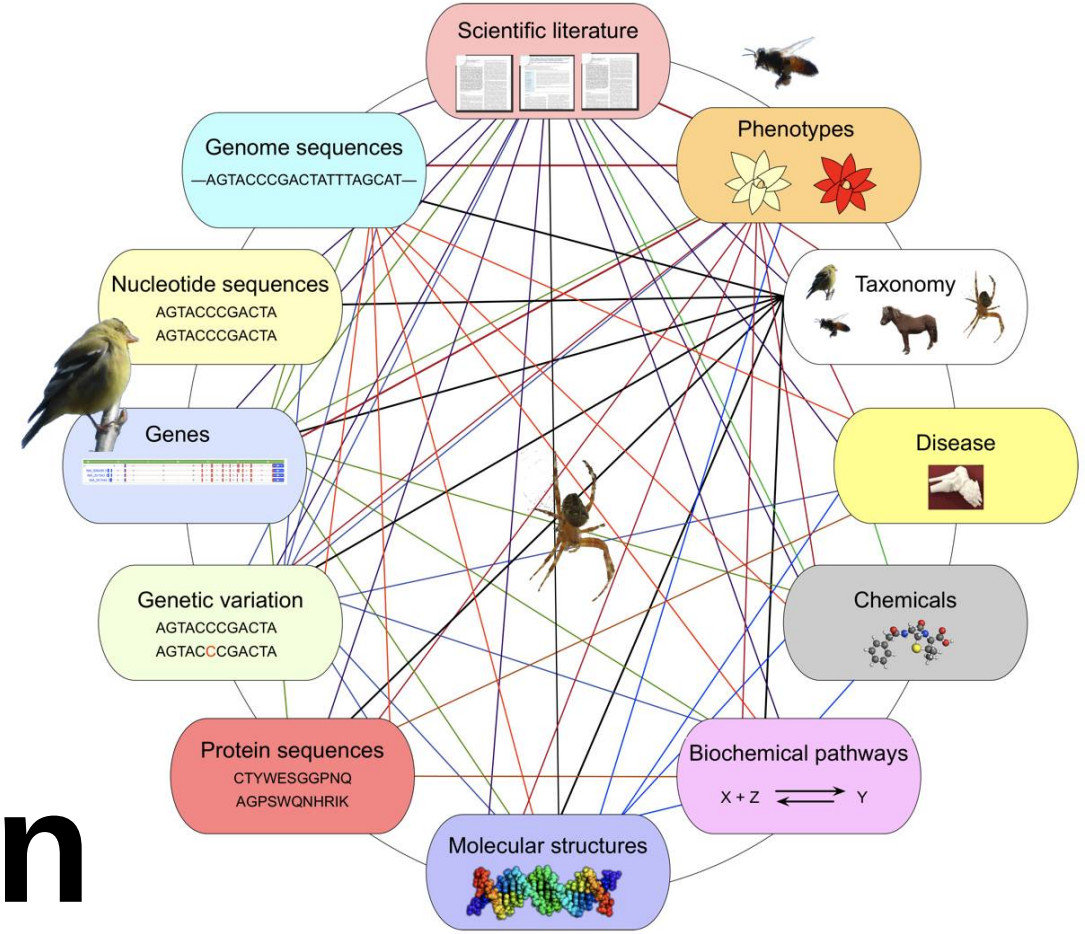
Biology Databases And Data Manipulation

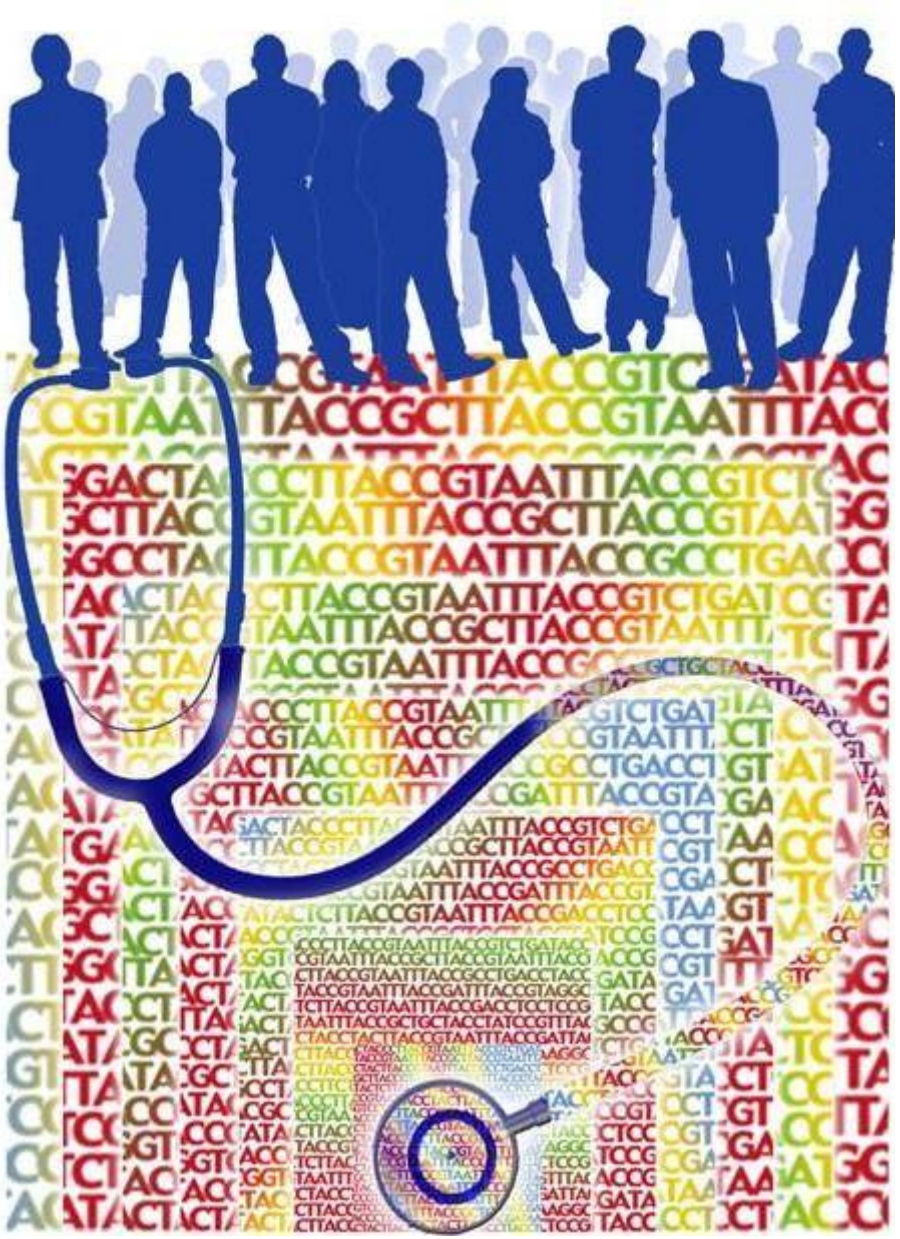


Sandra Porter, PhD

Digital World Biology LLC

V. Courtier-Orgogozo

Human genetic diversity



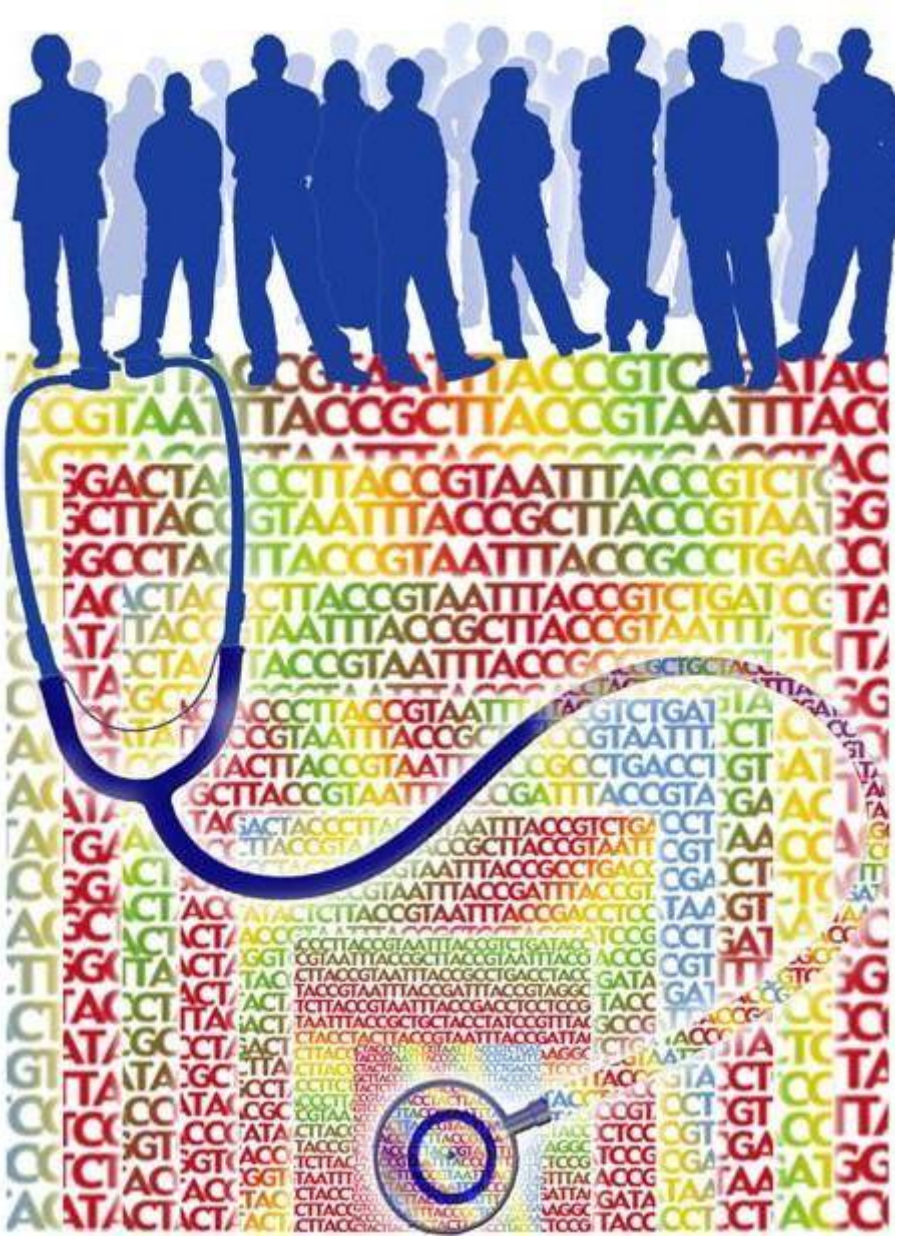
Genome size: 2.9 Gb
Gene number: 25 000
(1% of coding sequences)

In one individual:
~100 new mutations compared to his parents
~20 lethal mutations (heterozygous)

Genetic difference between two humans?

Genetic differences between humans and chimps?

Human genetic diversity



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

In one individual:
~100 new mutations compared to his parents
~20 lethal mutations (heterozygous)

Genetic difference between two humans?

~0.1%

Genetic differences between humans and chimps?

~4% (<1% for coding sequences)


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

Manipulate DNA sequences in silico

Manipulate DNA sequences in silico

Store
Store/retrieve data
Annotate

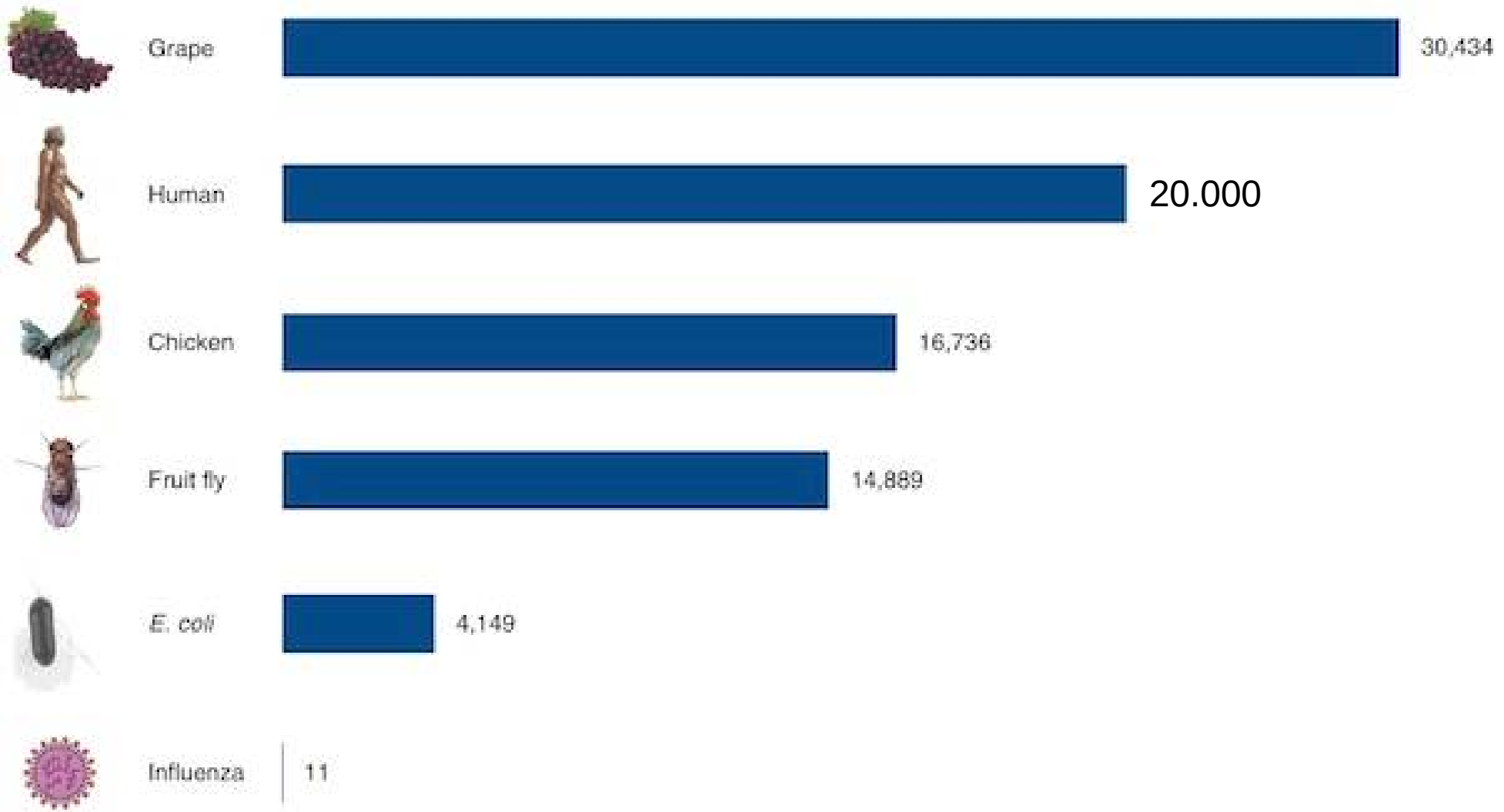
Assemble

Compare
BLAST
Align
Phylogenetic tree

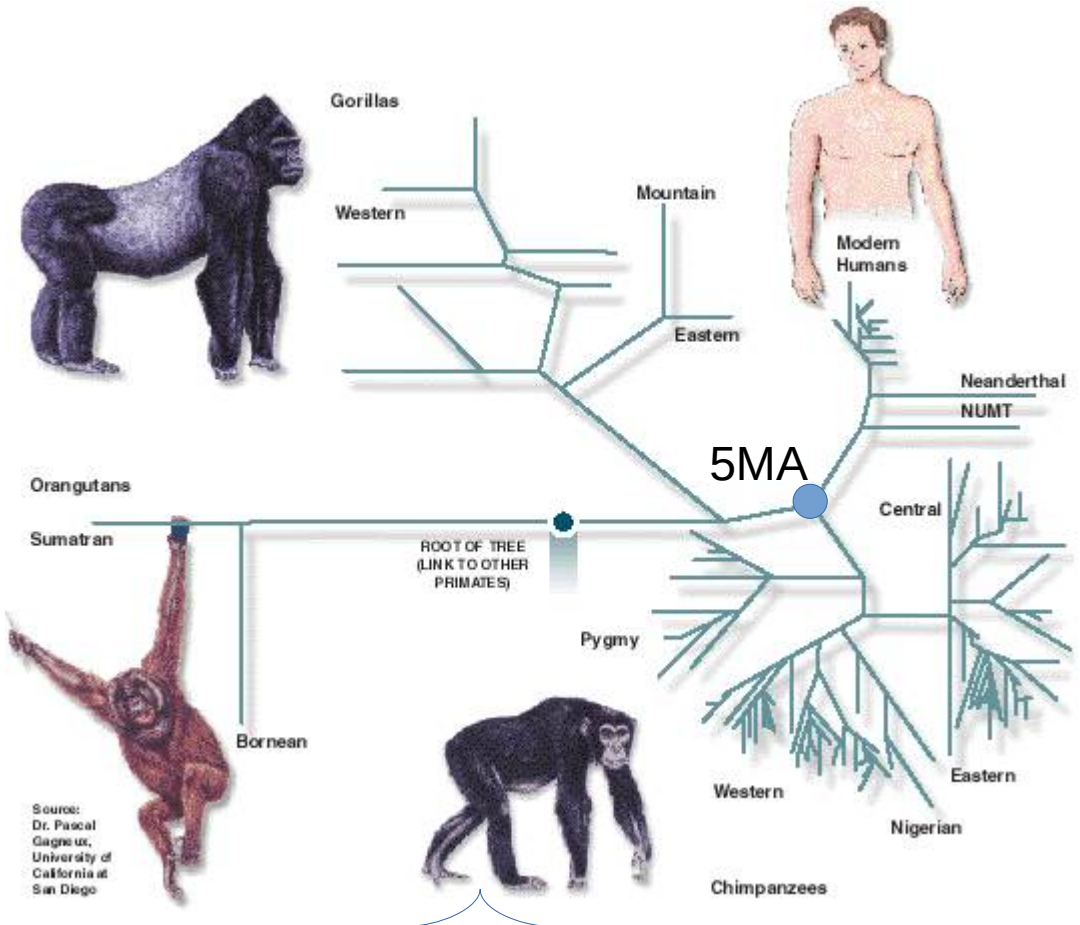
etc.

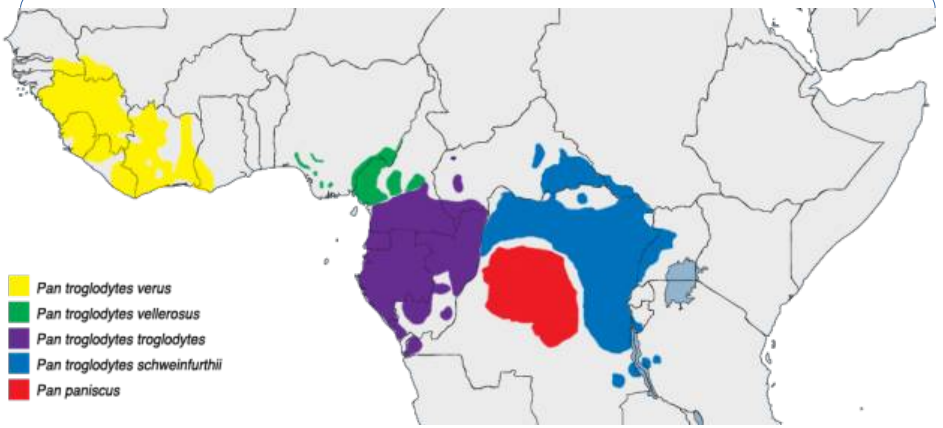
Nombre de gènes et complexité phénotypique

blé >94.000



L'homme et les autres primates





Taille du génome : 2,9 Gb = $2{,}9.10^9$ pb

Différence entre deux humains

~0.1%

Différence entre homme et chimpanzé

~5% (~1% for coding sequences)

Différence entre homme et Néanderthal

~0.13%

Différence entre deux chimpanzés

<250,000 chimpanzees, <110,000 gorillas,

<60,000 orangutans, and <50,000 bonobos

Peau, cheveux, santé…

Voici ce que Neandertal nous a vraiment légué

Si notre génome a conservé la trace de notre cousin, comment s'incarne cet héritage au niveau de nos gènes? Deux études américaines viennent de répondre. Avec quelques surprises à la clé…

Par **Emilie Rauscher**

En 2010, les rumeurs de leur union avaient défrayé la chronique scientifique… Le scandale était venu des laboratoires de paléogénétique du Max-Planck Institut (Allemagne), qui avaient fait d'une pierre deux coups : ils lisaient pour la première fois l'ADN de notre cousin perdu, Neandertal, et, par ricochet, en découvraient des traces dans notre propre génome, pourtant estampillé *sapiens*. La conséquence d'une relation forcément sexuelle.

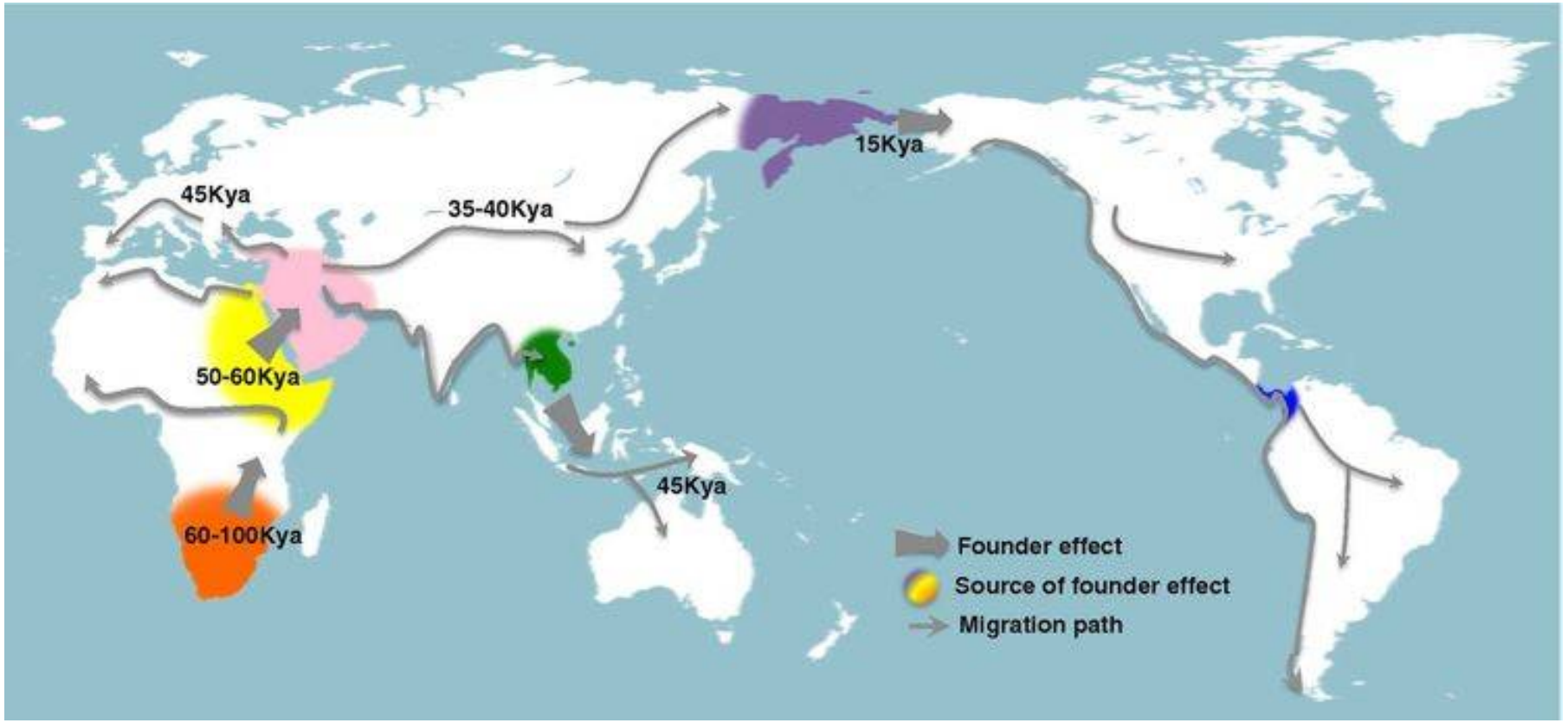
D'autres travaux avaient suivi, pour approfondir le sujet : il y avait eu relations, certes, mais il y a 47 000 à 65 000 ans, à notre sortie d'Afrique, et l'hypothèse d'une orgie généralisée était écartée puisque seul un faible pourcentage d'ADN néandertalien survit en nous.

Assez toutefois pour aiguiser la curiosité des biologistes : de lui à nous, qu'est-ce qui était passé ? *"Jusqu'ici, nous cherchions plutôt à confirmer qu'il y avait bien eu échange… Maintenant, nous traquons ces fragments à travers tout notre génome !"*, s'enthousiasme Joshua Akey, généticien à l'université de Washington.

Chronologie

Le dernier ancêtre commun à Neandertal et *Homo sapiens* est vieux de 500 000 ans. Puis, ces deux groupes ont évolué chacun de leur côté : en Europe et en Asie pour le premier, en Afrique pour le second. Si bien que quand ils se sont "retrouvés" il y a 40 000 ans au Moyen-Orient, ils avaient accumulé de nombreuses différences.

B. BOURGEOIS

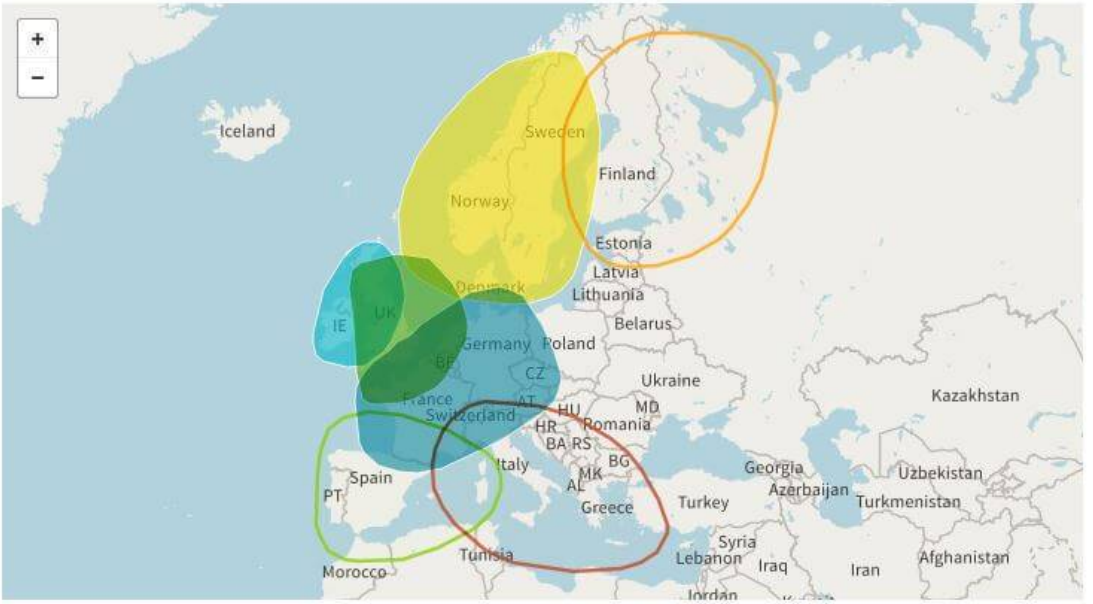


Afrique:
Plus grande diversité génétique

Séquences de Néanderthal: ~1-4% de notre génome, partout sauf en Afrique

Ethnicity estimate for Elaine Clark

REGION	APPROXIMATE AMOUNT
Asia	< 1%
Trace Regions	< 1%
Europe	99%
Europe West	40%
Ireland	21%
Great Britain	17%
Scandinavia	11%
Trace Regions	10%
Finland/Northwest Russia	5%
Italy/Greece	4%
Iberian Peninsula	< 1%



04-ancestry-com-o0mVUu2kRcs.mp4

Creating a phylogenetic tree

Model based approaches (computer intensive)

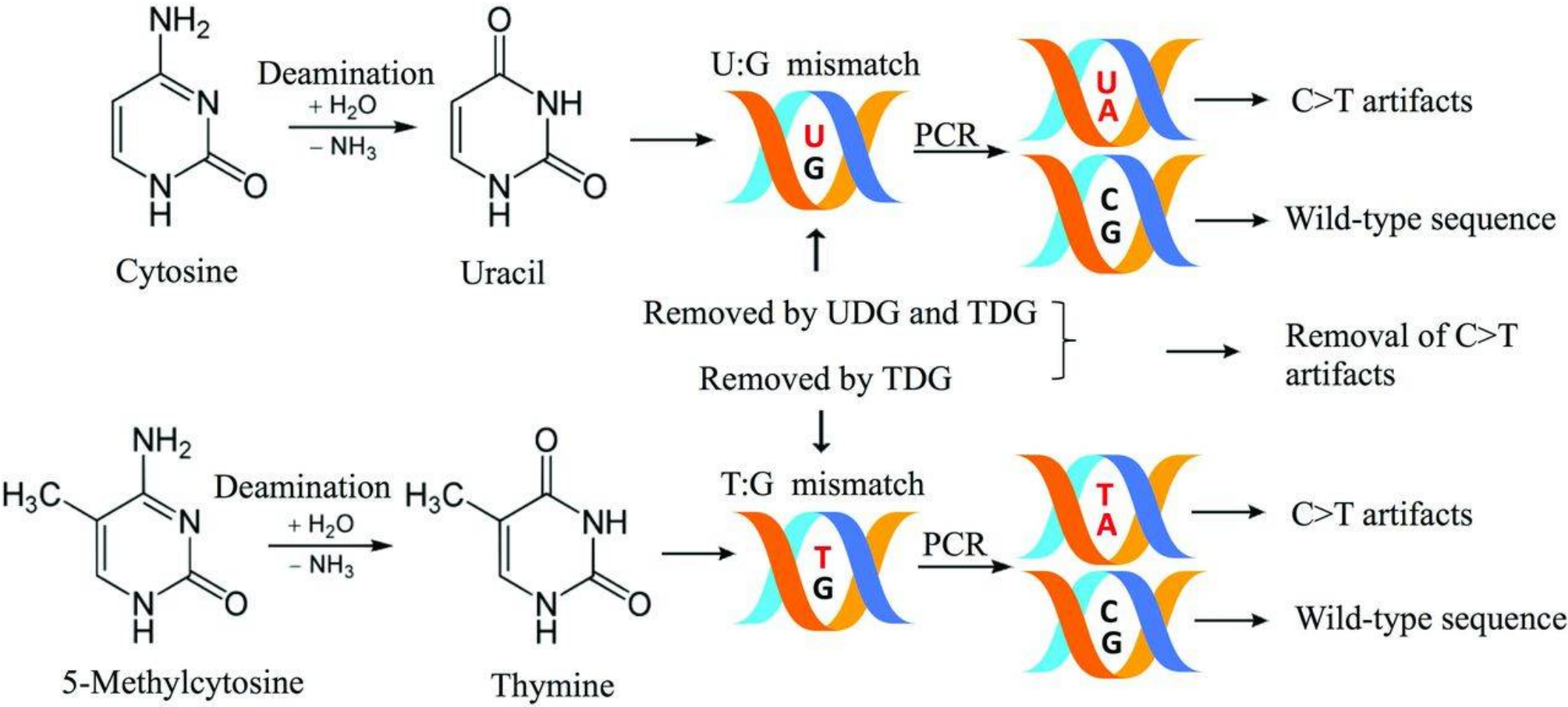
Assume a specific model (potentially very complicated) for the probability of evolutionary events occurring and then compare trees using these models.

Maximum likelihood: compute the probability of seeing the data we observe given the tree we are assuming. Pick tree with highest probability.

P(data|tree)

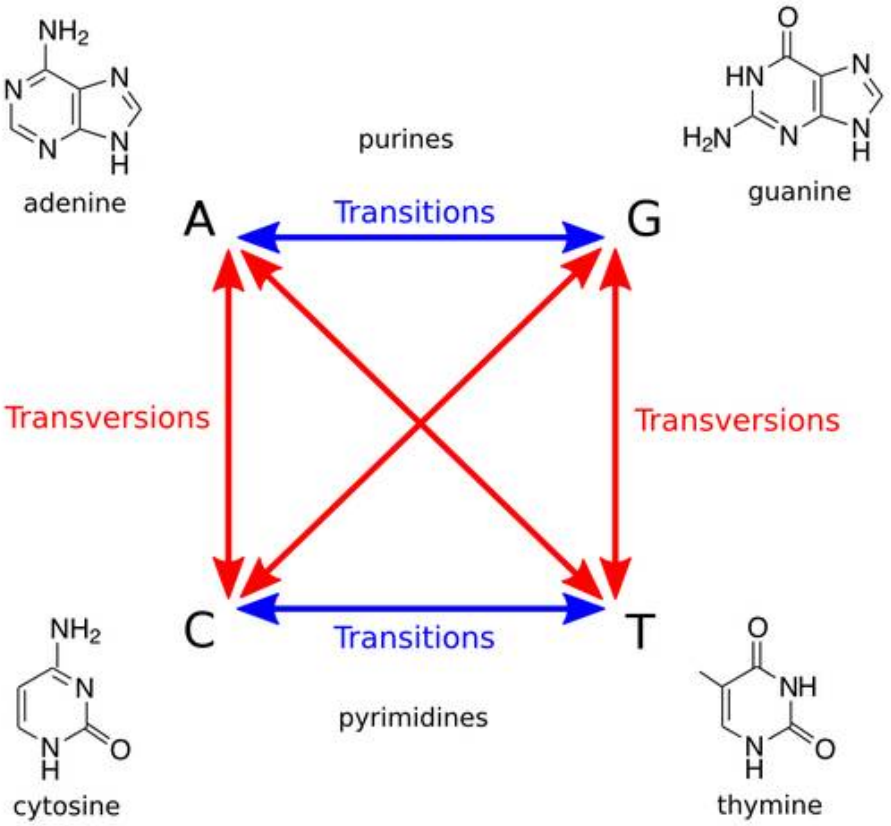
Bayesian: compute the probability of getting the tree we assume given the data we have. Pick tree with highest probability.

The most common mutation is C>T



UDG (uracile-ADN glycosylase) removes the uracil from U:G mismatches.
TDG (thymine ADN glycosylase) removes the uracil from U:G and thymine from T:G mismatches.

Do et al. 2017



IUPAC nucleotide code

Symbol	Bases	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T (or U)	T (or U)	Thymine (or Uracil)
W	**A** or **T**	Weak
S	**C** or **G**	Strong
M	**A** or **C**	aMino
K	**G** or **T**	Keto
R	**A** or **G**	puRine
Y	**C** or **T**	pYrimidine
B	C or G or T	not A (B comes after A)
D	A or G or T	not C (D comes after C)
H	A or C or T	not G (H comes after G)
V	A or C or G	not T (V comes after T and U)
N	any base	any Nucleotide (not a gap)

The standard statistical technique to determine the validity of a tree involves resampling our data to see how robust the tree is.

Bootstrapping is resampling (with replacement) to create a new data set.

- Pick n characters from the set (may duplicate some and omit others)
- Conceptually simulates robustness to gathering even more data

Jacknifing is subsampling (without replacement) to create a smaller data set.

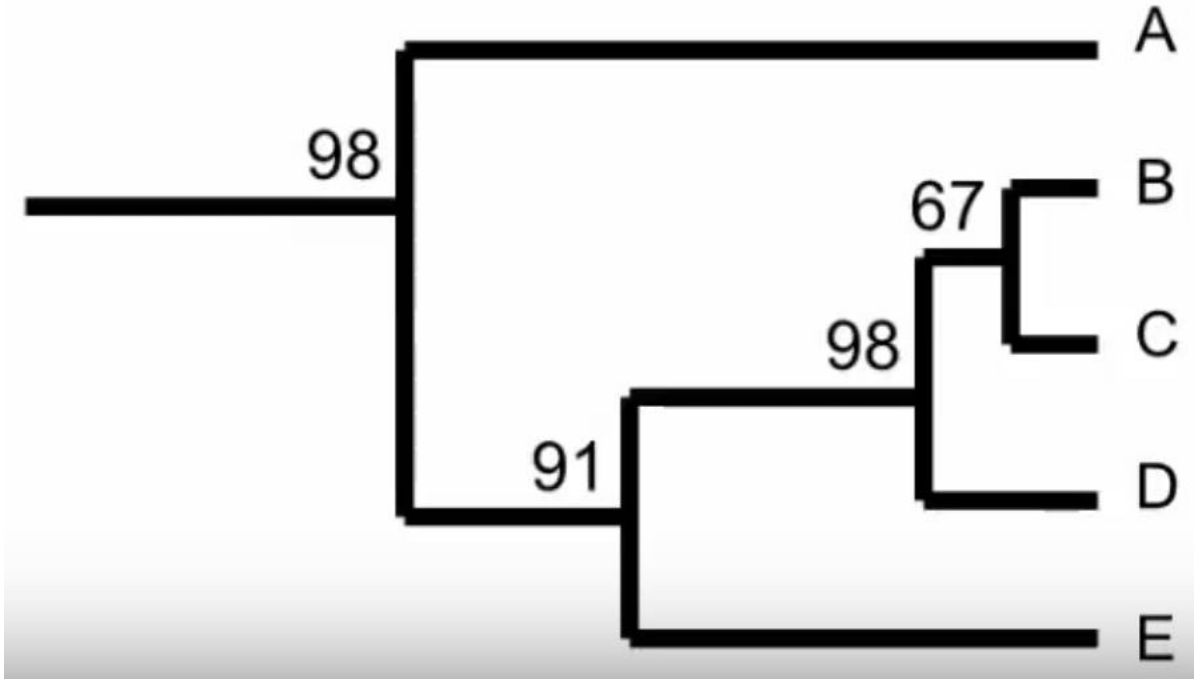
- Pick n characters from the set (will omit some)
- Conceptually simulates robustness to having gathered less data

Bootstrapping is more common.

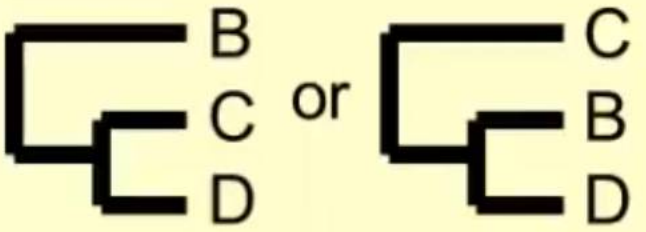
Bootstrap values for maximum likelihood trees

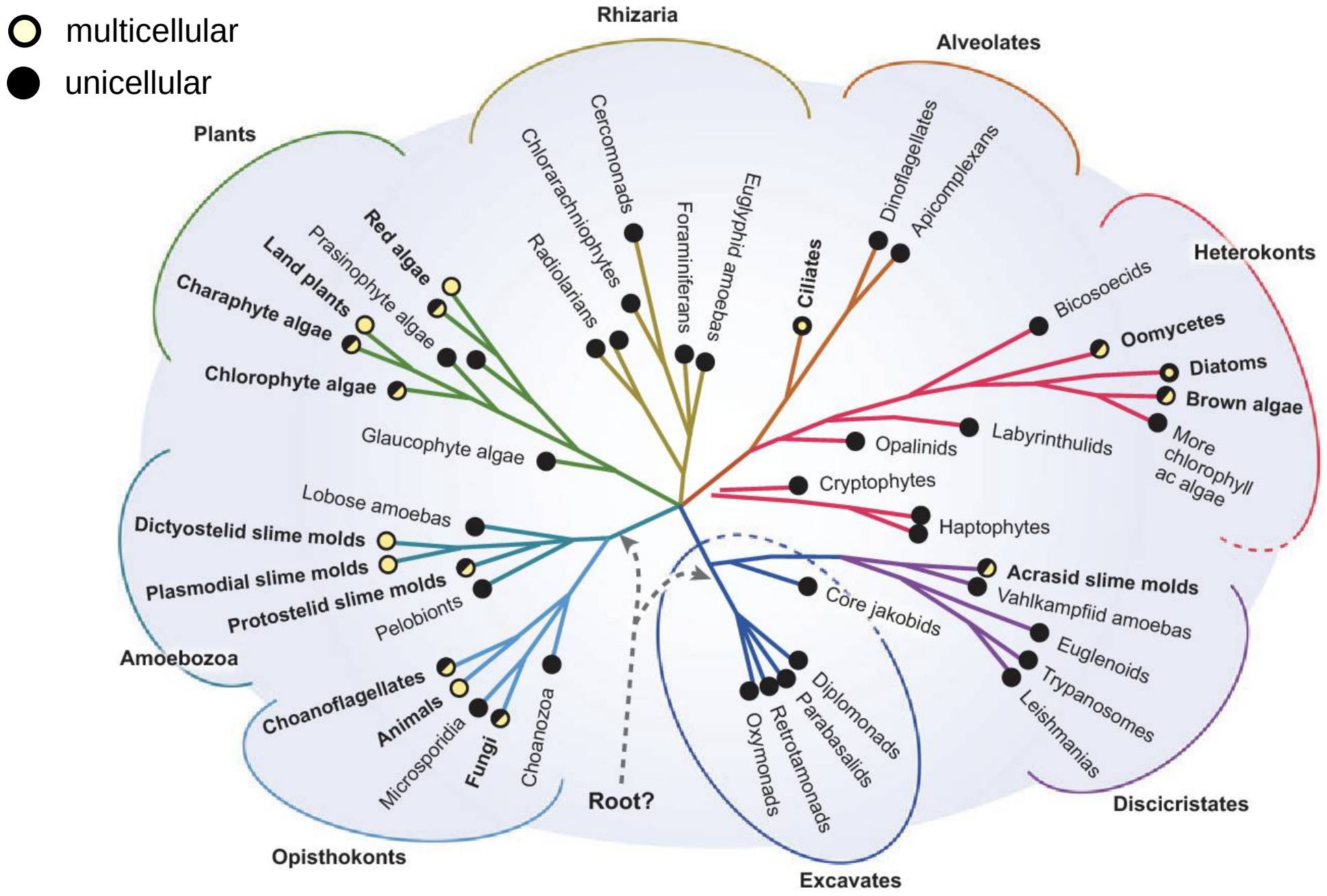
In practice the tree with the highest criteria score is shown and bootstrapping values are show for each node.

These values typically indicate the number of times out of 100 bootstraps that the branch was the same way in the best tree.



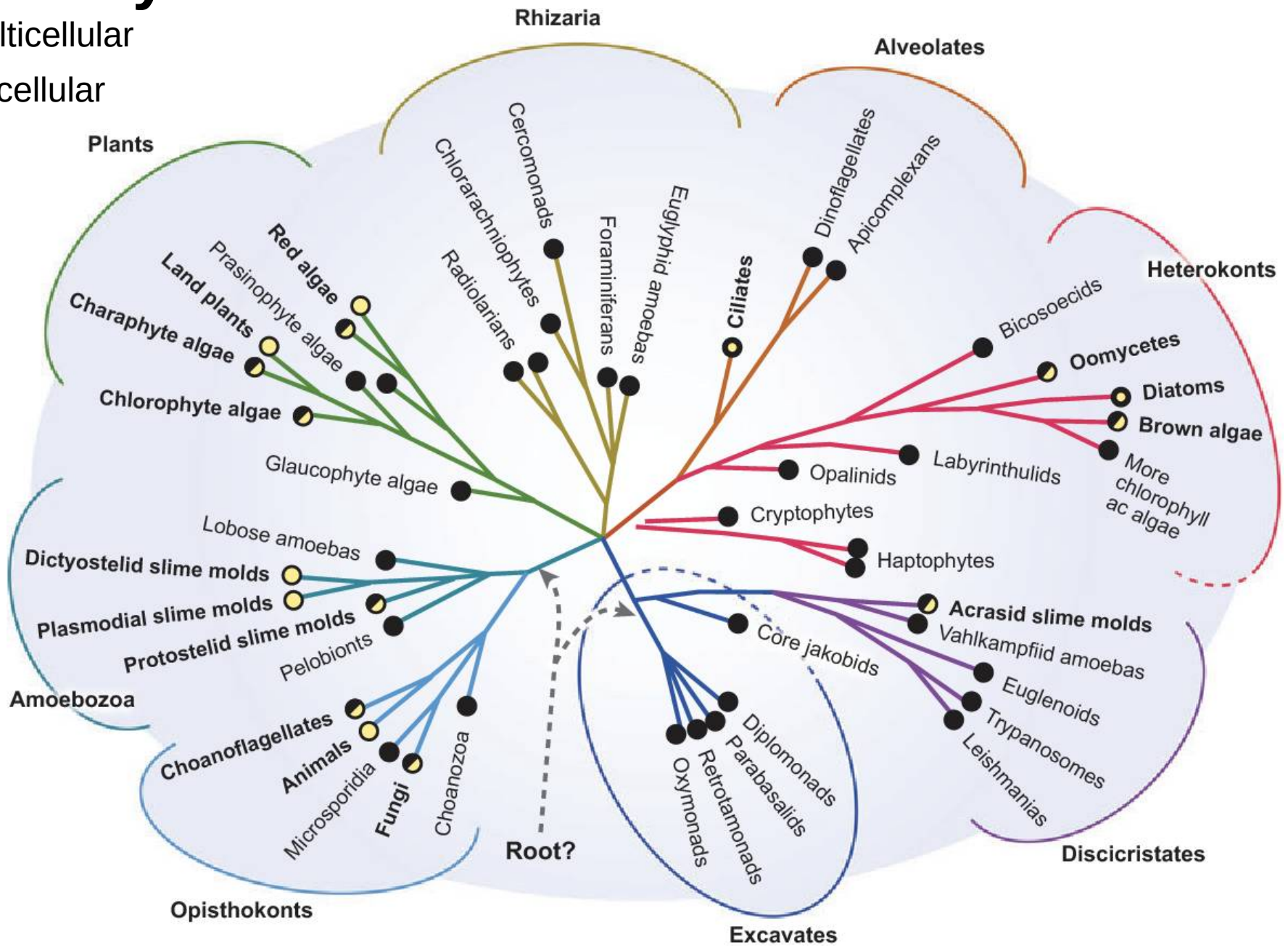
The 67 means that about 1/3 of the time




multicellular
unicellular
Rhizaria
Alveolates
Plants
Heterokonts
Amoebozoa
Opisthokonts
Excavates
Discicristates
Root?
Cercomonads
Chlorarachniophytes
Radiolarians
Foraminiferans
Euglyphid amoebas
Ciliates
Dinoflagellates
Apicomplexans
Red algae
Prasinophyte algae
Land plants
Charaphyte algae
Chlorophyte algae
Glaucophyte algae
Bicosoecids
Oomycetes
Diatoms
Brown algae
More chlorophyll ac algae
Labyrinthulids
Opalinids
Cryptophytes
Haptophytes
Lobose amoebas
Dictyostelid slime molds
Plasmodial slime molds
Protostelid slime molds
Pelobionts
Choanoflagellates
Animals
Microsporidia
Fungi
Choanozoa
Core jakobids
Acrasid slime molds
Vahlkampfiid amoebas
Euglenoids
Trypanosomes
Leishmanias
Diplomonads
Parabasalids
Retrotamonads
Oxymonads

Multicellularity evolved multiple times in Eukaryotes

- multicellular
- unicellular

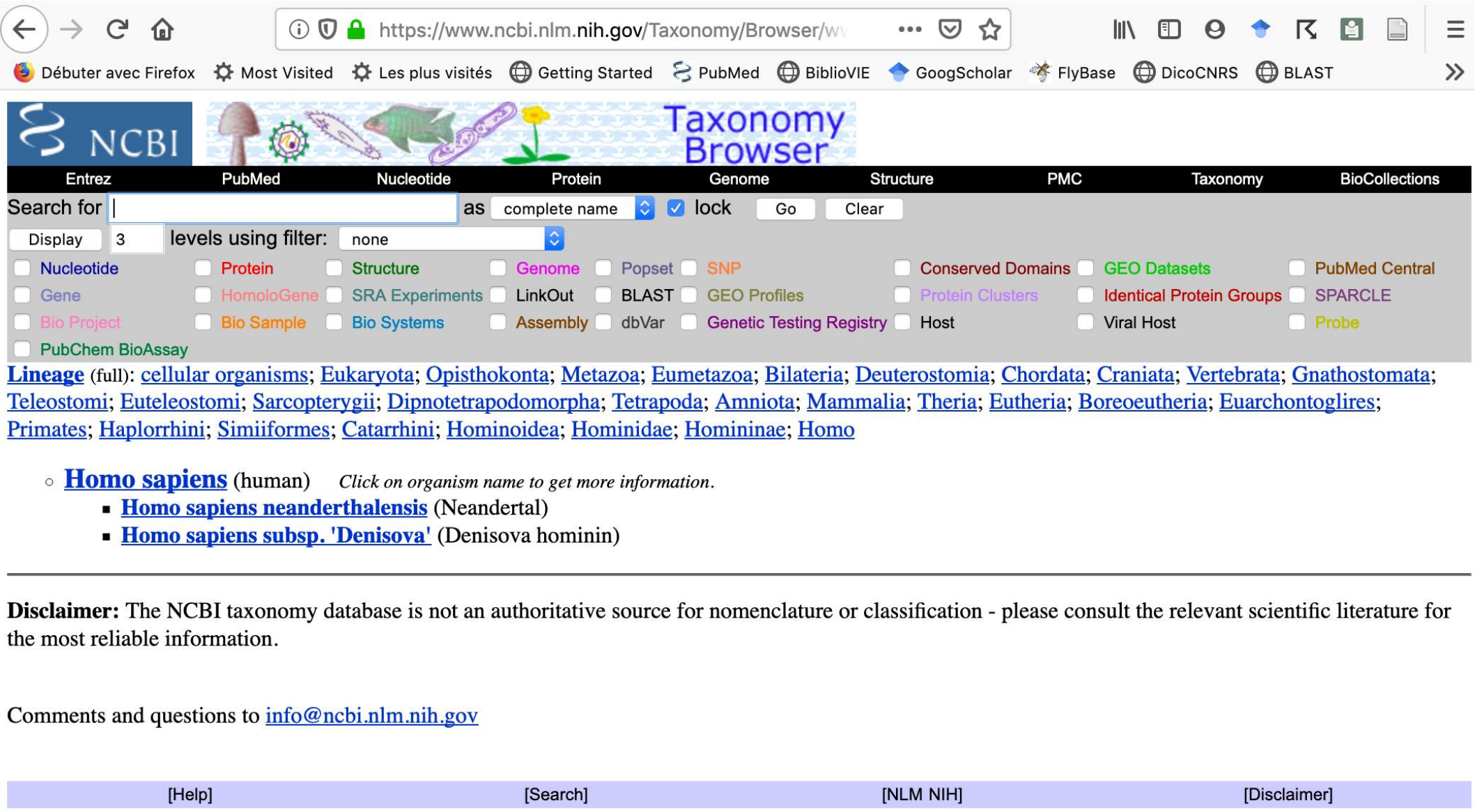


Taxonomy

Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Primates
Suborder:	Haplorhini
Infraorder:	Simiiformes
Family:	Hominidae
Subfamily:	Homininae
Tribe:	Hominini
Genus:	Homo
Species:	*Homo sapiens*

Linnaeus binomial name

https://www.ncbi.nlm.nih.gov/Taxonomy/



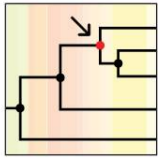
www.timetree.org

TimeTree is a public knowledge-base for information on the evolutionary timescale of life. Data from thousands of published studies are assembled into a searchable tree of life scaled to time. Three search modes are possible:

- **NODE TIME** - to find the divergence time of two species or higher taxa
- **TIMELINE** - to drill back through time and find evolutionary branches from the perspective of a single species
- **TIMETREE** - to build a timetree of a group of species or custom list

TIMEPANELS showing events in geological time and astronomical history are provided for comparison with timelines and timetrees. Results can be exported in different formats for additional analyses and publication.

NODE TIME



GET DIVERGENCE TIME FOR A PAIR OF TAXA

Specify 2 Taxon Names

Taxon 1:

Taxon 2:

Clear Search

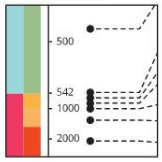
Resolve Ambiguity

Taxon 1:

Taxon 2:

Show Time

TIMELINE



GET AN EVOLUTIONARY TIMELINE FOR A TAXON

Specify a Taxon Name

Taxon:

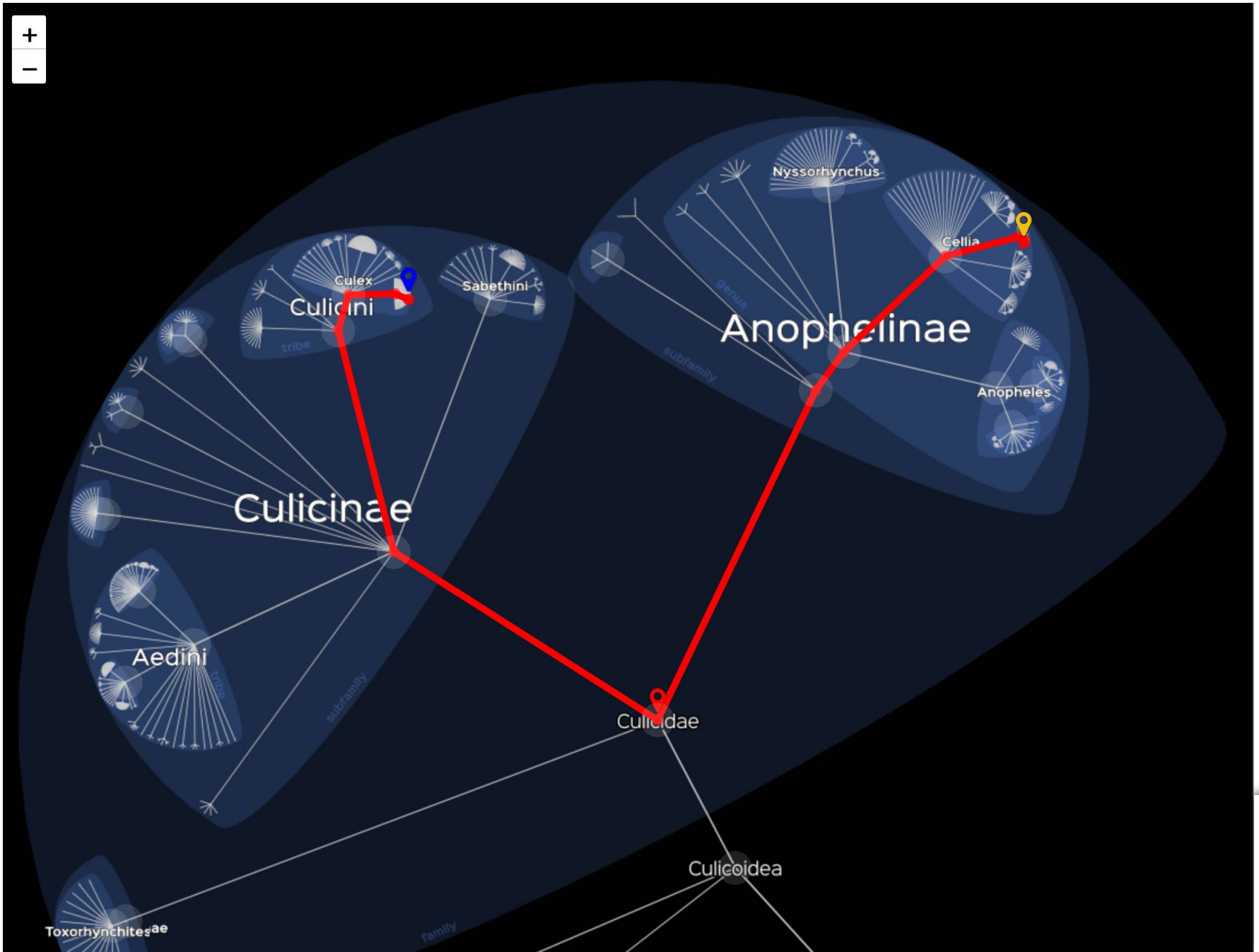
Clear Search

Resolve Ambiguity

Taxon

Show Timeline

http://lifemap.univ-lyon1.fr



temporary page
tolweb.org

Hominidae

Humans, great apes, and their extinct relatives



Phylogeny in part from Purvis (1995).

Containing group: Catarrhini

Other Names for Hominidae

Humans, great apes, and their extinct relatives

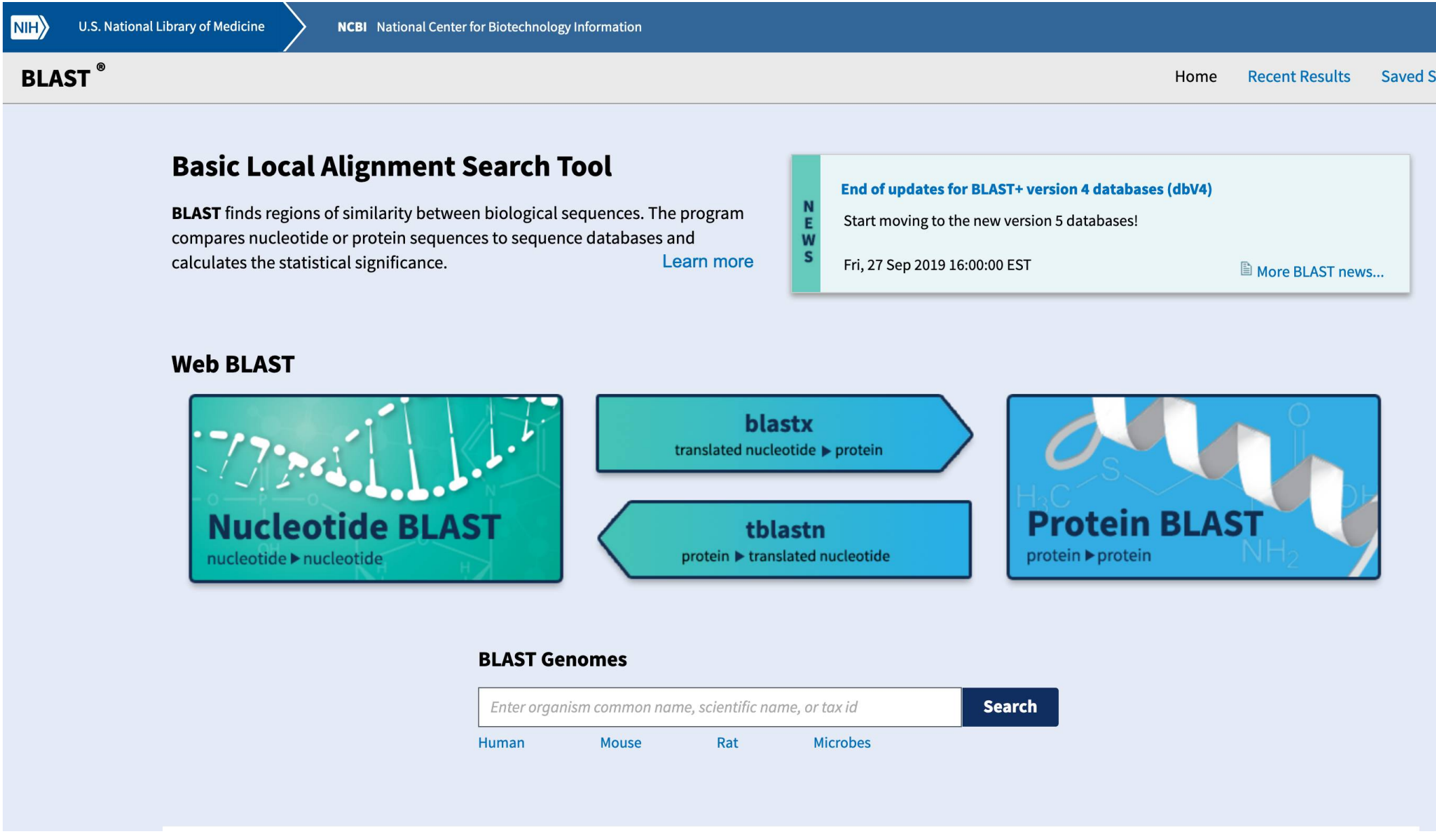
References

Hominidae
page content
articles & notes
treehouses
collections
people
options

Explore Other Groups

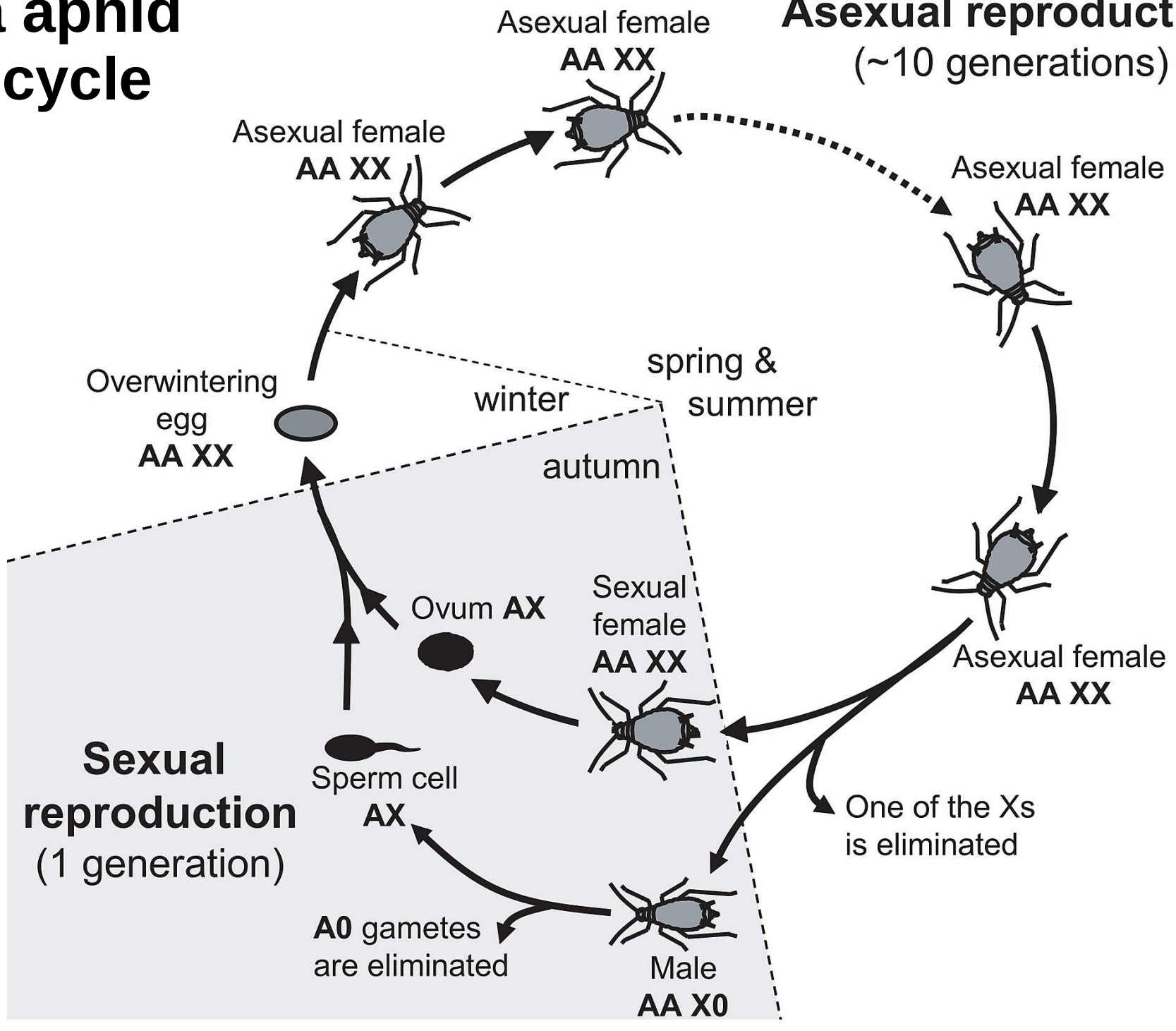
other Catarrhini
containing groups
subgroups
random page

https://blast.ncbi.nlm.nih.gov/Blast.cgi



À vous de jouer !

Pea aphid life cycle



Program TBLASTN Citation

Database nt See details

Query ID lcl|Query_96735

Description None

Molecule type amino acid

Query Length 602

Other reports

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to

E value to

Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments

Download | Manage Columns | Show 100

select all *7 sequences selected* GenBank Graphics

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase (LOC100161104), transcript variant X1, mRNA	249	249	98%	2e-71	28.43%	XM_001943135.5
PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase (LOC100161104), transcript variant X2, mRNA	249	249	98%	6e-71	28.43%	XM_016801018.2
PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase (LOC100574964), mRNA	245	245	99%	1e-70	28.87%	XM_003241620.4
PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase-like (LOC100571777), mRNA	213	213	83%	2e-61	30.53%	XM_008185904.3
PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase (LOC100159332), mRNA	206	206	99%	2e-56	26.70%	XM_001950752.5
Acyrthosiphon pisum ACYPI005179 mRNA, clone: 38I5, complete cds, full-insert cDNA sequence based on the ESTs (5'-EST:EX621862, 3	102	188	53%	6e-41	34.84%	AK340189.1
PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase-like (LOC103308018), mRNA	86.3	86.3	55%	1e-17	25.85%	XM_008180624.2

Descriptions | Graphic Summary | **Alignments** | Taxonomy

Alignment view: Pairwise | Download

7 sequences selected

Download | GenBank | Graphics | Next | Previous | Descriptions

PREDICTED: Acyrthosiphon pisum bifunctional lycopene cyclase/phytoene synthase (LOC100161104), transcript variant X1, mRNA

Sequence ID: XM_001943135.5 **Length:** 2952 **Number of Matches:** 1

Range 1: 191 to 1924 GenBank | Graphics | Next Match | Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
249 bits(635)	2e-71	Compositional matrix adjust.	174/612(28%)	300/612(49%)	54/612(8%)	+2

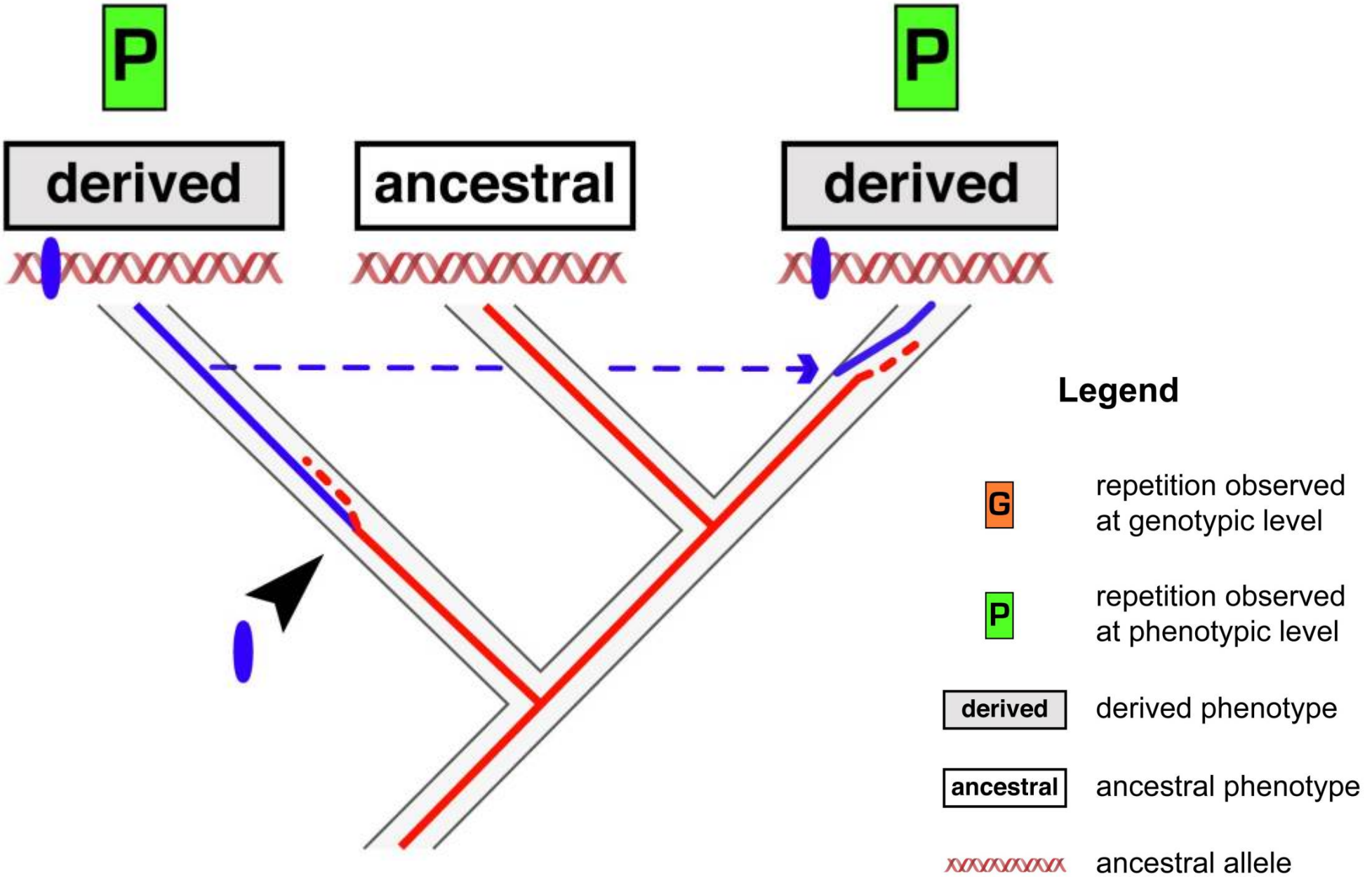
```
Query  1     MYDYAFVHLKFTVPAAVLLTAIAYPILNRIHLIQTGFLVVVAFTAALPWDAYLIKHKVWS  60
             M  Y  VH  +T+P   +L  I +P ++R+ L + GF+  +AF    PWD Y+I H  W
Sbjct  191   MLTYIDVHFIYTLPVVAVLALITWPFISRLELFKIGFVCTMAFVYTTPWDNYIIFHNAWM  370

Query  61    YPPEAIVGPRLLG-IPFEELFFFVIQTYITALVYILFNKPVLHALHLNNQQNPPAWMRVV  119
             Y P+ I+   ++G +P EE  FFVIQT +T+L  ++F +      + N  +     +R +
Sbjct  371   YKPKNILA--VIGYVPVEEYMFFVIQTLMTSLWALVFTRWSPACFNFNFNKTSYTLIRWI  544

Query  120   KVTGQVVLVALSVWGWNAAQVHQETSYLGLILVWACPFLLAIWTLAGRFILSLPWYATVL  179
              +   + LV  ++ G+N A   + T YLG I+ W+CP ++ +W  AG + +     ++ +
Sbjct  545   PI---LALVMTTIQGYNIAVPGKNTFYLGCIMWWSCPVIMFLWYGAGNYFVKKS-TSSAI  712

Query  180   PMFLPTFYLWAVDEFALHRGTWSIGSGTKLDFCLFGKLDIEEATFFLVTNMLIVGGMAAF  239
              + +PT YL  VD  AL    W I   T L+  +   L  EE  FFL+TN++IV G  AF
Sbjct  713   AVIVPTLYLCWVDRIALKDDVWHINEKTSLNIFVVDDLPFEECLFFLITNVIIVLGGMAF  892

Query  240   DQYLAVIYAFPTLFPKVNRYPTTHMLLQSRLINTSRYDLER----IEGLREAVERLRLKS  295
             D+   +   +  FP   RY ++      ++   R + +     +  +R+ +  L+  S
Sbjct  893   DKSYGLADTYTFEFPL--RYSSSWKYYSQQMQQFVRAECDMSPSPVNDIRQCLNVLKRAS  1066

Query  296   RSFYLANSLFSGRLRIDLILLYSFCRLADDLVDDAKSRREVLSWTAKL-NHFLDLHYKDA  354
             +SF +A+ +F   +R+ LI+LY+FCR+ DD++ D++ +  V     KL   F+D  + D
Sbjct  1067  KSFNVASLVFPAGVRLHLIILYAFCRVTDDMI-DSEPKVGVKKQKLKLIETFIDELFADR  1243

Query  355   DATED------PKKKAERIDAY--IKTAFPPCAYQALHLLPTHILPPKPLYDLIKGFEMD  406
              A  D      P+K   + + Y    T      ++A+  +  + LP KP Y+L+ G+  D
Sbjct  1244  SADYDVKTSMTPRKPEVKWEQYRLDLTDEELSCFRAISRISFY-LPRKPFYELLDGYRWD  1420
```

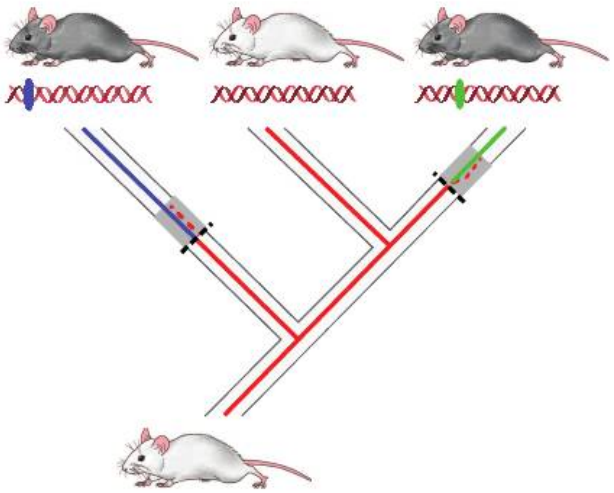
Related Information

Gene - associated gene details

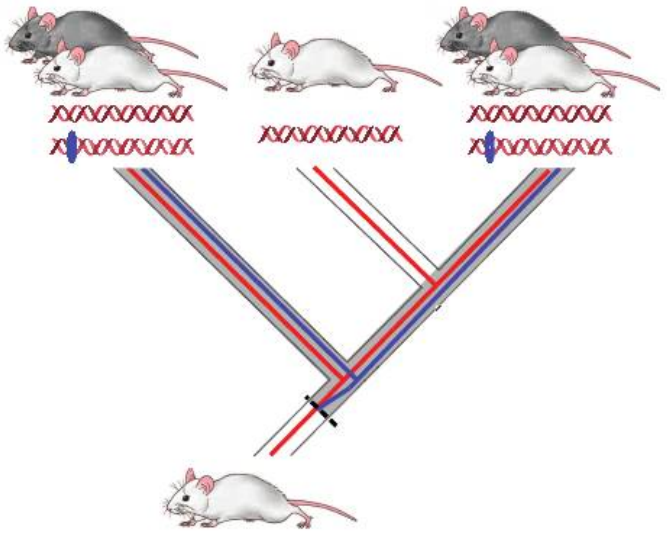
Genome Data Viewer - aligned genomic context

Lateral transfer

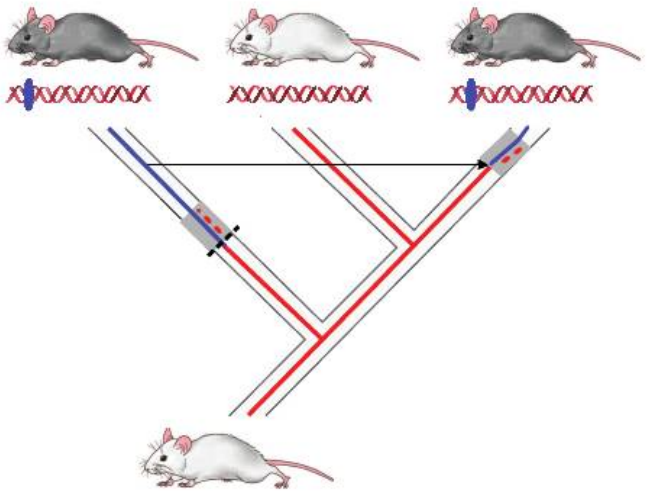




Genetic convergence



Ancestral polymorphism



Lateral transfer

OÙ ILS SE SONT CROISÉS

- Pelage : **brun**
- Taille : **1,70 à 2,80 m** *(tête et corps)*
- Membres : **griffes non rétractiles**
- Régime alimentaire : **omnivore**
- Lieu de vie : **forêts, zones côtières, montagnes**
- Période d'accouplement : **mai à juillet**

- Pelage : **blanc**
- Taille : **1,80 à 3 m** *(tête et corps)*
- Membres : **doigts partiellement palmés**
- Régime alimentaire : **carnivore**
- Lieu de vie : **banquise**
- Période d'accouplement : **avril à juin**

Grizzly ou ours brun

Ours polaire

Pizzly ou grolar **Le pizzly**, encore appelé grolar ou prizzly, **a été trouvé sur l'île de Banks**, au nord-ouest du Canada, **en 2006.**

Brown Bears

1. Mainland Alaska ♀
2. ABC Islands ♀

Polar bears

3. West Hudson Bay ♀
4. West Hudson Bay ♂
5. North Beaufort Sea ♂
6. Lancaster Sound ♂
7. South Beaufort Sea ♂
8. Wrangel Island ♀
9. Chukchi Sea ♂

Cahill et al 2013

The enigmatic ABC Islands brown bears likely derive from a population of polar bears likely stranded by the receding ice at the end of the last glacial period. Since then, male brown bear migration onto the island has gradually converted these bears into an admixed population whose phenotype and genotype are principally brown bear. This process of genome erosion and conversion may be a common outcome when climate change or other forces cause a population to become isolated and then overrun by species with which it can hybridize.

Cahill et al 2013

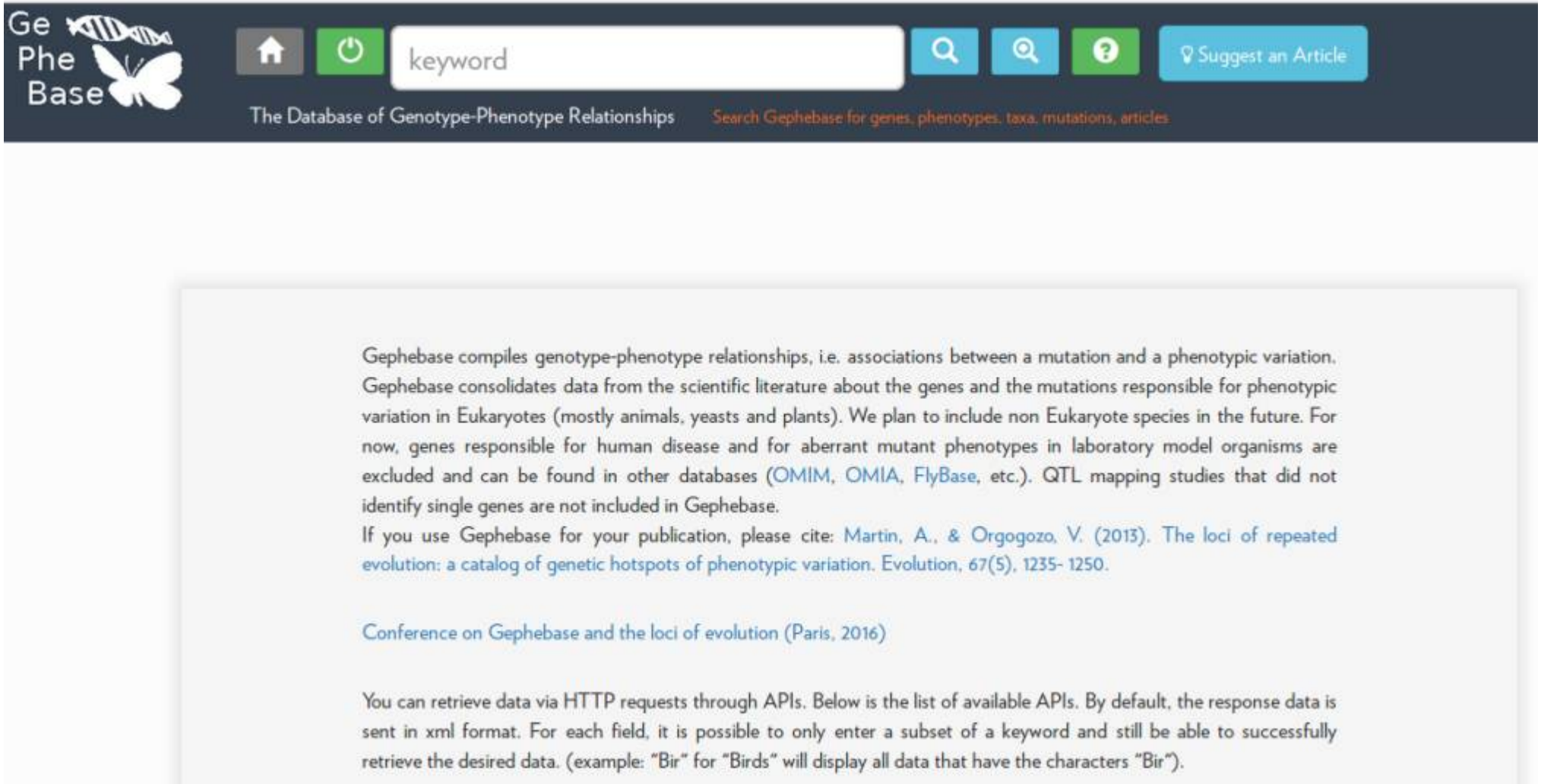
The genes and the mutations responsible for phenotypic evolution

---

Evolution repeats itself

www.Gephebase.org
>2000 entries

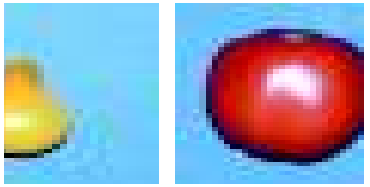
@gephebase



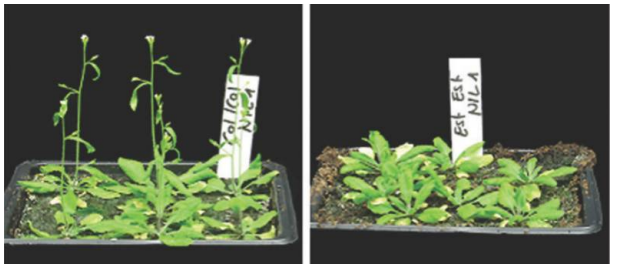
Orgogozo et al Nucleic Acid Research 2019

Wrinkled seed: TE insertion
(Bhattacharyya 1990)

OVATE coding region
(Liu 2002)

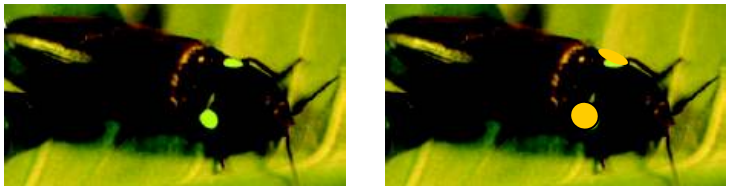


FRI coding region
(Johanson 2000)



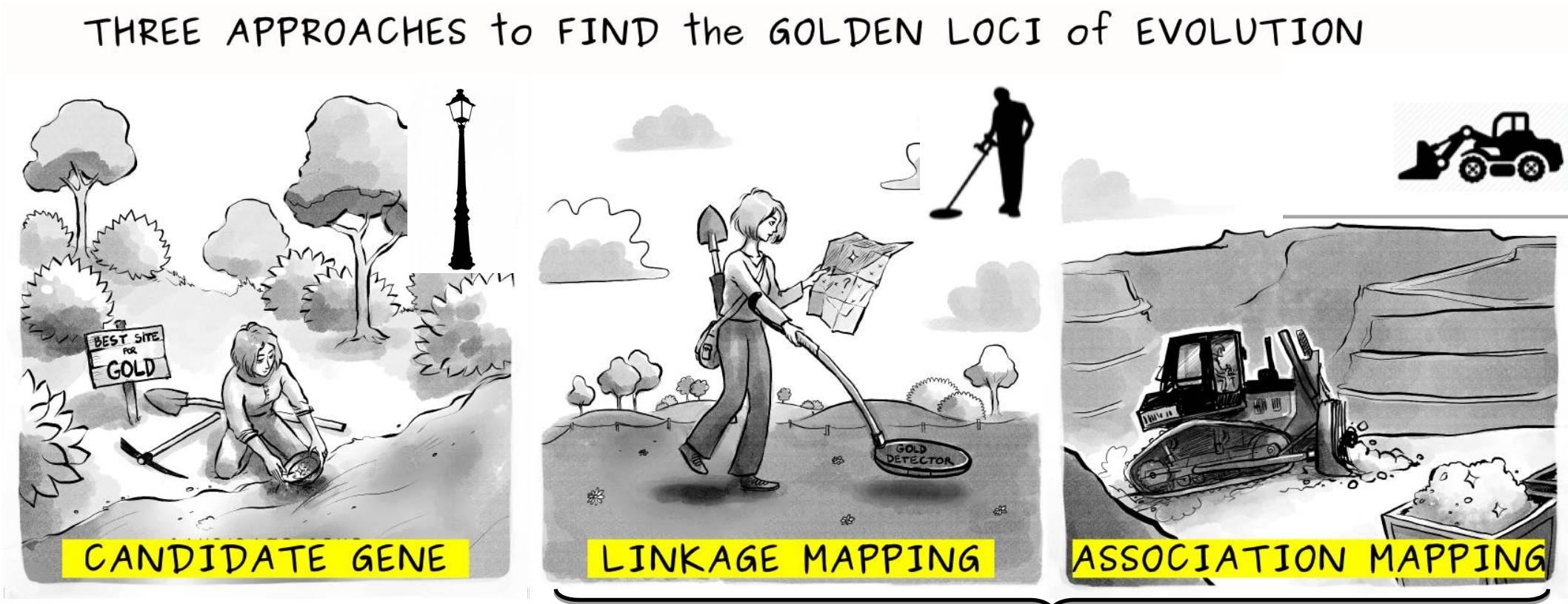
myostatin coding region
(Grobet 1997)

luciferase coding region
(Stolz 2003)



anthocyanin-2 coding region
(Quattrocchio 1999)

Mc1r coding region
(Eizirik 2003)



REVERSE GENETICS

From genes to traits

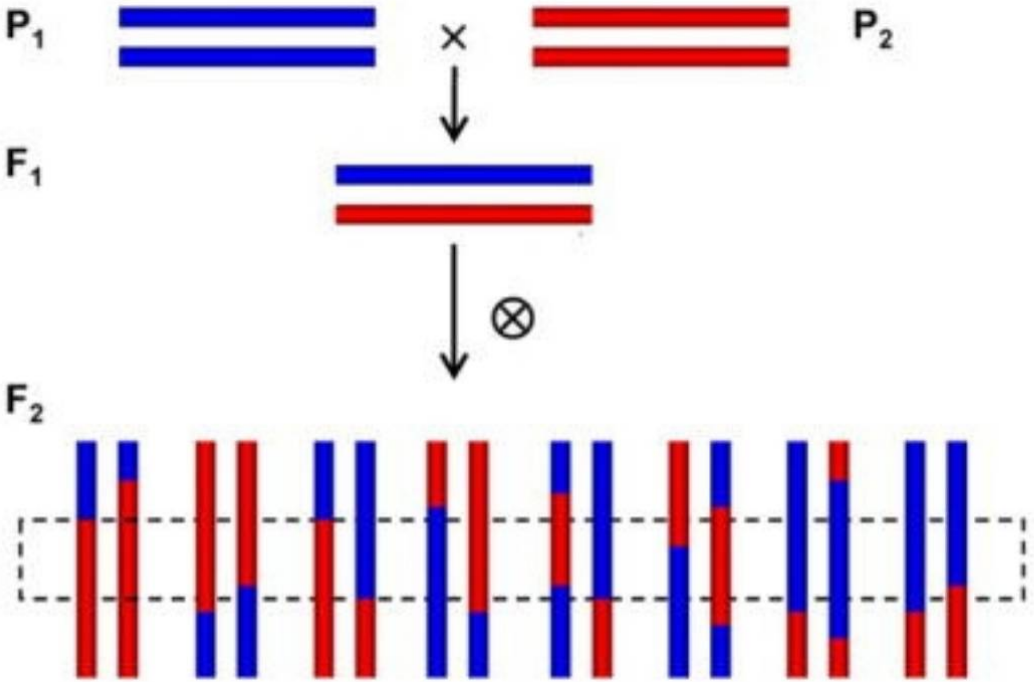
FORWARD GENETICS

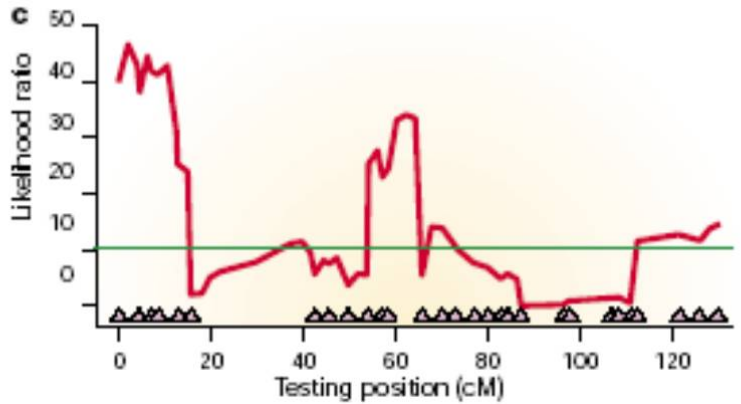
From traits to genes

Little Ascertainment Bias, but Micro-Evolution only

Linkage Mapping

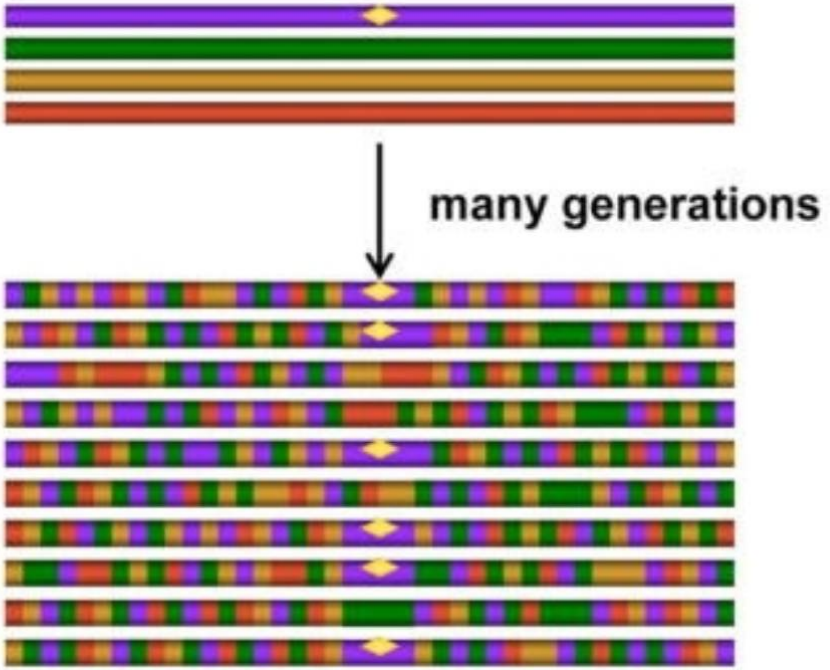
Crosses in the lab

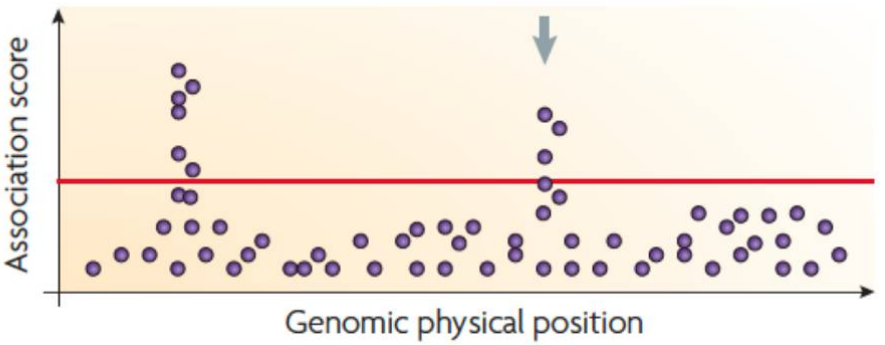




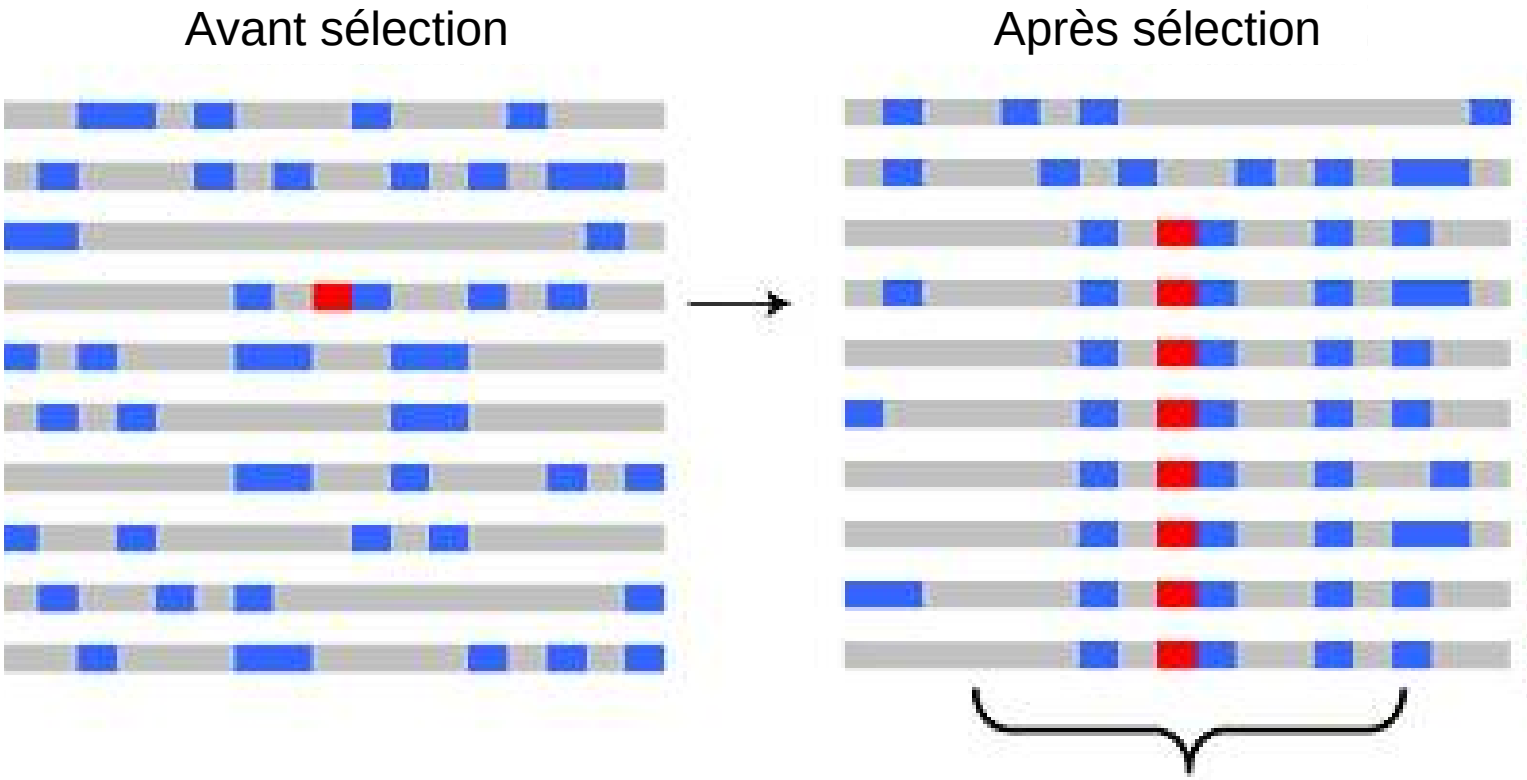
Association Mapping

Past crosses in natural populations





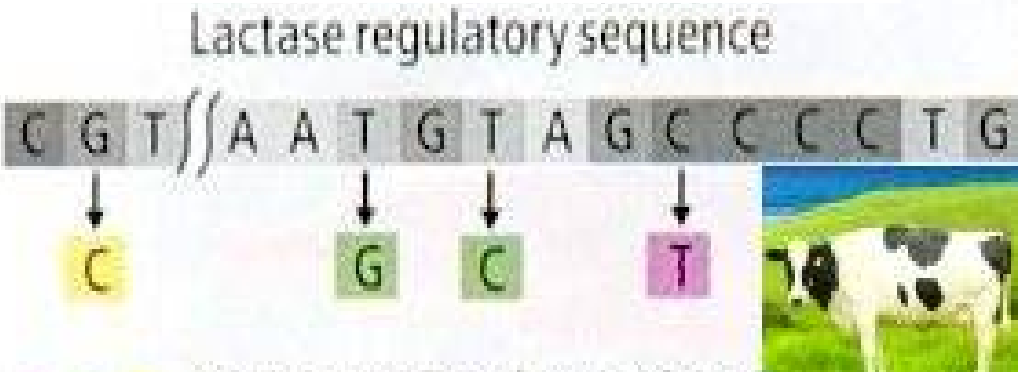
Des événements de sélection récente laissent des traces dans les génomes



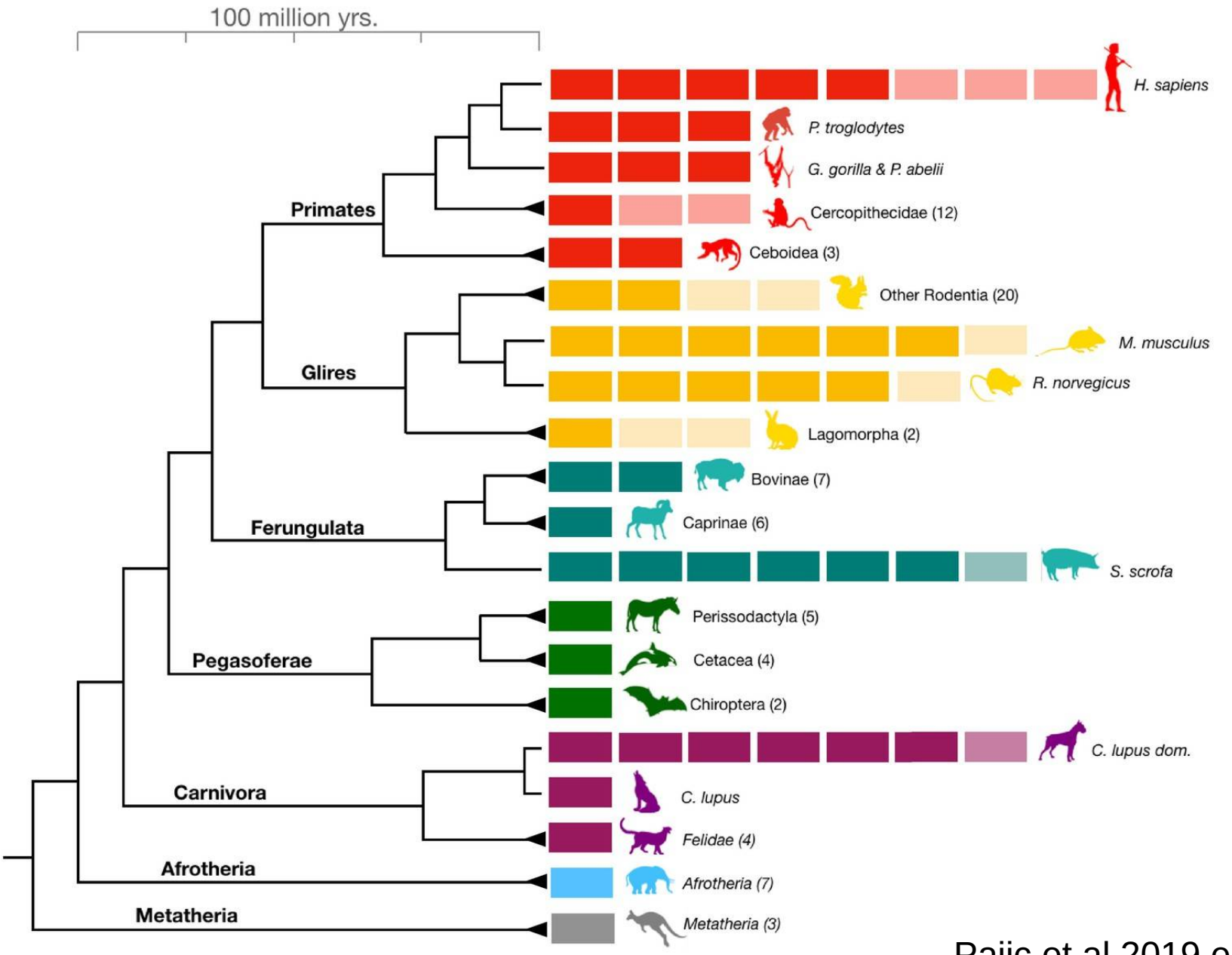
Ex: *lactase*, *CCR5-Δ32* et résistance au VIH

Repeated evolution in the *lactase* gene

Séquence régulatrice du gène *lactase*

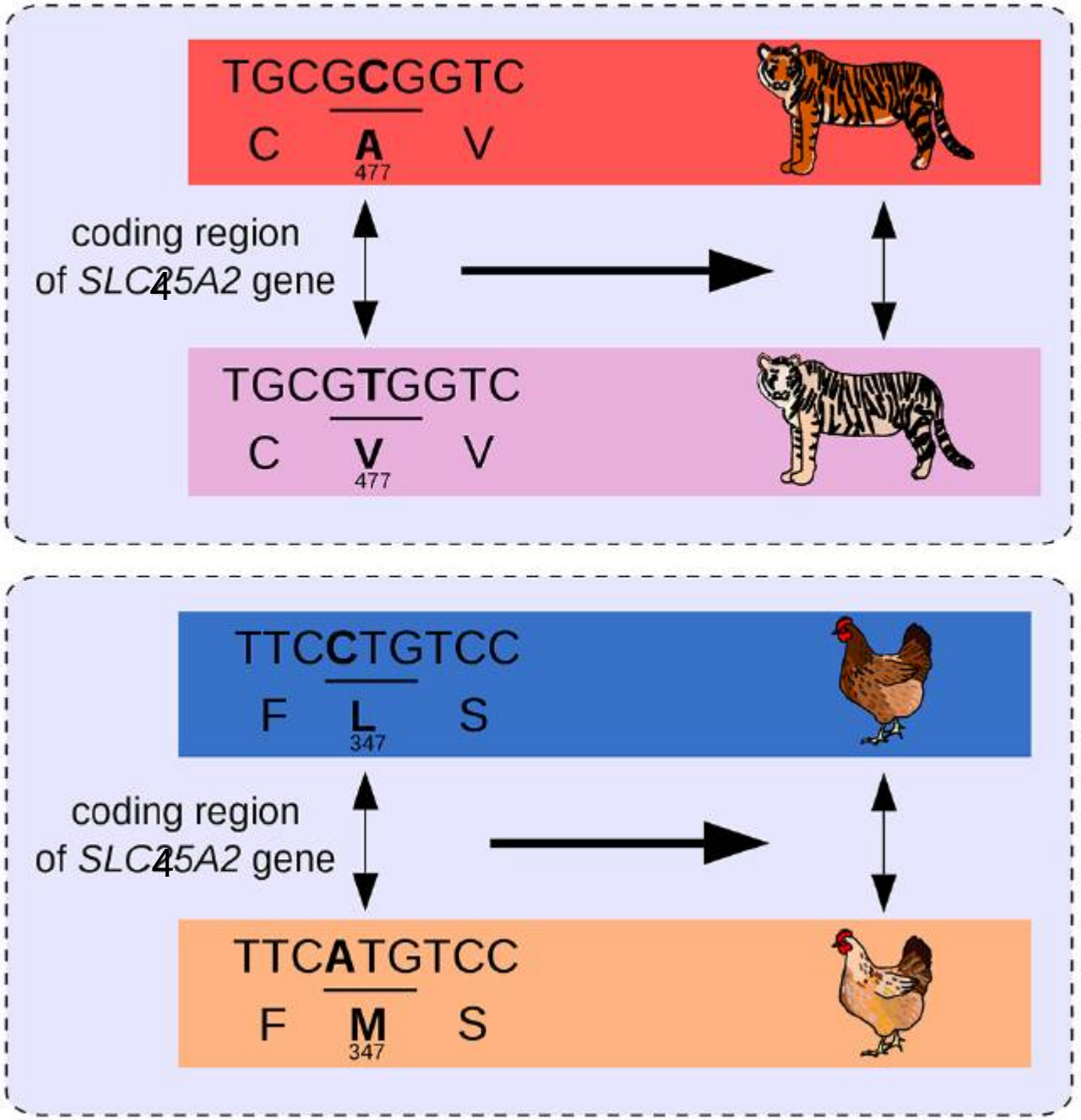


Digestion de l'amidon
Duplications répétées du gène amylase



Pajic et al 2019 eLife

Repeated evolution



Also in:
Humans
Horses
Quails
Chickens
Mice
Pigeons

Orgogozo et al 2015 Frontiers Genetics

Repeated evolution via the same amino acid change

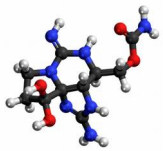
clam

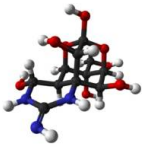
garter snake

Nav1.4
X E945D

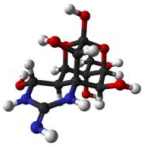
Nav1.4 sodium channel
X E945D

Saxitoxin

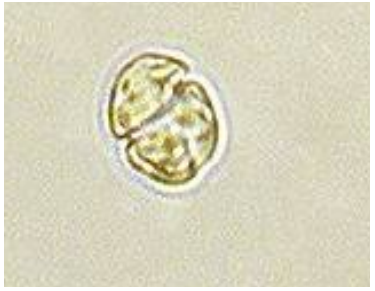




Tetrodotoxin



Nav1.4
E945D
X



toxic plancton

toxic newt

pufferfish (fugu)

Bricelj 2005
Geffeney 2005
Venkatesh 2005

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)

.. the phenotypes (convergent evolution)

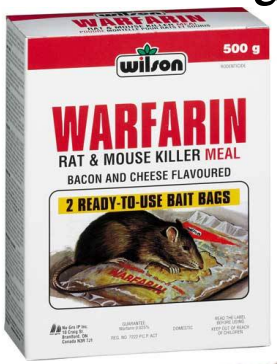
Ex : thylacines and dogs, water chevrotain

Orgogozo 2014 Interface Focus

From random processes can emerge predictability

Many unpredictable processes at a low level

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



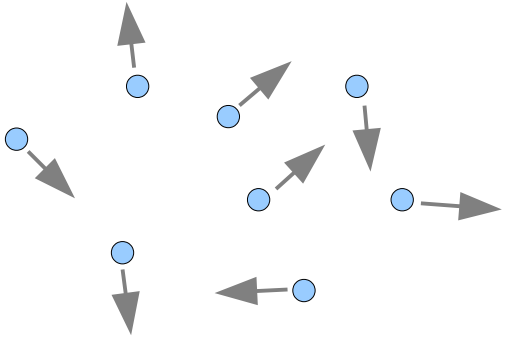
Predictable Evolution at the genetic level

coding mutation in *VKORC1* gene

Orgogozo 2014 Interface Focus

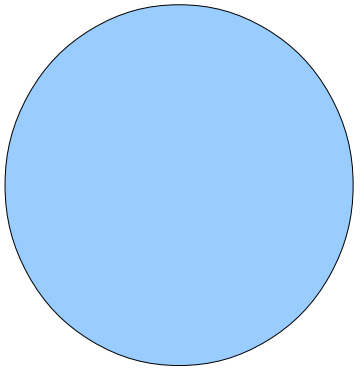
From random processes can emerge predictability

Microscopic world

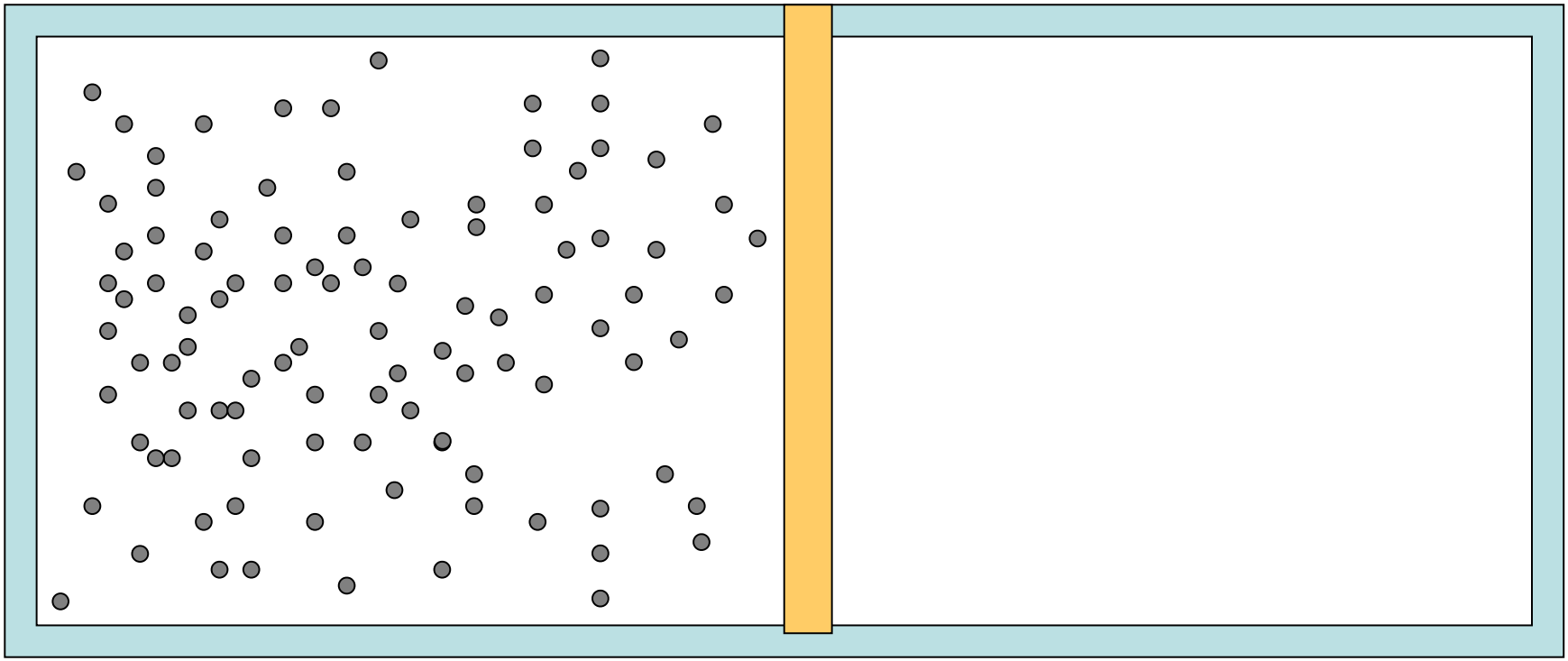


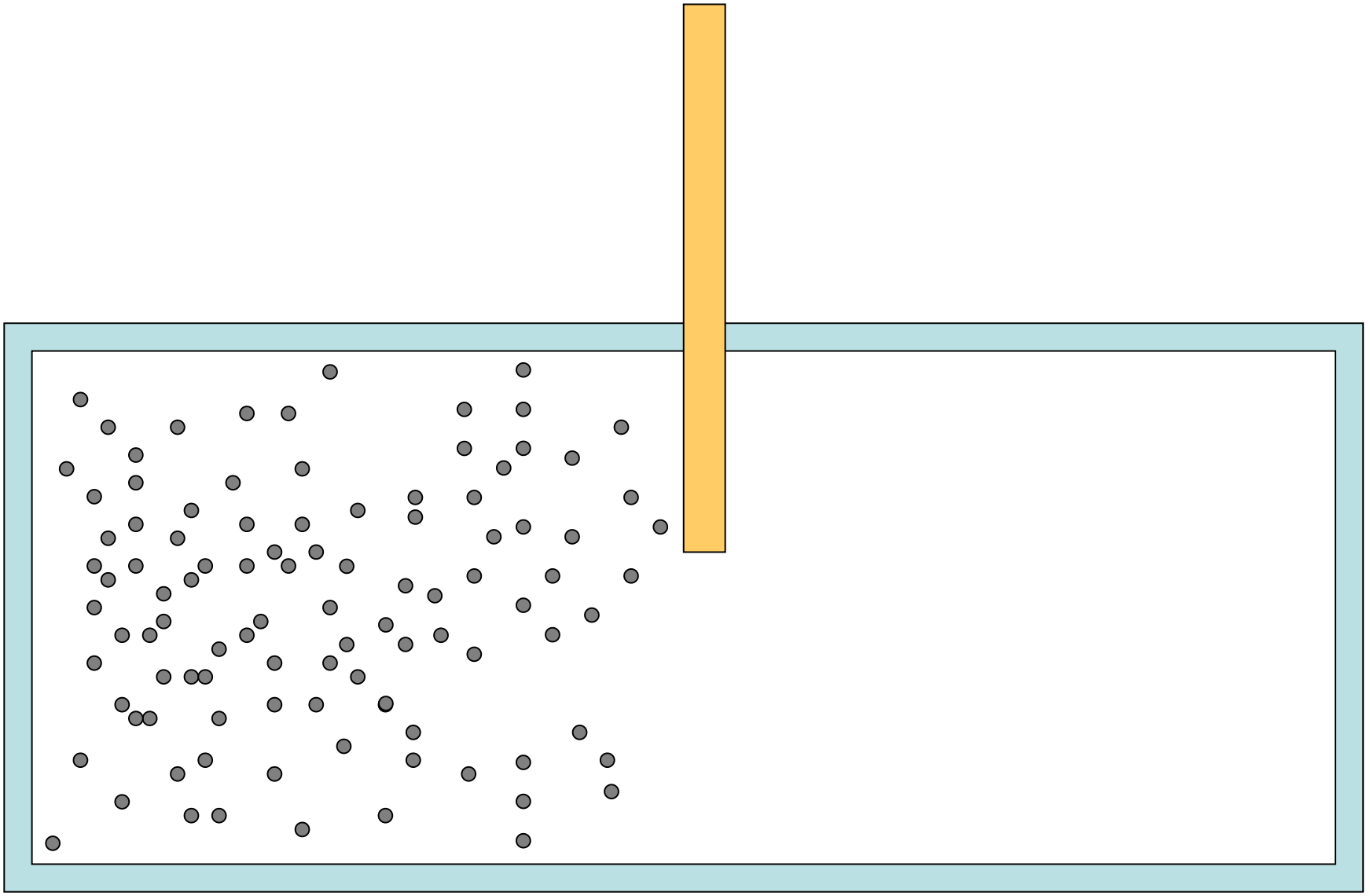
Position, mass, velocity of each particle

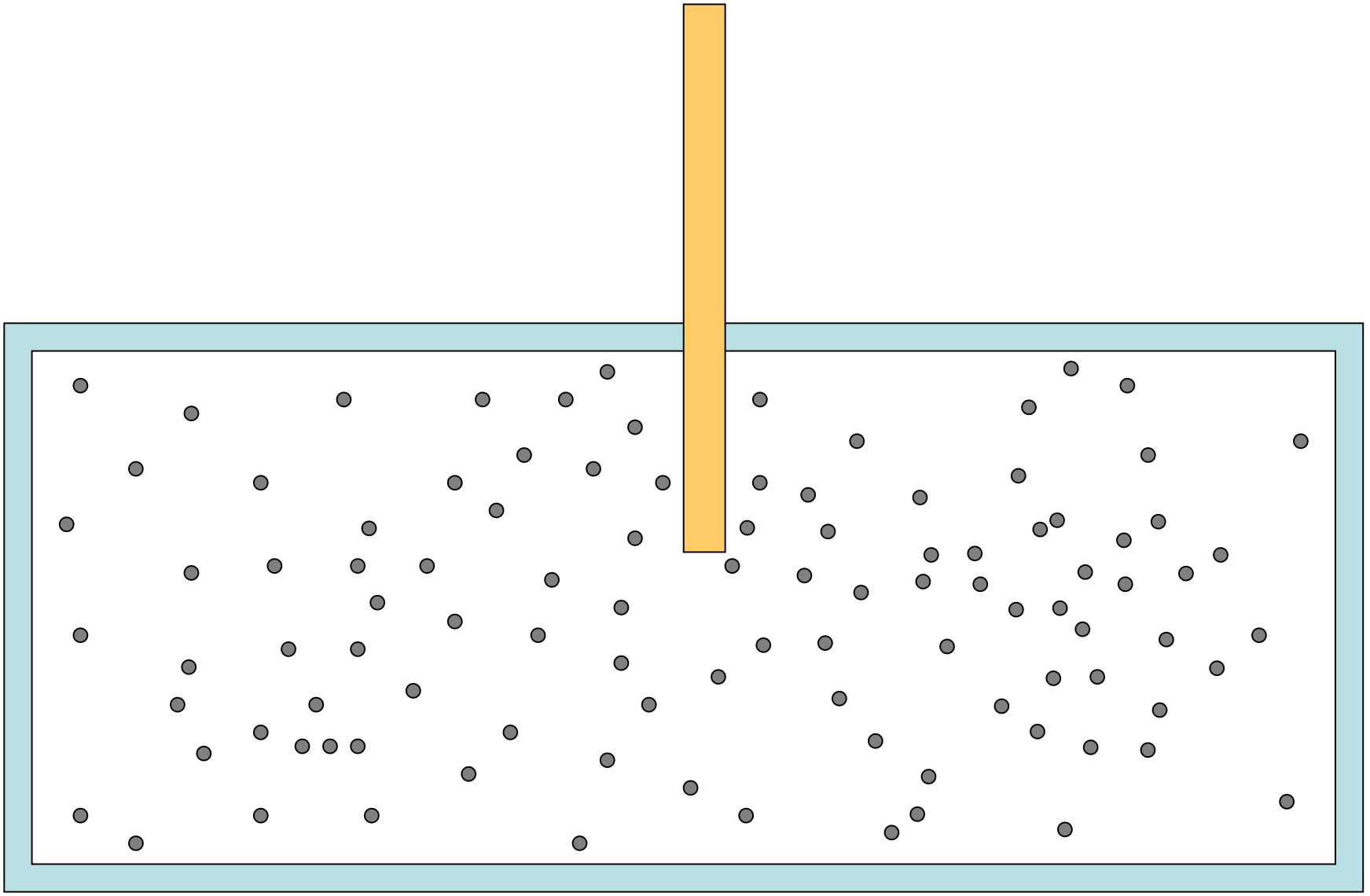
Macroscopic world



Pressure, Volume, Temperature,
Number of moles





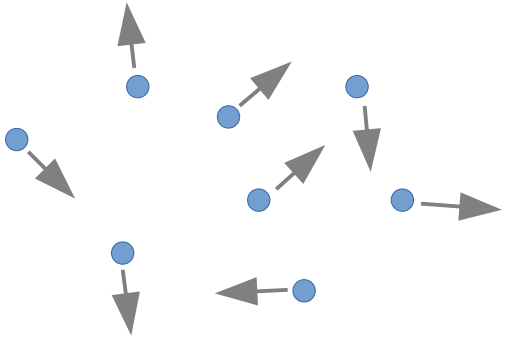


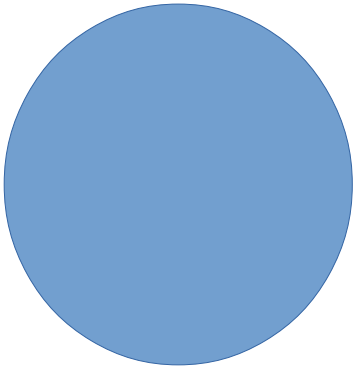
After a few seconds

Predictable evolution at a higher level?

Many unpredictable processes at a low level → Predictable Evolution at higher level?

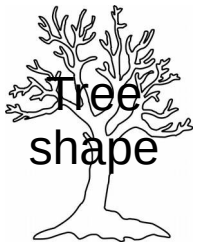
Microscopic world → Macroscopic world

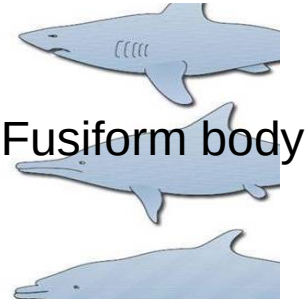




Position, mass, velocity of each particle

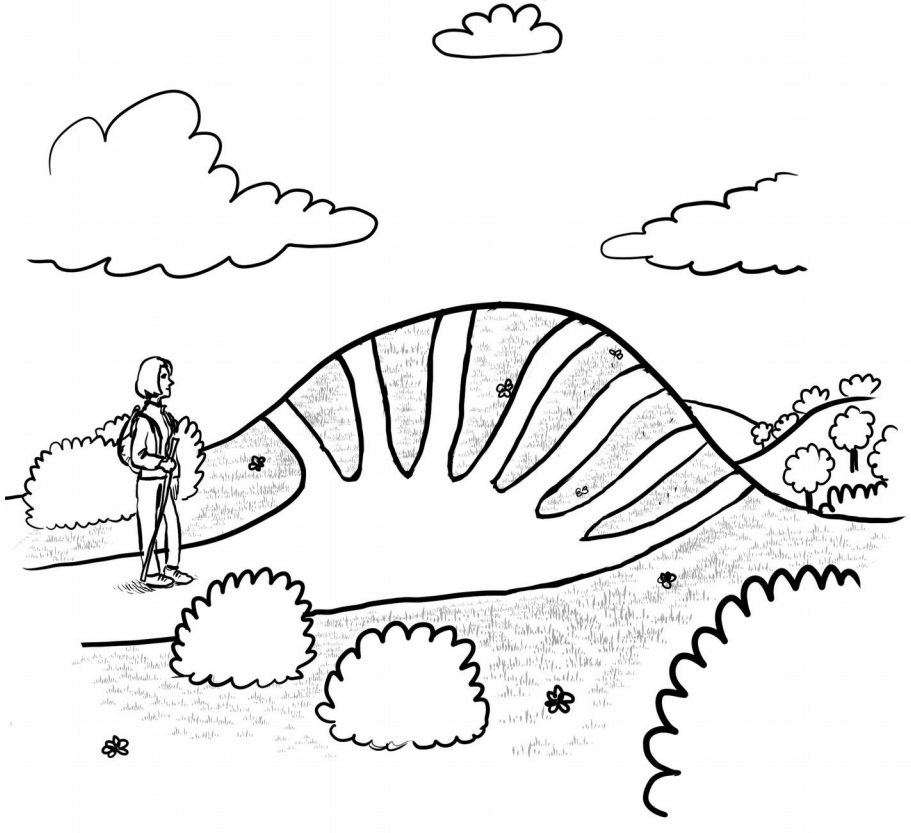
Pressure, Volume, Temperature, Number of moles

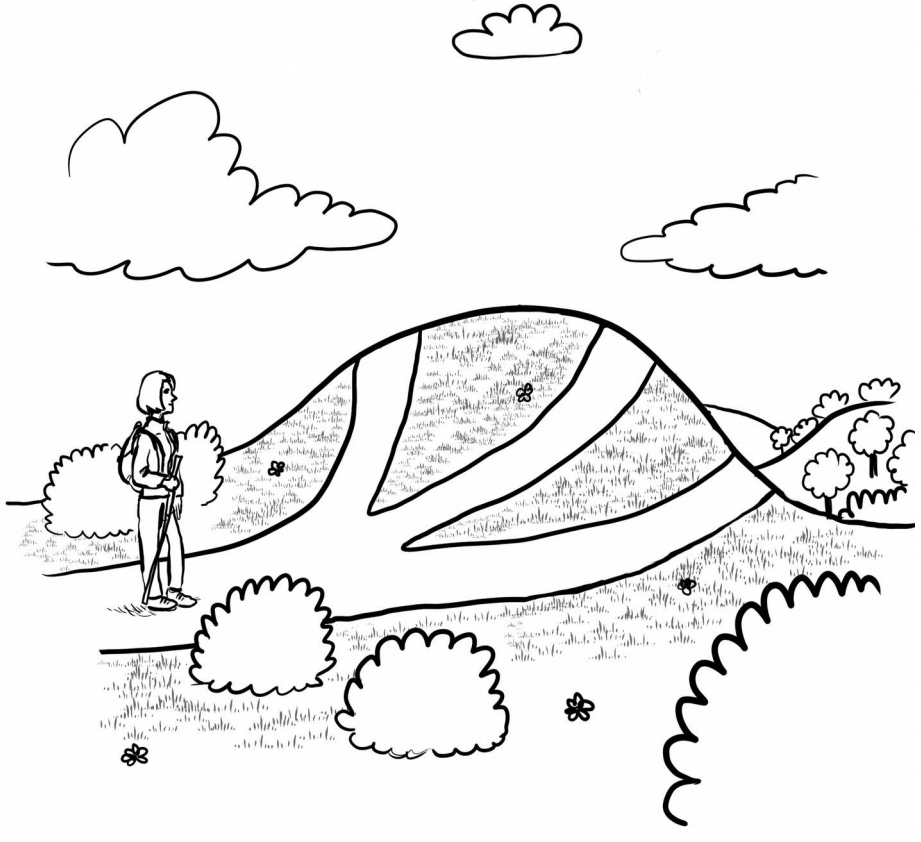




Key: finding the relevant concepts for predictive evolution

A small number of genetic solutions for a given phenotypic change



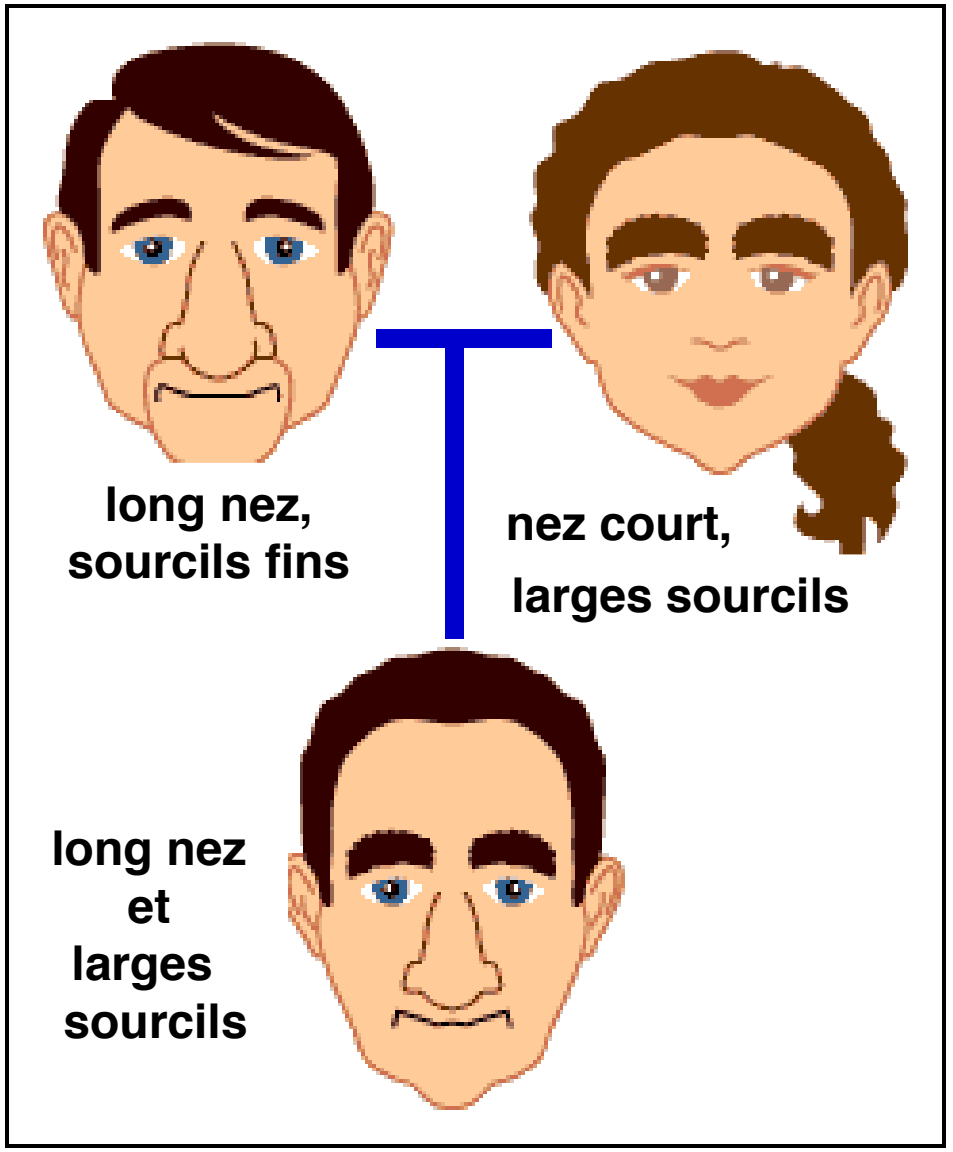


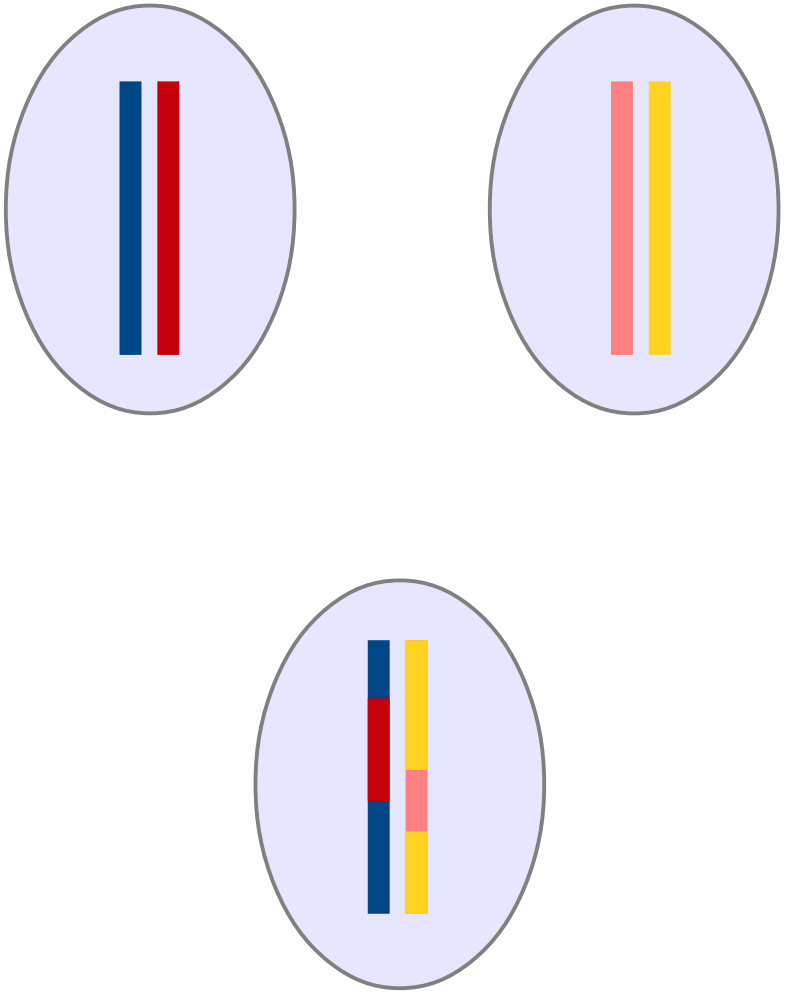
Conclusion: on ne peut pas prévoir dans quelle direction se fera l'évolution, il y a trop d'événements impossibles à prévoir qui interviennent.

Michael DuBow

We are mosaics

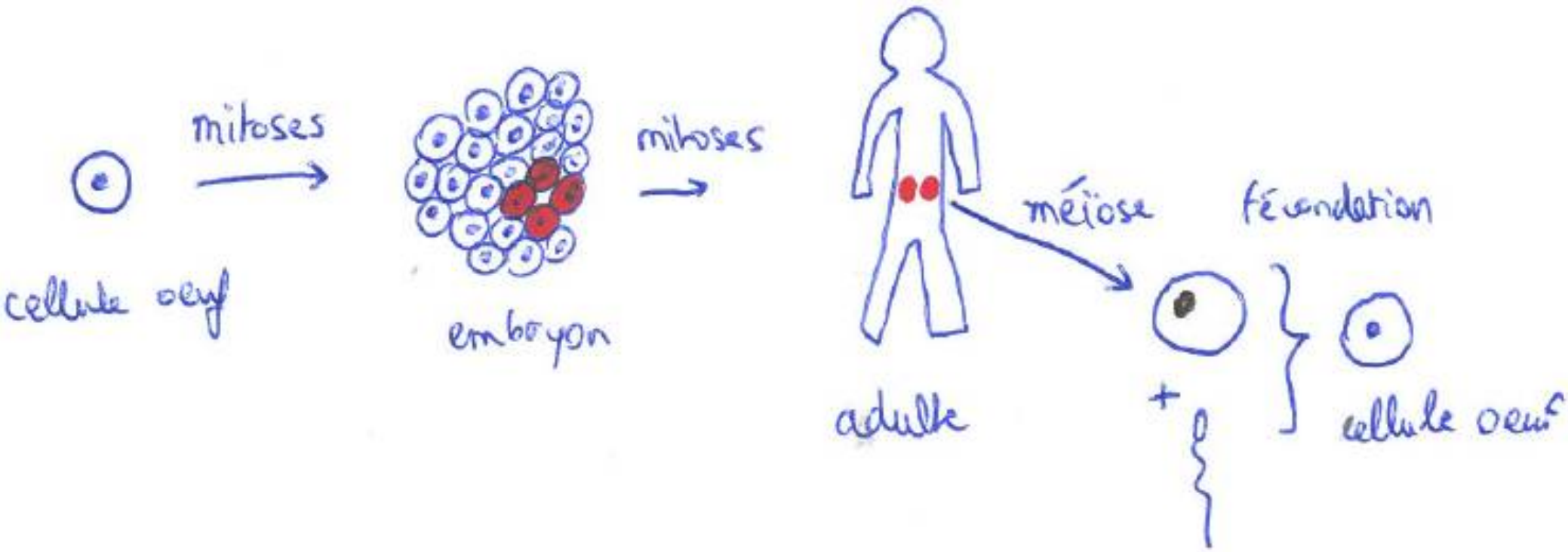
Assortiment des chromosomes du père et de la mère





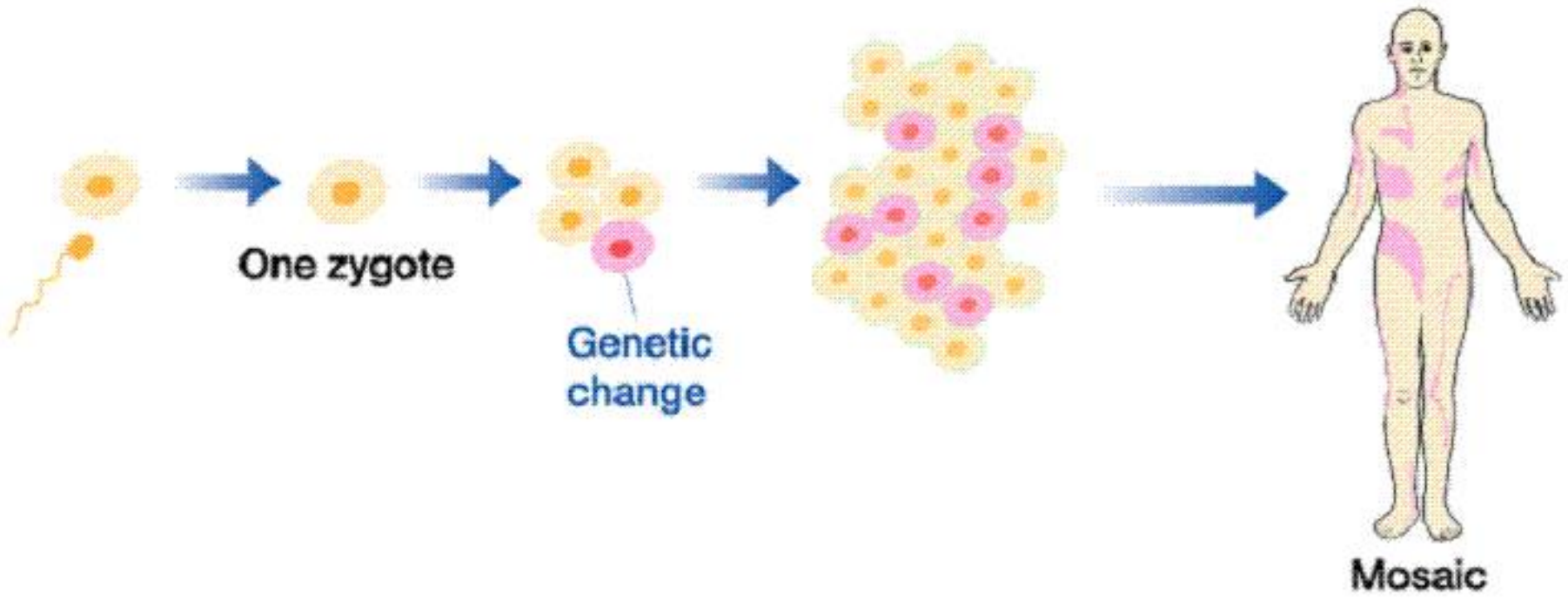
http://evolution.berkeley.edu

Le génome au cours des générations

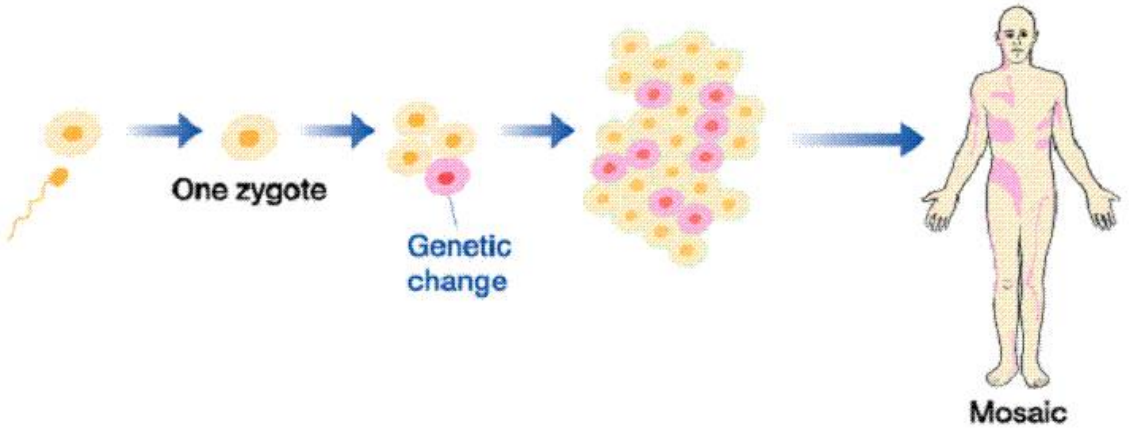


Lignée germinale

Lignée somatique

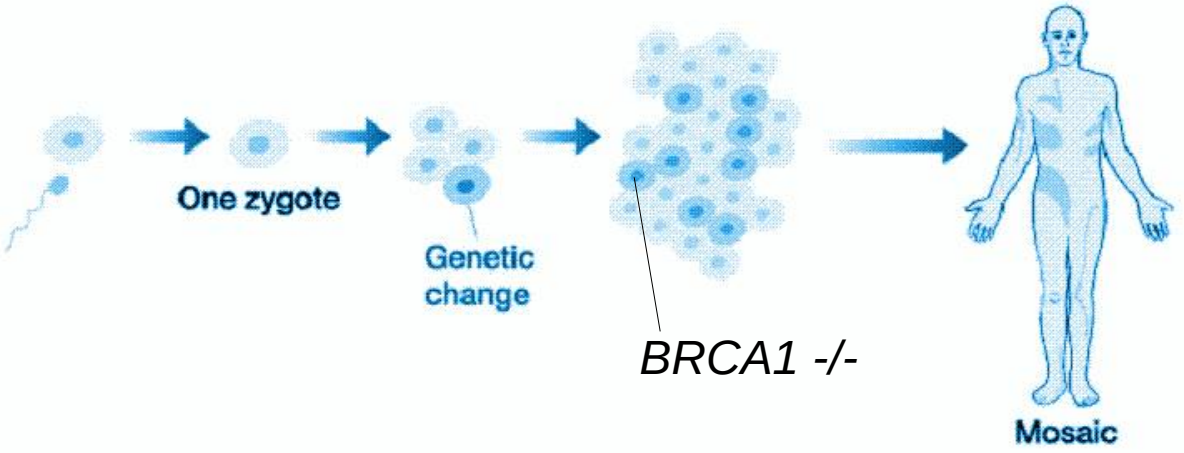

One zygote
Genetic change
Mosaic

femme
BRCA1 +/+



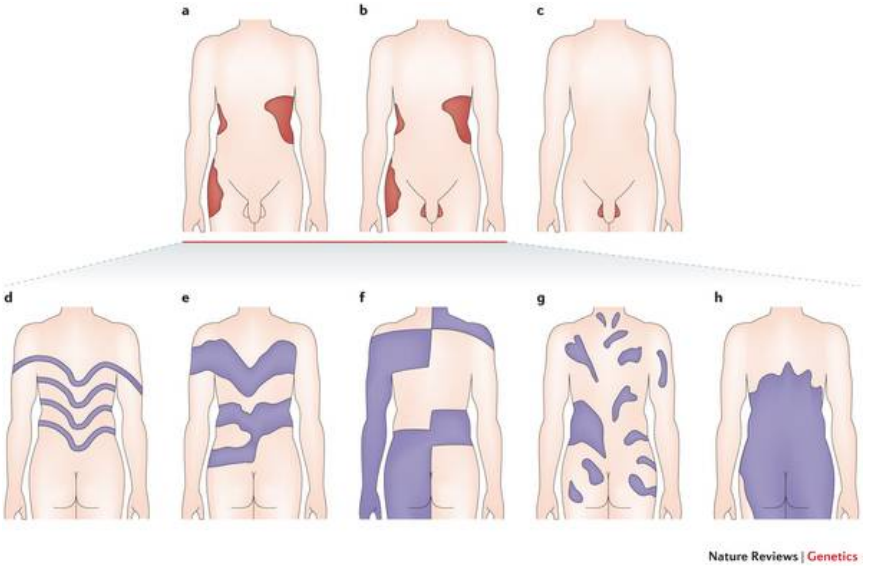
10% de chance de développer un cancer du sein durant sa vie

femme
BRCA1 +/-



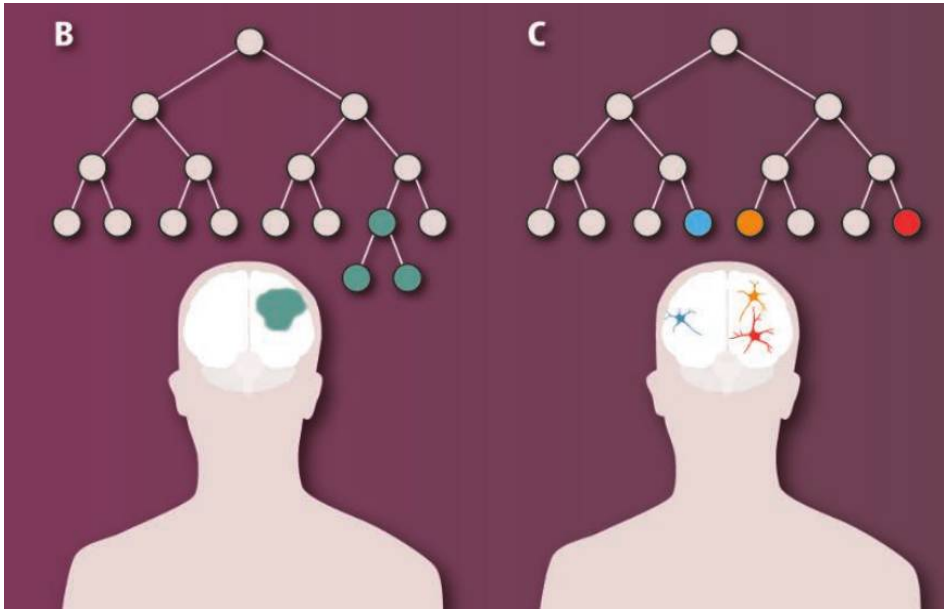
45% de chance de développer un cancer du sein avant 70 ans
Les cellules de ce cancer sont *BRCA1 -/-*

Mosaïque somatique

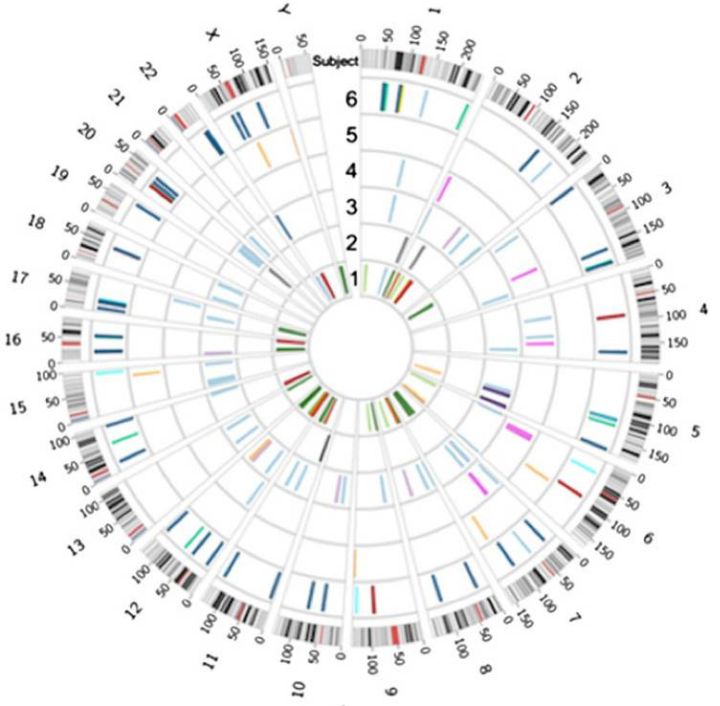


73 somatic CNVs in 11 tissues of six persons

Dans l'hippocampe et le noyau caudé de 3 individus :
7,743 somatic L1 insertions, 13,692 somatic Alu insertions and 1,350 SVA insertions



Baillie 2011 Nature

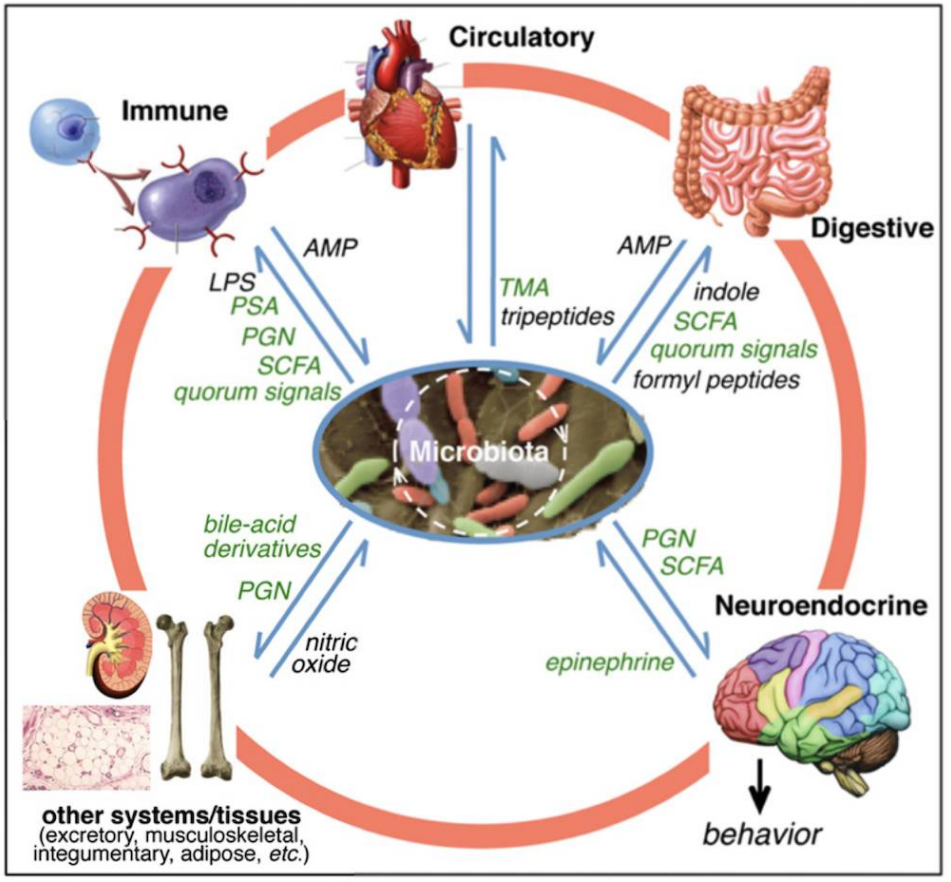


O'Huallachain 2012 PNAS

Holobiont perspective: Physiology

Slide from S. Gilbert

Symbionts are intimately involved in regulating our normal physiological functions

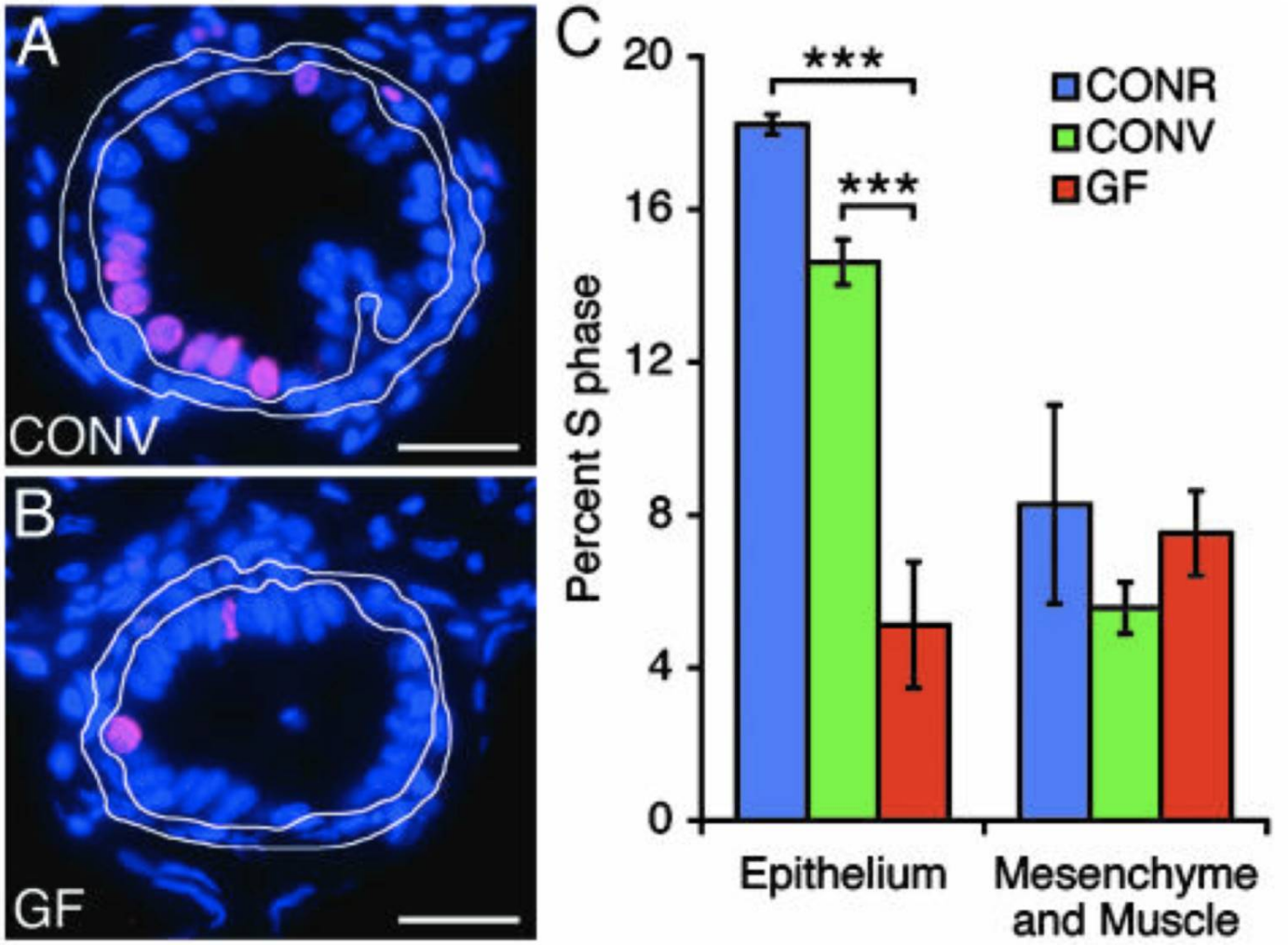


McFall-Ngai et al . 2013. PNAS 110: 3229 -3236

- Approximately 70% of human resting body temperature is the result of gut microbe metabolism (Henry 2005; Rosenberg and Zilber-Rosenberg, 2016).

Microbial Symbionts Induce Epithelial Cell Formation in Zebrafish

Slide from S. Gilbert



A, B. 6d conventional (A) and germ-free (B) intestine stained for cells (blue) and dividing cells (magenta). Mesenchyme and muscles outlined in white
Rawls et al. 2004. *PNAS* 101: 4596 – 4601. (Symbionts work through beta-catenin stabilization)

INTERFACE FOCUS

Replaying the tape of life in the twenty-first century

2015

rsfs.royalsocietypublishing.org

Virginie Orgogozo

The "Mendelian Gene" and the "Molecular Gene": Two Relevant Concepts of Genetic Units

2016

V. Orgogozo*,[1], A.E. Peluffo*, B. Morizot[†]

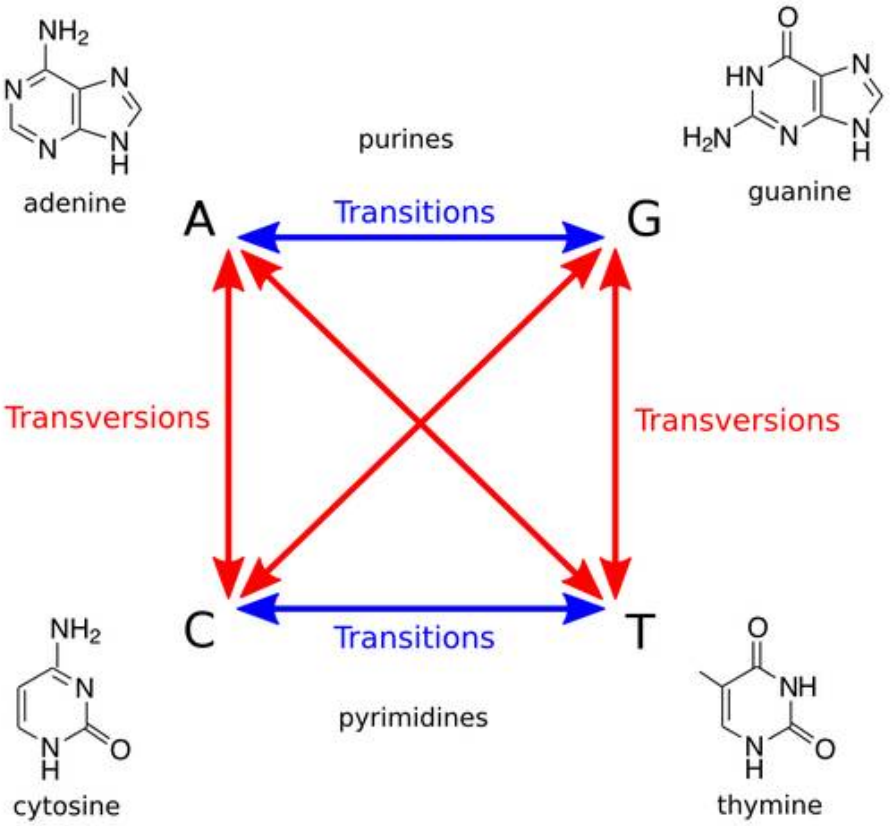
A SYMBIOTIC VIEW OF LIFE: WE HAVE NEVER BEEN INDIVIDUALS

2012

SCOTT F. GILBERT , J. SAPP and A. TAUBER

À vous de jouer !

	U	C	A	G	
U	UUU, UUC } Phe - F UUA, UUG } Leu - L	UCU, UCC, UCA, UCG } Ser - S	UAU, UAC } Tyr - Y UAA stop UAG stop	UGU, UGC } Cys - C UGA stop UGG Trp - W	U C A G
C	CUU, CUC, CUA, CUG } Leu - L	CCU, CCC, CCA, CCG } Pro - P	CAU, CAC } His - H CAA, CAG } Gln - Q	CGU, CGC, CGA, CGG } Arg - R	U C A G
A	AUU, AUC, AUA } Ile - I AUG Met - M start	ACU, ACC, ACA, ACG } Thr - T	AAU, AAC } Asn - N AAA, AAG } Lys - K	AGU, AGC } Ser - S AGA, AGG } Arg - R	U C A G
G	GUU, GUC, GUA, GUG } Val - V	GCU, GCC, GCA, GCG } Ala - A	GAU, GAC } Asp - D GAA, GAG } Glu - E	GGU, GGC, GGA, GGG } Gly - G	U C A G



IUPAC nucleotide code

Symbol	Bases	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T (or U)	T (or U)	Thymine (or Uracil)
W	**A** or **T**	Weak
S	**C** or **G**	Strong
M	**A** or **C**	aMino
K	**G** or **T**	Keto
R	**A** or **G**	puRine
Y	**C** or **T**	pYrimidine
B	C or G or T	not A (B comes after A)
D	A or G or T	not C (D comes after C)
H	A or C or T	not G (H comes after G)
V	A or C or G	not T (V comes after T and U)
N	any base	any Nucleotide (not a gap)

Specialized loci in the genome

Proteins that interact with external molecules

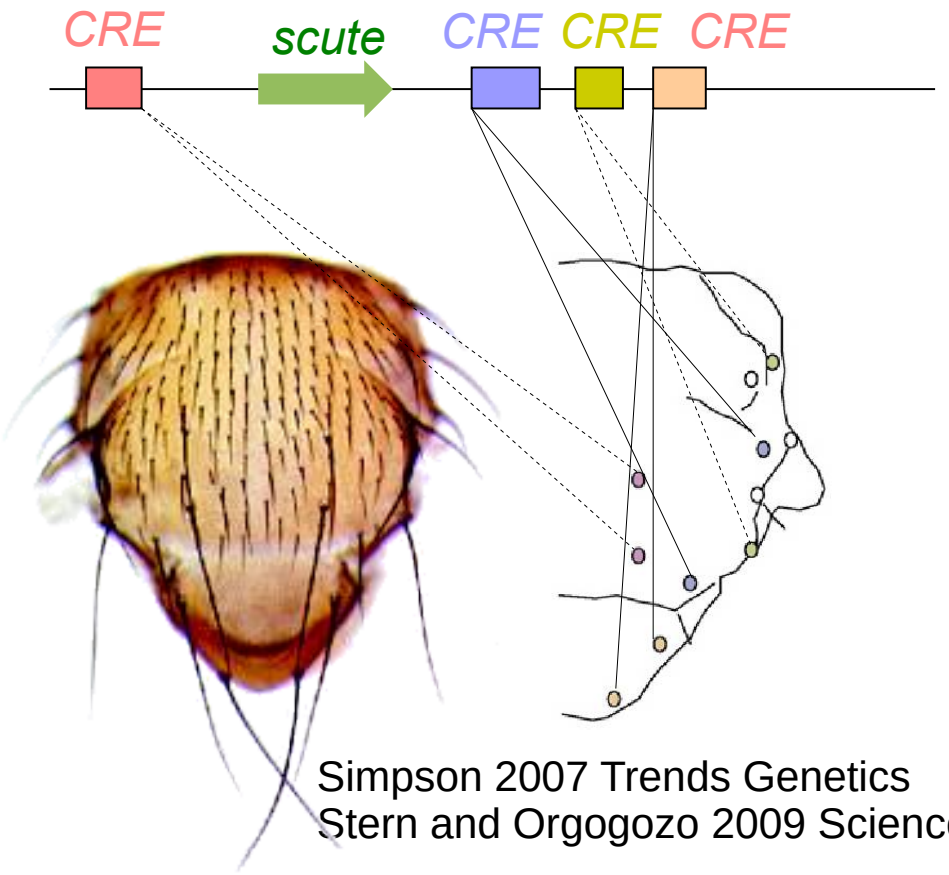
oxygen, photons, insecticide, cholesterol...

Specialized loci in the genome

Proteins that interact with external molecules

oxygen, photons, insecticide, cholesterol...

Cis-regulatory elements of “developmental switch genes”



Simpson 2007 Trends Genetics
Stern and Orgogozo 2009 Science