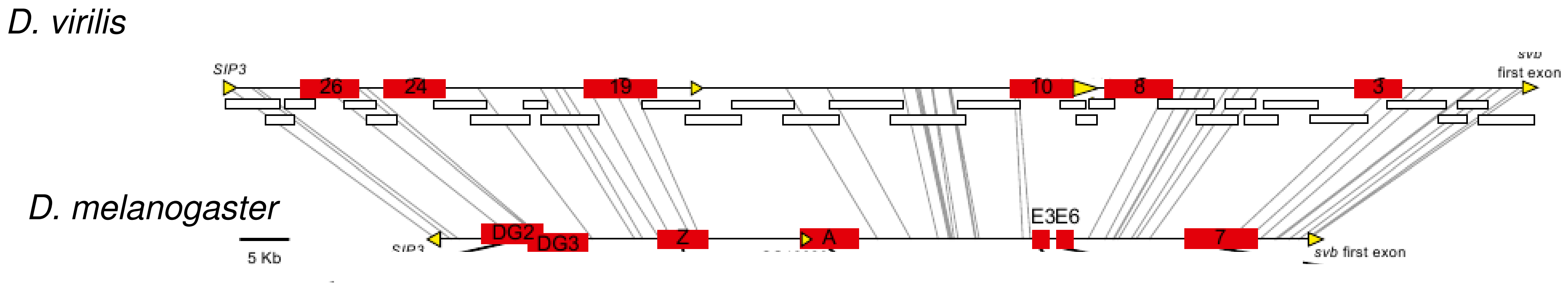


McGregor, Orgogozo et al. 2007 Nature
Frankel et al 2010 Nature
Frankel et al 2011 Nature

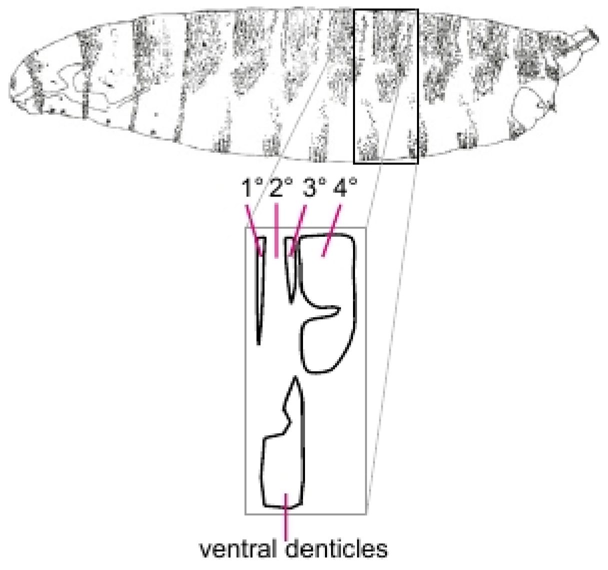
Convergent evolution in *D. ezoana*



Frankel, Wangl and Stern 2012 PNAS

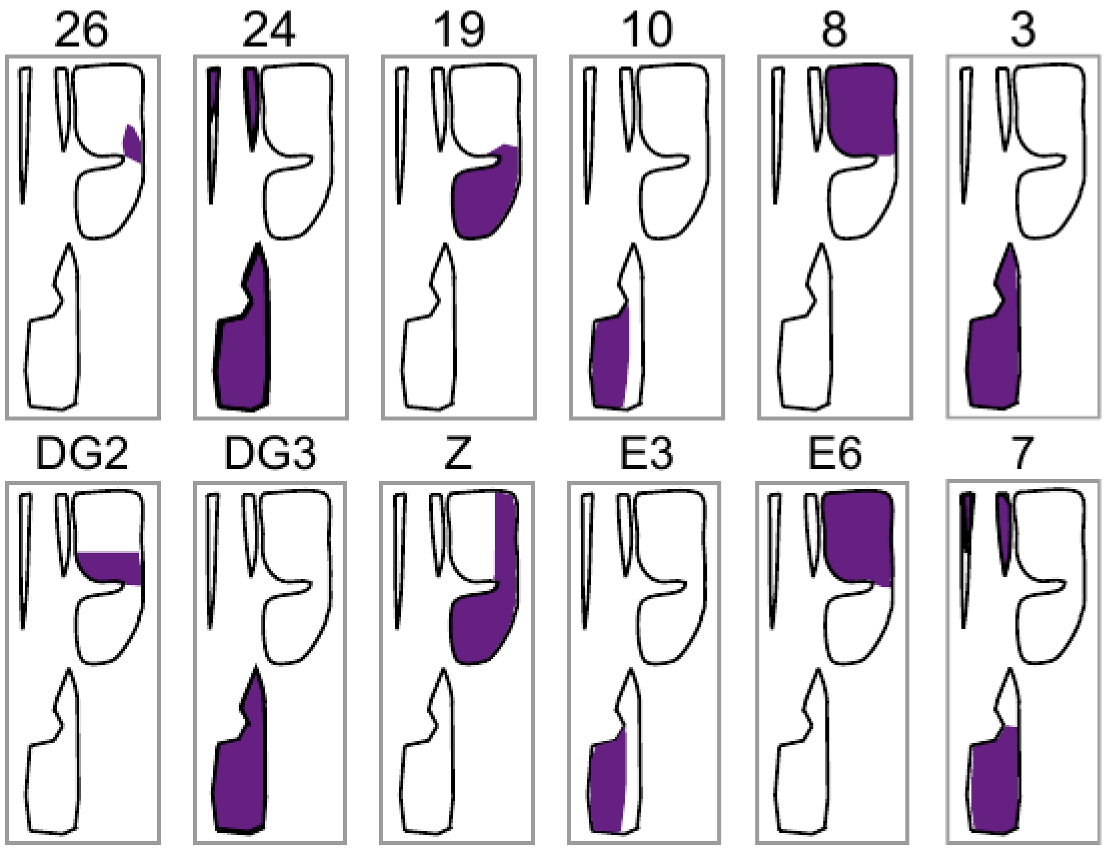


Links = 30-bp-conserved sequences



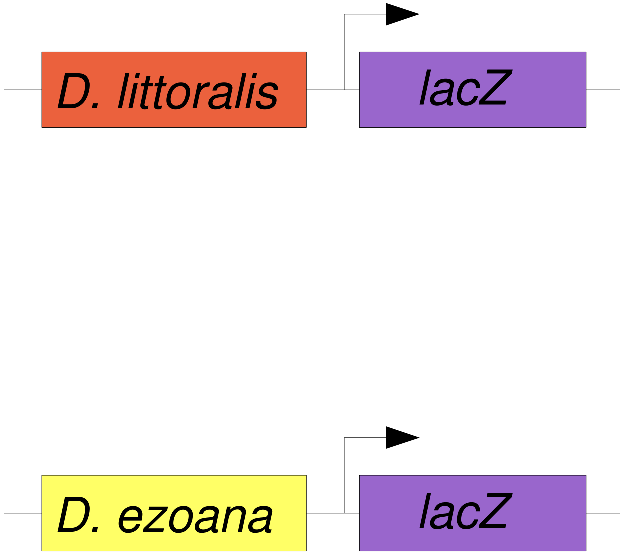
D. virilis enhancer constructs

D. melanogaster enhancer constructs



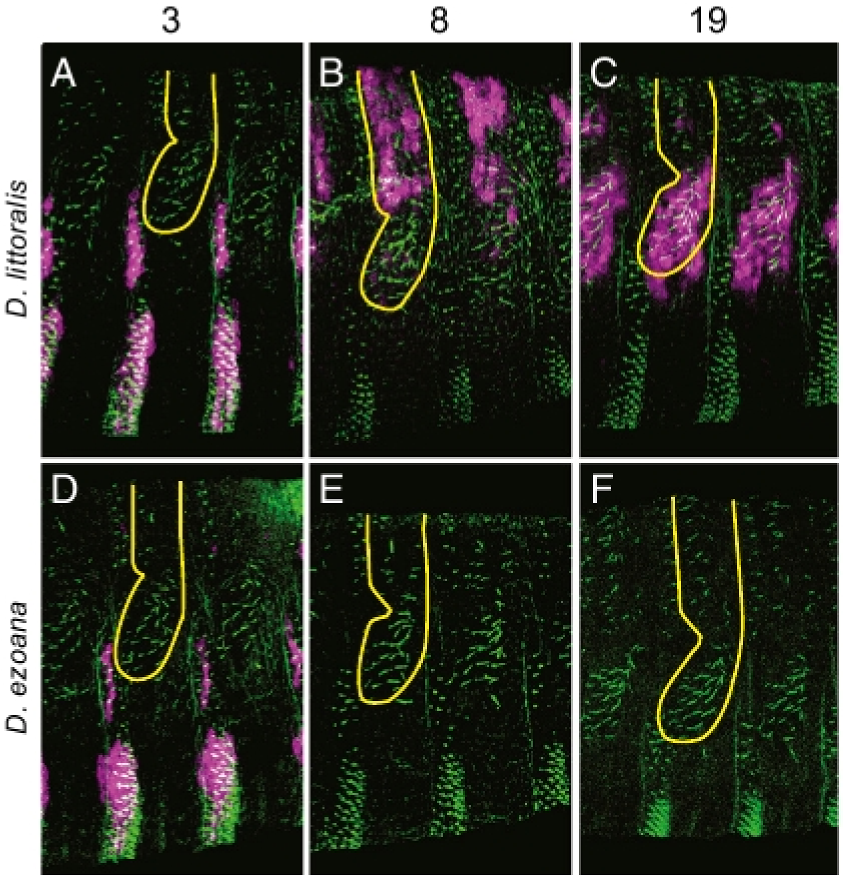
At least two cis-regulatory mutations

Reporter genes tested in *D. virilis*



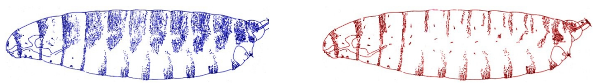
D. littoralis → *lacZ*

D. ezoana → *lacZ*



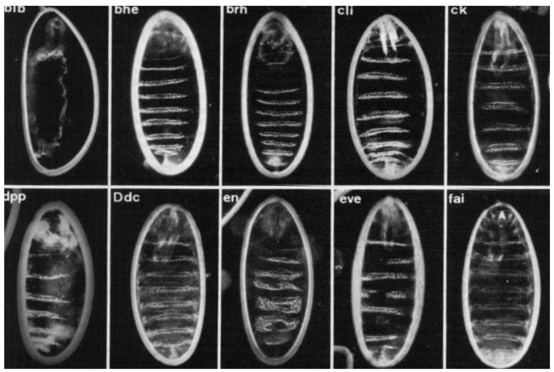
svb is
a hotspot gene
for evolution

EVOLUTION



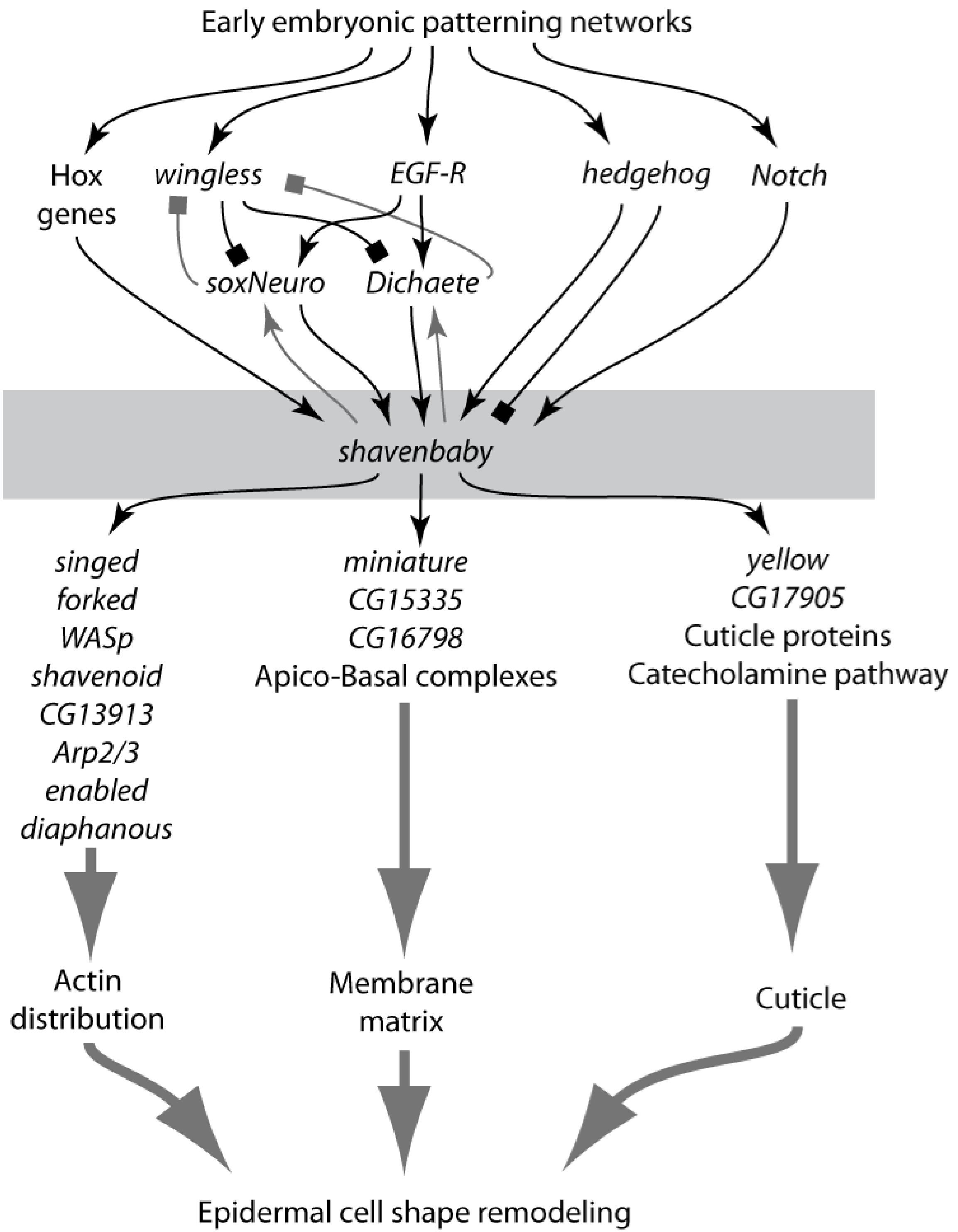
A single gene

MUTAGENESIS



(Nüsslein-Volhard et Weischaus)

~100 genes



Early embryonic patterning networks

Hox genes

wingless

EGF-R

hedgehog

Notch

soxNeuro

Dichaete

shavenbaby

singed
forked
WASp
shavenoid
CG13913
Arp2/3
enabled
diaphanous

miniature
CG15335
CG16798
Apico-Basal complexes

yellow
CG17905
Cuticle proteins
Catecholamine pathway

Actin distribution

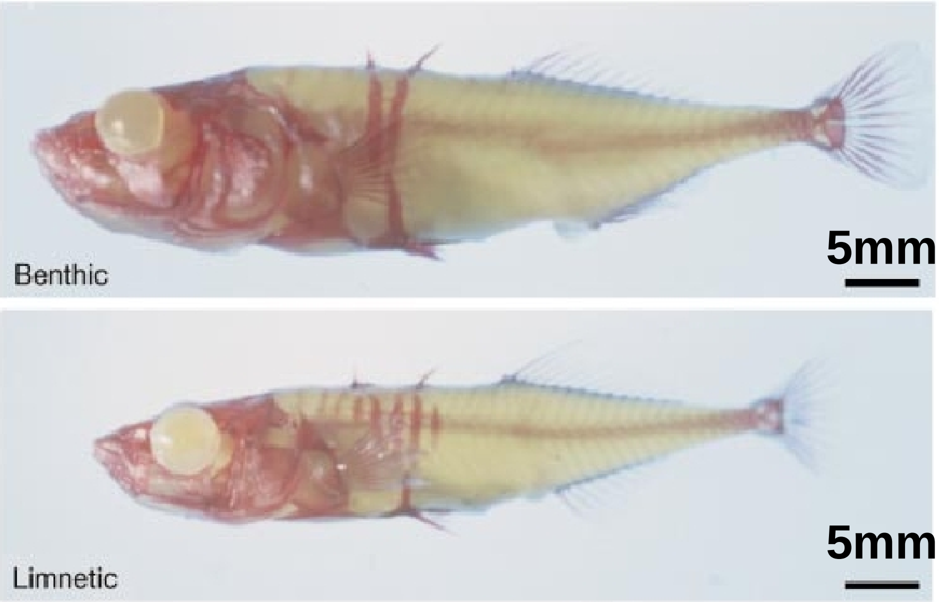
Membrane matrix

Cuticle

Epidermal cell shape remodeling

(Stern et Orgogozo, 2008 Evolution – Stern et Orgogozo, 2009 Science)

Pelvic reduction in sticklebacks



marine

Paxton Lake, Canada

Gasterosteus aculeatus

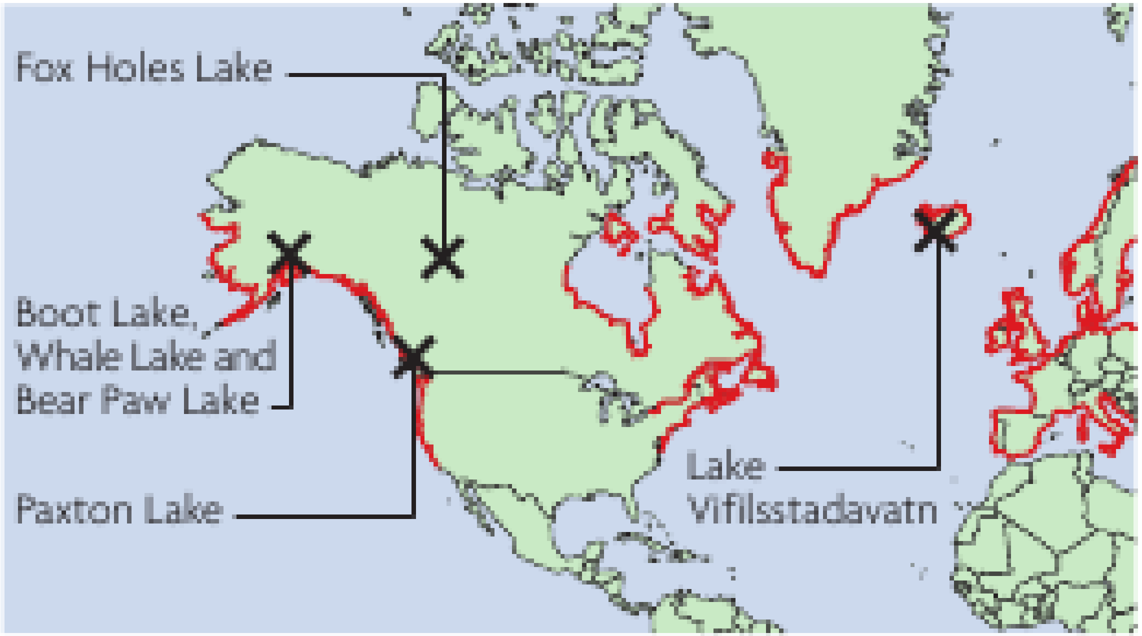
(Peichel et al., 2001 ; Shapiro et al, 2004 ; Chan et al. 2010)

Marine fishes with robust pelvis = ancestral

Freshwater fishes with reduced pelvic structures = derived, independently at least 20 times

- limited calcium availability
- absence of gape-limited predatory fishes
- predation by grasping insects

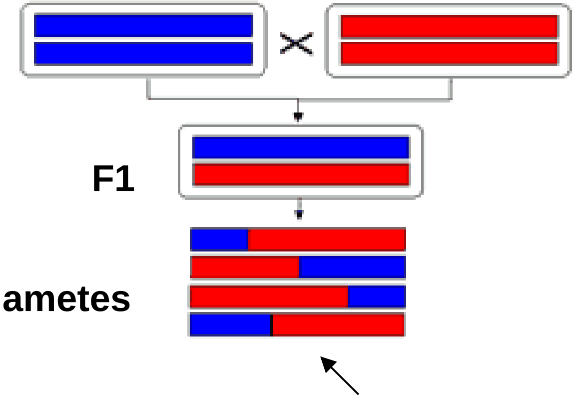
Last glacier retreat = 10 000 – 20 000 years ago



QTL mapping

lake

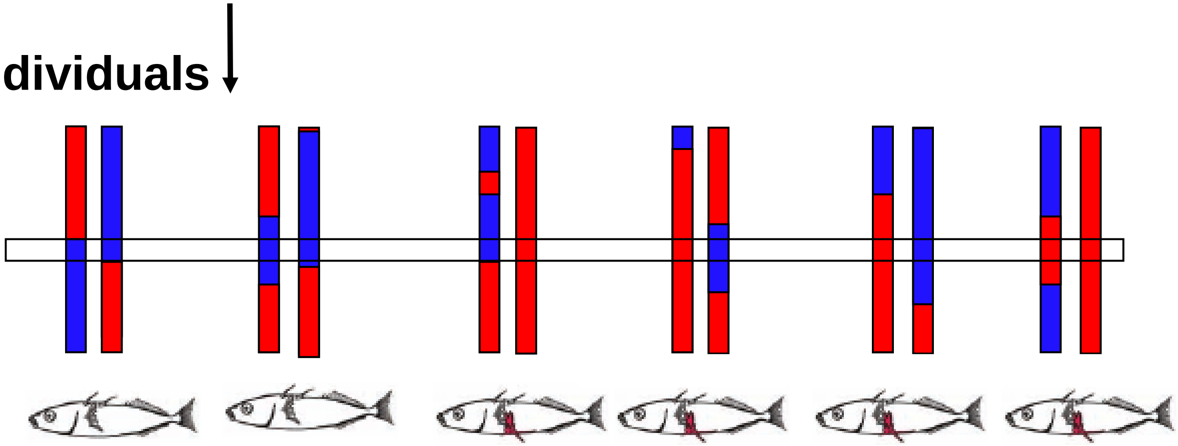
marine



parents

F1

gametes



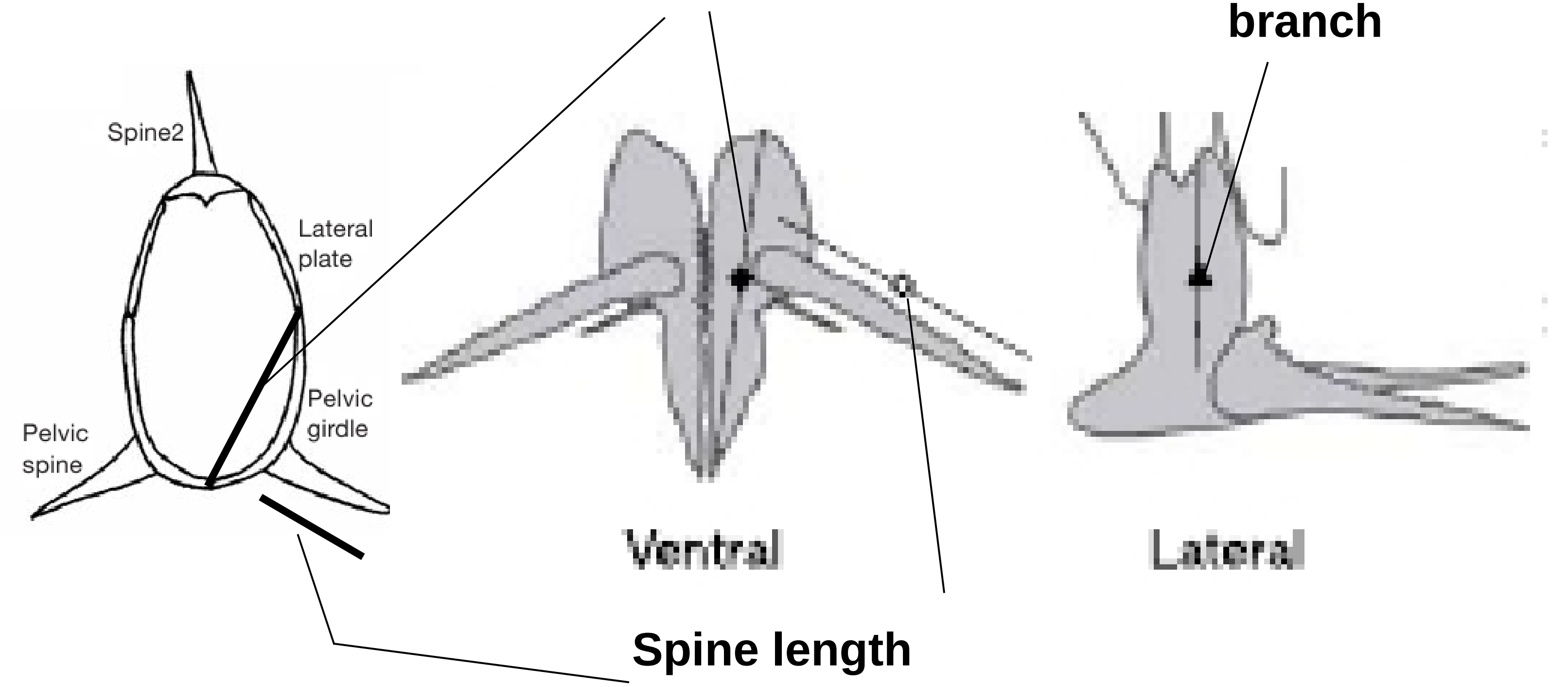
375 F2 individuals

QTL?

phenotype

(Shapiro et al., 2004)

Quantitative measurement of the phenotype



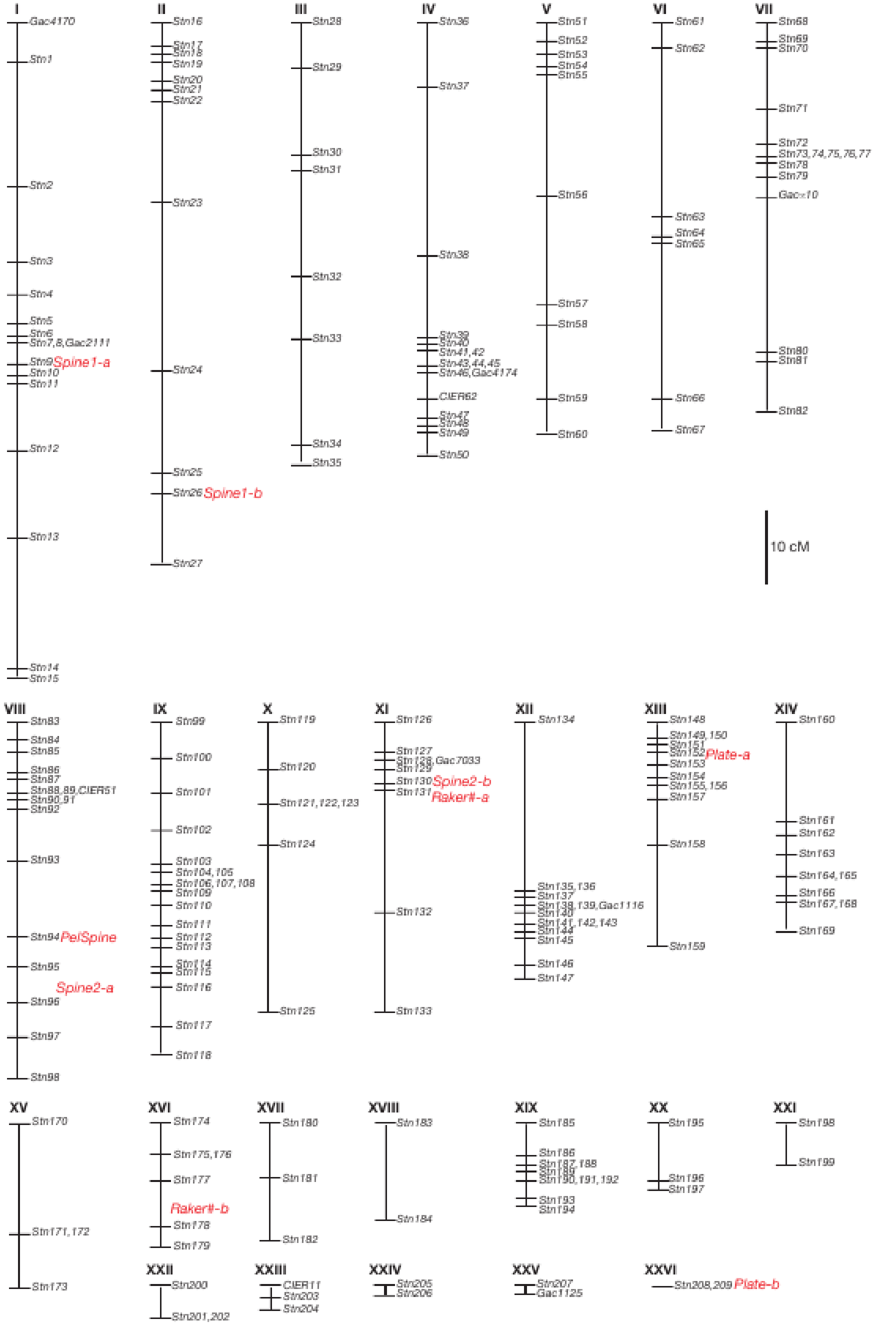
Length of pelvic girdle

Height of ascending branch

Spine length

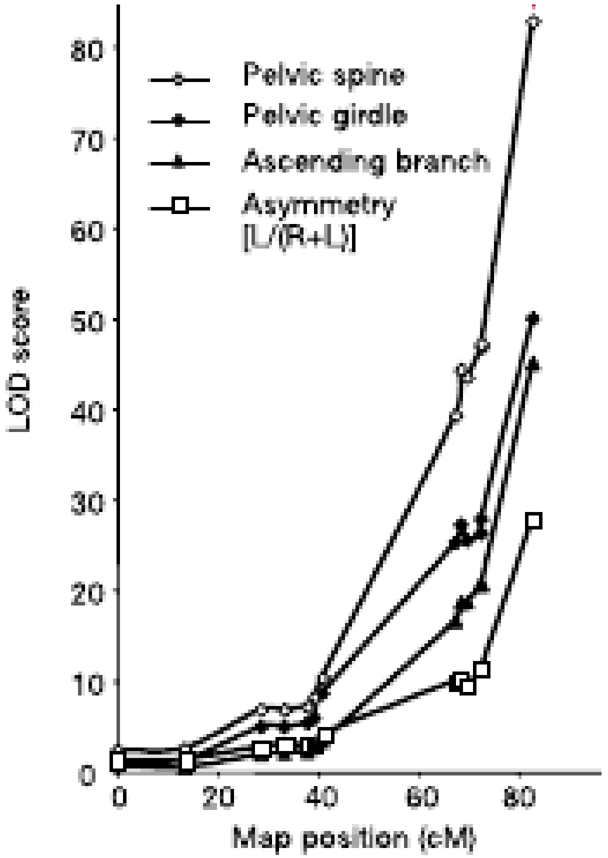
1000 microsatellite markers

26 linkage groups



(Peichel et al., 2001)

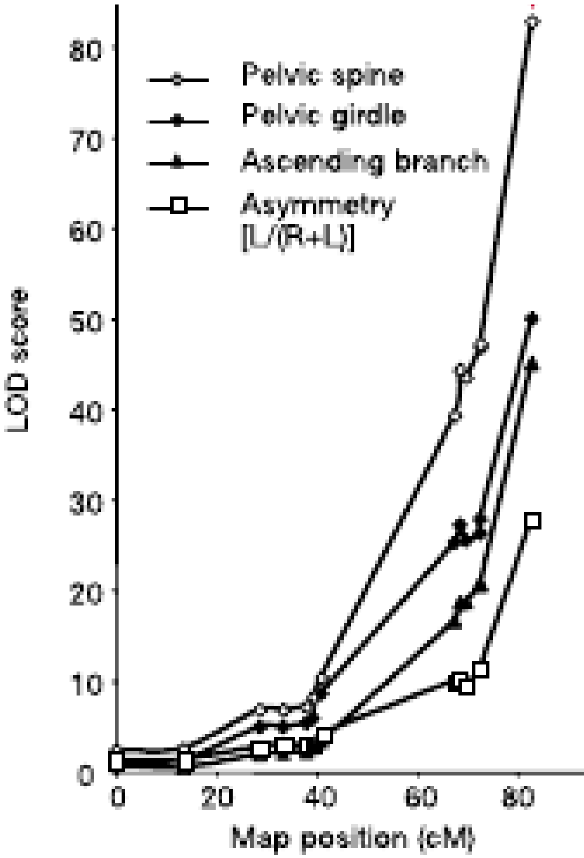
One major locus at the end of linkage group 7

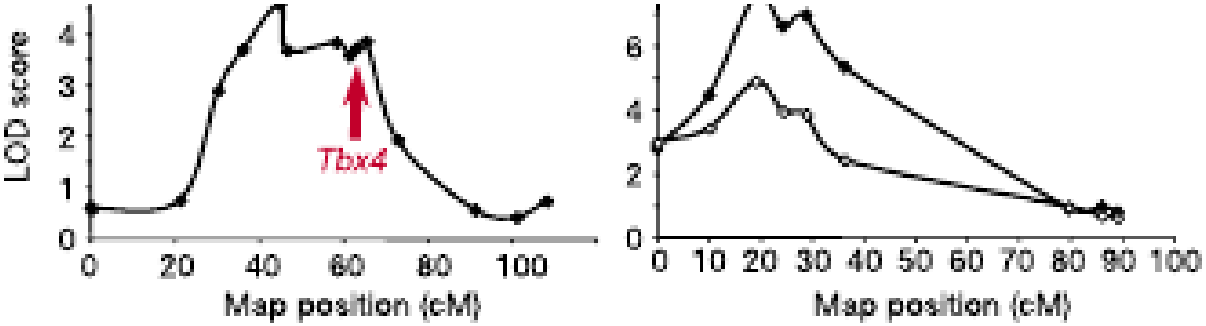


Major locus responsible for 65% of the variance

One major locus at the end of linkage group 7

A few minor loci





Major locus responsible for 65% of the variance

Three candidate genes: Pitx1, Pitx2, Tbx4

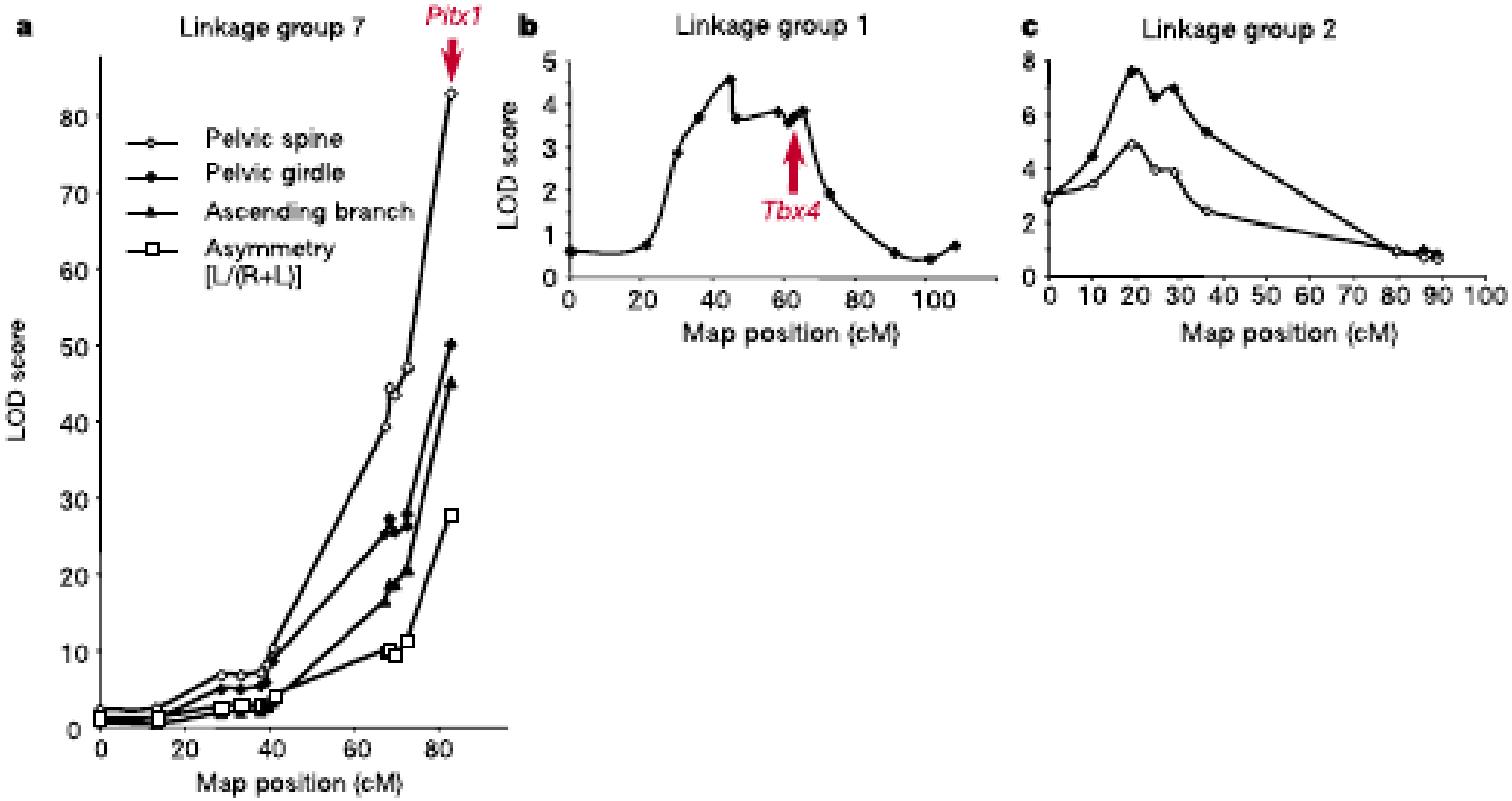
Screen of a BAC library (Bacterial Artificial Chromosomes)
containing 100-350-kb fragments of stickleback genomic DNA

Partial sequencing and generation of new markers

Genotyping of 375 F2 individuals

One major locus at the end of linkage group 7

A few minor loci



Pitx1, responsible for the phenotypic change?

Pitx1 **null mutations in mice**
(pelvis reduction, stronger on right side)

QTL mapping

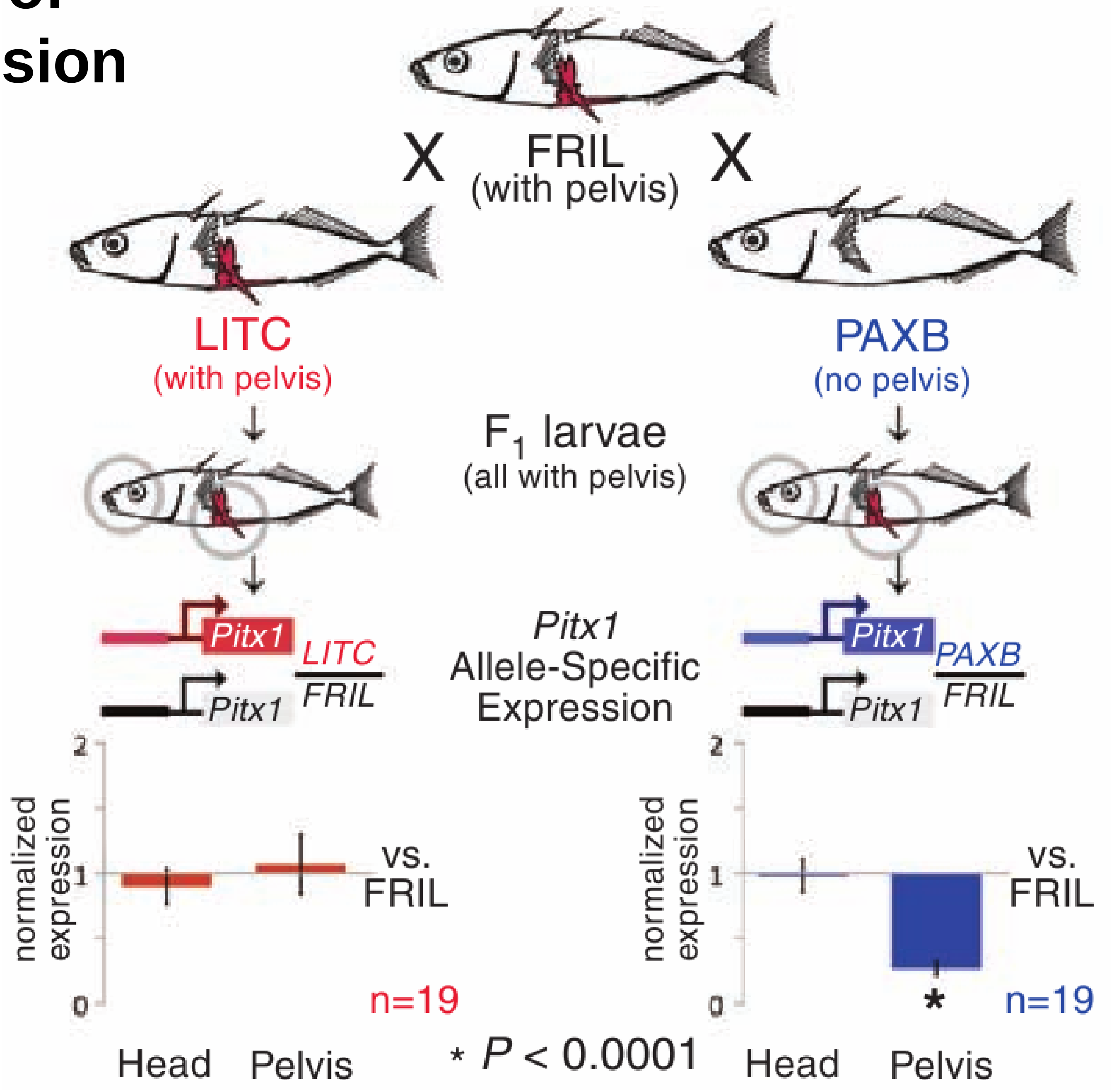
Same coding sequence in lake and marine forms

Pitx1 **expressed at stage 29 in marine individuals but not in**
marine individuals

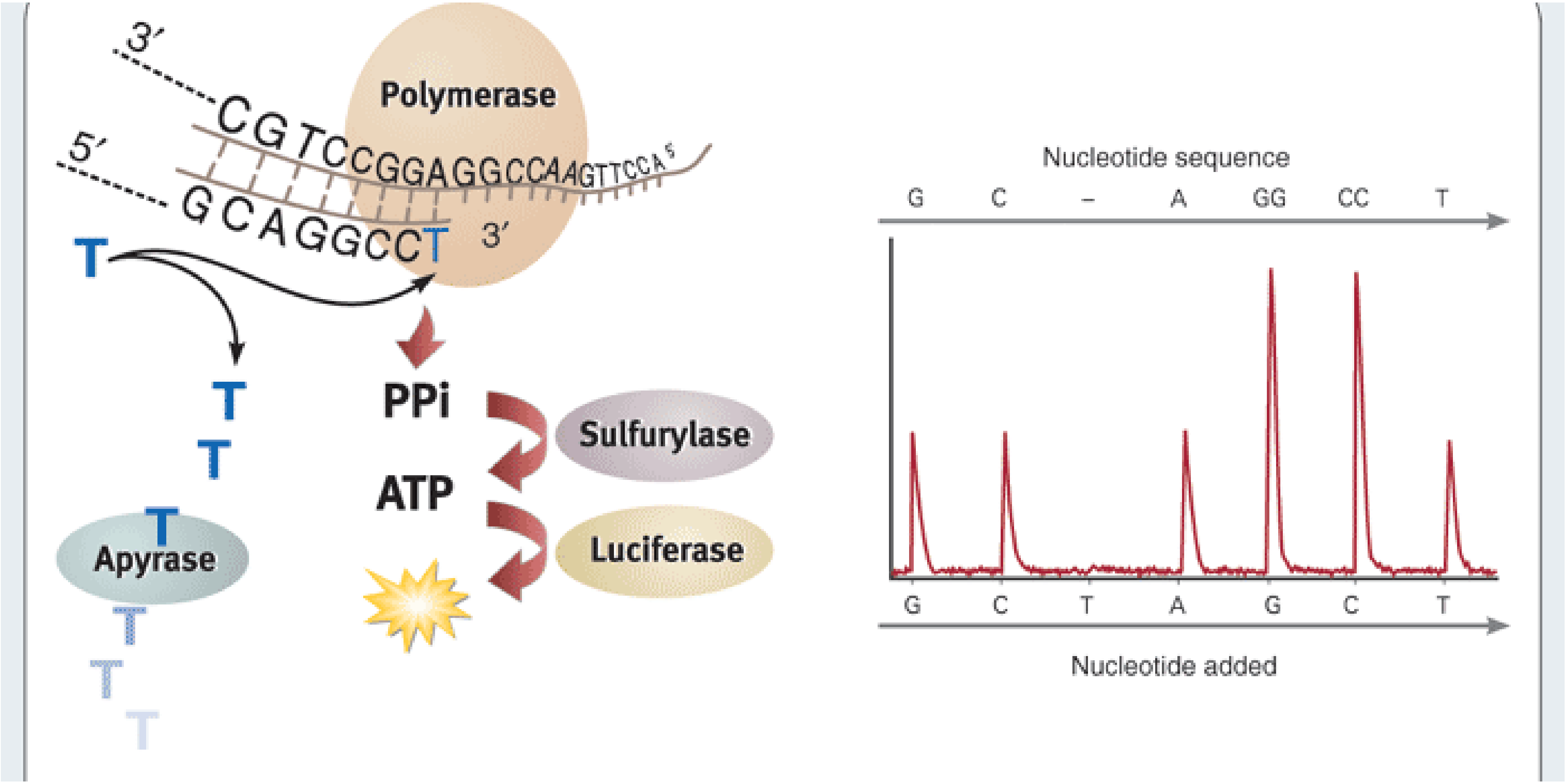
BUT

The decrease in *Pitx1* expression levels might have evolved due to mutations in an upstream regulatory gene

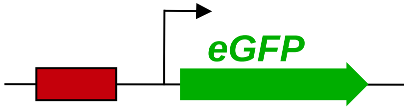
Comparison of allele expression in hybrids

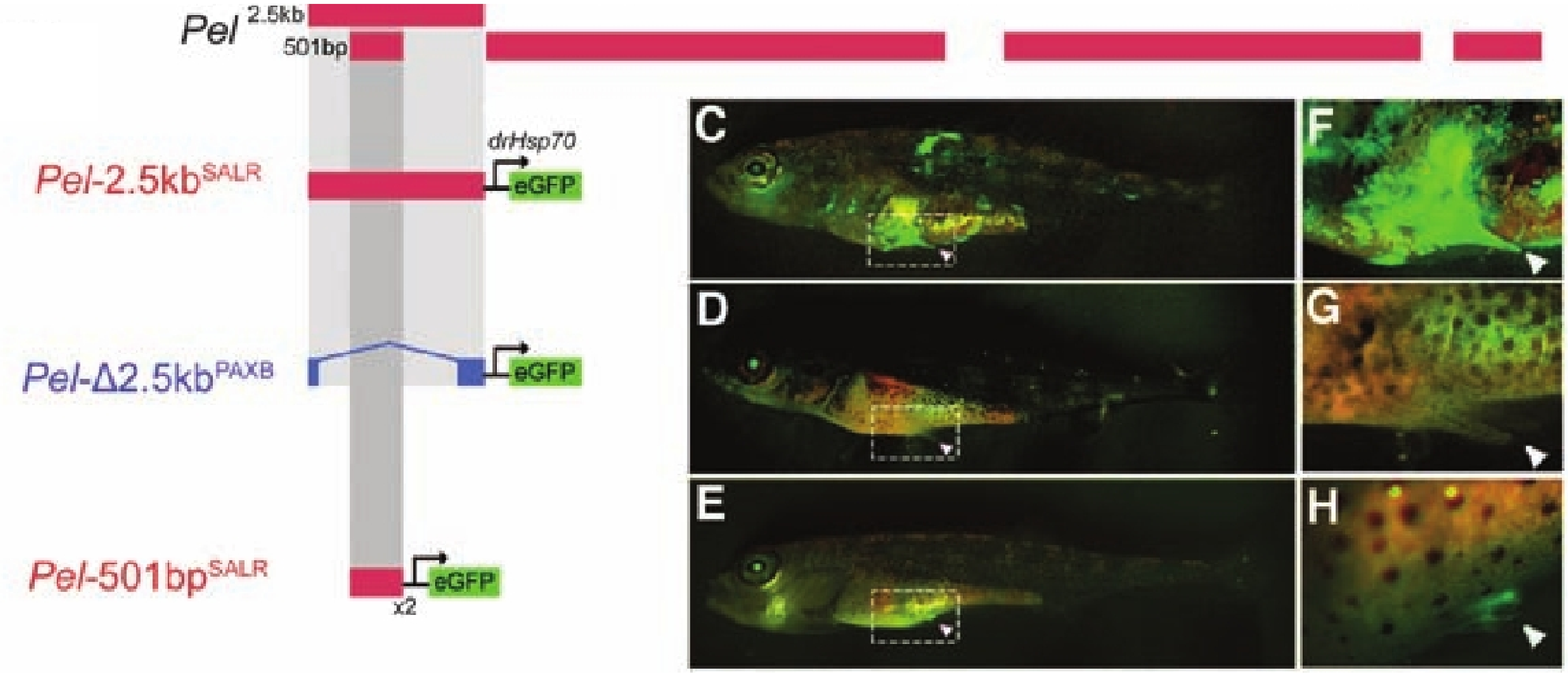


What is pyrosequencing?

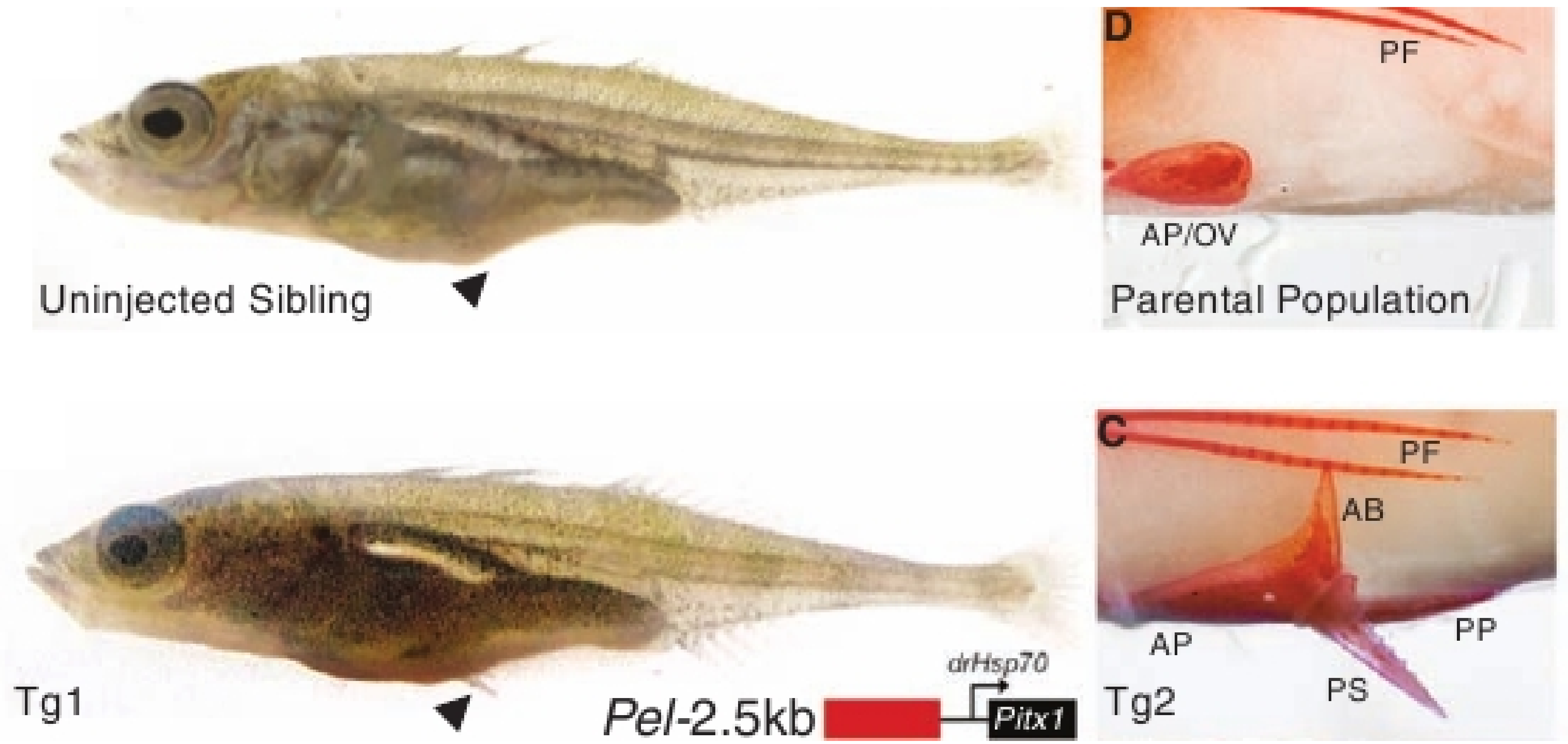


Test of *Pitx1* cis-regulatory regions





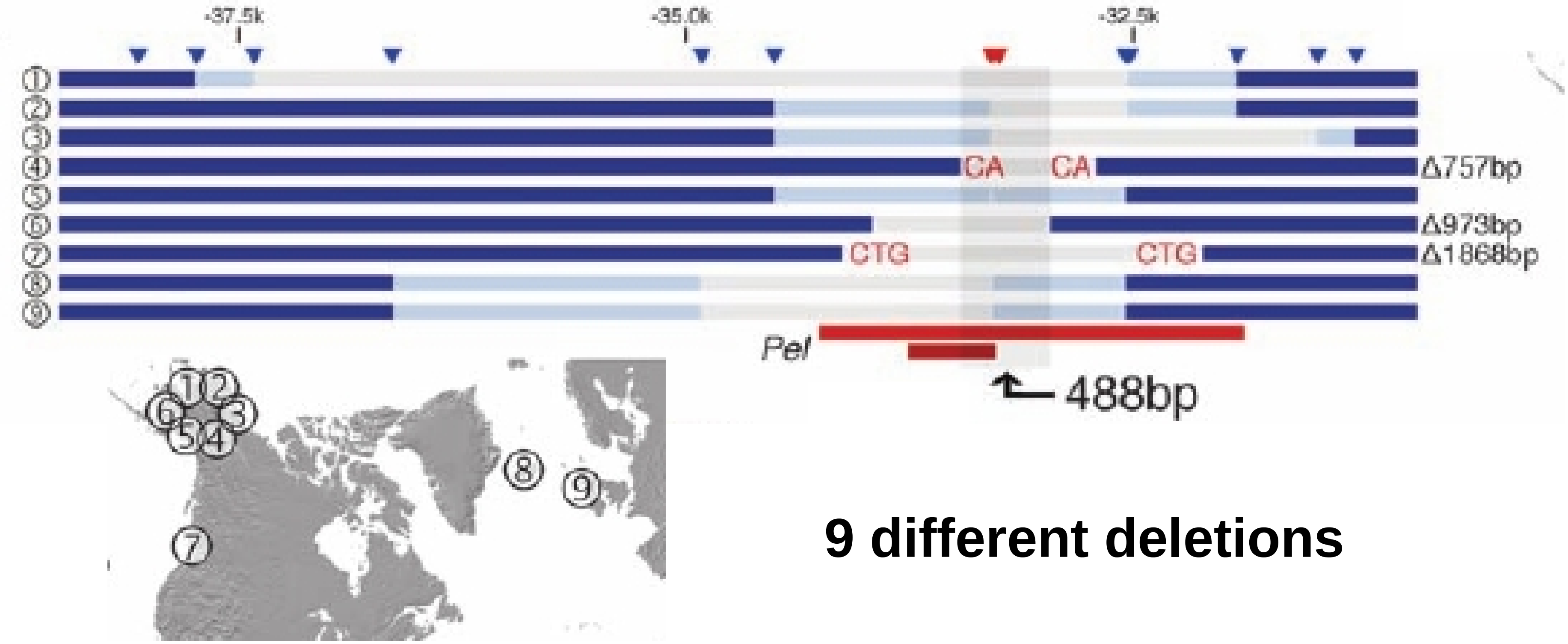
Rescue of a pelvis in freshwater individuals



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

Region sequenced in two lake pop[ulations: a 2-kb deletion in one and a 757-bp deletion in the other one

SNP genotyping in 13 populations with reduced pelvis and in 21 populations with complete pelvis



9 different deletions

Two types of appraoches

Genetic mapping	Candidate gene
no a priori, fewer biais	Based on an a priori idea
long and tedious	can be fast and efficient
rarely ends with identification of the gene	
only with strains/species which produce fertile hybrids	will only find known genes

<u>In both cases, genes with small effect are more difficult to identify</u>

Methods to identify the genes and the mutations responsible for phenotypic evolution

Genetic
which chromosome (ex: autosomal versus sex)
QTL mapping
Genetic association studies
Complementation tests

General biology
General knowledge of the genes involved in the phenotype
Similarity with a known phenotype
Correlation with a change in gene expression level/pattern

Final test of protein activity
in vitro in *E. coli*, by transgenesis in the studied species or the closest model organism (ex: *beta-defensin* of dogs tested in mouse)

Final test of cis-regulatory regions
- **with reporter constructs, transgenesis, comparison of both regions**
- **comparison of allele expression levels in hybrids (pyrosequencing)**

Genes versus Environment

Causes of these differences ?

Genetic

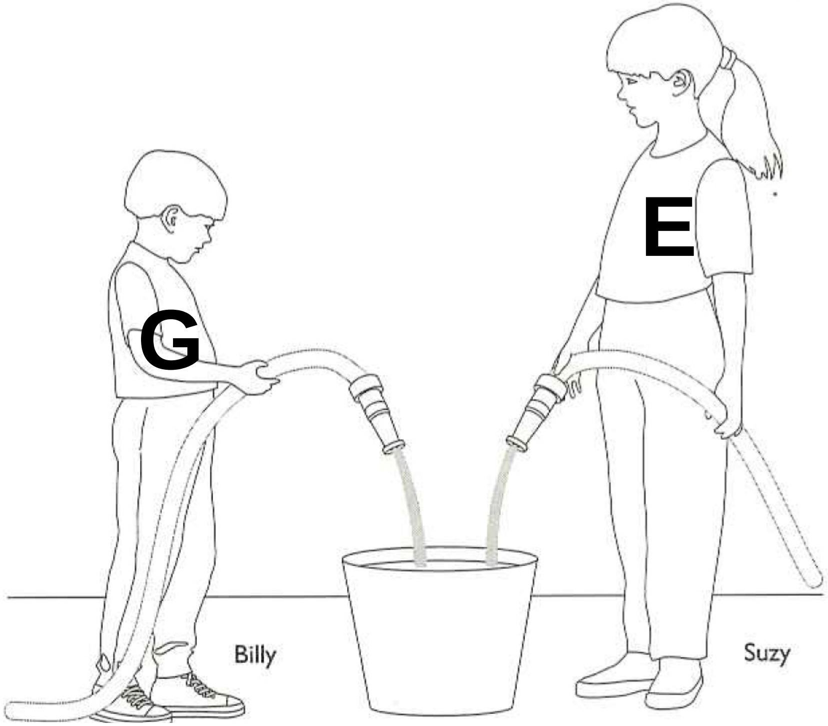
Environment

Phenotype = G + E + GxE

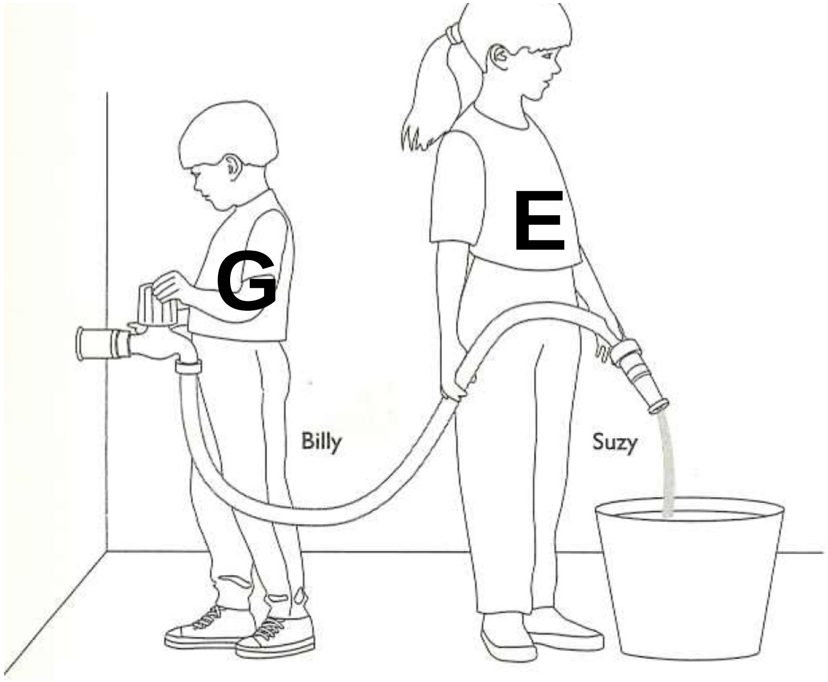
The Siamese cat
An example of GxE

Mutation in *tyrosinase*
Heat-sensitive enzyme
No production of melanin in warm body parts

Contributions of the genotype (G) and the environment (E) to phenotypic variation

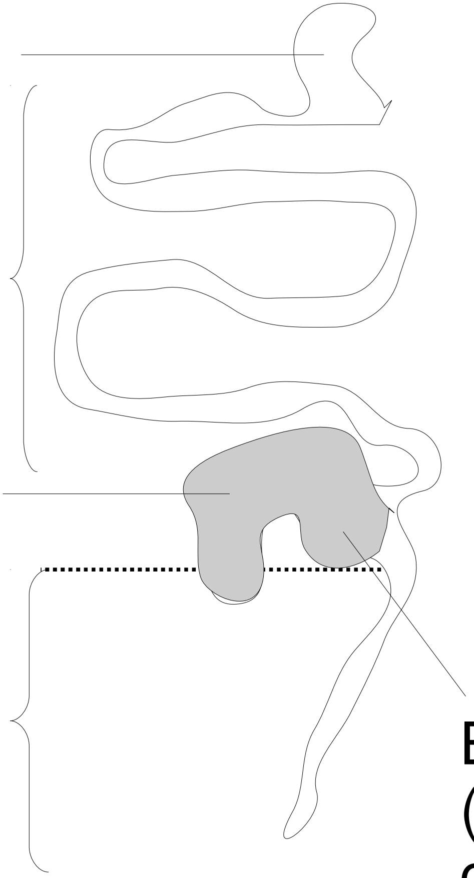


GxE



From Fox Keller (2010)

Mouse caecum development
An other example of GxE



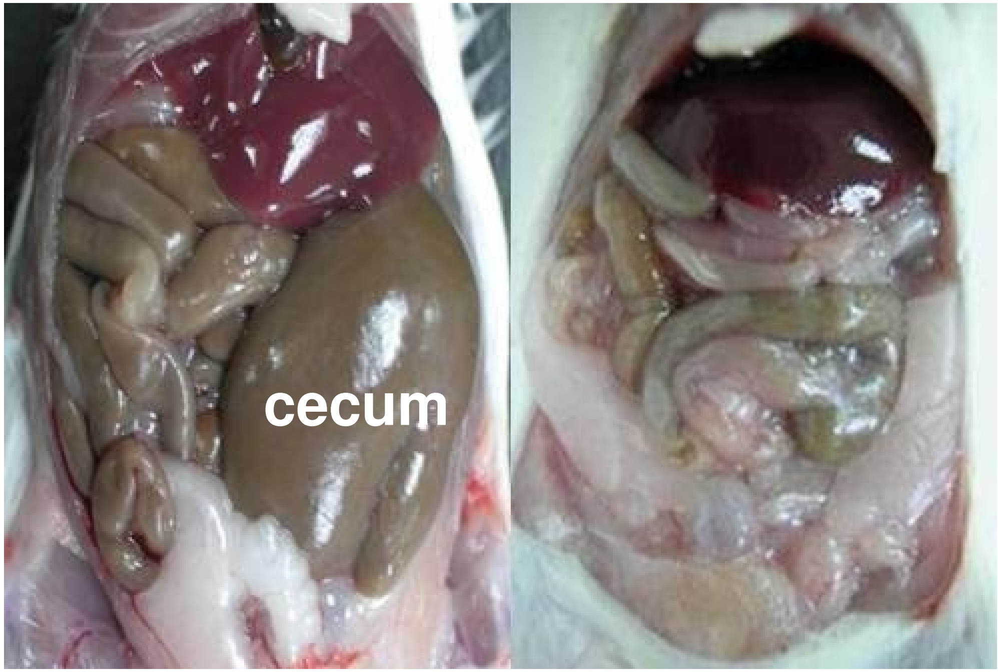
stomach

small intestine

caecum

large intestine (colon)

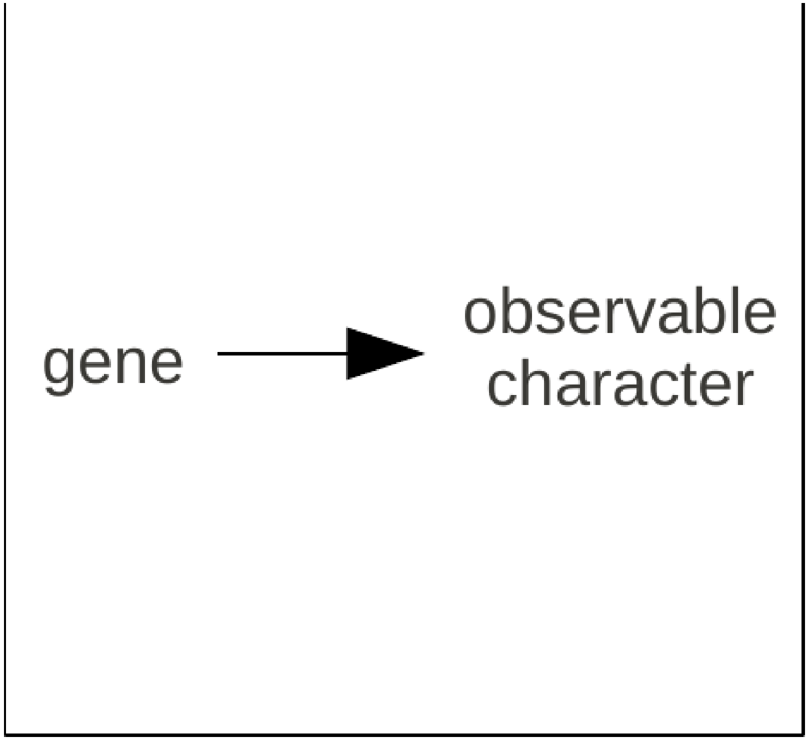
Bacteria (digest cellulose, produce vitamine K)



germfree

normal

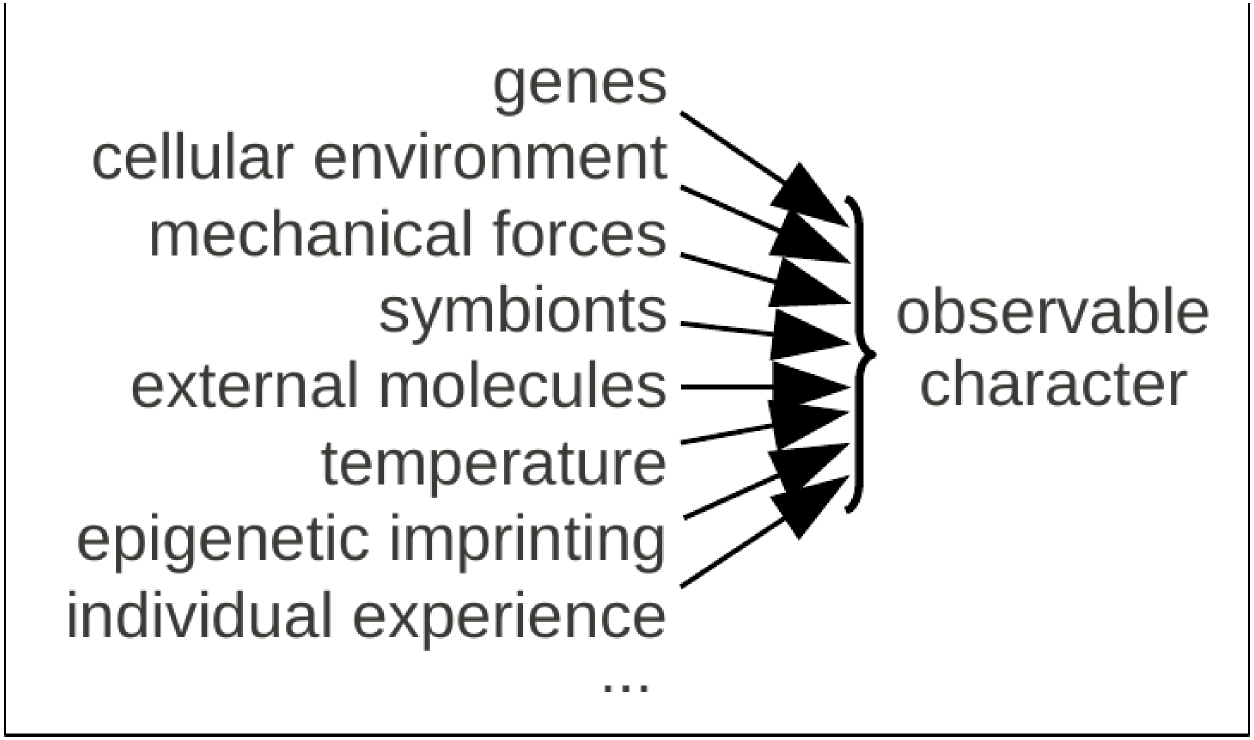
Classical genetic reductionism



gene

observable character

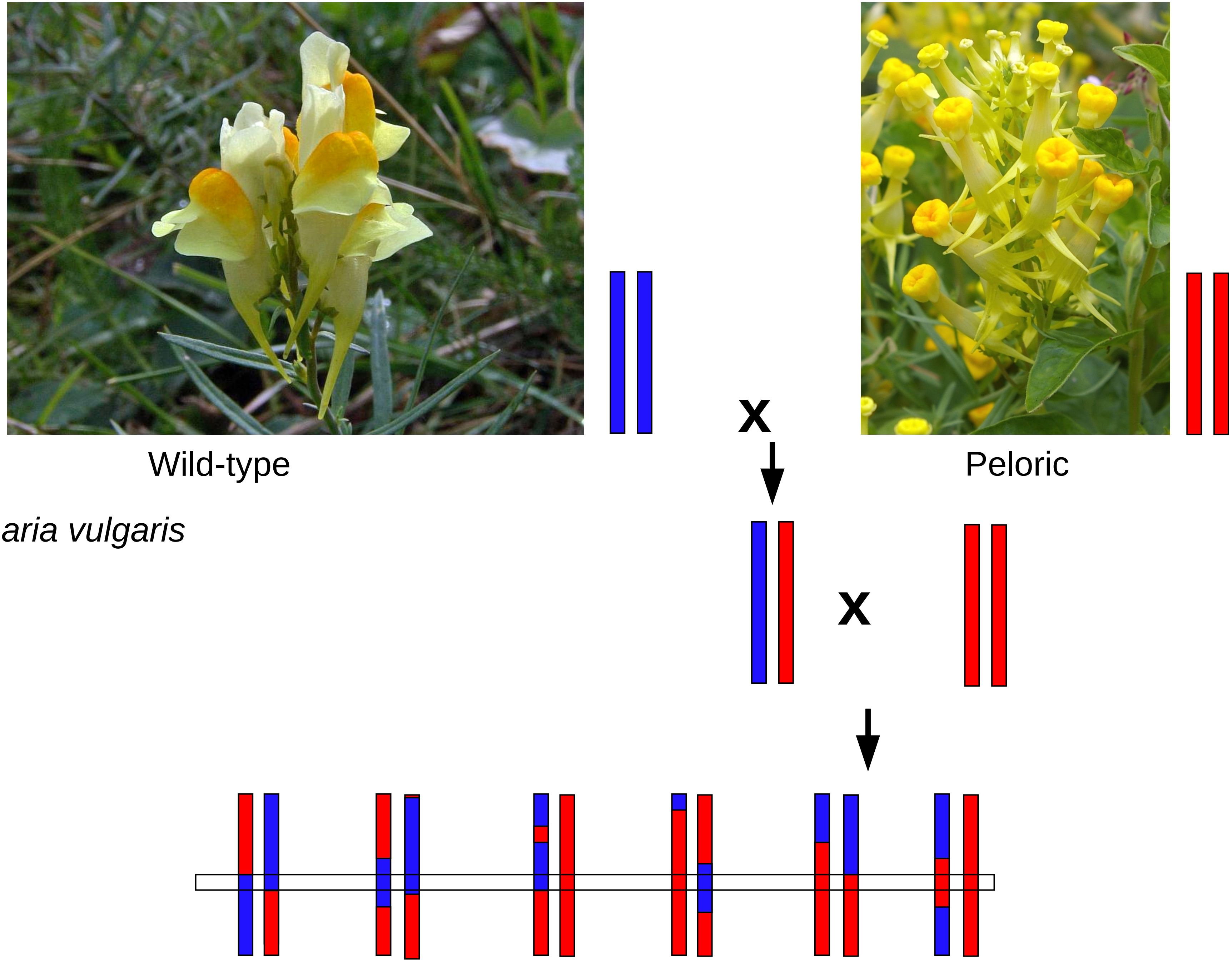
Integrative approach



genes
cellular environment
mechanical forces
symbionts
external molecules
temperature
epigenetic imprinting
individual experience
...

observable character

Orgogozo et al., submitted



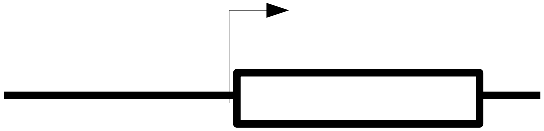
Linaria vulgaris

Cubas 1999 Nature

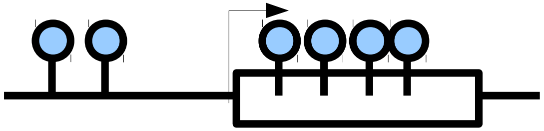
An epimutation

Wild-type

Peloric



CYCLOIDEA



CYCLOIDEA

Methylated DNA

Presence of
CYCLOIDEA
proteins

Absence of
CYCLOIDEA
proteins

Causes of phenotypic differences ?

Heritable

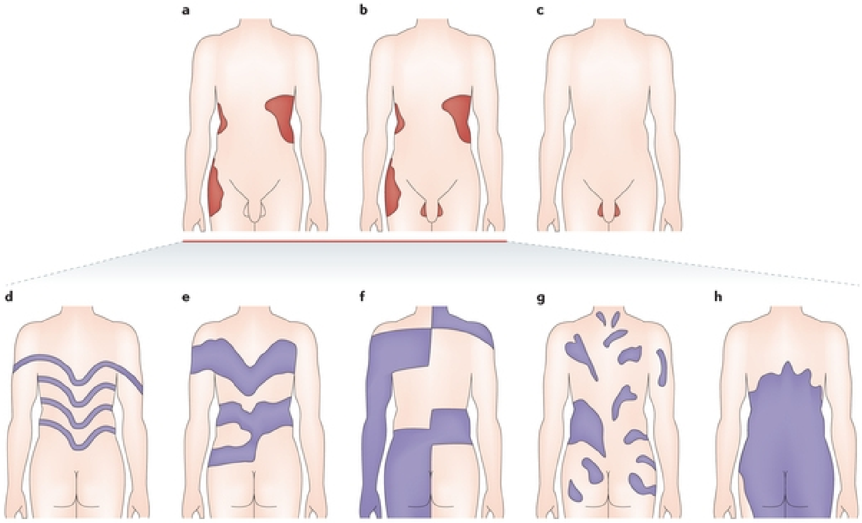
Non heritable

Phenotype = H + NH + HxNH

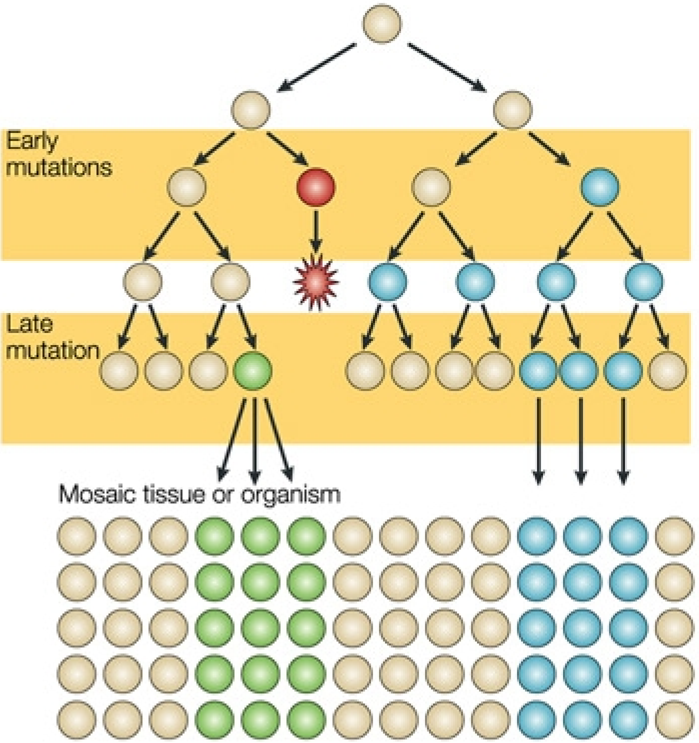
Like GxE but not always (Exceptions: méthylation de l'ADN, microbiome du tube disgestif, langage, accent, culture, style de vie, parental care, effet maternel...)

The importance of stochasticity

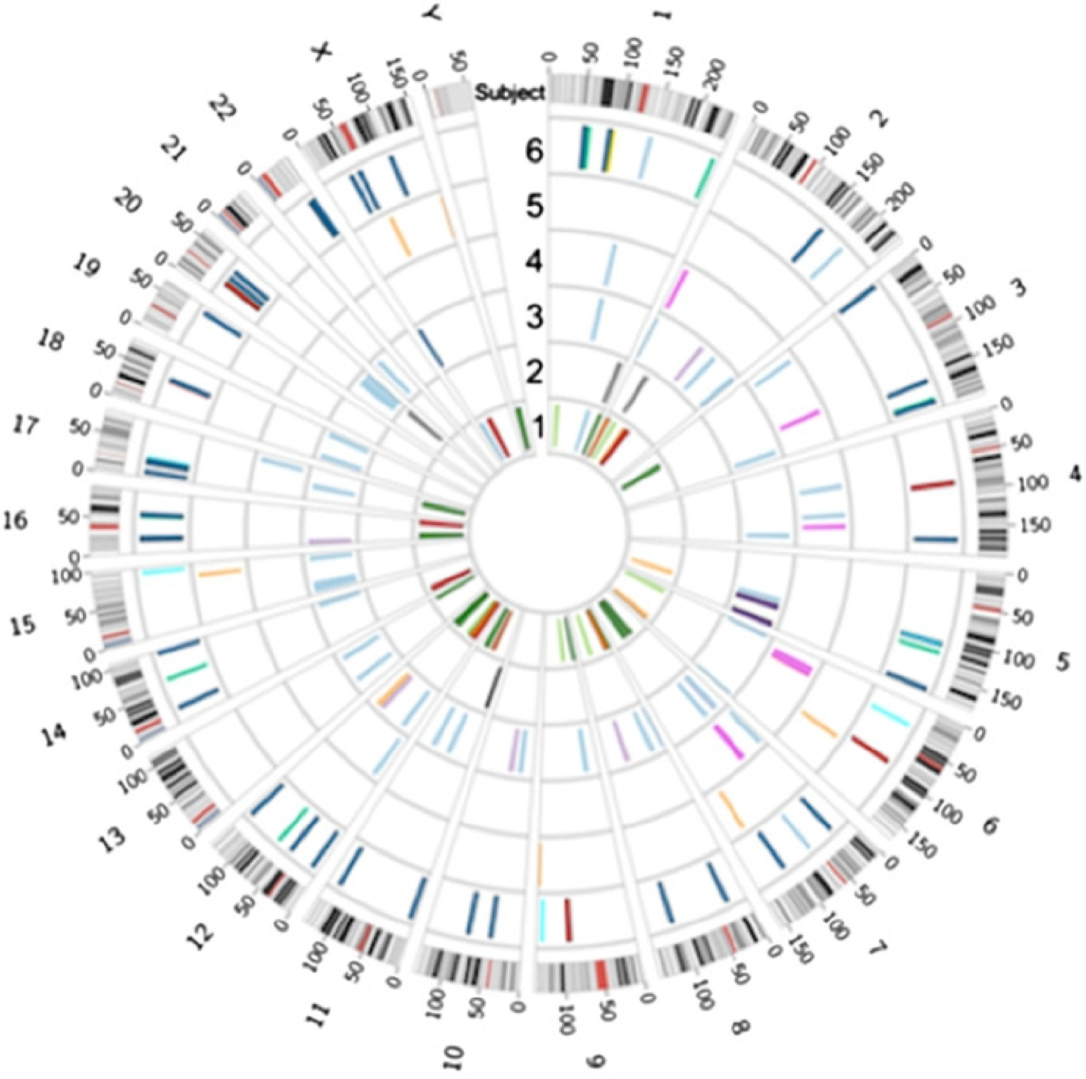
Somatic mosaicism



Nature Reviews | Genetics

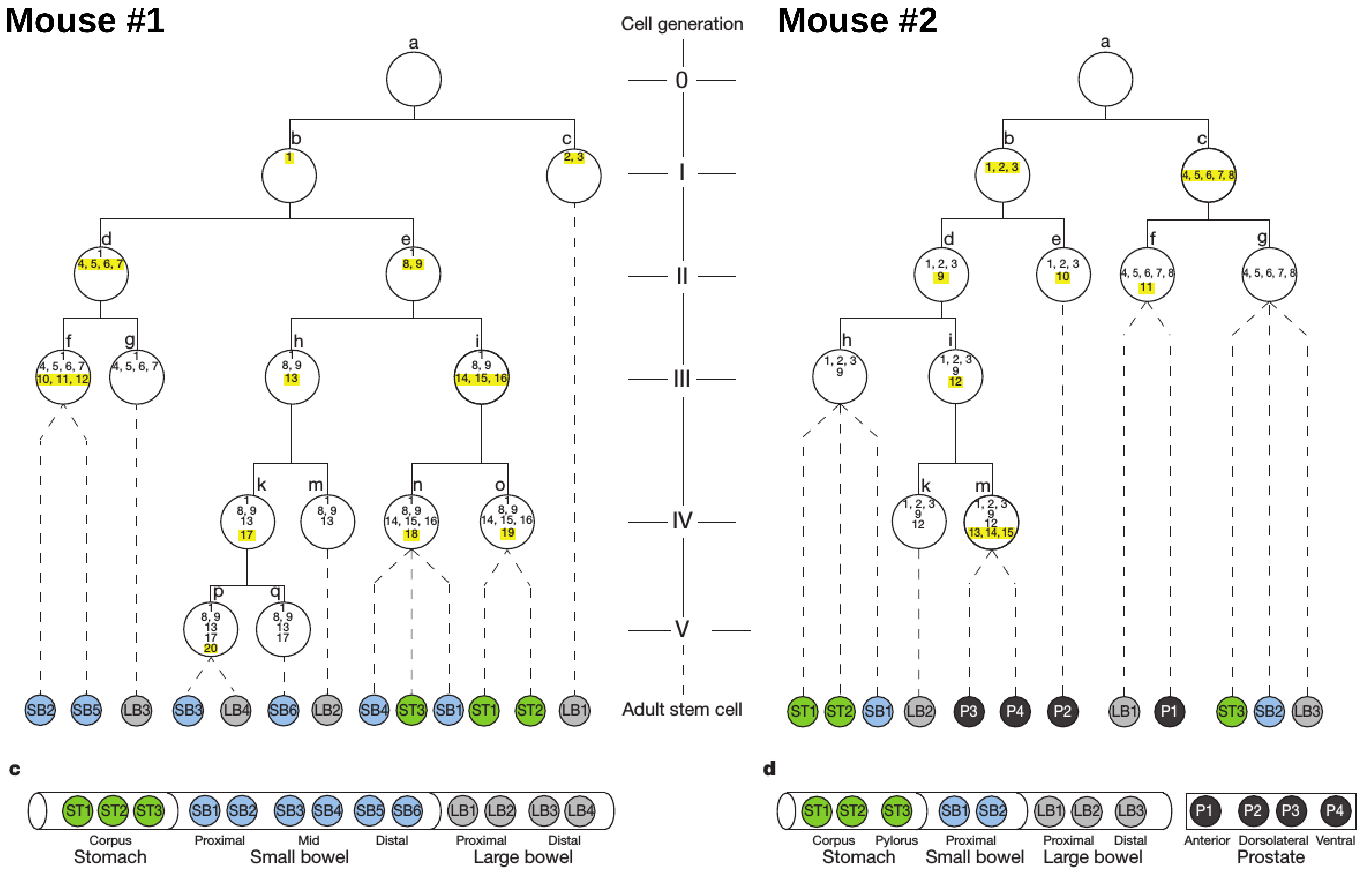


73 somatic CNVs in 11 tissues of six persons



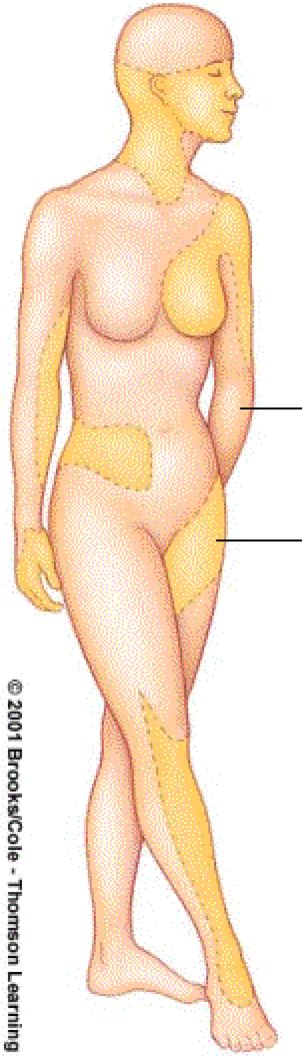
O'Huallachain 2012 PNAS

Somatic mosaicism
used to reconstruct cell lineages



Behjati 2014 Nature

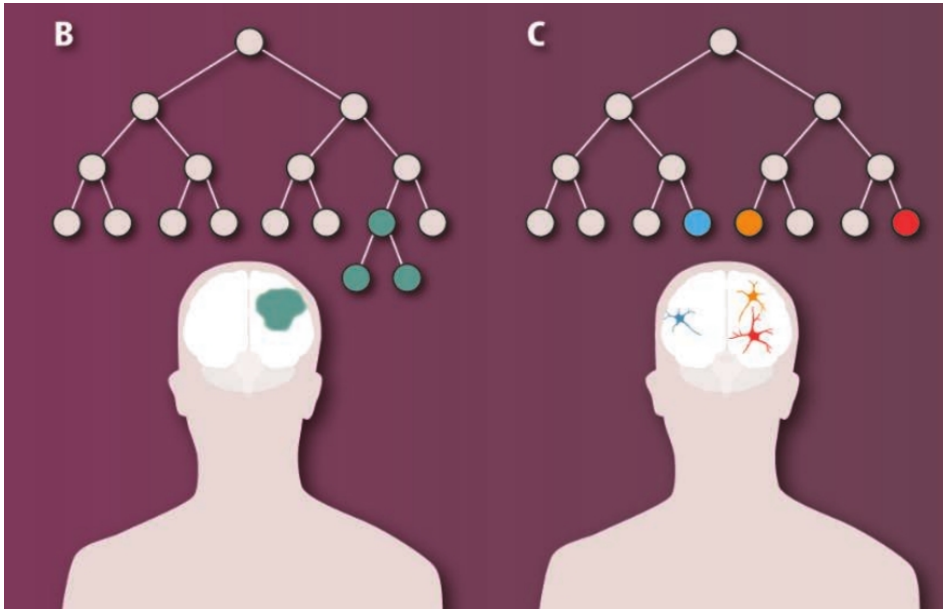
Female mosaicism :
X inactivation pattern



unaffected skin (X chromosome with recessive allele was condensed; its allele is inactivated. The dominant allele on other X chromosome is being expressed in this tissue.)

affected skin with no normal sweat glands (In this tissue, the X chromosome with dominant allele has been condensed. The recessive allele on the other X chromosome is being transcribed.)

Somatic transposition in human brain



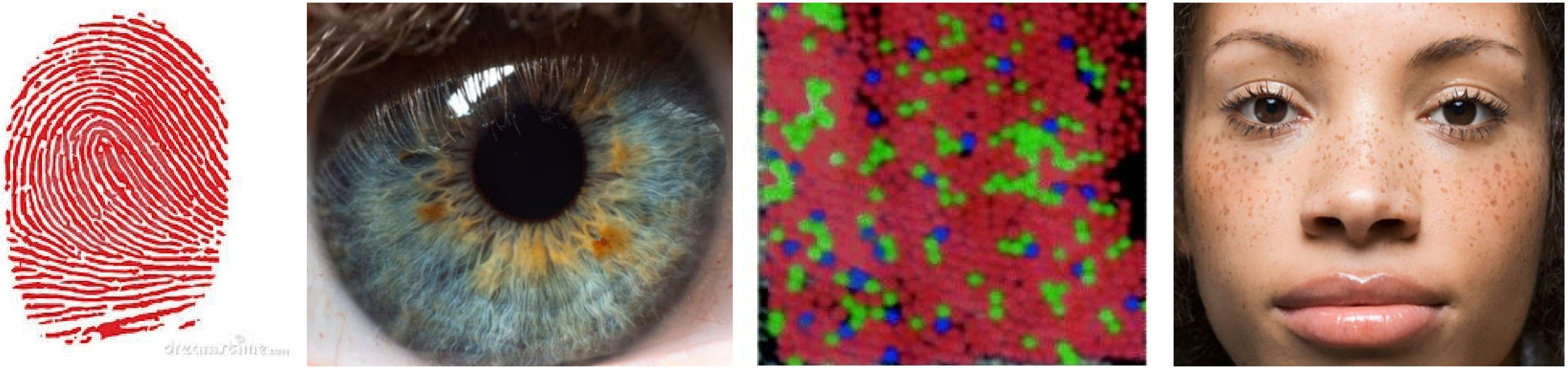
In three individuals:

in the hippocampus and caudate nucleus
7,743 somatic L1 insertions, 13,692 somatic Alu insertions and 1,350 SVA insertions

Baillie 2011 Nature

Developmental noise

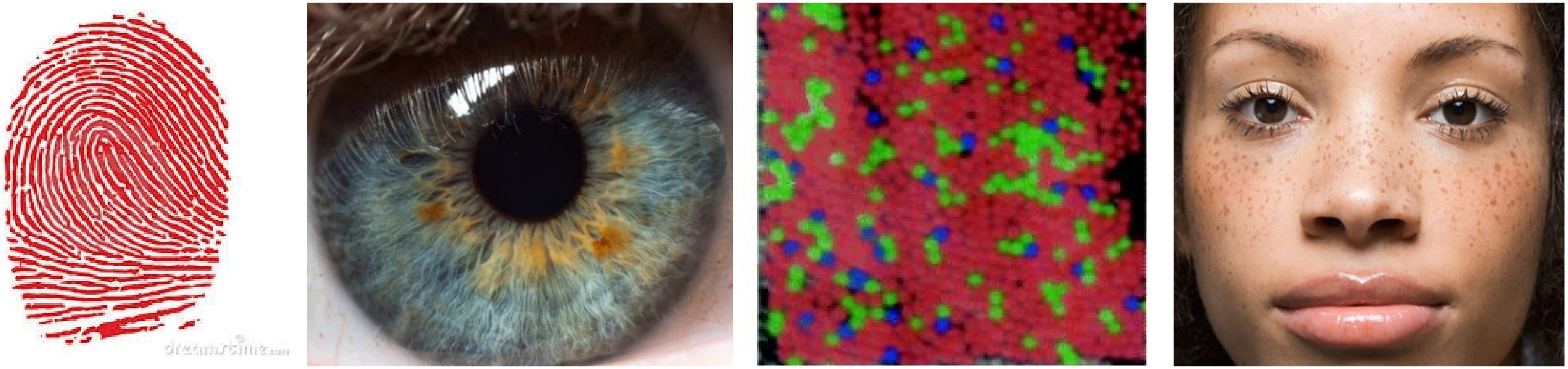
Differences between left and right sides of the body



ear shape, neuron connectivity, olfactory receptor gene expression, X inactivation pattern, organ cell number and size...

Developmental noise

Differences between left and right sides of the body



ear shape, neuron connectivity, olfactory receptor gene expression, X inactivation pattern, organ cell number and size...

Differences between twins

immune system cells, gait, arms crossing, voice, heart beat, brain waves...

Some can be attributed to variation in the number of determinant molecules

During terminal differentiation of mouse 3T3-L1 pre-adipocytes, individual TF abundance differs dramatically (from ~250 to >300,000 copies per nucleus) and the dynamic range can vary up to fivefold during differentiation.

Simicevic 2013 Nature

Causes of phenotypic differences ?

Genetic

Epigenetic

Environment

Stochasticity

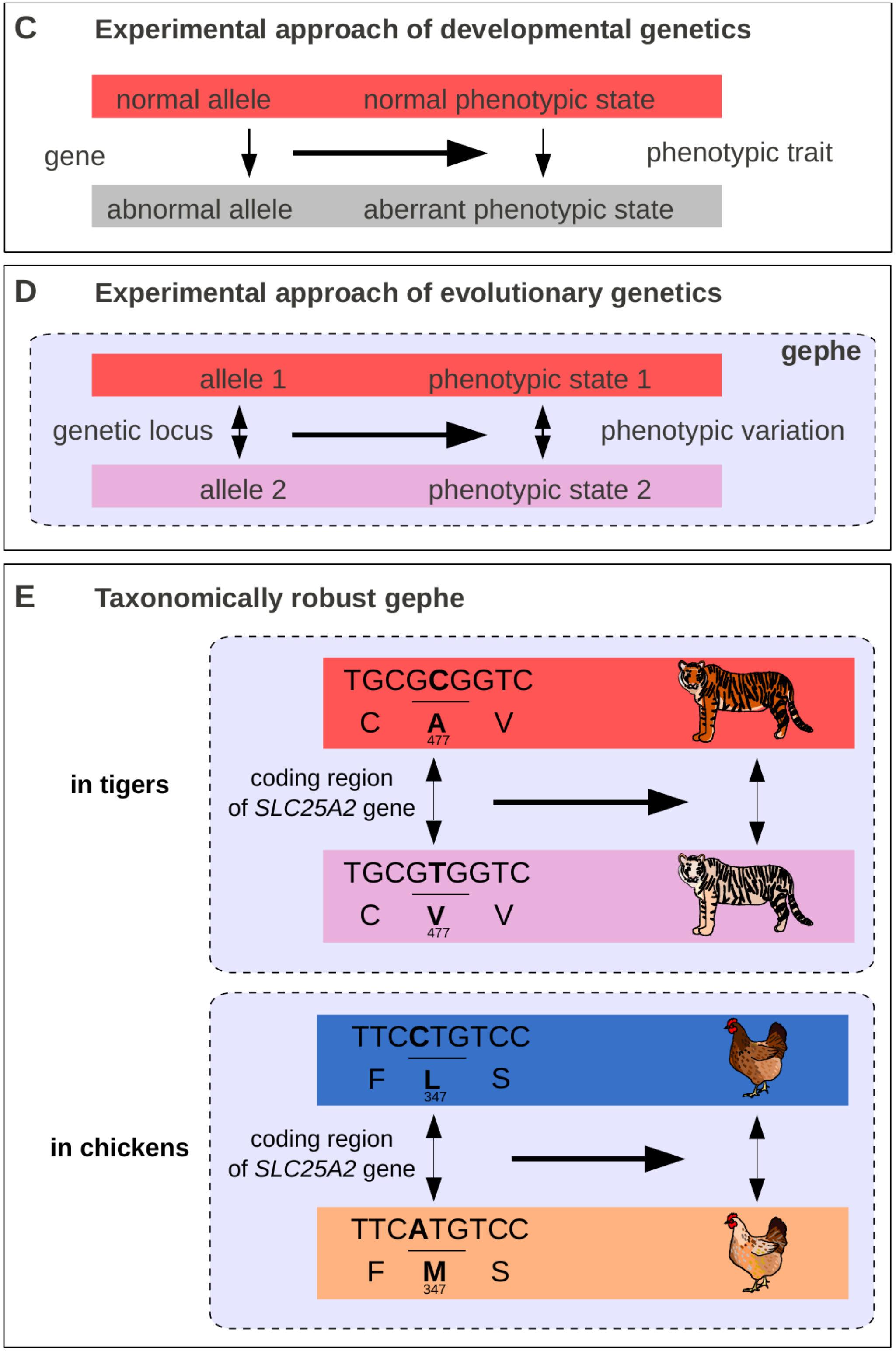
Transmitted

Determistic

Interactions

The differential view of phenotype-genotype relationships

The differential view of phenotype-genotype relationships



C **Experimental approach of developmental genetics**

normal allele normal phenotypic state

gene → phenotypic trait

abnormal allele aberrant phenotypic state

D **Experimental approach of evolutionary genetics**

gephe

allele 1 phenotypic state 1

genetic locus → phenotypic variation

allele 2 phenotypic state 2

E **Taxonomically robust gephe**

in tigers

TGCG**C**GGTC
C **A** V
477

coding region of *SLC25A2* gene →

TGCG**T**GGTC
C **V** V
477

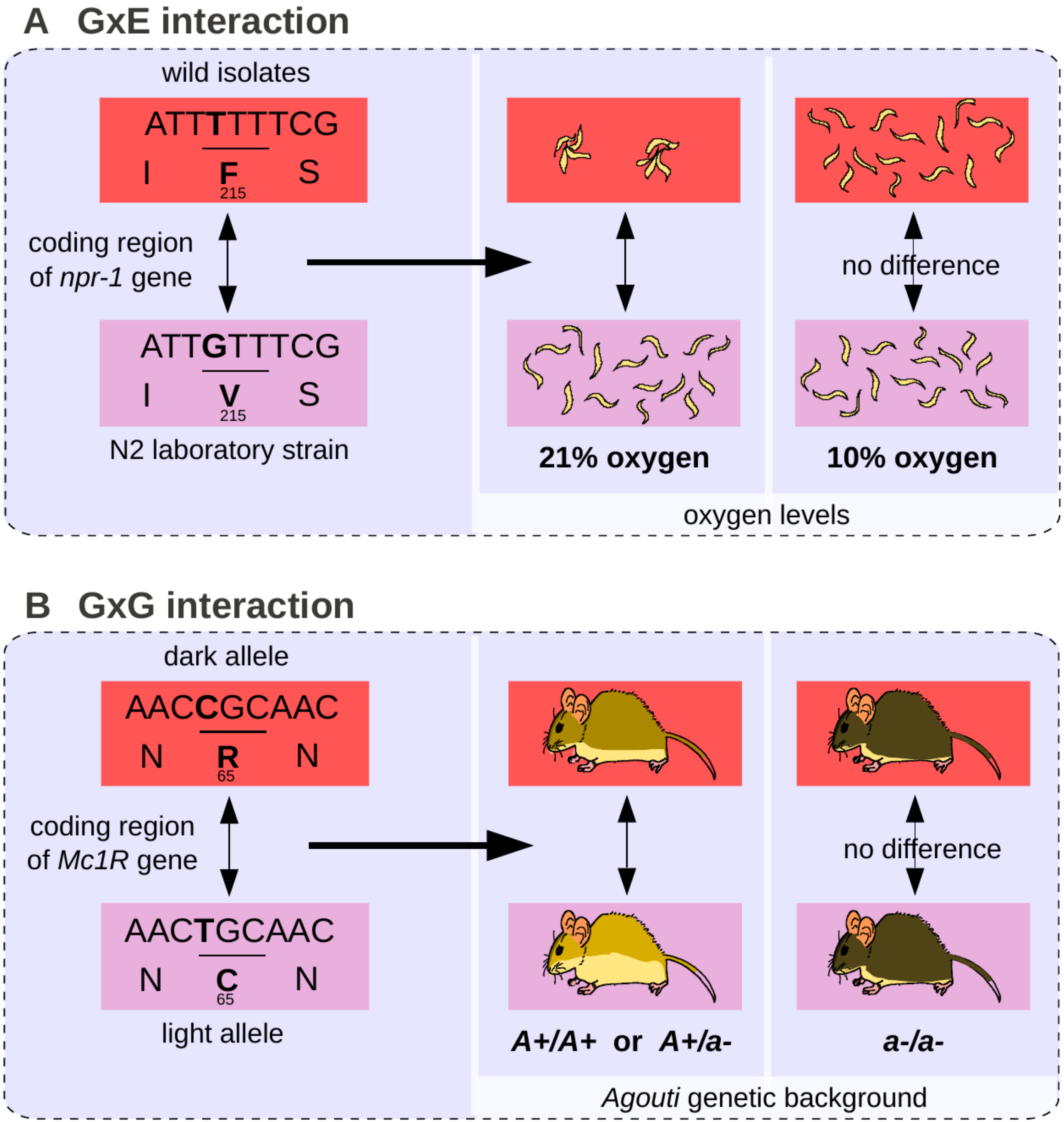
in chickens

TTC**C**TGTCC
F **L** S
347

coding region of *SLC25A2* gene →

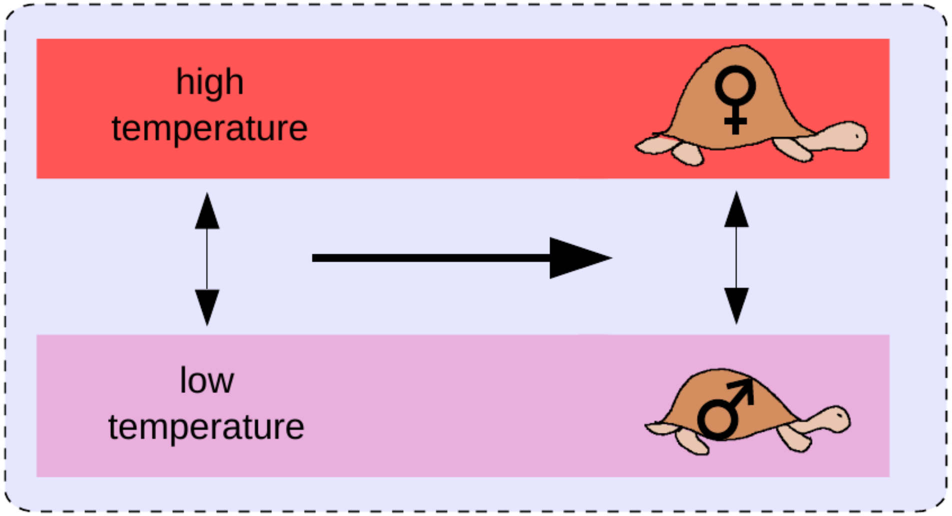
TTC**A**TGTCC
F **M** S
347

Orgogozo et al., submitted

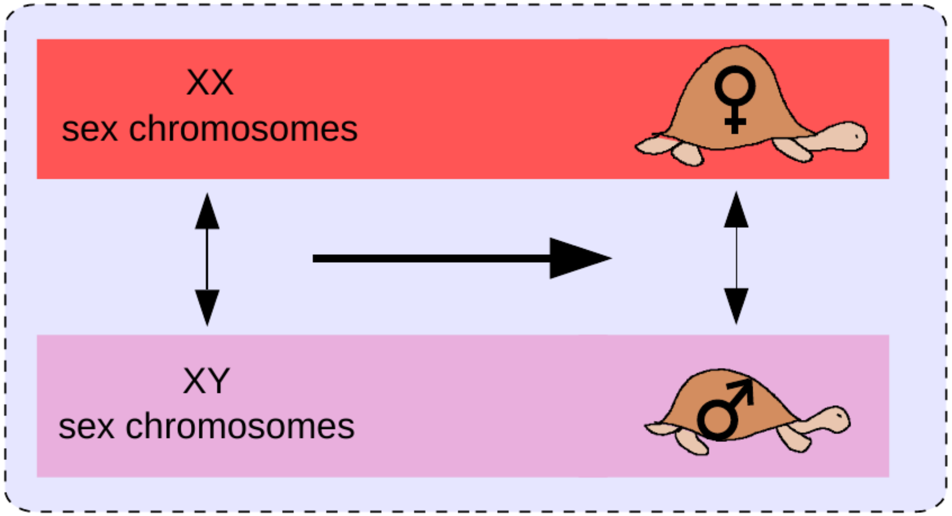

A GxE interaction
wild isolates
ATTTTTTCG
I F S
215
coding region of npr-1 gene
ATTGTTTCG
I V S
215
N2 laboratory strain
no difference
21% oxygen
10% oxygen
oxygen levels
B GxG interaction
dark allele
AACCGCAAC
N R N
65
coding region of Mc1R gene
AACTGCAAC
N C N
65
light allele
no difference
A+/A+ or A+/a-
a-/a-
Agouti genetic background

Comparing G and E effects

A enphe

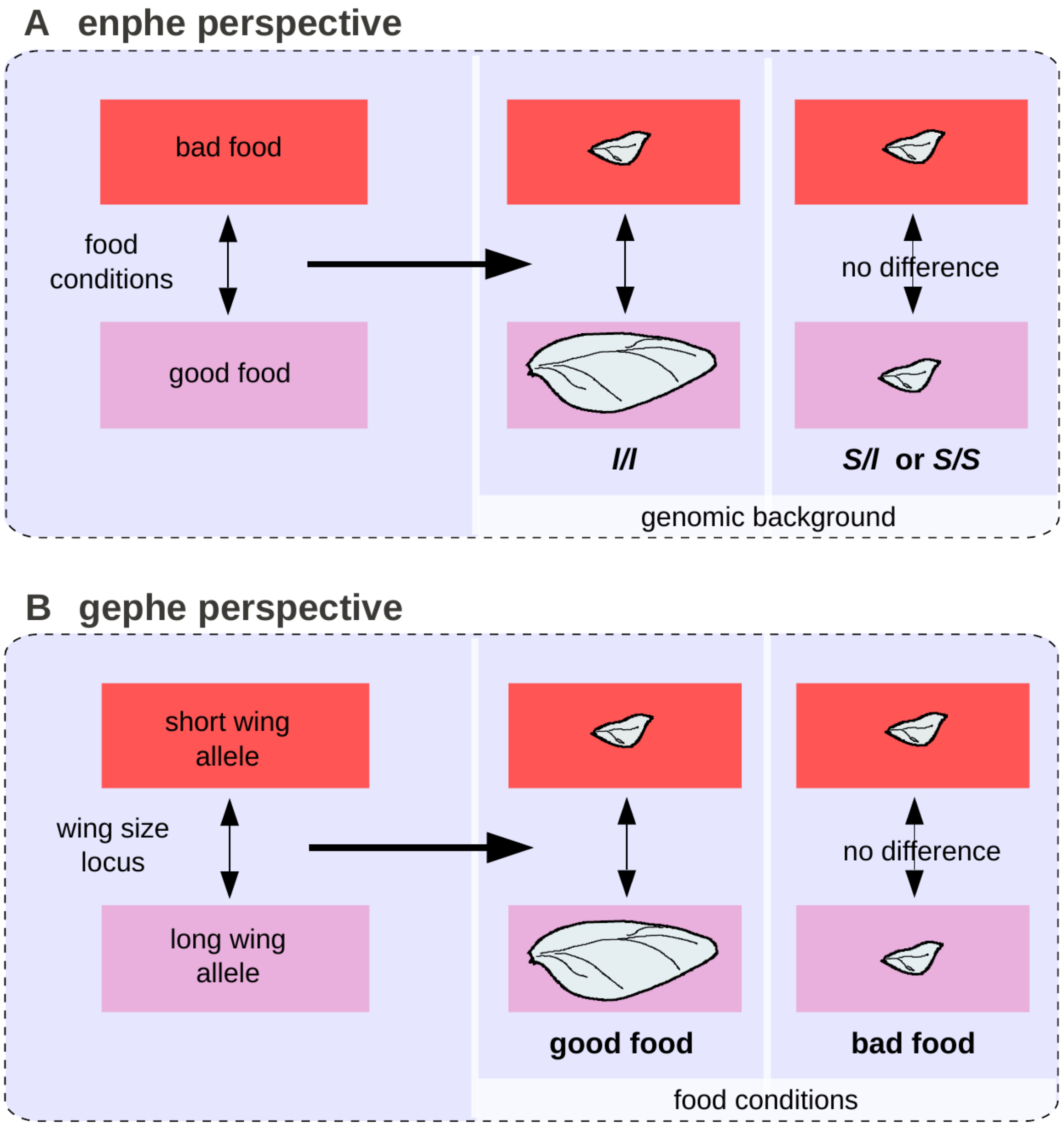


B gephe



Intermingled G and E effects

Calathus melanocephalus

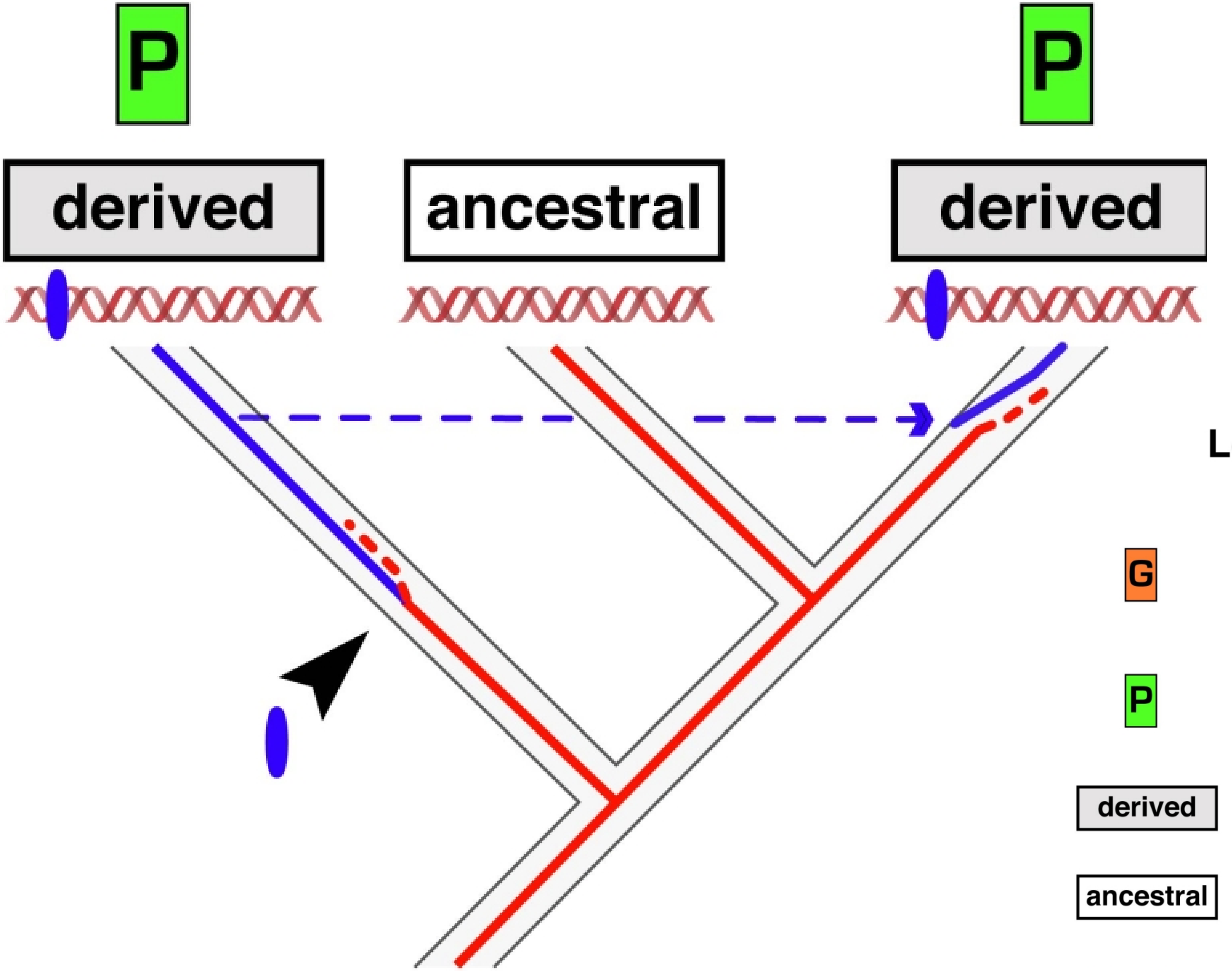


What makes us different?
Where do we come from?

(1)

Exchange of genetic material and genetic plagiarism

Lateral transfer



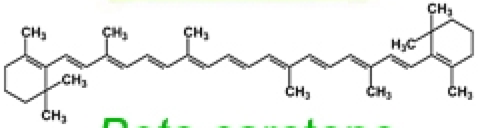
Legend

G	repetition observed at genotypic level
P	repetition observed at phenotypic level
derived	derived phenotype
ancestral	ancestral phenotype
XXXXXXXXX	ancestral allele

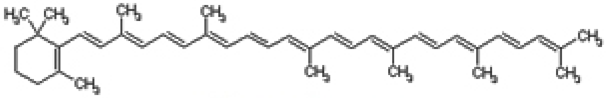
Genetic plagiarism of body color in aphids

Green morph

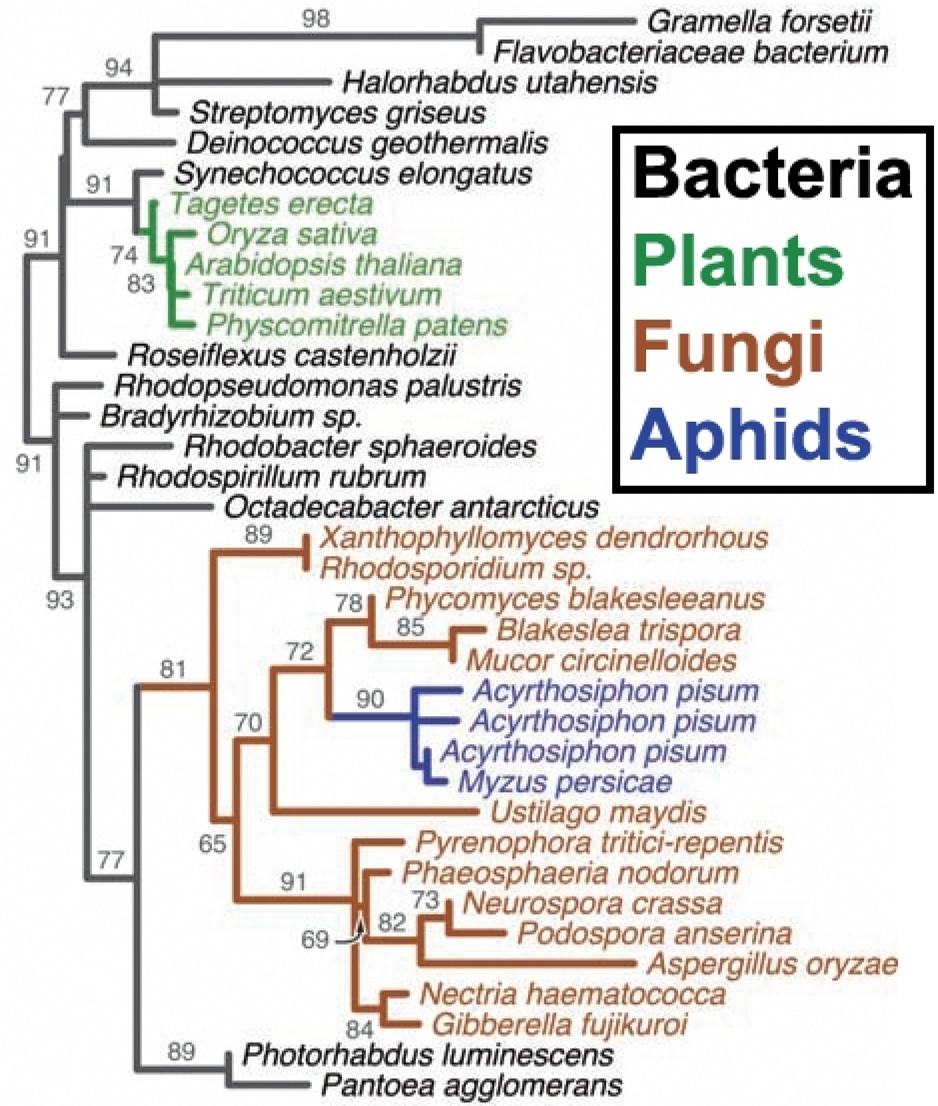
Red morph



Beta-carotene



Torulene

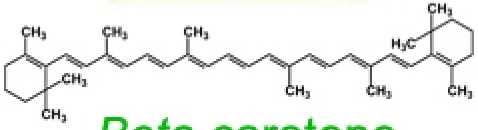


Moran and Jarvik 2010 Science

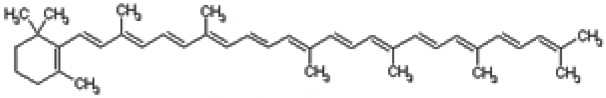
Genetic plagiarism of body color in aphids

Green morph

Red morph



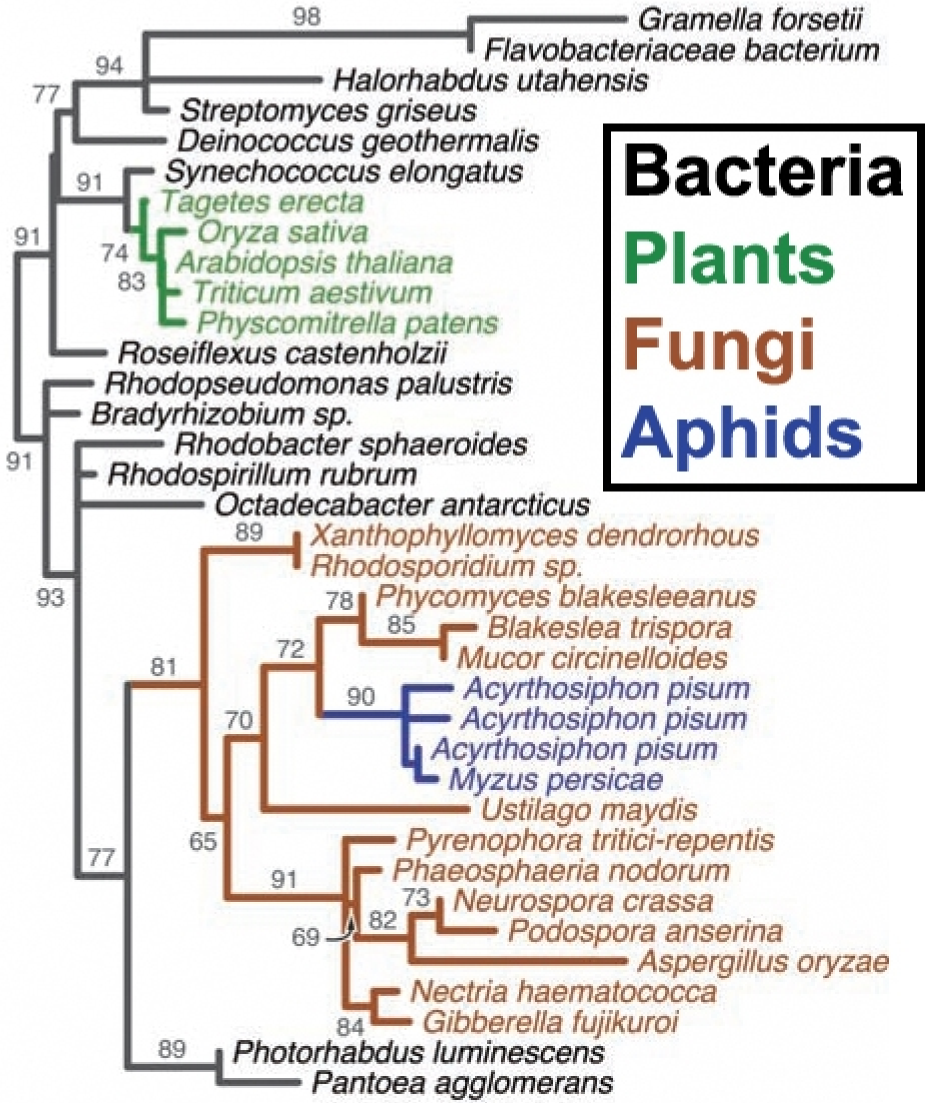
Beta-carotene



Torulene

Carotenoid biosynthesis enzymes transferred also from fungi to the two-spotted spider mite

(Altincicek 2012)

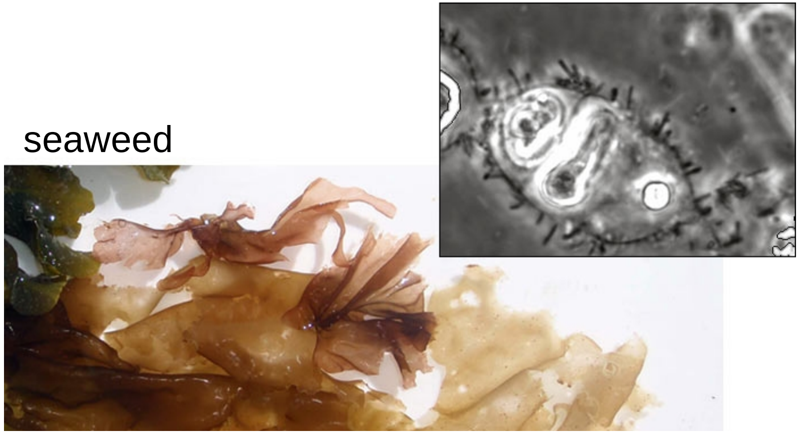


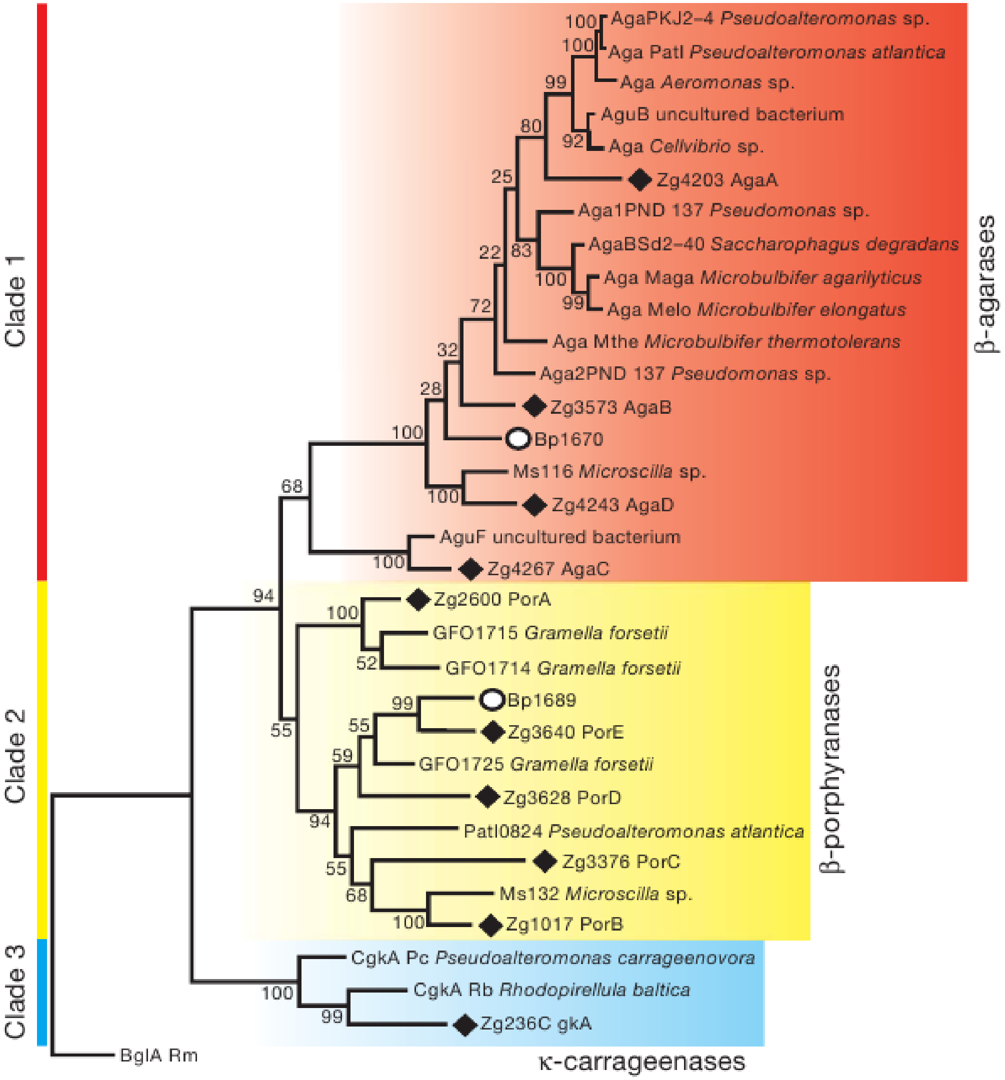
Moran and Jarvik 2010 Science

We are what we eat
Seaweed digestion in Japanese people

◆ *Zobellia galactanivorans*

seaweed

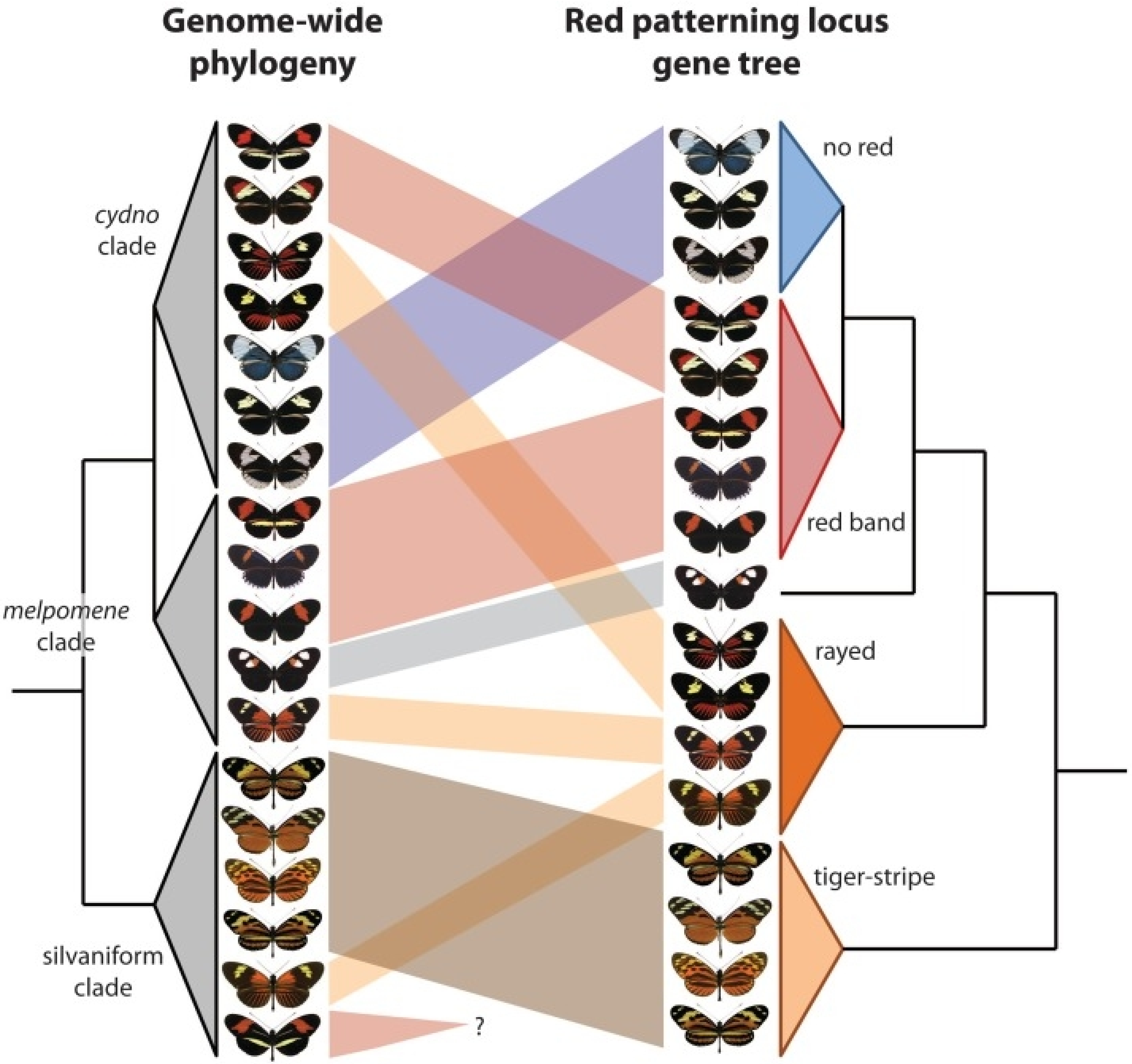




○ *Bacteroides plebeius*
from Japanese microbiome only!

Hehemann et al 2010 Nature

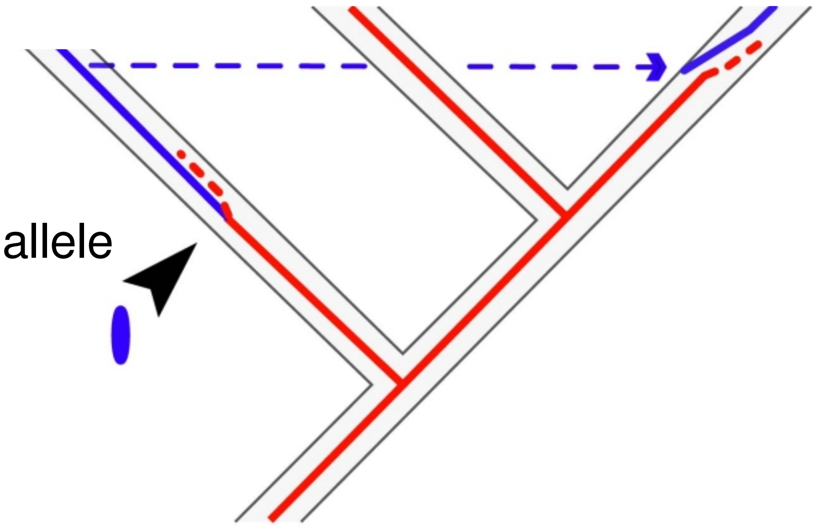
Genetic plagiarism of the red wing color pattern



(Reed 2011, Pardo-Diaz 2012, Heliconius Consortium 2012)

Genetic plagiarism of immune resistance genes

Siberia to Southeast Asia



modern humans

Europe

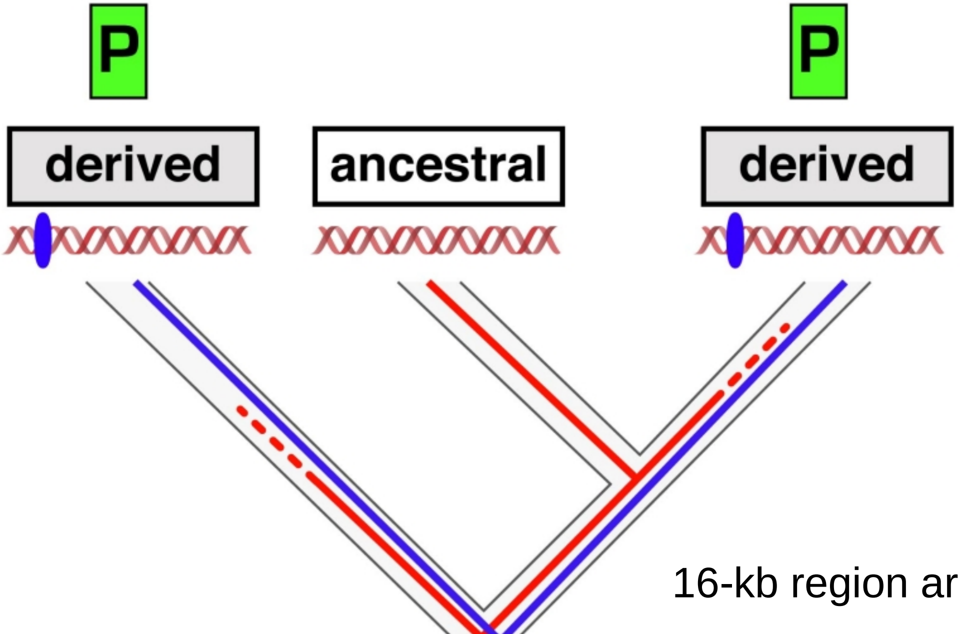
Western Asia

Denisovans

B*73 HLA allele

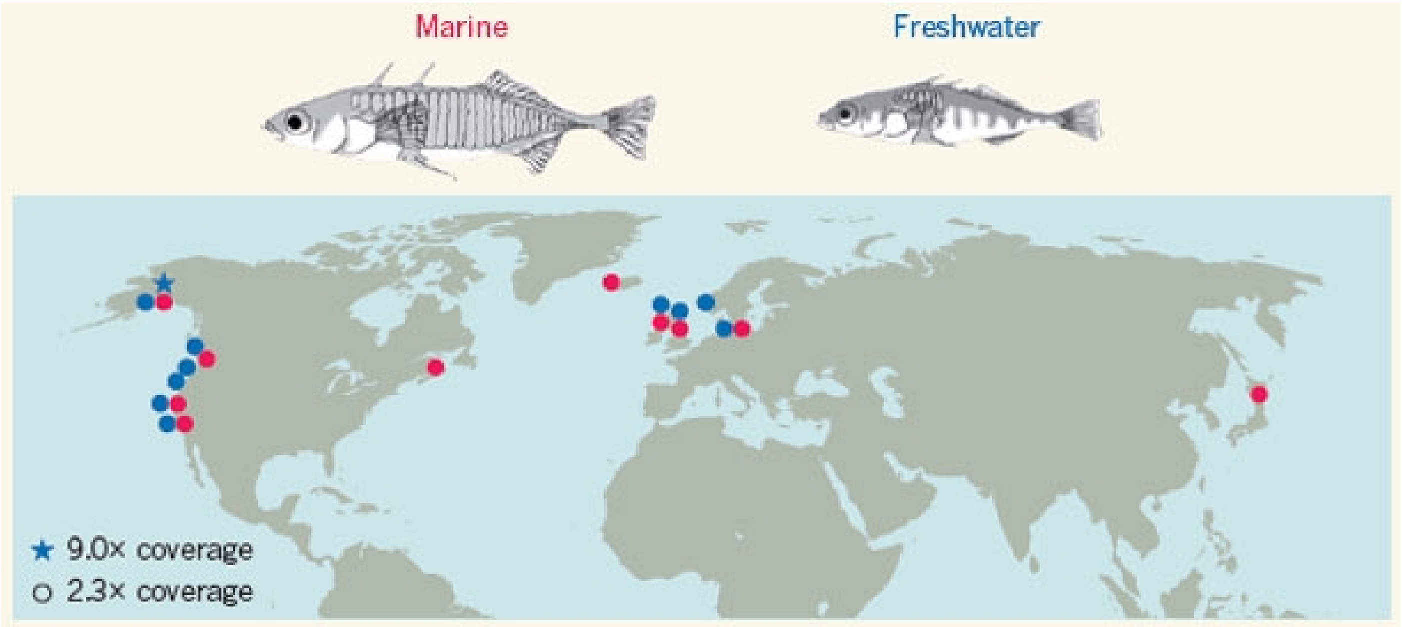
Abi-Rached et al 2011 Nature

Repeated evolution via ancestral polymorphisms



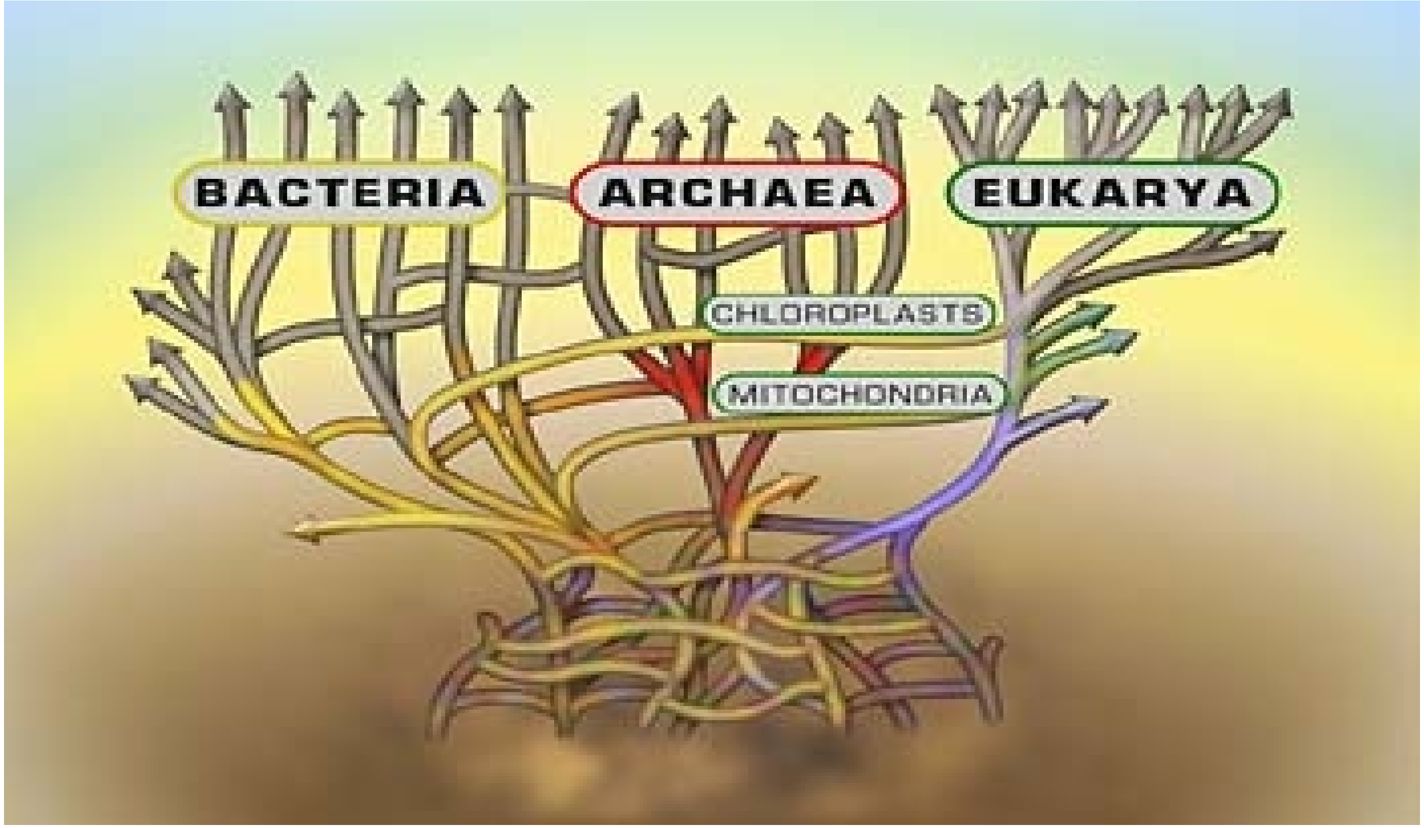
Incomplete lineage sorting
Standing genetic variation

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



Armor plate
(Colosimo 2005
Jones 2012)

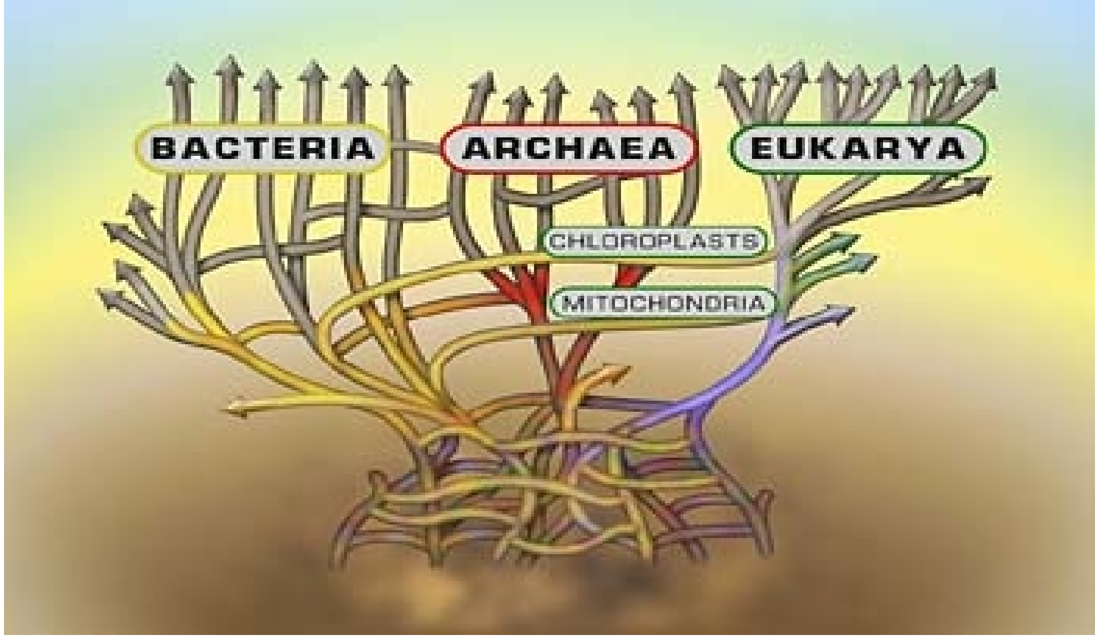
The tree of life is not a tree



Even more interconnections!

Exchange of metabolites throughout the tree of life

Arginine
Histidine
Isoleucine
Leucine
Lysine
Methionine
Phenylalanine
Threonine
Valine
Ascorbic acid (Vitamin C)
Biotin (Vitamin H)
Folic acid (Vitamin M)
Riboflavin (Vitamin B2)
Thiamine (Vitamine B1)
Cobalamin (Vitamin B12)



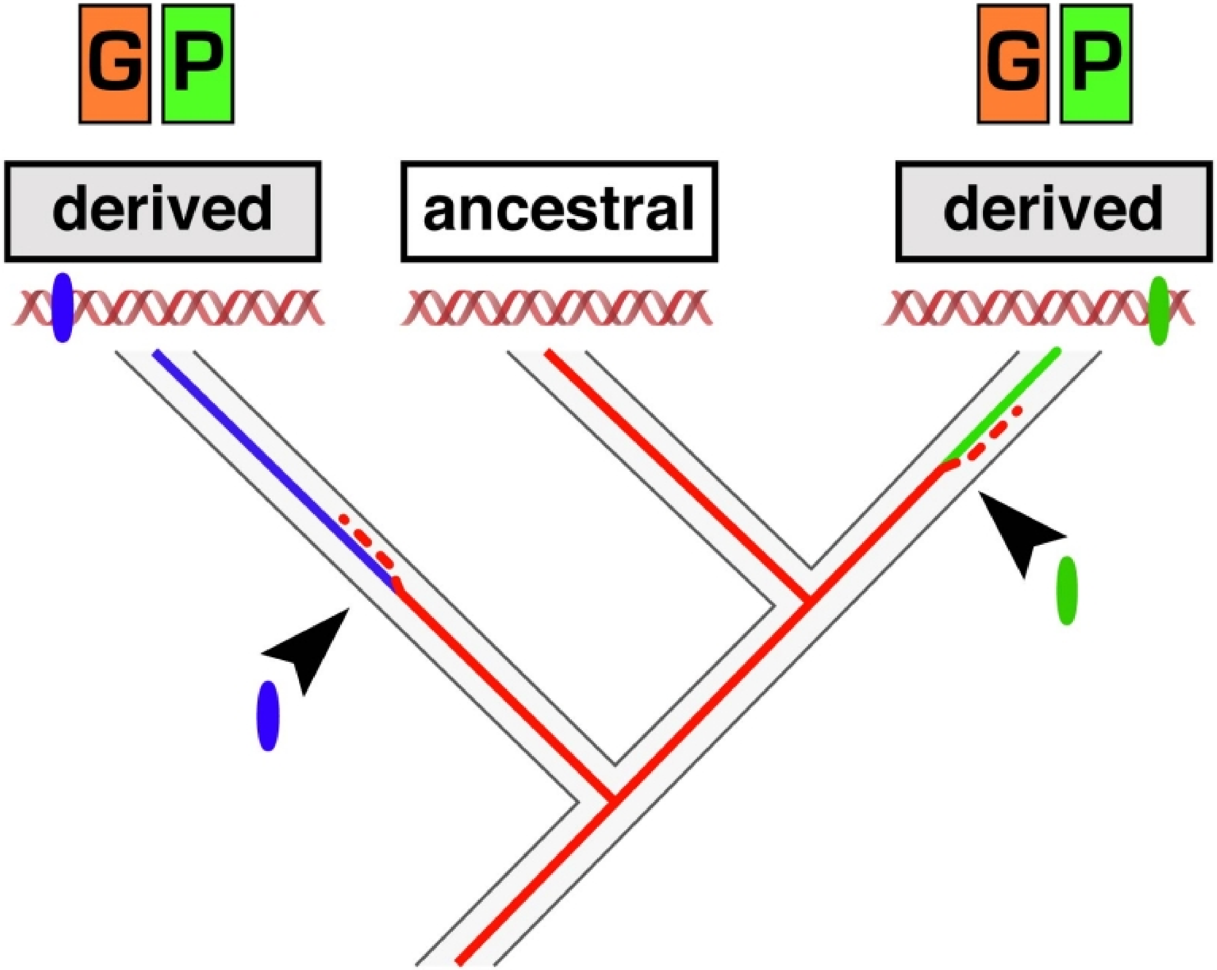
Even more interconnections!

(Romero et al., 2005 Genome Biology)

(2)

There is a limited set of genetic paths to evolution

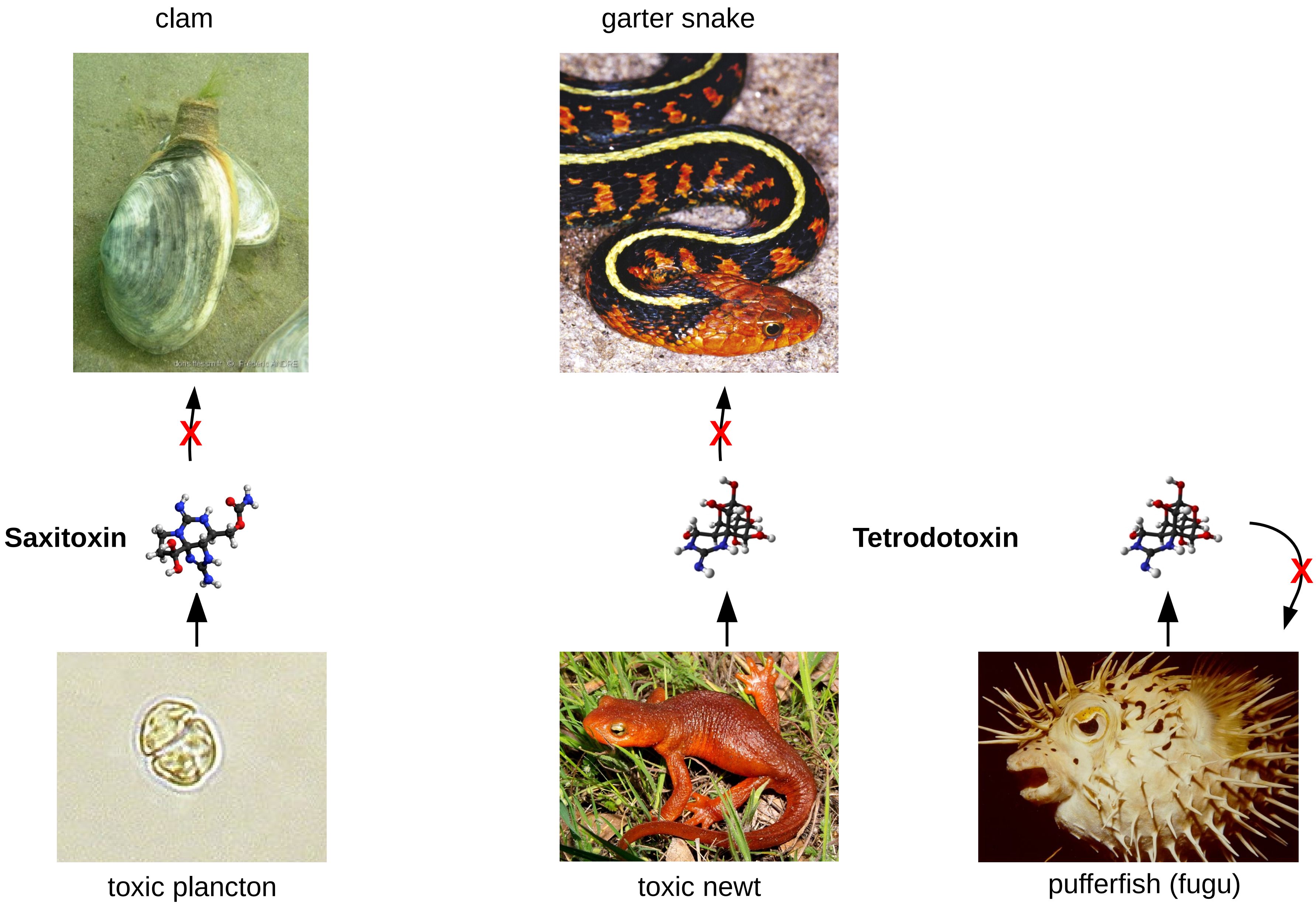
Repeated evolution sensu stricto



Legend

G	repetition observed at genotypic level
P	repetition observed at phenotypic level
derived	derived phenotype
ancestral	ancestral phenotype
	ancestral allele

Repeated evolution



clam

garter snake

Saxitoxin

Tetrodotoxin

toxic plancton

toxic newt

pufferfish (fugu)

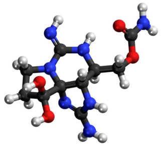
Bricelj 2005
Geffeney 2005
Venkatesh 2005

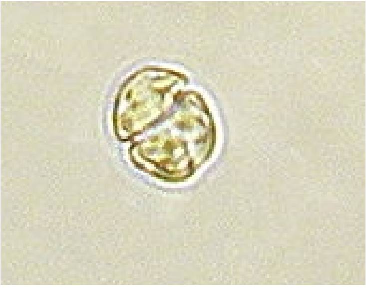
Repeated evolution
via the same amino acid change

clam

Nav1.4
X E945D

Saxitoxin

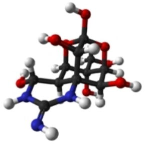




toxic plancton

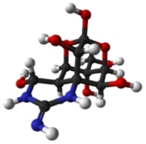
garter snake

Nav1.4 sodium channel
X E945D



Tetrodotoxin

toxic newt

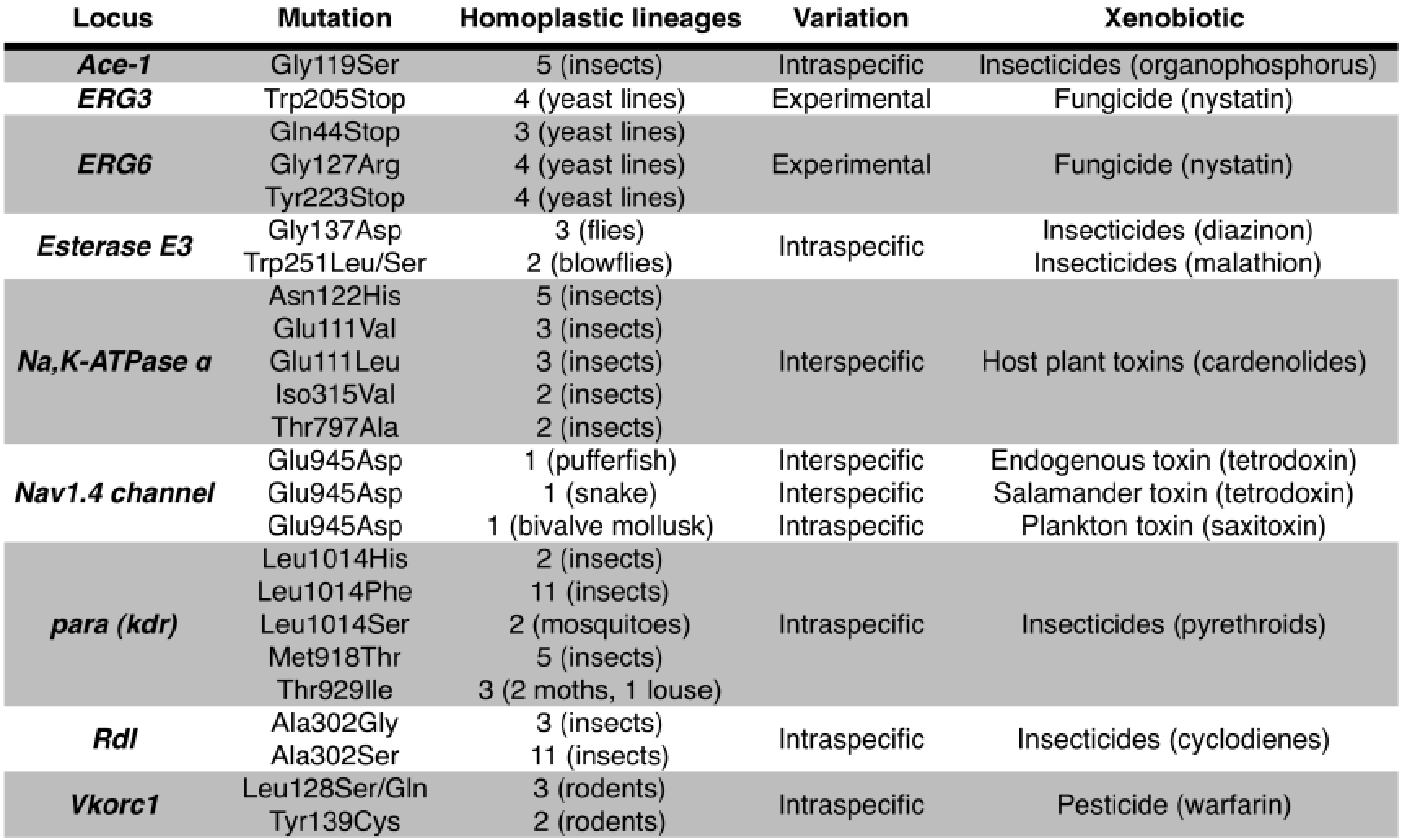


Nav1.4
E945D
X

pufferfish (fugu)

Bricelj 2005
Geffeney 2005
Venkatesh 2005

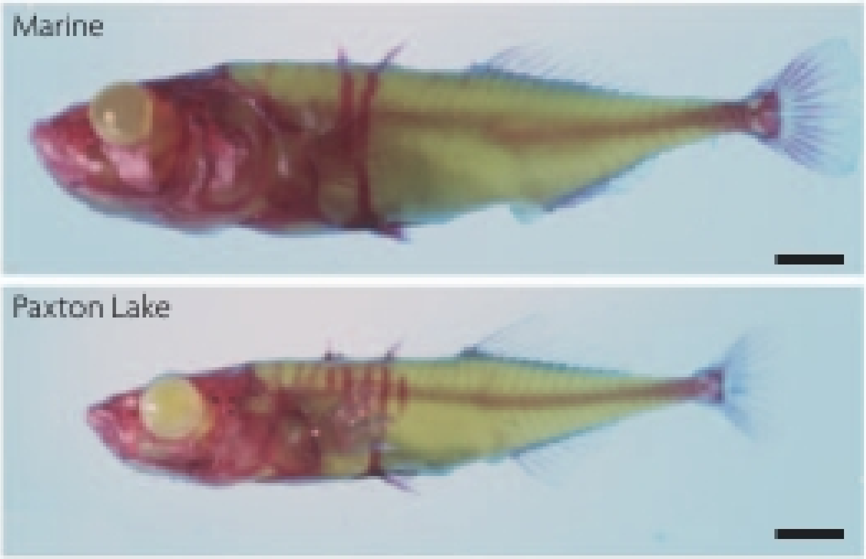
Repeated evolution via the same amino acid change

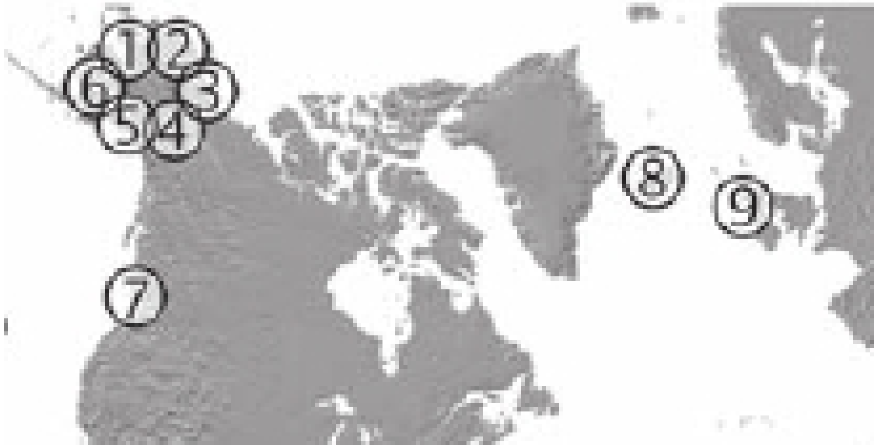


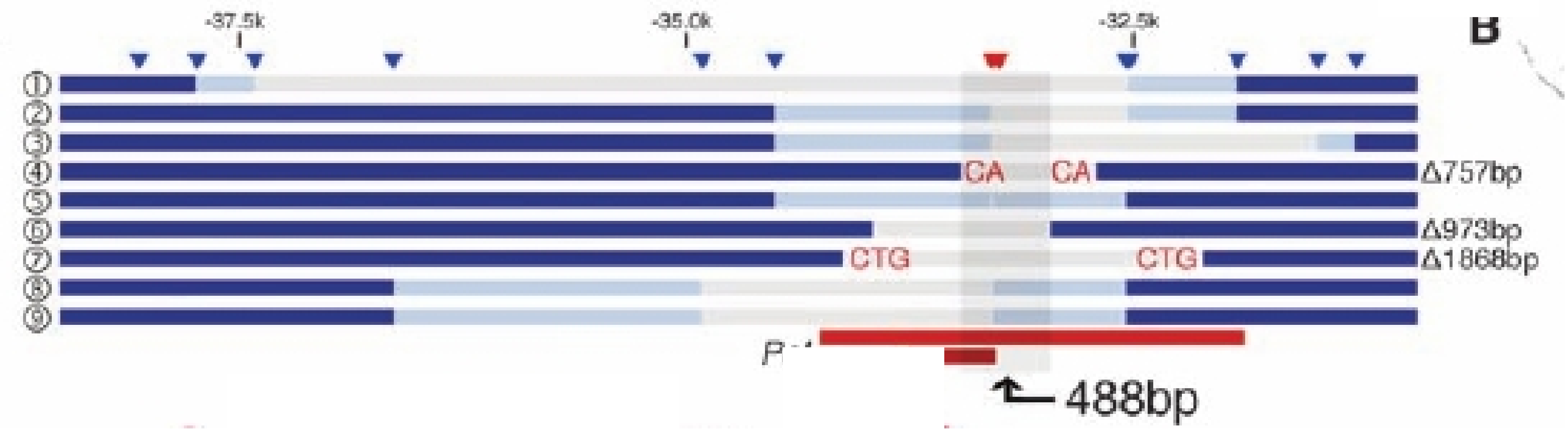
Locus	Mutation	Homoplastic lineages	Variation	Xenobiotic
Ace-1	Gly119Ser	5 (insects)	Intraspecific	Insecticides (organophosphorus)
ERG3	Trp205Stop	4 (yeast lines)	Experimental	Fungicide (nystatin)
ERG6	Gln44Stop Gly127Arg Tyr223Stop	3 (yeast lines) 4 (yeast lines) 4 (yeast lines)	Experimental	Fungicide (nystatin)
Esterase E3	Gly137Asp Trp251Leu/Ser	3 (flies) 2 (blowflies)	Intraspecific	Insecticides (diazinon) Insecticides (malathion)
Na,K-ATPase α	Asn122His Glu111Val Glu111Leu Iso315Val Thr797Ala	5 (insects) 3 (insects) 3 (insects) 2 (insects) 2 (insects)	Interspecific	Host plant toxins (cardenolides)
Nav1.4 channel	Glu945Asp Glu945Asp Glu945Asp	1 (pufferfish) 1 (snake) 1 (bivalve mollusk)	Interspecific Interspecific Intraspecific	Endogenous toxin (tetrodoxin) Salamander toxin (tetrodoxin) Plankton toxin (saxitoxin)
para (kdr)	Leu1014His Leu1014Phe Leu1014Ser Met918Thr Thr929Ile	2 (insects) 11 (insects) 2 (mosquitoes) 5 (insects) 3 (2 moths, 1 louse)	Intraspecific	Insecticides (pyrethroids)
Rdl	Ala302Gly Ala302Ser	3 (insects) 11 (insects)	Intraspecific	Insecticides (cyclodienes)
Vkorc1	Leu128Ser/Gln Tyr139Cys	3 (rodents) 2 (rodents)	Intraspecific	Pesticide (warfarin)

Resistance to xenobiotics

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

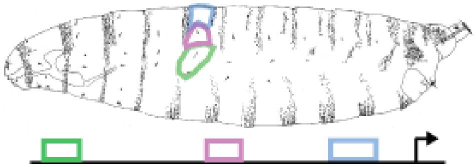






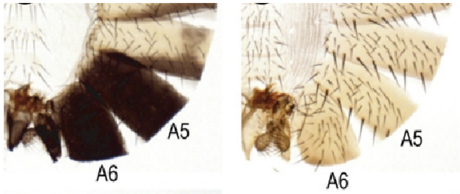
Accumulation of evolutionary-relevant mutations at the same locus

≥ 6 mutations in *svb*



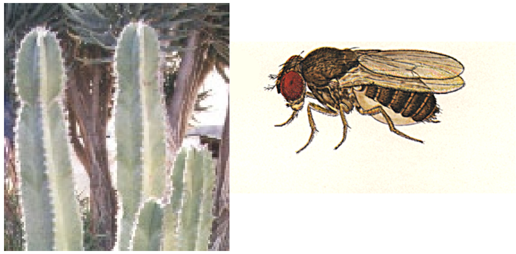
(McGregor, Orgogozo et al. 2007 Nature)

3 mutations in *tan*



(Jeong et al., 2008 Cell)

2-4 mutations in *nvd*

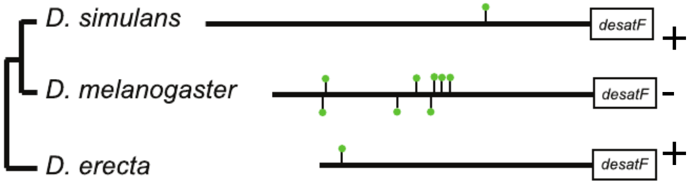


(Lang et al. 2012 Science)

5 mutations in *ebony*

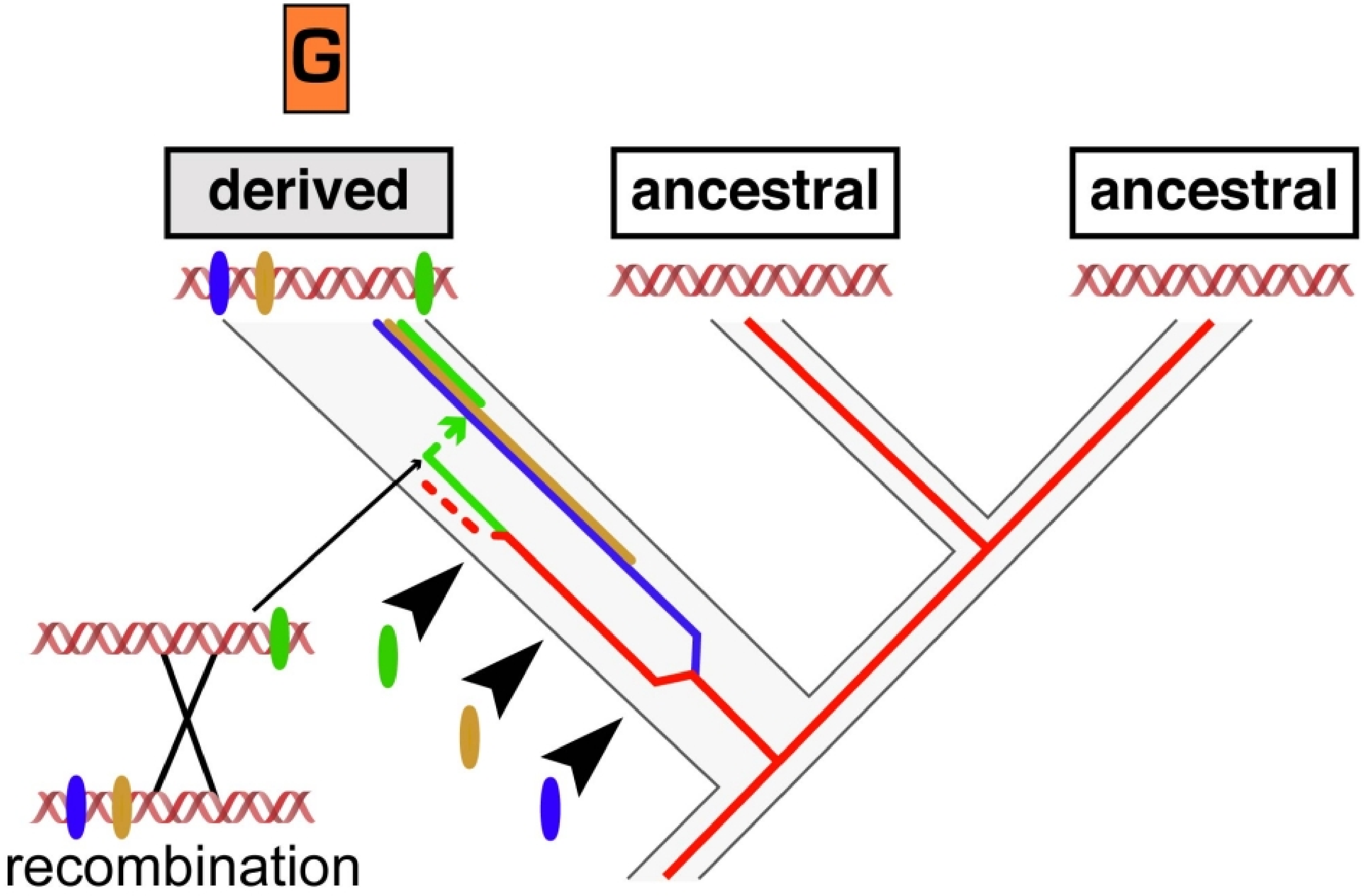
(Rebeiz et al., 2009 Cell)

3 deletions in *desatF*

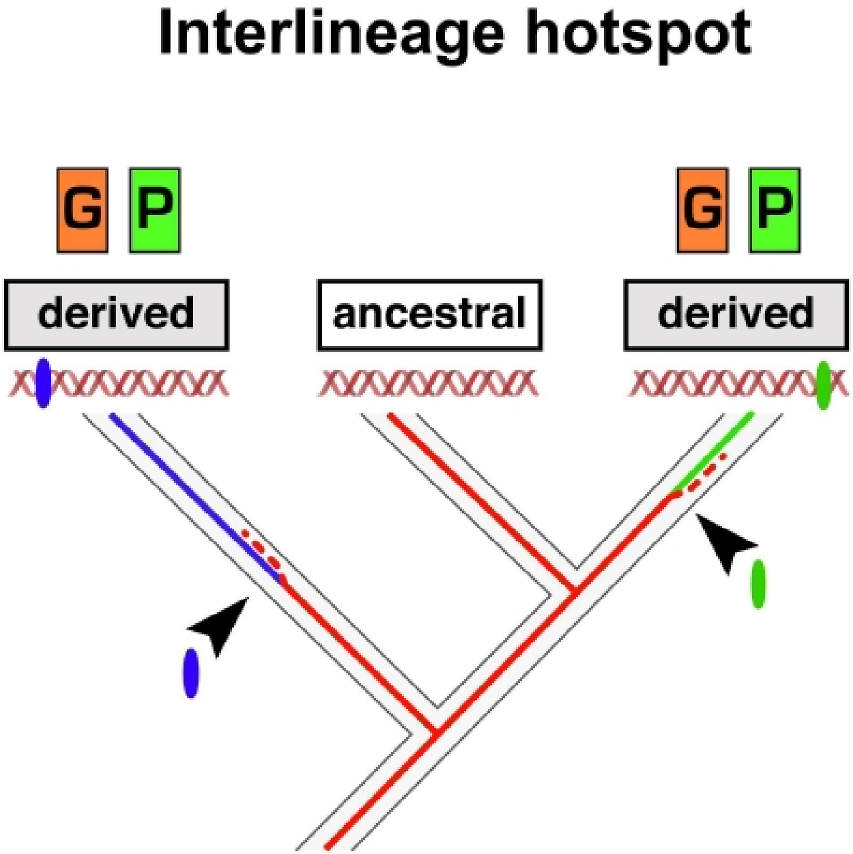


(Shirangi et al., 2009 PloS Biol)

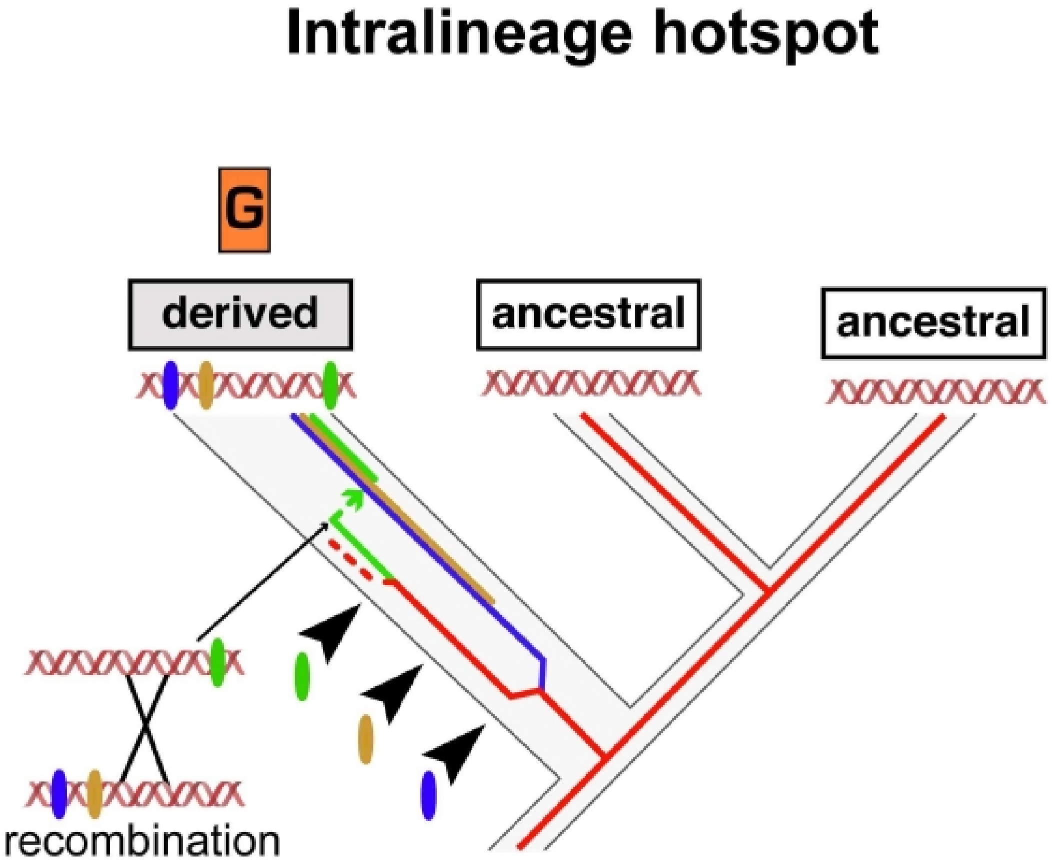
Intralineage hotspot



Hotspot genes: preferred targets of evolution



Interlineage hotspot



Intralineage hotspot

Martin and Orgogozo,2013

(2)

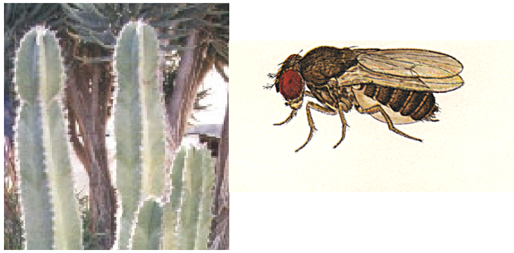
Why is the set of genetic paths limited?

There are specialized genes in a genome

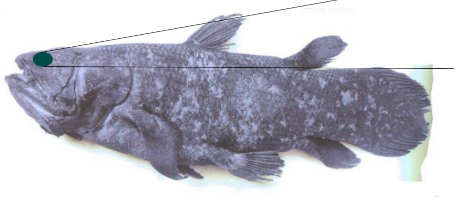
Steroid hormone biosynthesis

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

(2)

Why is the set of genetic paths limited?

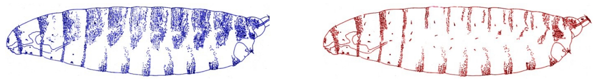
→ **genes with specialized functions**

→ ***But what about phenotypes involving multifunctional genes?***

Evolution appears to use a restricted set of all possible paths

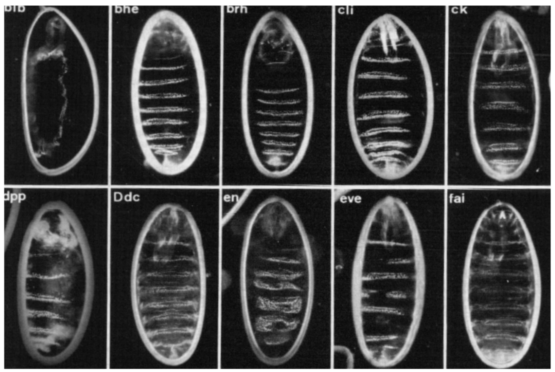
Changes in trichome pattern

EVOLUTION



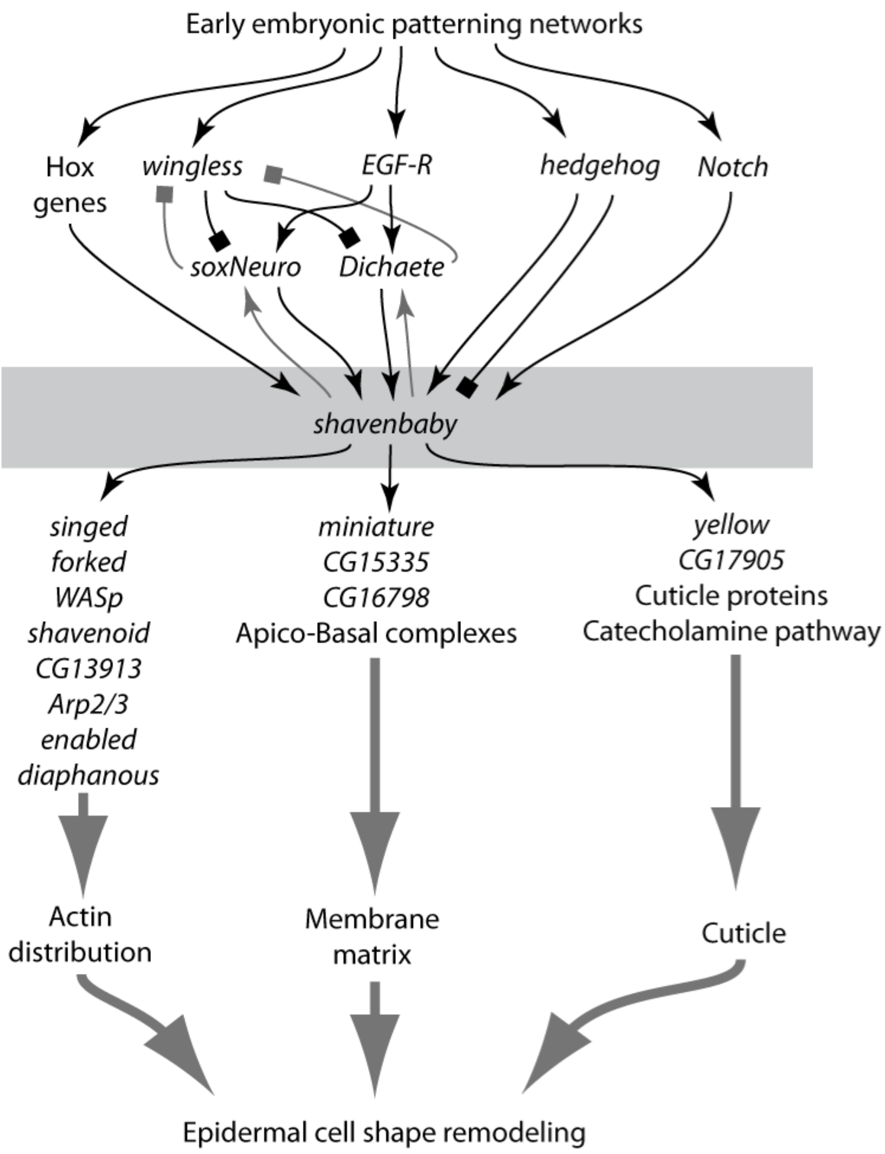
A single gene

MUTAGENESIS



(Nüsslein-Volhard et Weischaus)

~100 genes



Early embryonic patterning networks

Hox genes

wingless

EGF-R

hedgehog

Notch

soxNeuro

Dichaete

shavenbaby

singed
forked
WASp
shavenoid
CG13913
Arp2/3
enabled
diaphanous

miniature
CG15335
CG16798
Apico-Basal complexes

yellow
CG17905
Cuticle proteins
Catecholamine pathway

Actin distribution

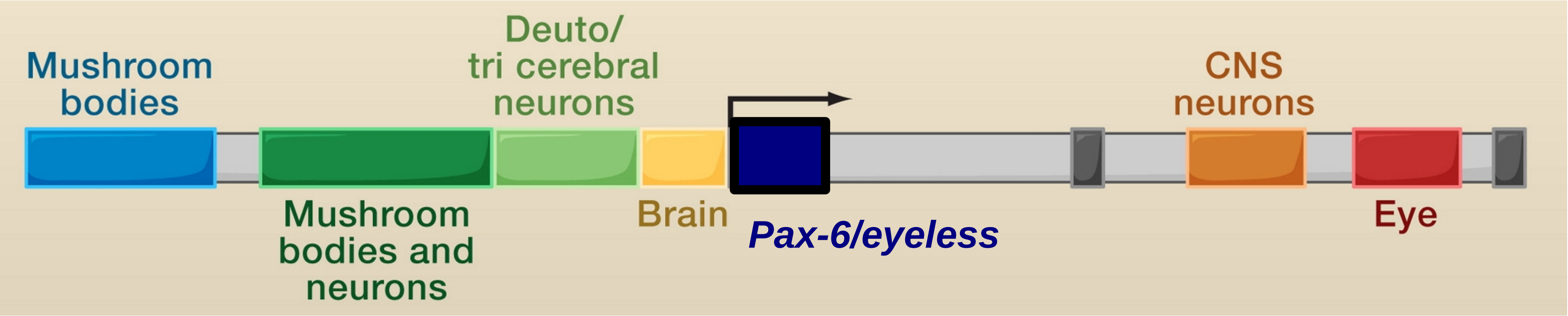
Membrane matrix

Cuticle

Epidermal cell shape remodeling

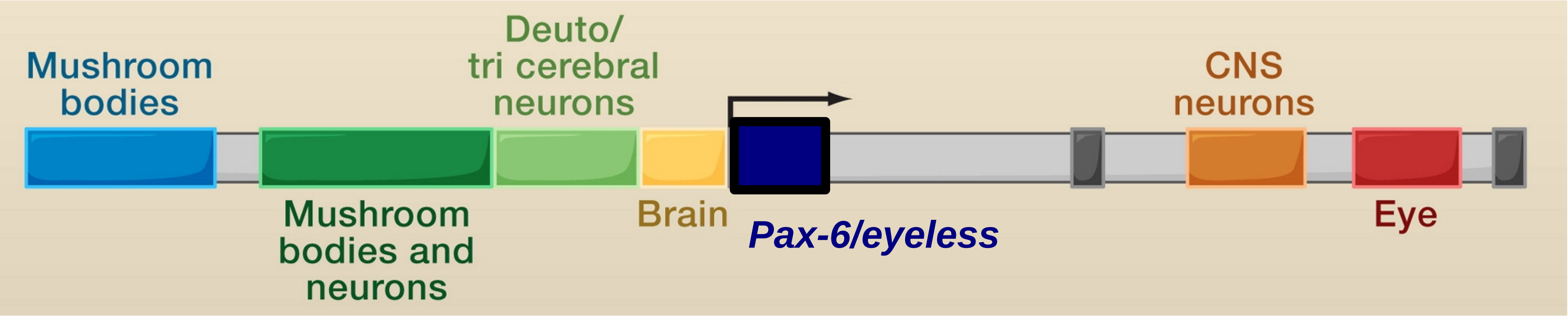
(Stern et Orgogozo, 2008 Evolution – Stern et Orgogozo, 2009 Science)

There are specialized loci within multifunctional genes in a genome



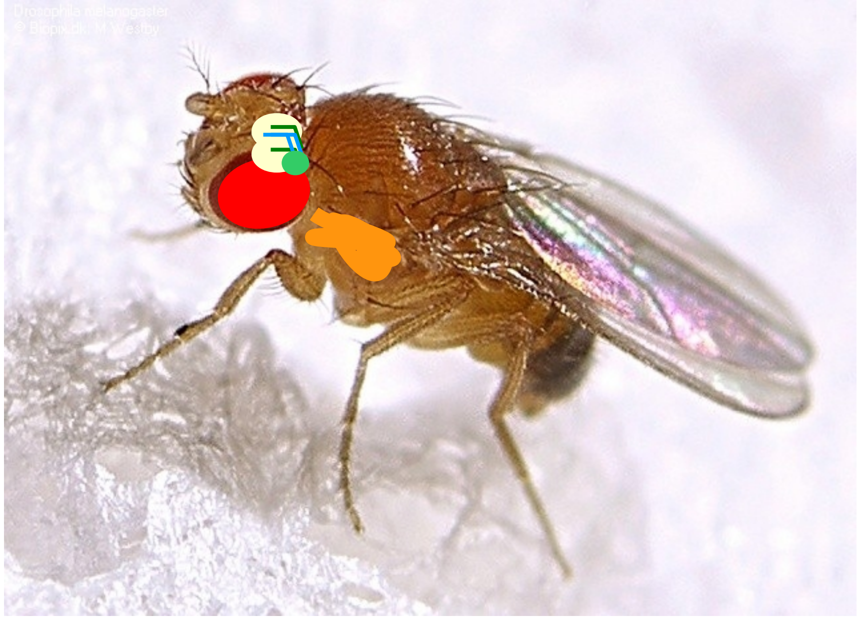
Carroll 2008

There are specialized loci within multifunctional genes in a genome



Carroll 2008

The modularity of cis-regulatory elements is reflected in the modularity of body parts



(2)

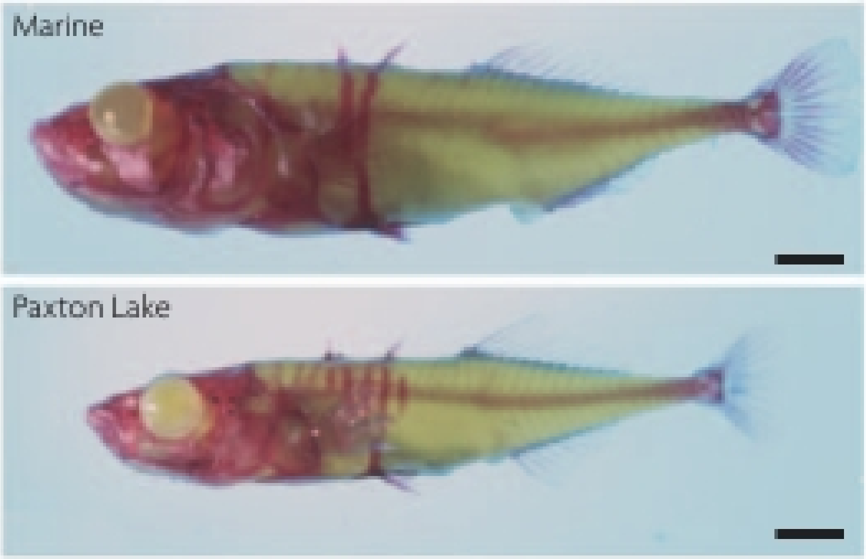
Why is the set of genetic paths limited?

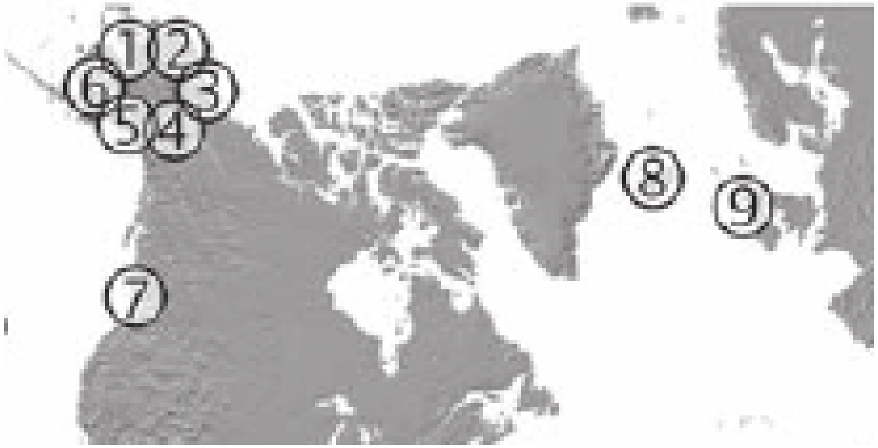
→ genes with specialized functions

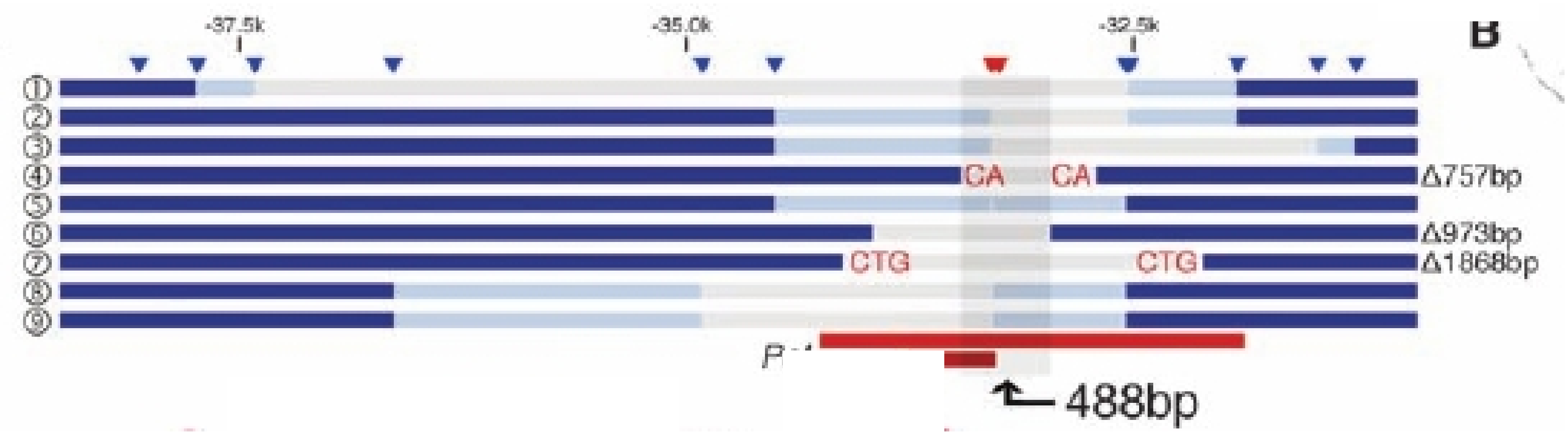
→ multifunctional genes with specialized regions

→ mutational bias

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







(1)

Intense exchange and genetic plagiarism.

(2)

There is a limited set of genetic paths to evolution.

(1)

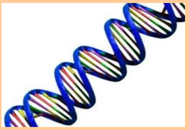
Intense exchange and genetic plagiarism.

(2)

There is a limited set
of genetic paths to evolution.
Genetic paths to evolution are conserved.

What makes us different?

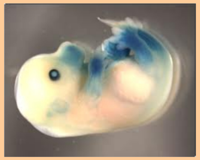
Genetics
DNA structure, universal genetic code
Mutations = changes in DNA sequence (+ epimutations)



Genomics
total gene number unrelated to phenotypic complexity
conserved genes in distinct species

```
CCTCCTCCATACCCAAA
TGGATGGTACGGCATTC
TTGAATCATCAAAGCTT
AGAGCGGGGGAATCGAC
CTCCTCCATACCCAAAT
```

Developmental biology
1990s: toolkit genes (set of genes with similar functions in multiple species)
Importance of cis-regulatory mutations for morphological evolution



Evolutionary genetics
similar phenotypic differences due to mutations in hotspot genes

SIMILARITIES

Where do we come from?

1859: Darwin and Wallace. Theory of evolution.
1920-40s: population genetics (evolution = change in allele frequencies)
Evolutionary genetics: intense exchange of genetic material and plagiarism

What makes us different?
Where do we come from?

INDIVIDUALITY

SIMILARITIES

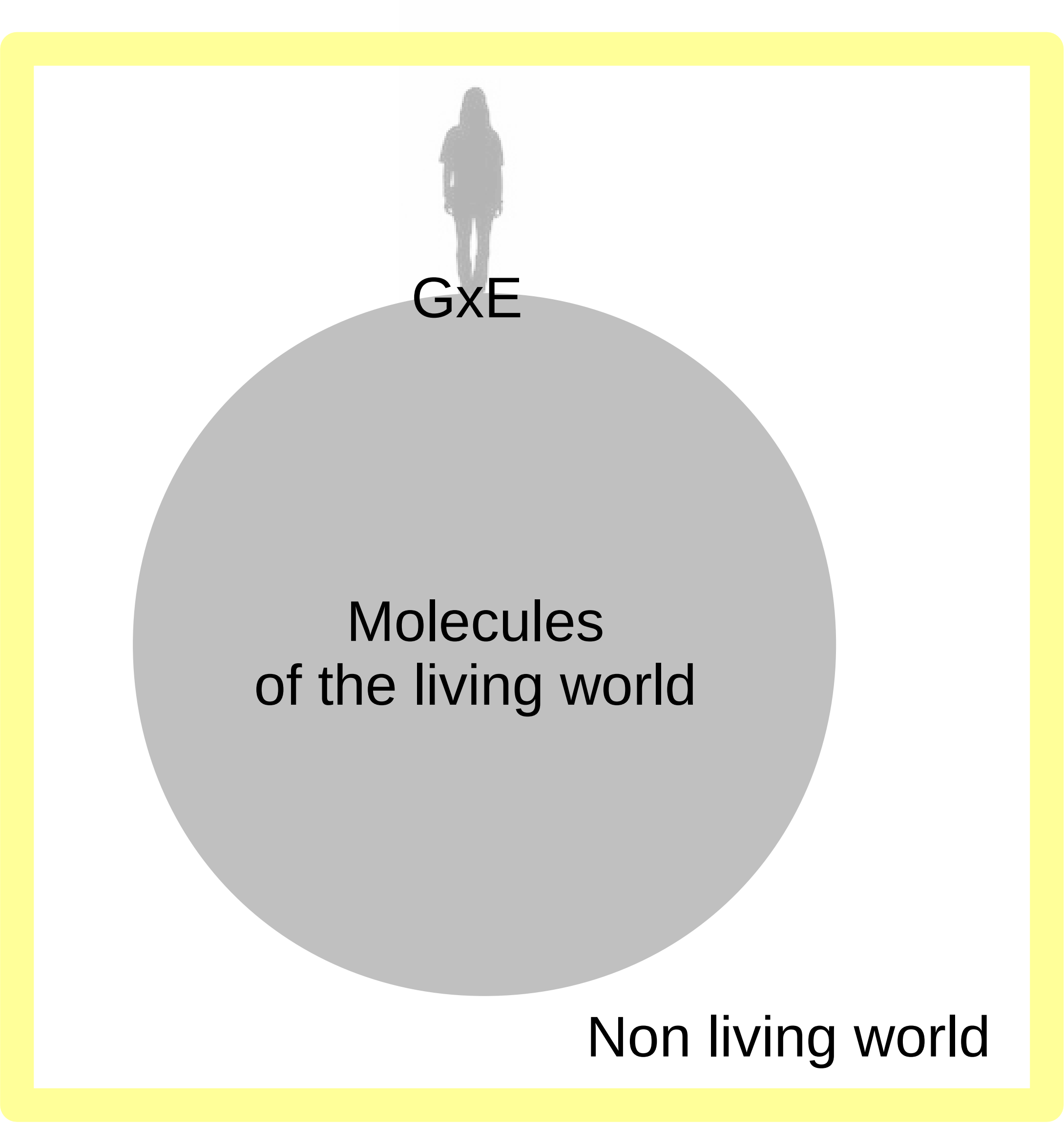
SHARED ORIGIN

SHARED MOLECULES

What makes us different?
Where do we come from?

INDIVIDUALITY

a combination of widespread phenotypic elements and novelties



GxE

Molecules
of the living world

Non living world

SIMILARITIES

SHARED ORIGIN

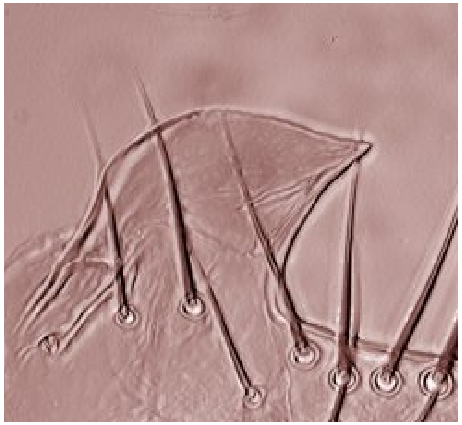
SHARED MOLECULES

Welcome to my lab!

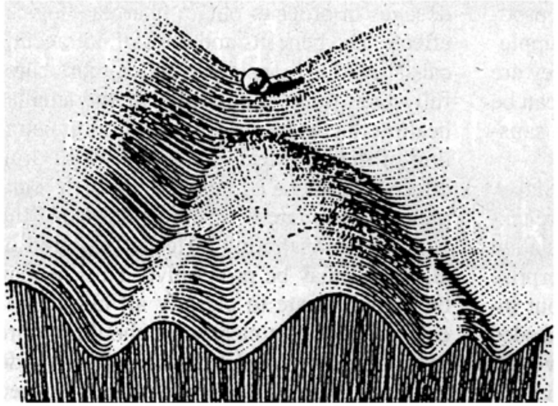
How two species adapted independently to the same ecological niche

How novelties arise

How precise organ shapes evolve



How new phenotypes become stable



Summary

The importance of DNA
The genes and the mutations responsible for phenotypic variation
Examples: *svb in flies, pitx1* in sticklebacks

Genes versus environment

The importance of stochasticity

The differential view of phenotype-genotype relationships

Our origin
Exchanges of genetic material and metabolites
Limited set of genetic paths to evolution

www.virginie.orgogozo.org
Institut Jacques Monod, Paris