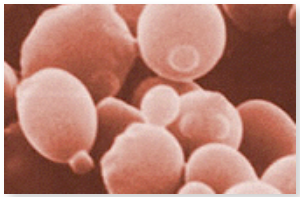
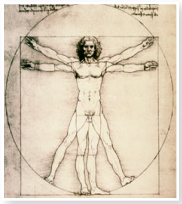
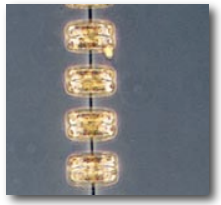
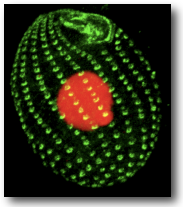


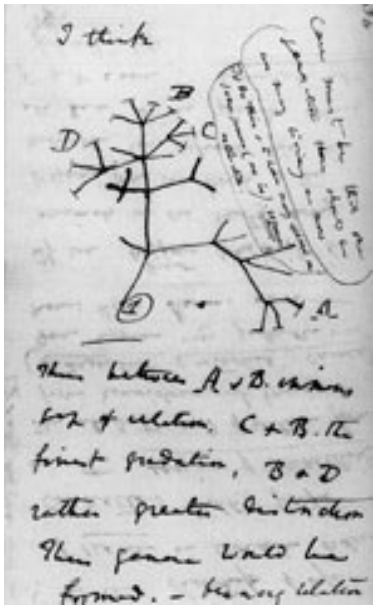




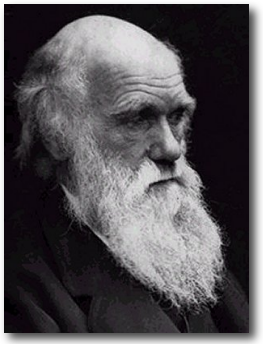


Evolution des génomes eucaryotes:

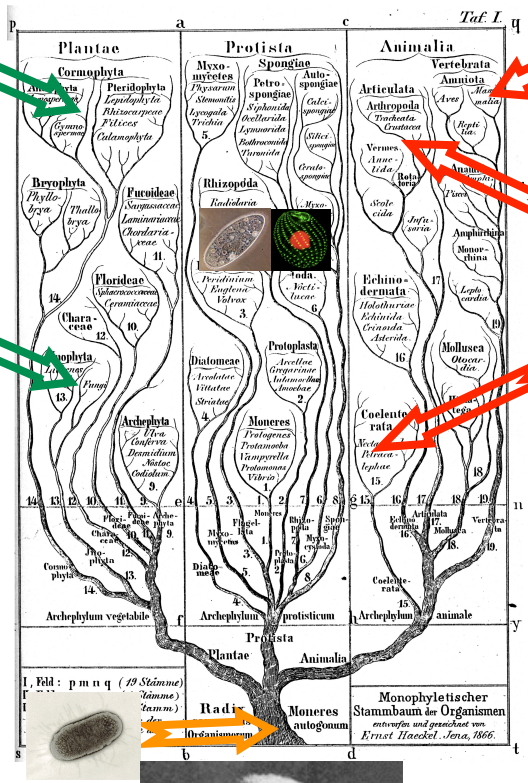
exemple des levures et autres unicellulaires

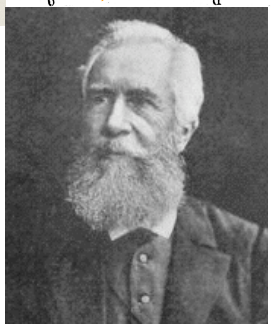
Cours BIM-PHYG
4 décembre 2014
Bernard Dujon





Charles Darwin
Reproduction avec
changement **graduel**



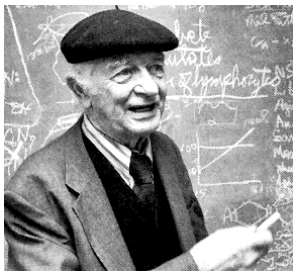


Ernst Haeckel, 1866
Phylogénie / ontogénie

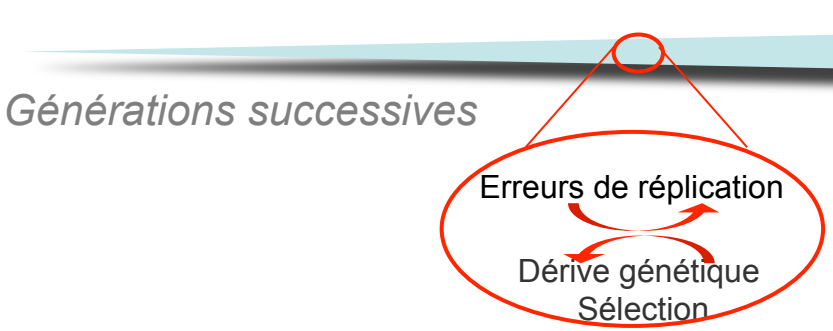
1- Les organismes vivants n'évoluent pas, ils se reproduisent. Ce sont les populations qui évoluent.
2- La génomique moderne est en train de bouleverser nos conception sur l'évolution.

1- Les séquences des protéines homologues divergent

```
Q96J94/555-847   IVVCLLSS.NRKDK.YDAIKKYLCTDCPTPSQCVVARTLGKQ.....QT.......VMAIATKIALQMNCKMGGE...LWRVD
Q23415/40-350    KFAFVITD.DSITH.LHKKYKALEQKSMMVIQDMKISKANSV.....VK.....DGKRLTLENVINKTNMKLGGLN..YTVSD
O74957/500-799   YLFFILDK.NSPEP.YGSIKRVCNTMLGVPSQCAISKHILQS..............KPQYCANLGMKINVKVGGINC.SLIPK
O62275/594-924   TFVFIITD.DSITT.LHQRYKMIEKDTKMIVQDMKLSKALSV..........INAGKRLTLENVINKTNVKLGGSN..YVFVD
Q58717/426-699   CFALIIGKEKYKDNDYYEILKKQLFDLKIISQNILWENWRKD.....DK........GYMTNNLLIQIMGKLGIK...YFILD
Q9VKM1/538-829   LILCLVPN.DNAER.YSSIKKRGYVDRAVPTQVVTLKTTKNR.....SL........MSIATKIAIQLNCKLGYT...PWMIE
Q21691/673-1001  TIVFGIIA.EKRPD.MHDILKYFEEKLGQQTIQISSETADKF.....MR....DHGGKQTIDNVIRKLNPKCGGTNFLIDVPE
O16386/548-847   QLIMFITK..SMNN.YHTEIKCLEQEFDLLTQDIRFETAVKL.....AQ.......QQNTRKNIIYKTNMKLGGLN..YELRS
Q9XGW1/625-946   LLLAILPD.NNGSL.YGDLKRICETELGLISQCCLTKHVFKI..............SKQYLANVSLKINVKMGGRNT.VLVDA
Q17567/397-708   MLVVMLAD.DNKTR.YDSLKKYLCVECPIPNQCVNLRTLAGK.....SKDGGENKNLGSIVLKIVLQMICKTGGA...LWKVN
Q19645/674-996   PFVLFISD..DVPN.IHECLKFEERMSDIPTQHVLLKNVKKMRDNIEKKSQGGRRAYDLTLDNIVMKANIKCGGLN..YT.AD
Q09249/650-977   DILVGIAR.EKKPD.VHDILKYFEESIGLQTIQLCQQTVDKM.....MG....GQGGRQTIDNVMRKFNLKCGGTNFFVEIPN
Q21495/52-336    GIVLPTPRIFFRDG..........QETSLNNQSFRNPT.....................DFAQTGFFVDAKQQLGGLN..YVVNS
O28951/110-406   GIMLVLPE.YNTPL.YYKLKSYLINS..IPSQFMRYDILSNR.....NL........TFYVDNLLVQFVSKLGGK...PWILN
O16720/566-867   LIVVVLPG..KTPI.YAEVKRVGDTVLGIATQCVQAKNAIRT..............TPQTLSNLCLKMNVKLGGVNS.ILLPN
O76922/555-852   IVMVVMRS.PNEEK.YSCIKKRTCVDRPVPSQVVTLKVIAPR.....QQ...KPTGLMSIATKVVIQMNAKLMGA...PWQVV
Q9ZVD5/577-885   FILCVLPDKKNSDL.YGPWKKKNLTEFGIVTQCMAPTRQPND.................QYLTNLLLKINAKLGGLNS.MLSVE
O48771/542-860   FILCILPERKTSDI.YGPWKKICLTEEGIHTQCICPIKI.................SDQYLTNVLLKINSKLGGINS.LLGIE
O67434/419-694   LVIVFLEEYPKVDP.YKSFLLYDFVKRELLKKMIPSQVILNR.....TL...KNENLKFVLLNVAEQVLAKTGNIP..YKLKE
P34681/660-966   CIIVVLQS.KNSDI.YMTVKEQSDIVHGIMSQCVLMKNVSRP..............TPATCANIVLKLNMKMGGIN..SRIVA
O02095/574-878   QLLFFVVK..SRYN.YHQQIKALEQKYDVLTQEIRAETAEKV.....FR.......QPQTRLNIINKTNMKLGGLN..YAIGS
```



Linus Pauling

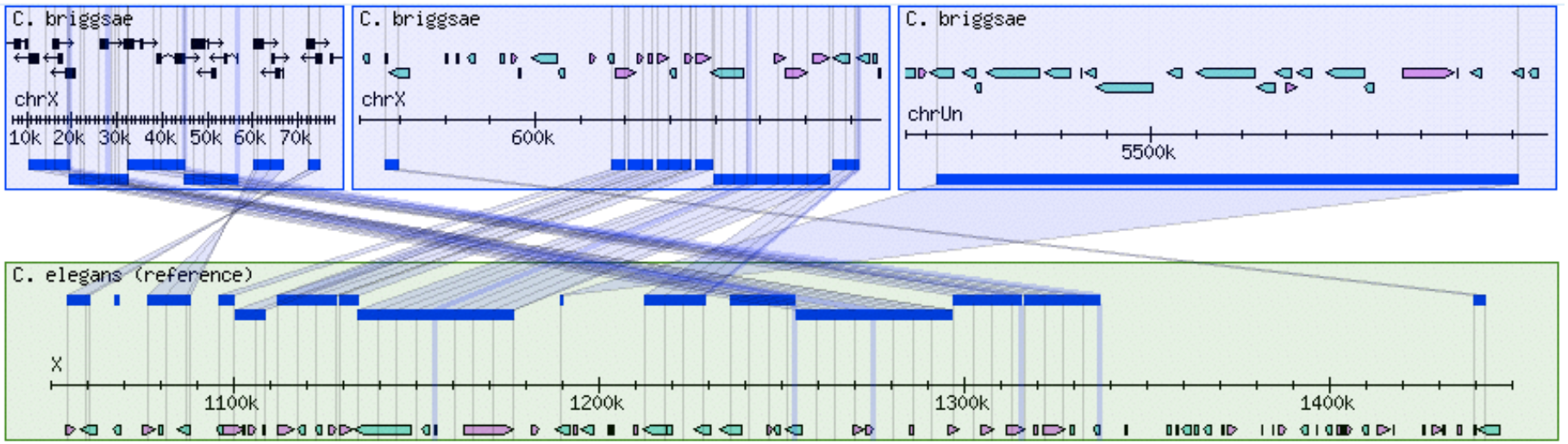


Horloges moléculaires

Phylogénies moléculaires

au sein d'une espèce: polymorphisme
entre espèces: divergence

2- Les génomes se réarrangent



Thomas H. Morgan

Générations successives

duplications

invesrsions, translocations

délétions

Perte de synténie et redondance

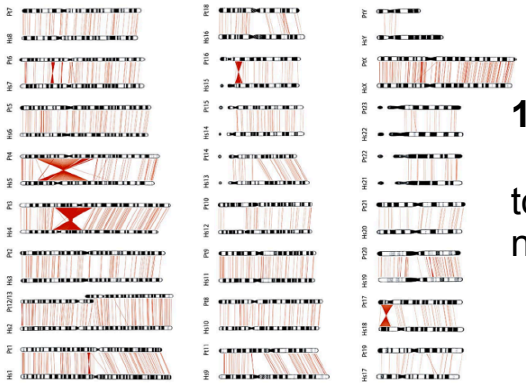
Variations structurelles

Evolution des cartes génétiques

au sein d'une espèce: polymorphisme structurel, CNV
entre espèces: perte de synténie

Point mutations and chromosomal alterations, two bases of genome evolution

2- Grandes altérations chromosomiques
(changements d'ordre et/ou nombre de copies des gènes)



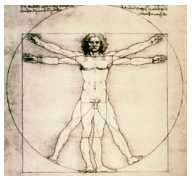
1576 inversions

total de **154** millions de nucléotides impliqués
(~ 5 %)

2 translocations

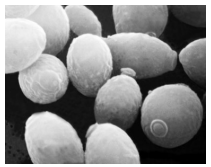
total de **680 000** nucléotides impliqués **(~ 6 %)**

1- Mutations « ponctuelles » (changement de séquence)

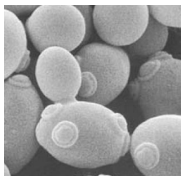


35 millions de SNPs
5 millions indels

~ 40 millions de nucléotides de divergence de séquence
(~ 1.3 %)



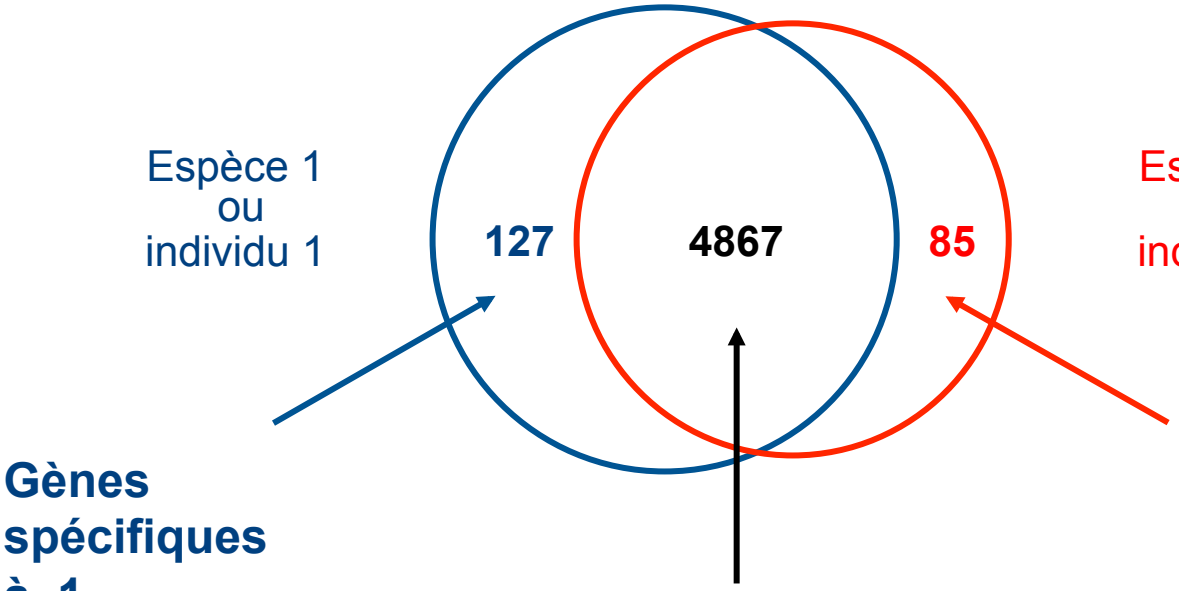
Saccharomyces cerevisiae

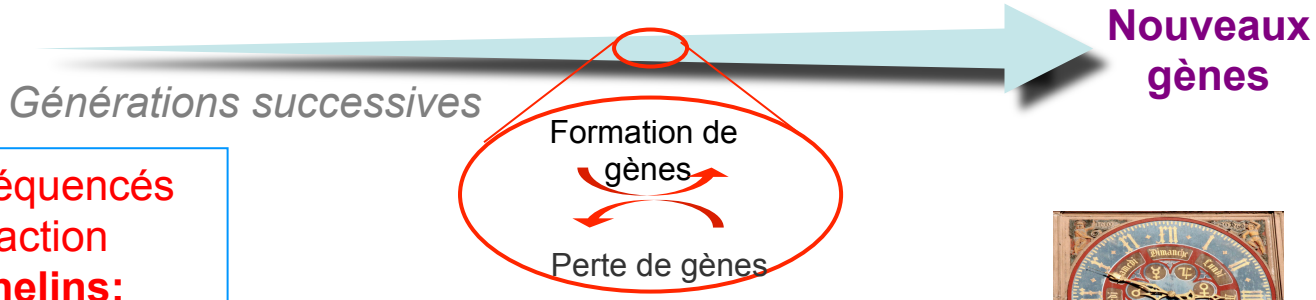


Saccharomyces uvarum

~ 4 millions de nucléotides de divergence de séquence
(~ 33 %)

3- Des gènes apparaissent et disparaissent

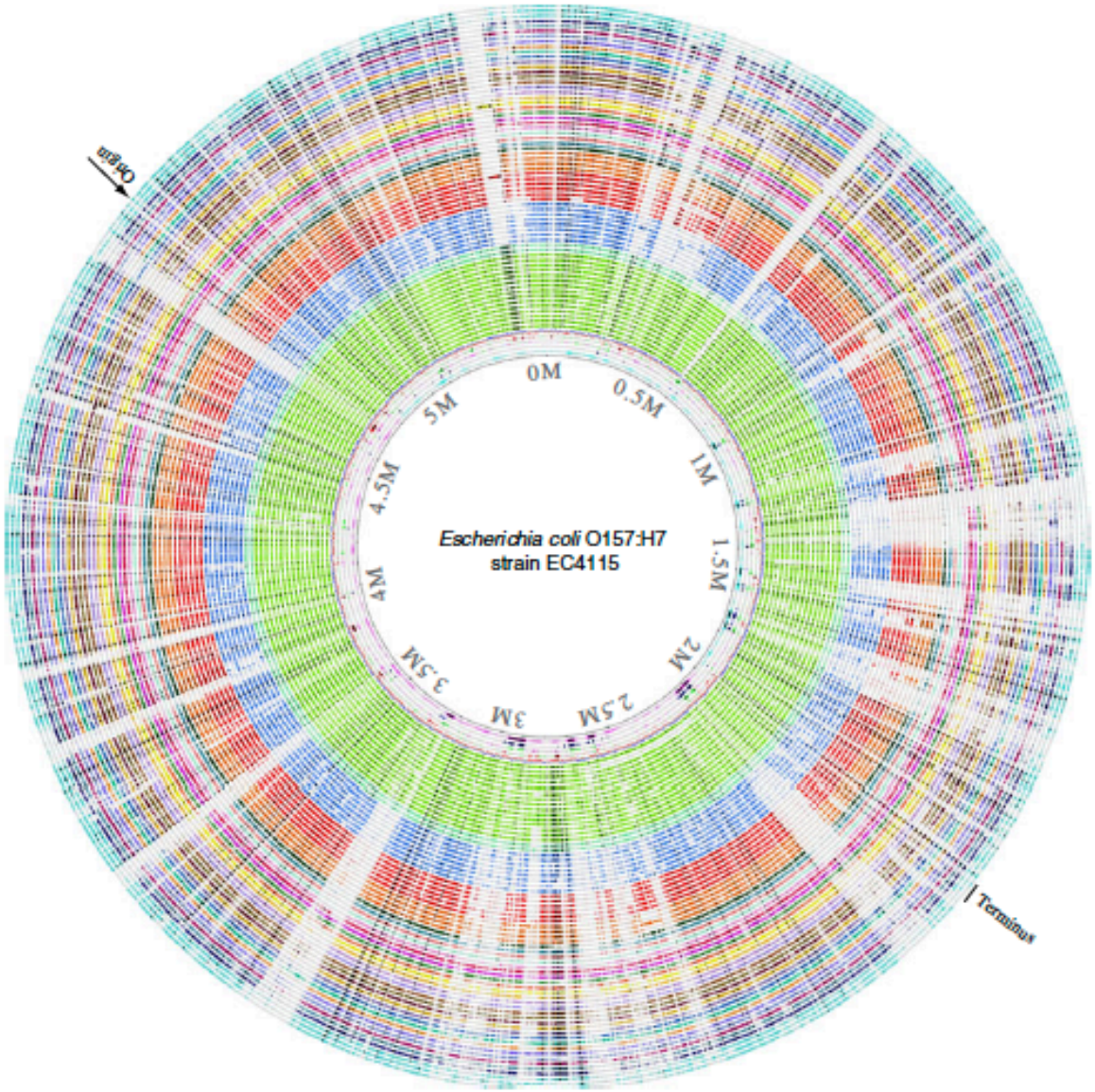




Dans tous les génomes séquencés à ce jour, on trouve une fraction importante de **gènes orphelins:**
- transferts horizontaux
- formation *de novo*

au sein d'une espèce: core-genome, pan-genome

Core-genome and *pan-genome*



core genome: essential cell functions

represents ~ **20 %** of each genome

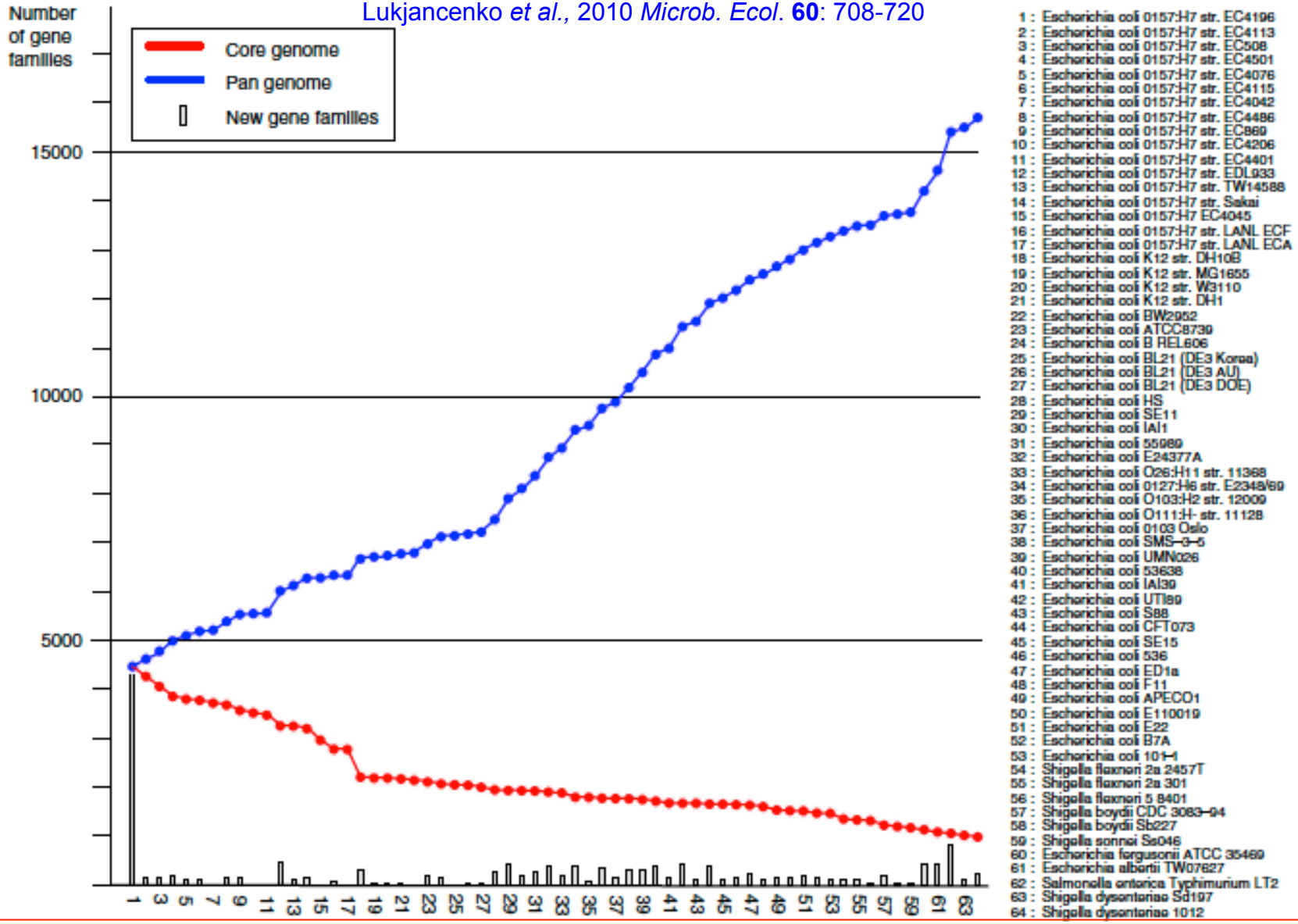
accessory genome: *fitness* and adaptation to different environments

Represents **80 %** of each genome

In *Bacteria*, accessiory genes tend to co-localize on the genome, forming **genomic islands**.

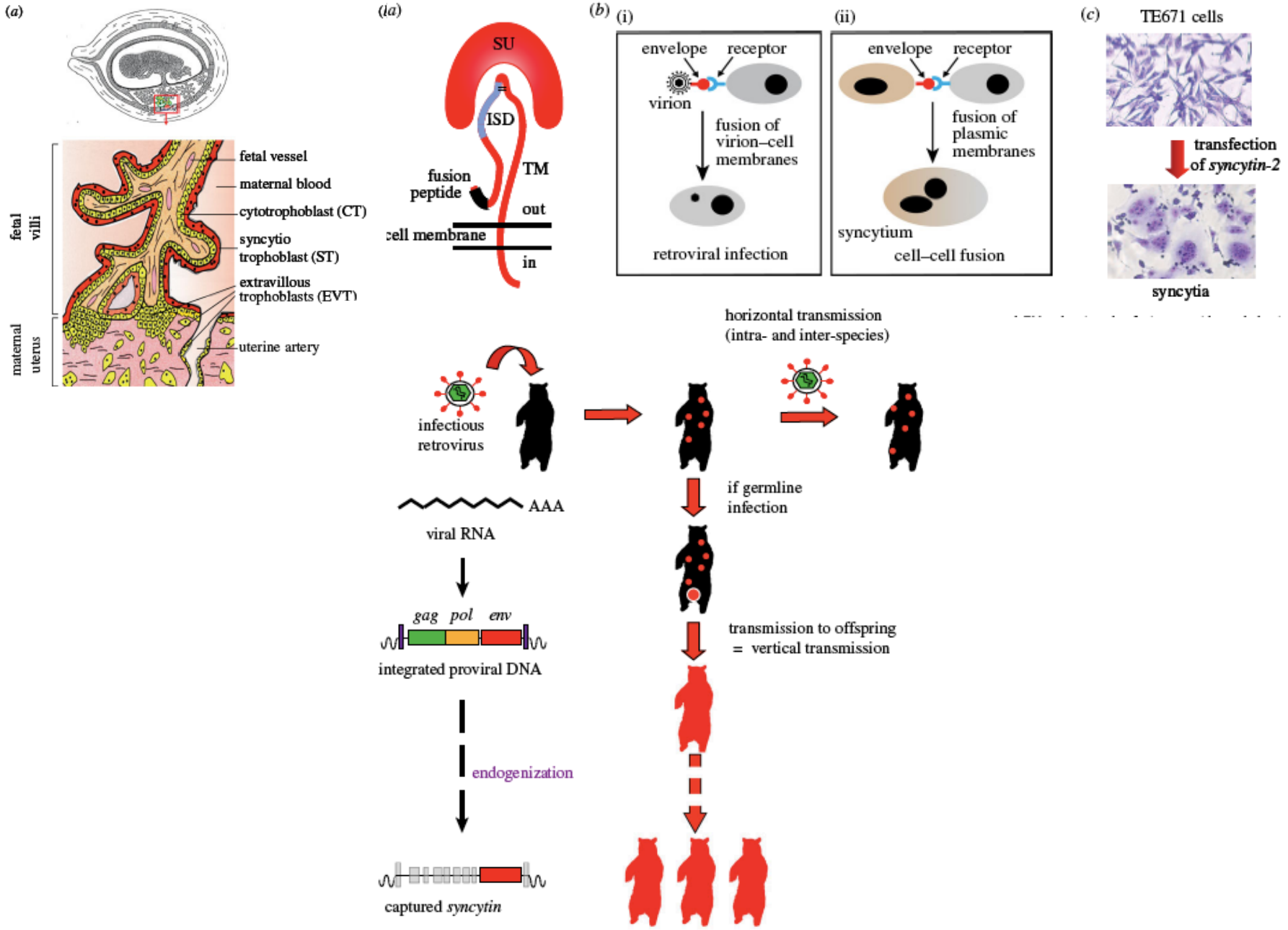
Core-genome and *pan-genome*

Lukjancenko *et al.*, 2010 *Microb. Ecol.* **60**: 708-720

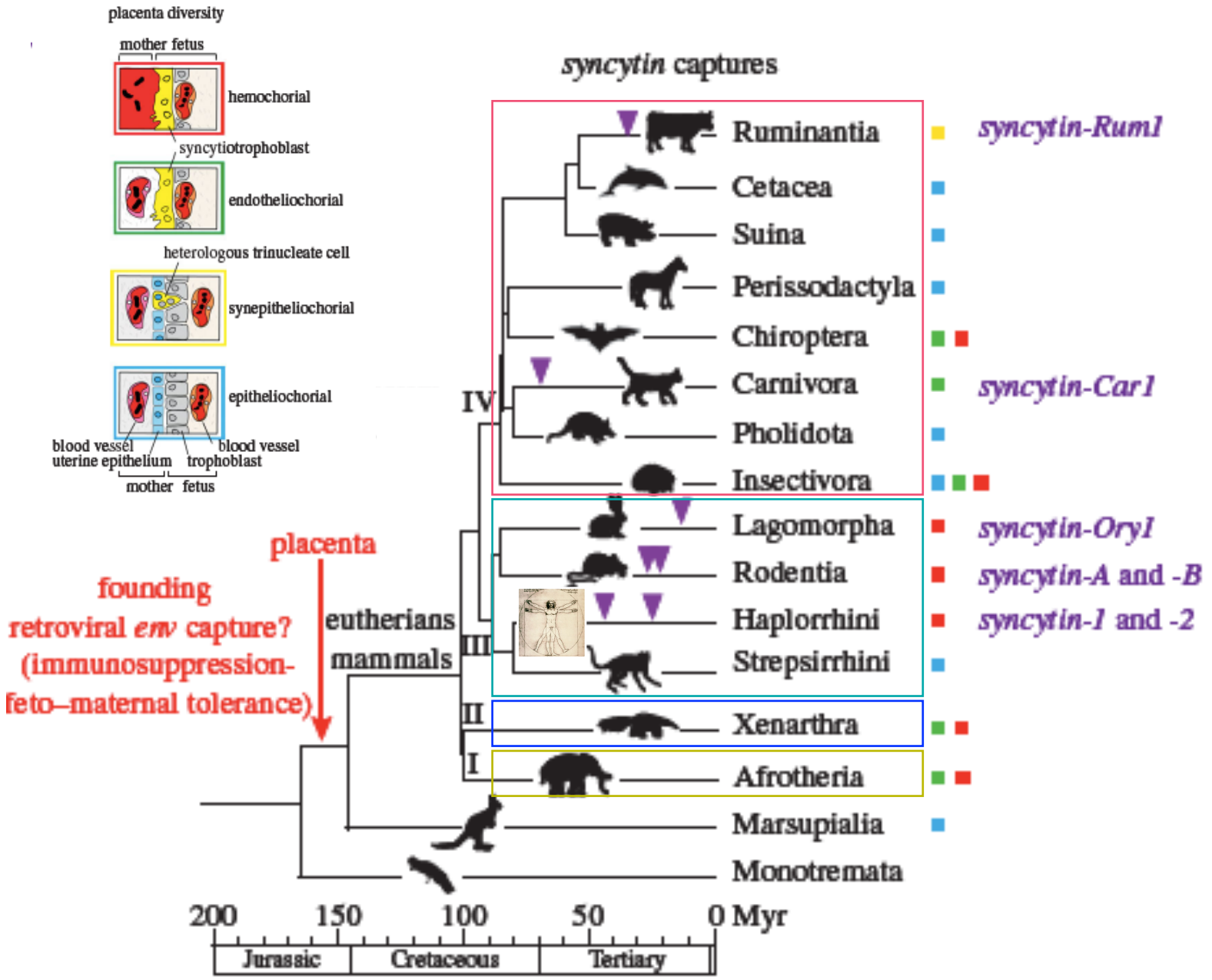


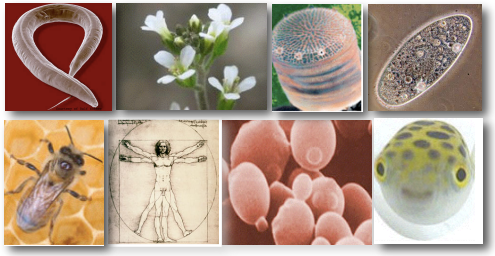
With 61 *E. coli* genomes sequenced, the *pan-genome* comprises 15 574 gene families of whic only 993 (6 %) are present in all isolates (*core-genome*).

Important example of foreign gene acquisition: syncytines



A founding event followed by multiple similar events

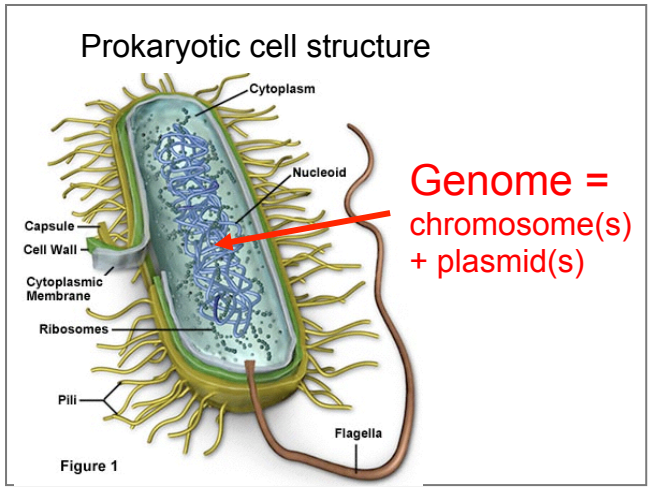




- Le monde des eucaryotes, vu par la génomique comparative

  Les mécanismes moléculaires de l'évolution des génomes eucaryotes

  La formation *de novo* de gènes, *frameshifts* et ARNs



Genome =
chromosome(s)
+ plasmid(s)

Eukaryotic cells

Genome =
nuclear chromosomes (+ plasmids)
+ mitochondrial genome (+ plasmids)
+ chloroplastic genome

Multiple membranes: cell compartments

Mitochondrial genome →

Chloroplastic genome →

Nuclear genome →

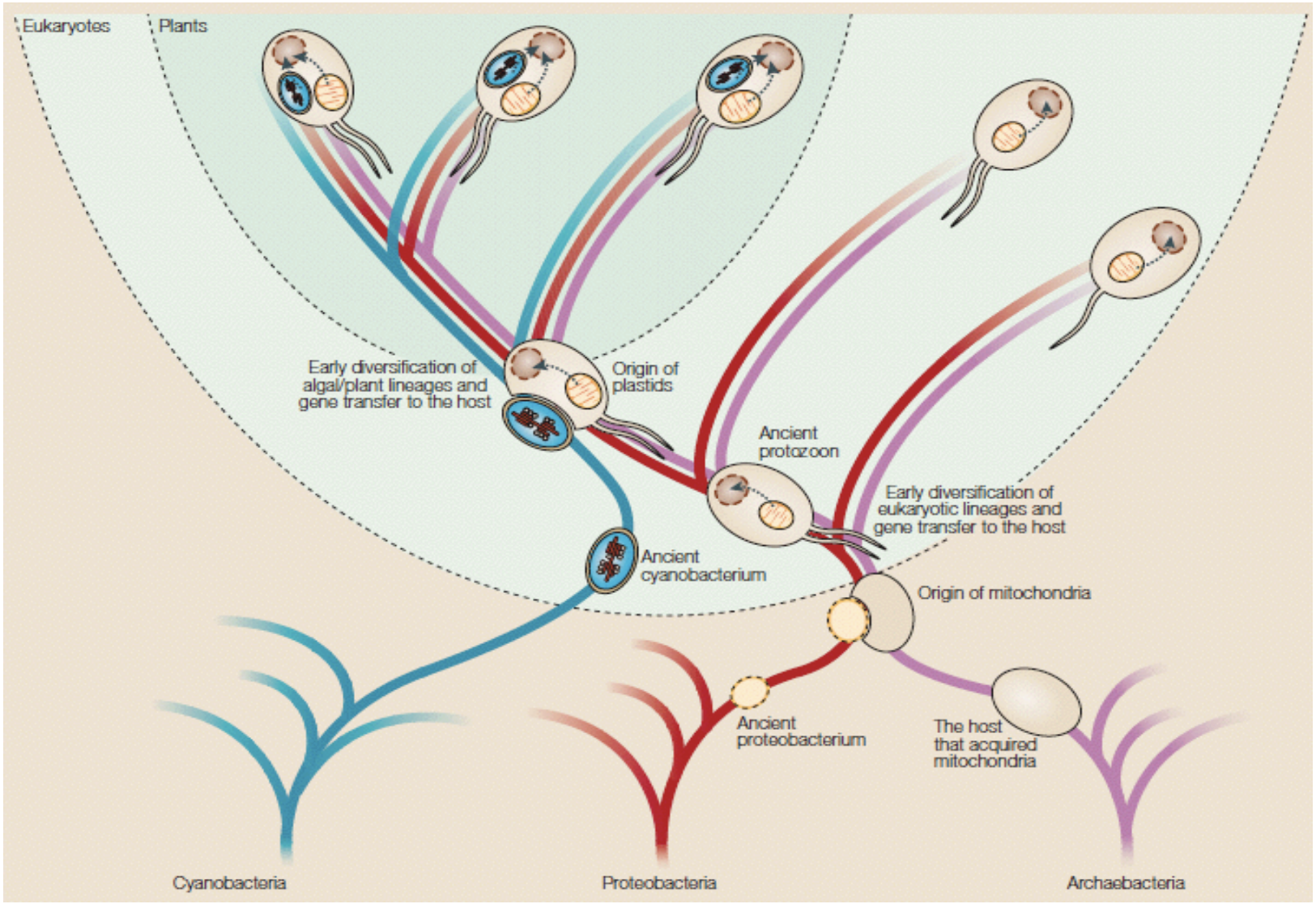
PLANT CELL
- Peroxisome
- Mitochondrion
- Golgi complex
- Microfilament
- Vacuole
- Microtubule
- Chloroplast
- Cytoplasm
- Ribosome
- Smooth endoplasmic reticulum
- Rough endoplasmic reticulum
- Plasma membrane
- Cell wall
- Nucleolus
- Nucleus

ANIMAL CELL
- Flagellum
- Nucleus ← Nuclear genome
- Nucleolus
- Golgi complex
- Cytoplasm
- Basal body
- Microfilament
- Lysosome
- Centrosome:
  - Centriole
  - Pericentriolar material
- Ribosome
- Microtubule
- Peroxisome
- Rough endoplasmic reticulum
- Mitochondrion ← Mitochondrial genome
- Smooth endoplasmic reticulum
- Plasma membrane

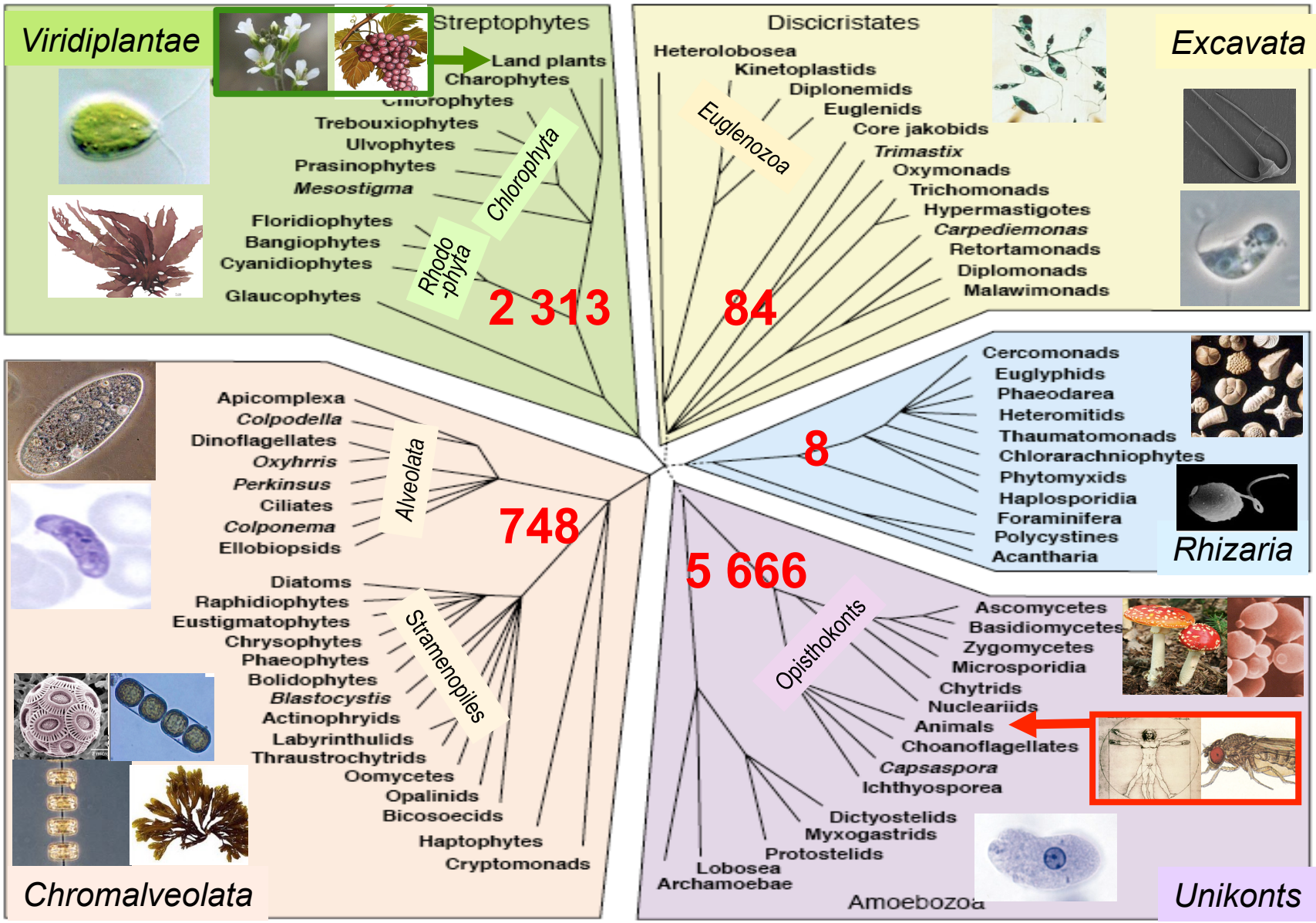
(a) Highly schematic diagram of a composite eukaryotic cell, half plant and half animal

Origin of eukaryotes: endosymbioses

Timmis *et al.* (2004) *Nature Reviews Genetics* **5**: 123-135

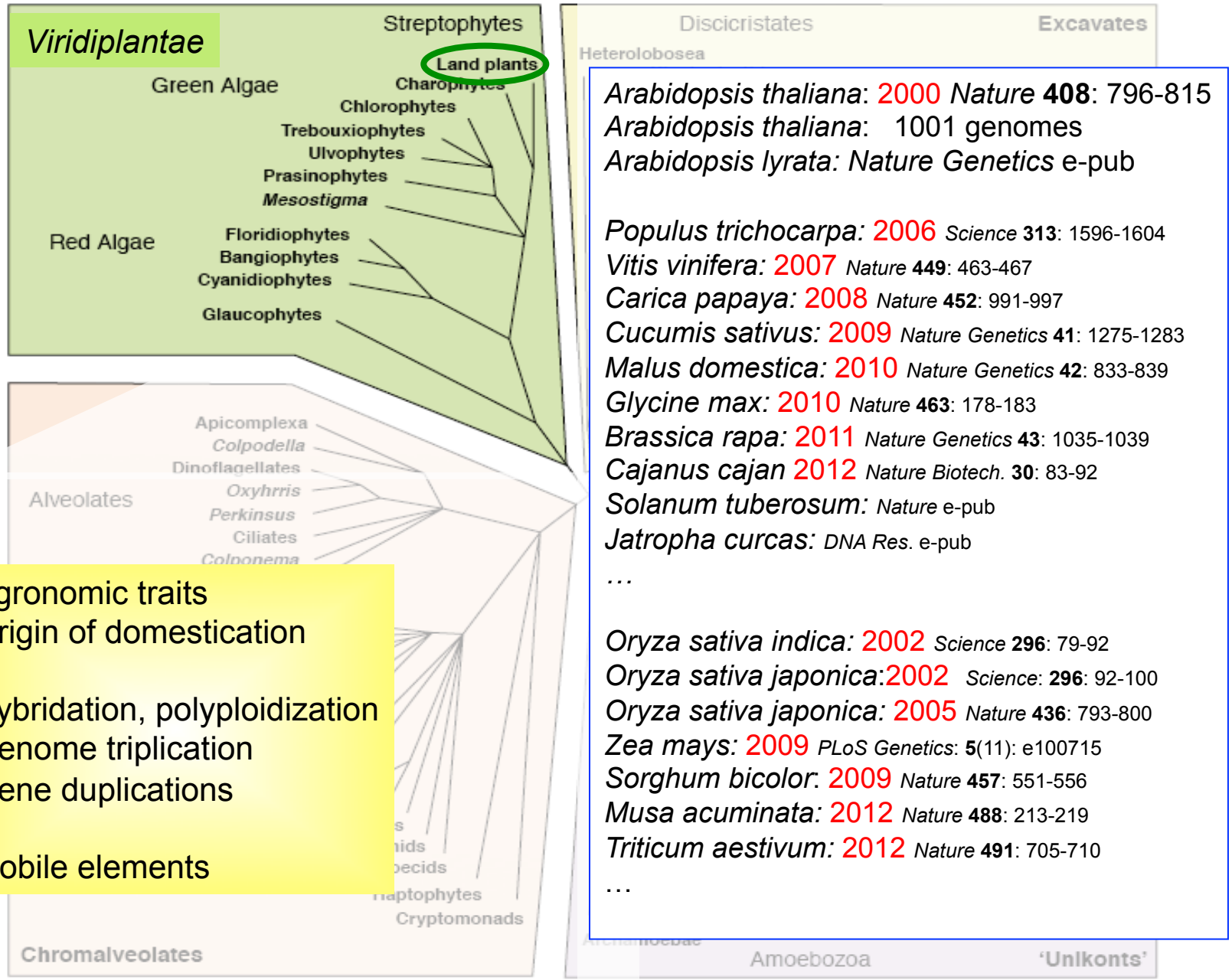


Genomics of the eukaryotes



Keeling *et al.*, 2005 *Trends in Ecology and Evolution*, **20**: 670-676

Genomes of *Streptophyta*



Agronomic traits
Origin of domestication

Hybridation, polyploidization
Genome triplication
Gene duplications

Mobile elements

Arabidopsis thaliana: 2000 *Nature* **408**: 796-815
Arabidopsis thaliana: 1001 genomes
Arabidopsis lyrata: Nature Genetics e-pub

Populus trichocarpa: 2006 *Science* **313**: 1596-1604
Vitis vinifera: 2007 *Nature* **449**: 463-467
Carica papaya: 2008 *Nature* **452**: 991-997
Cucumis sativus: 2009 *Nature Genetics* **41**: 1275-1283
Malus domestica: 2010 *Nature Genetics* **42**: 833-839
Glycine max: 2010 *Nature* **463**: 178-183
Brassica rapa: 2011 *Nature Genetics* **43**: 1035-1039
Cajanus cajan 2012 *Nature Biotech.* **30**: 83-92
Solanum tuberosum: *Nature* e-pub
Jatropha curcas: *DNA Res.* e-pub
…

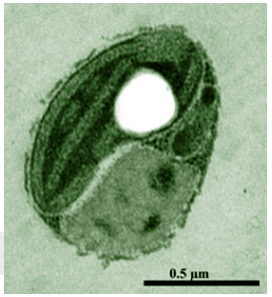
Oryza sativa indica: 2002 *Science* **296**: 79-92
Oryza sativa japonica:2002 *Science*: **296**: 92-100
Oryza sativa japonica: 2005 *Nature* **436**: 793-800
Zea mays: 2009 *PLoS Genetics*: **5**(11): e100715
Sorghum bicolor: 2009 *Nature* **457**: 551-556
Musa acuminata: 2012 *Nature* **488**: 213-219
Triticum aestivum: 2012 *Nature* **491**: 705-710
…

Genomes of unicellular eukaryotes: *Viridiplantae*

The genome of *Ostreococcus tauri*

Derelle *et al*., 2006 *PNAS* **103**: 11647-11652

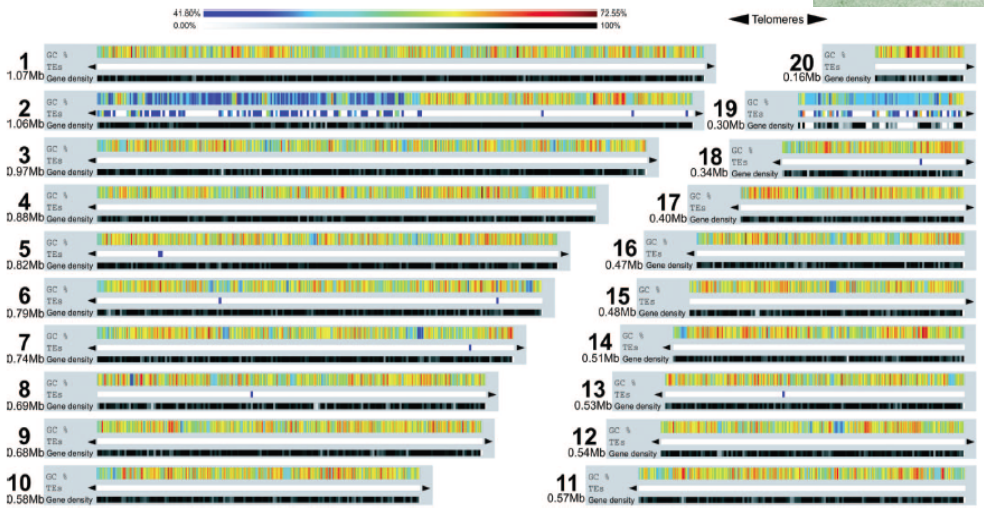
the smallest free living eukaryote



The genome of *Ostreococcus lucimarinus*

Palenik *et al.*, 2007 PNAS **104**: 7705-7710

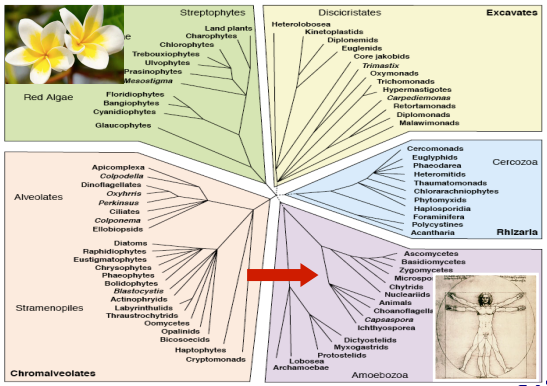
Compositional heterogeneity related to transposons

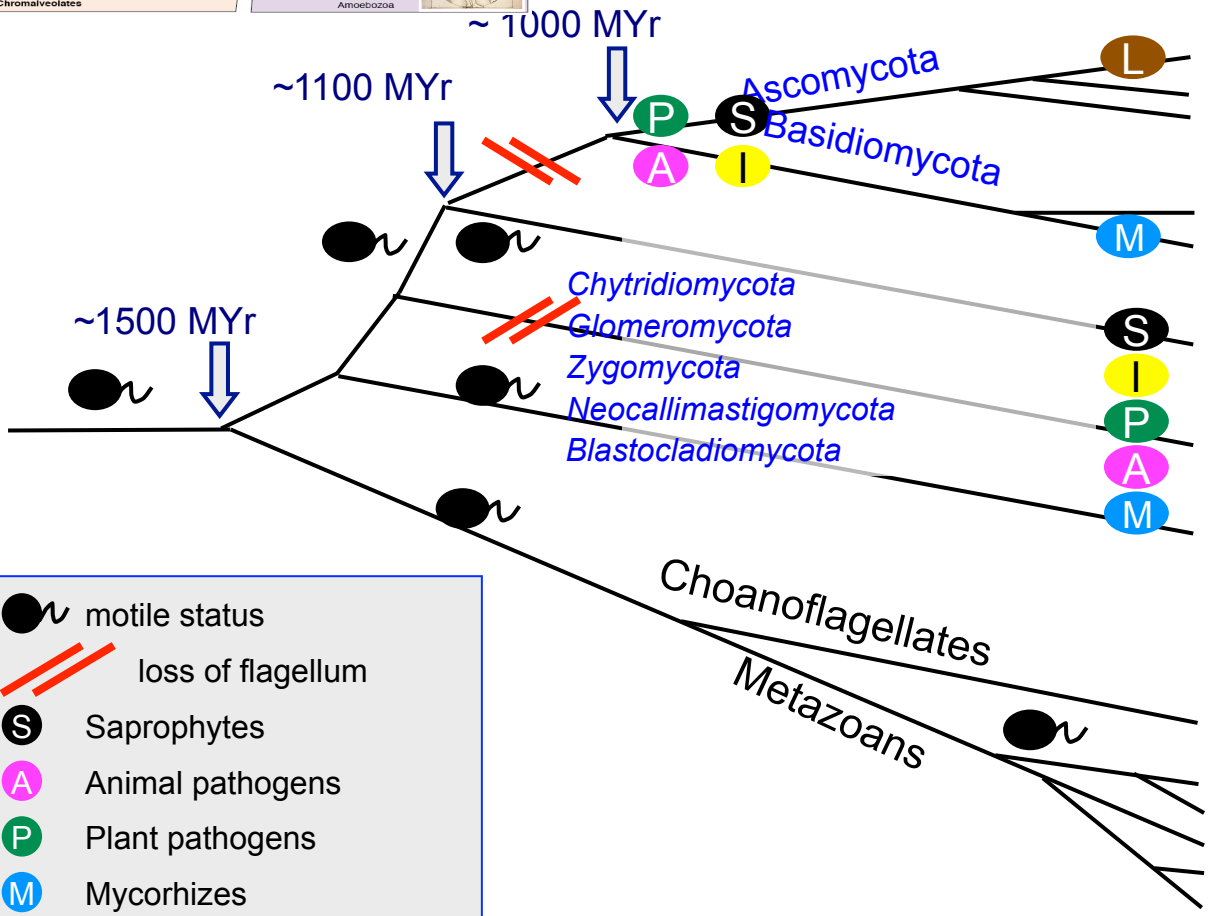


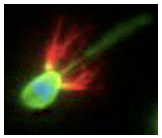
	O. lucimarinus	*O. tauri*
genome size (Mb)	13.2	12.6
chromosomes	21	20
protein-coding genes	7651	7892
split genes (%)	20	25

- Multiple mechanisms contribute to species divergence, act differently on different chromosomes.
- **Horizontal gene transfer** altering cell-surface characteristics.
- Numerous **gene fusions** 330 (*O.t.*), 348 (*O. l.*) of which 137 are common to both species.
- Numerous (20) genes for selenocysteine-containing proteins (TGA codons).

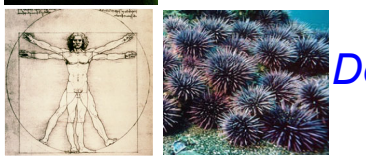
The world of Opisthokonts







Monosiga brevicolis

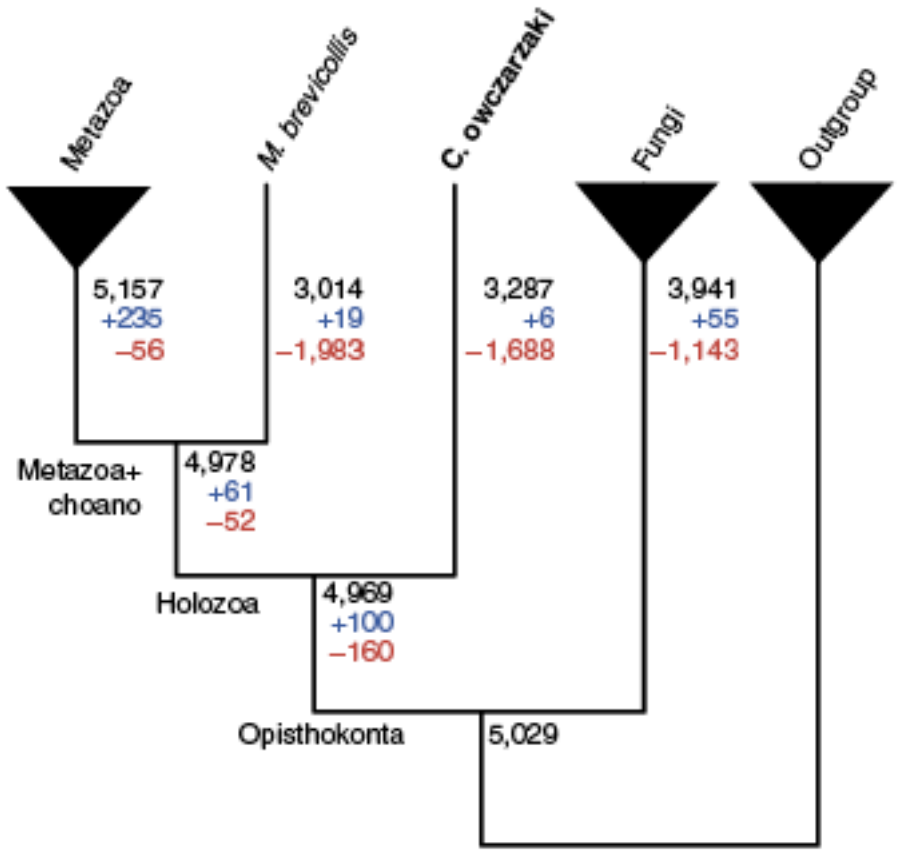


Deuterostomia

Ecdysozoa
Lophotrochozoa

The origin of metazoan protein domains

Total number of Pfam domains and numbers of domain gain **(+)** and loss (-) events inferred from Dollo parsimony



Gain and loss of protein domains within the Opisthokonta

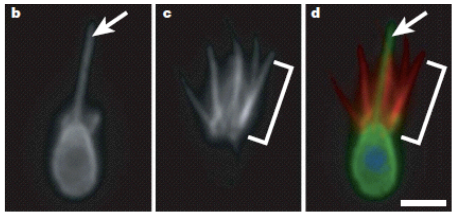
The origin of metazoan protein domains

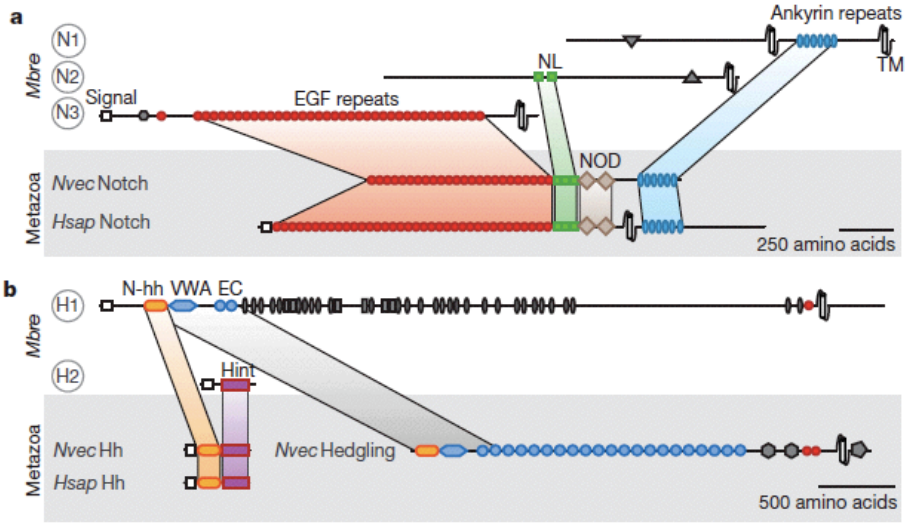
Monosiga brevicollis
(Choanoflagellate)

King *et al.*, 2008, *Nature* **451**: 783-788

41.6 Mb

9 171 protein-coding genes





Examples of protein domain fusions

Capsaspora owczarzaki
(Filasterean)

Suga *et al.*, 2013, *Nature Commun.* **4:** 2325

28.0 Mb

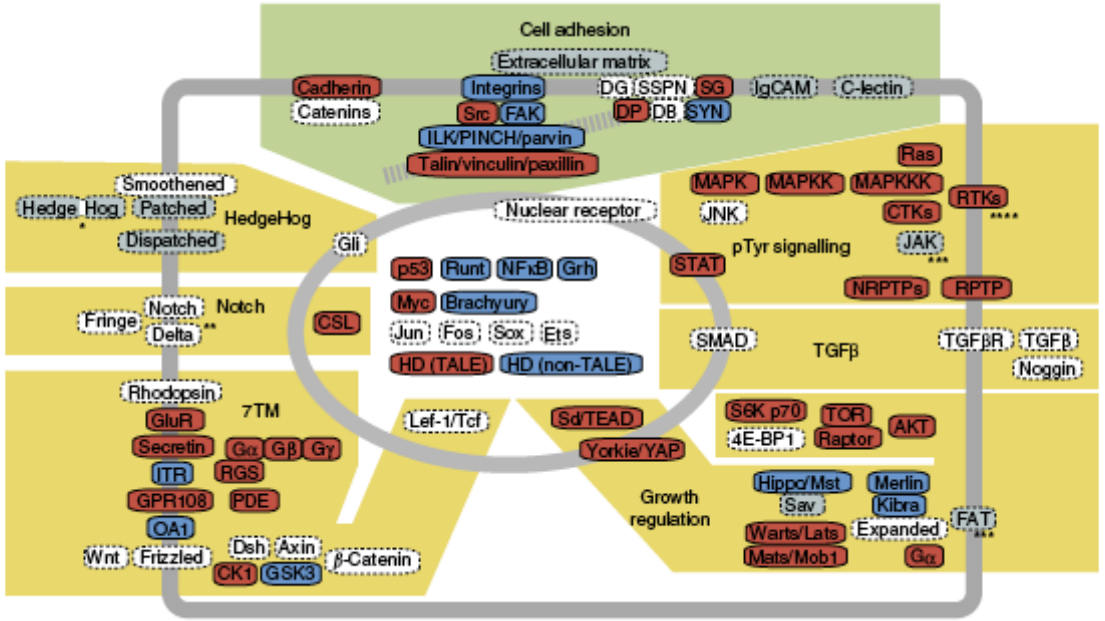
8 657 protein-coding genes

Present in *M. bre*. and *C. owc*.

Present in *C. owc*. only.

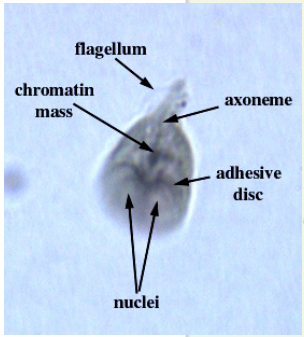
Present in *M. bre.* only.

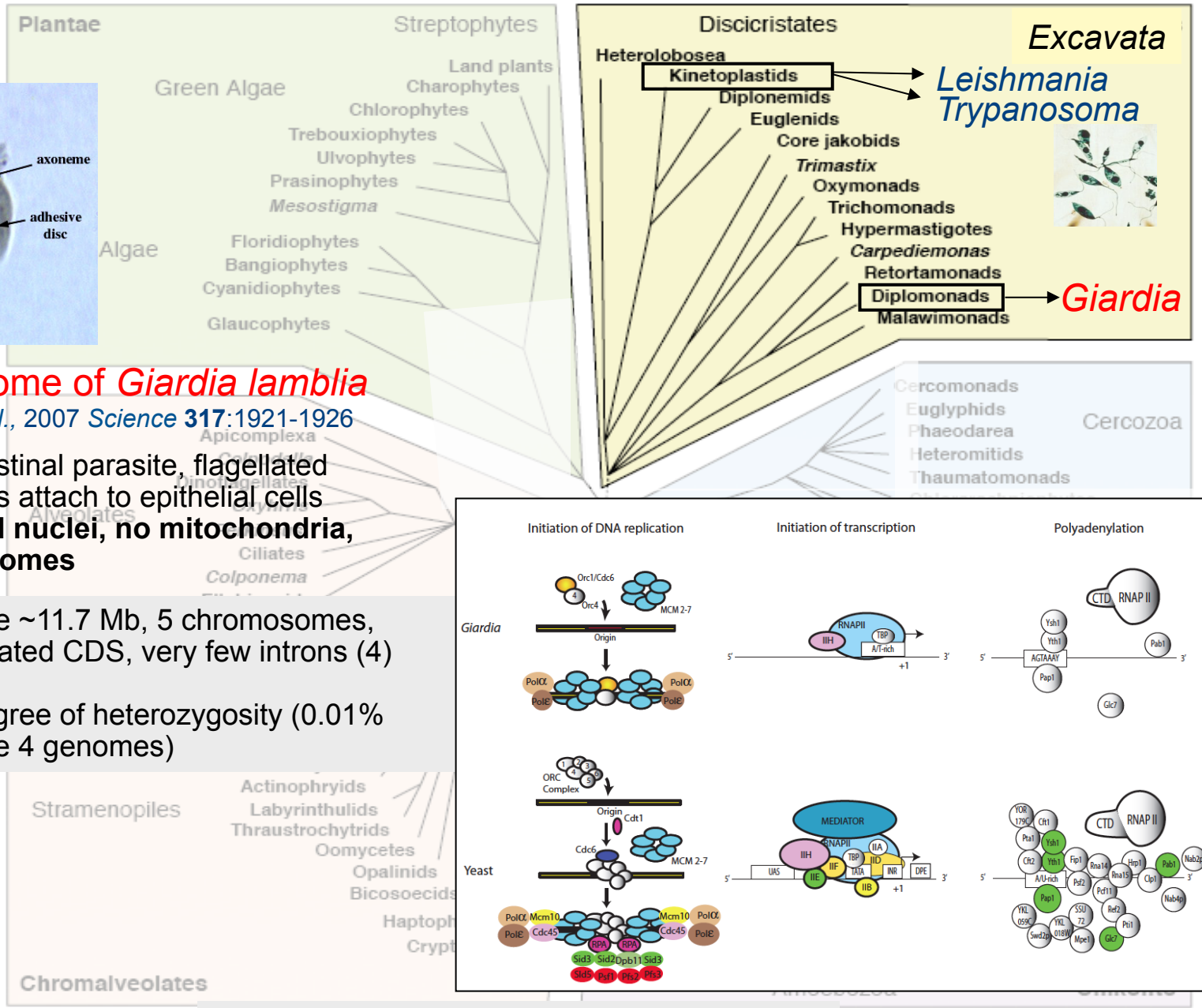
Absent in *M. bre.* and *C. owc*.



Traces of protein componenets of major metazoan cell adhesion complexes (green) and various signlling pathways (yellow).

Genomes of unicellular eukaryotes: *Excavata*





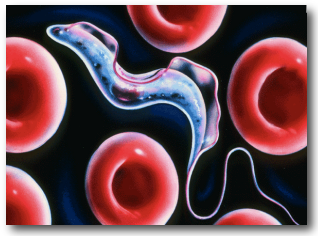
The genome of *Giardia lamblia*

Morrison *et al.*, 2007 *Science* **317**:1921-1926

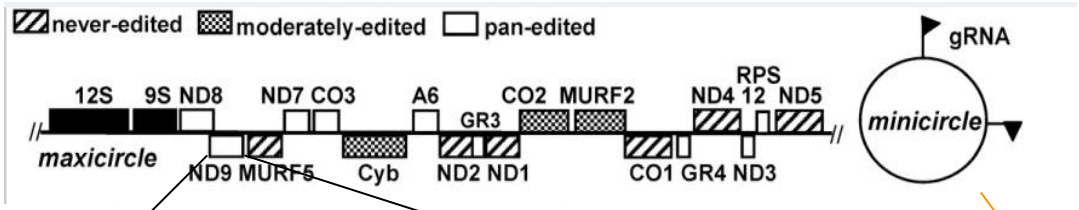
human intestinal parasite, flagellated trophozoites attach to epithelial cells
two diploid nuclei, no mitochondria, no peroxisomes

- Genome ~11.7 Mb, 5 chromosomes, 6470 annotated CDS, very few introns (4)
- Low degree of heterozygosity (0.01% between the 4 genomes)
- Evolution is often regressive

RNA editing



1983 Kinetoplasts of *Trypanosoma*



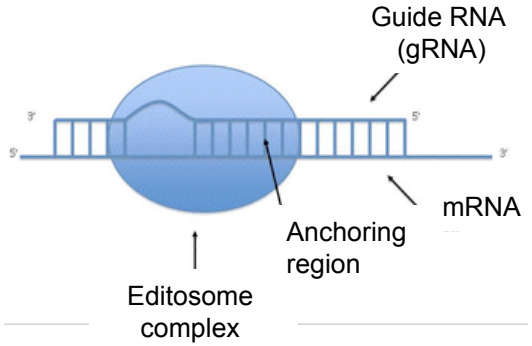
Guide RNAs

~~Ser Gly Glu Lys~~

AGCGGAGAAAAA

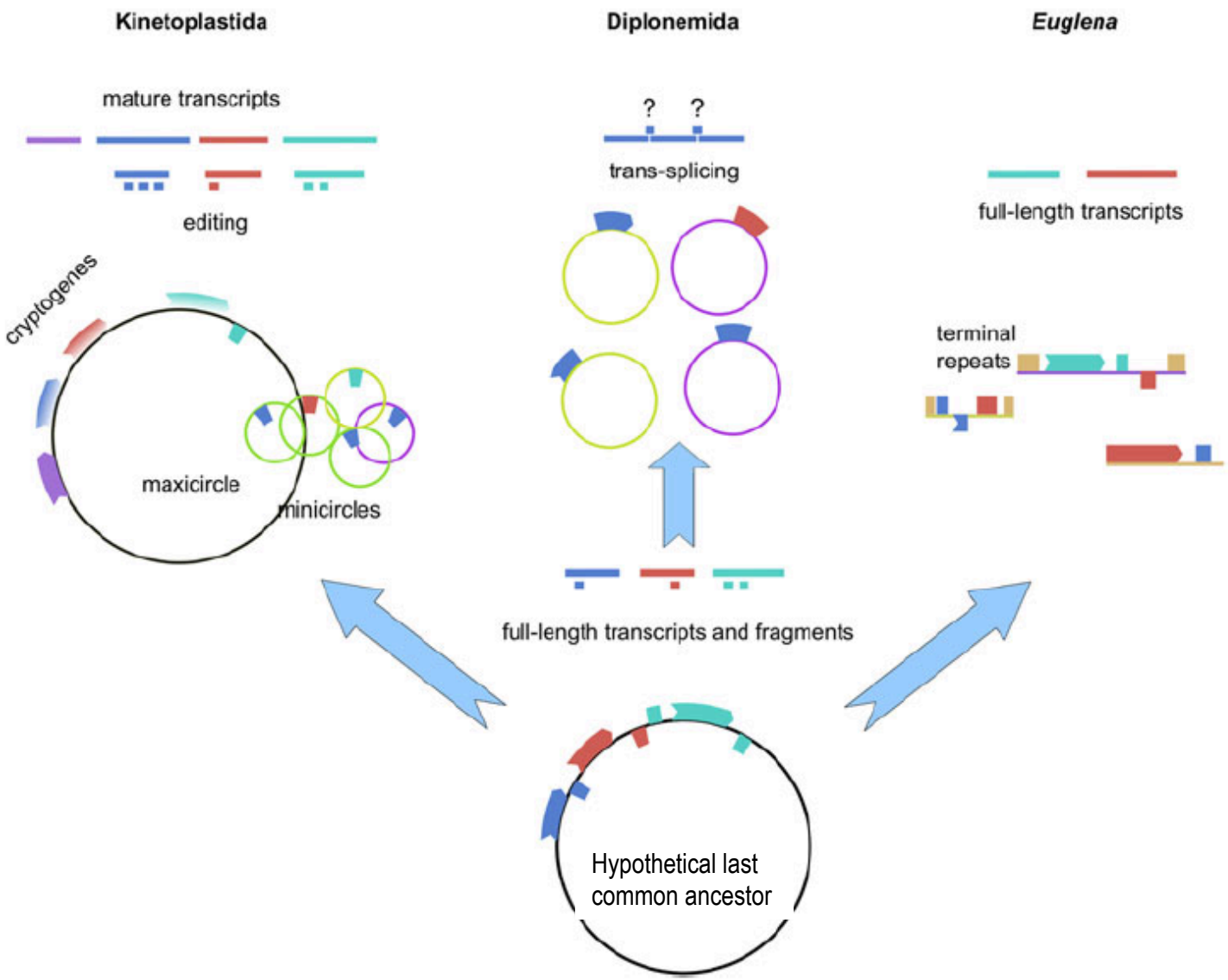
AuGuuuCGuuGuAGAuuuuuuAuuAuuuuuuuuuAuuA

Met Phe Arg Cys Arg Phe Leu Leu Phe Phe Leu Leu

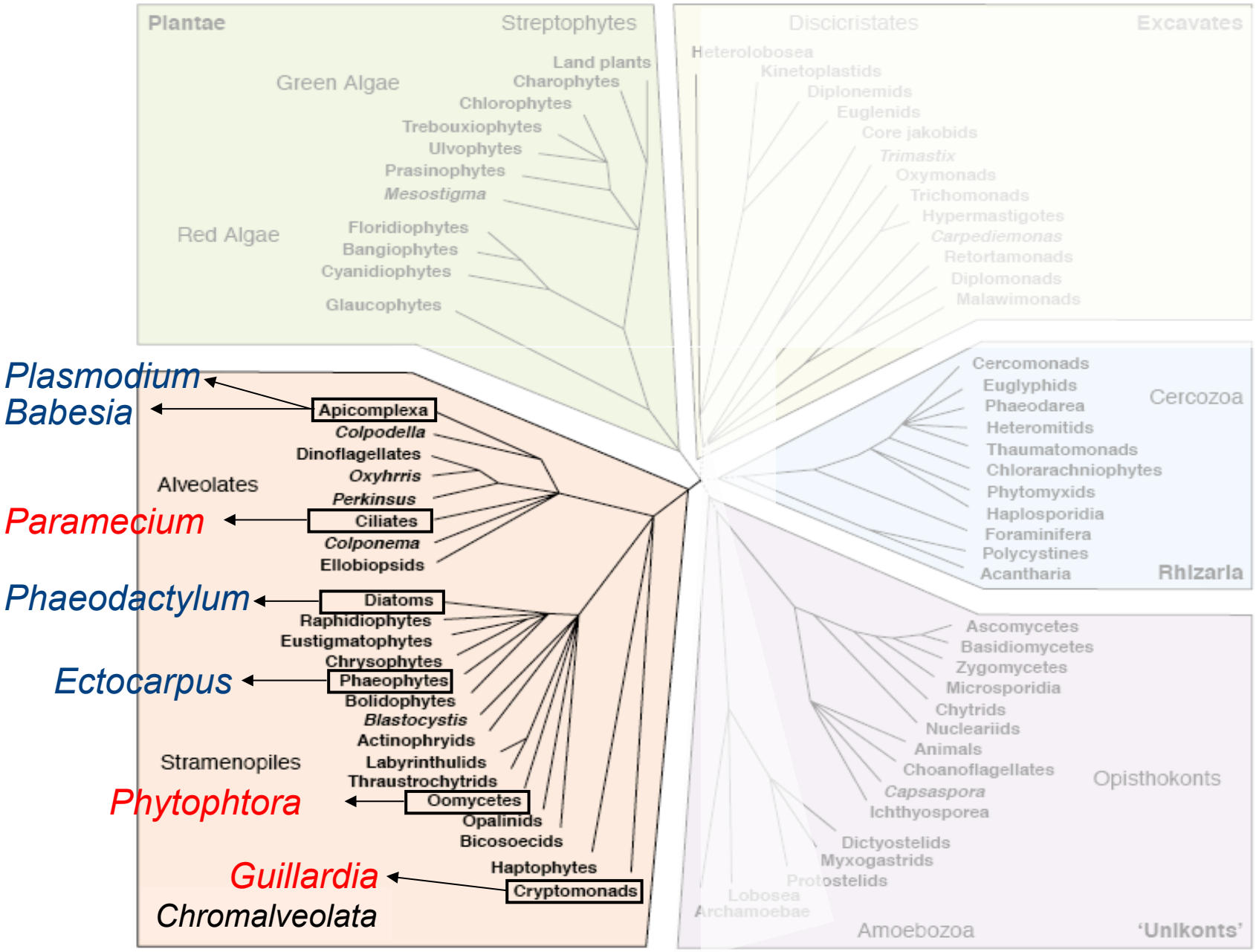


Germline mitochondrial genomes in *Euglenozoa*

Flegontov et al., (2011) Curr. Genet. **57:** 225-232



Genomes of unicellular eukaryotes: *Chromalveolata*



The genomes of *Phytophtora infestans (sojae* and *ramorum)*

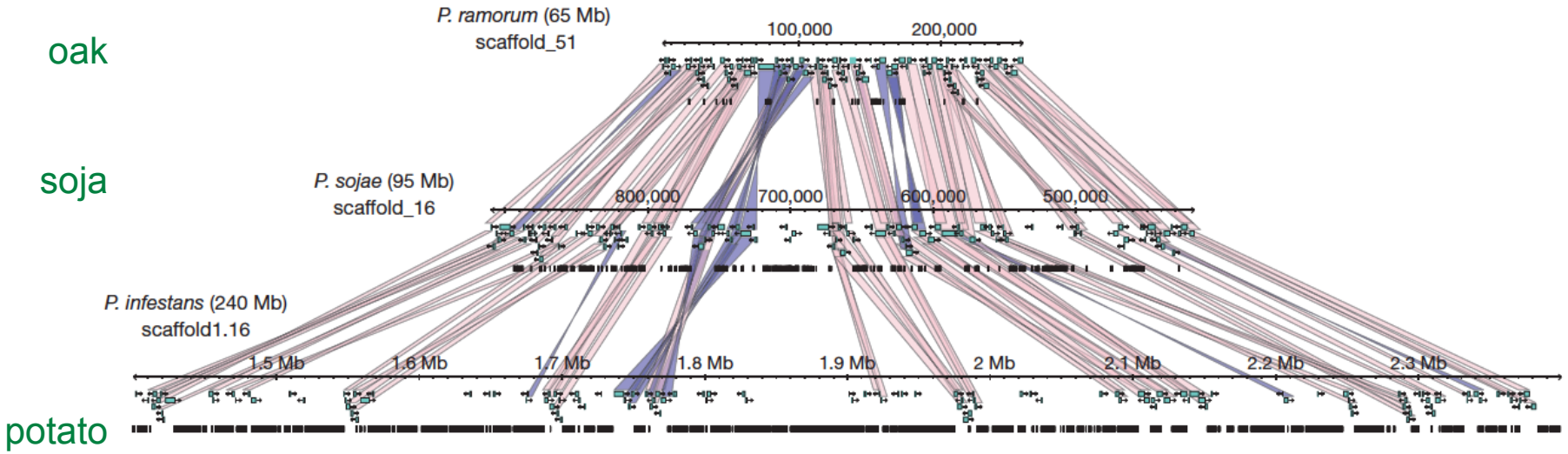
Haas *et al.*, 2009 *Nature* **461**: 393-398

	P. infestans	*P. sojae*	*P. ramorum*
Genome size (Mb)	240	95	65
Scaffolds	4921	1810	2576
Repeat (%)	*74*	*39*	*28*
Protein-coding genes	17797	16988	14451

- **Conserved syntenic blocks** containing most common genes and few repeated DNA are separated by regions of repeated DNA with low gene density and no conservation of gene order **--> dynamic genomes**

- **Rapidly evolving effector genes** in non-conserved regions (modular secreted proteins, major types: RXLR and Crinkler, targetted to plant cells and responsible for necrosis, mostly species-specific)

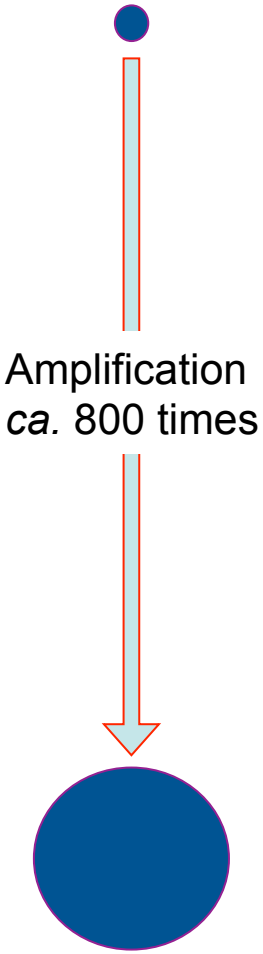
- Effector **gene family expansion** in *P. infestans* associated with numerous mobile elements (helitron)



Genomes of unicellular eukaryotes: *Chromalveolata*

The macronuclear genome of *Paramecium tetraurelia*

Aury *et al.*, (2006) *Nature* **444**: 171-178 (697 scaffolds, totalling 72 Mb)



Micronucleus (2n) genome size
ca. 100 Mb
> 50 chromosomes

Amplification
ca. 800 times

- Precise elimination of > 10 000 short, unique copy elements ---> **Reconstruction of functional genes**

RNA mediated process

- Imprecise elimination of transposable elements and other repeated sequences
---> **Chromosome fragmentation, *de novo* telomere addition, internal deletions**

Macronucleus (polyploid)
genome size *ca.* 75 Mb

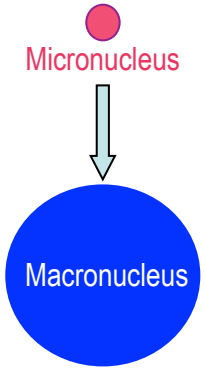
39 642 genes

Reconstructruction of a functional genome from the germline genome

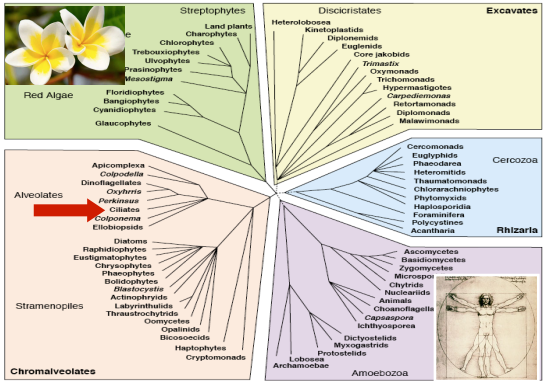
Aury *et al.*, (2006); Lepère G *et al.* (2008)

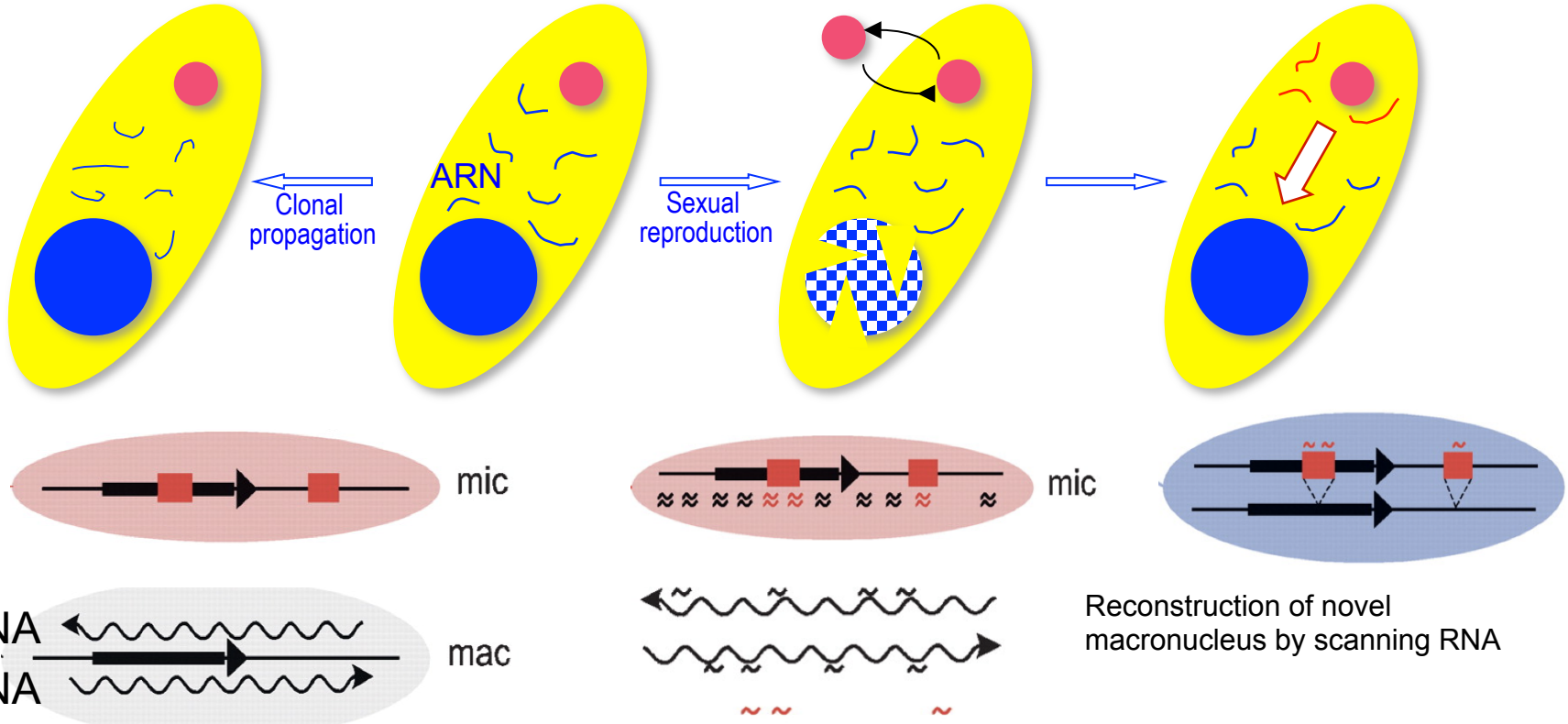
- Precise elimination of > 10 000 short, unique copy elements ---> **Reconstruction of functional genes** **RNA mediated process**

- Imprecise elimination of transposable elements and other repeated sequences ---> **Chromosome fragmentation, *de novo* telomere addition, internal deletions**



Paramecium tetraurelia

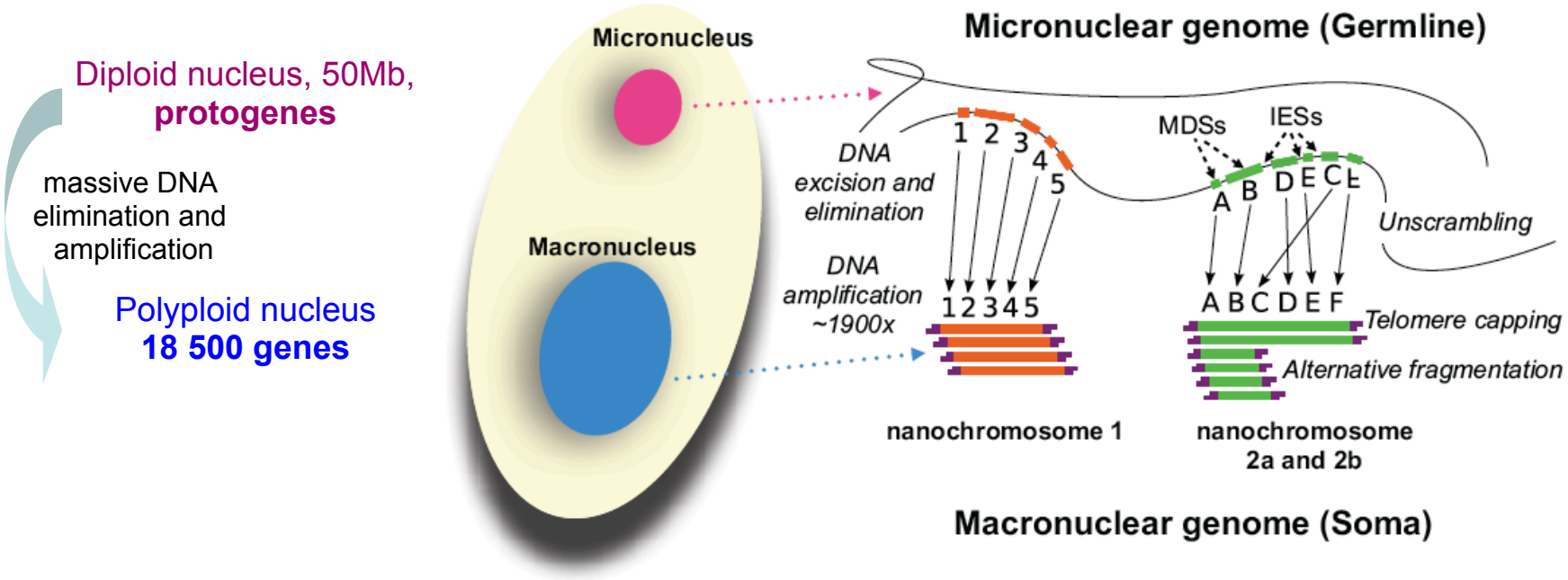




Reconstructruction of a functional genome from the germline genome

Oxytricha trifallax

Stuart et al., 2013 PLoS Biology **11**:1 e1001473



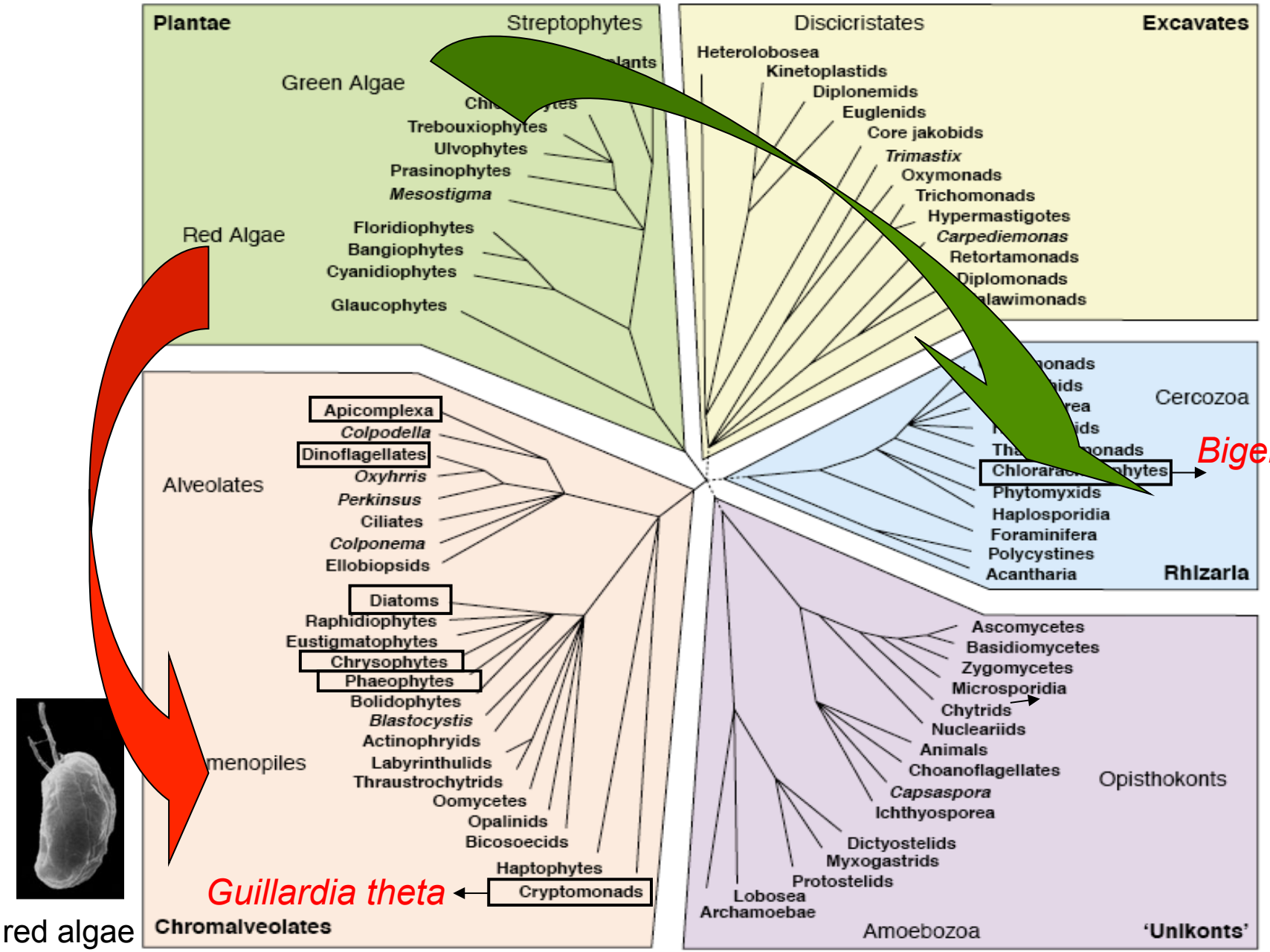
> 16 000 nanochromosomes (length 0,4 à 66 kb, mean 3,2 kb)

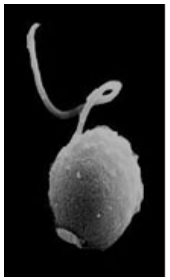
90 % correspond to a single gene (others to 2-8 genes)

Multiple isoforms, variable amplification (mean 2 000 copies) > polymorphism between individuals

de novo generation of genes by reassociation from protegene elements. millions of telomeres

Four membrane plastids and nucleomorphs (double endosymbiosis)





Bigelowiella natans

green algae endosym-bionte

Guillardia theta

red algae endosymbionte

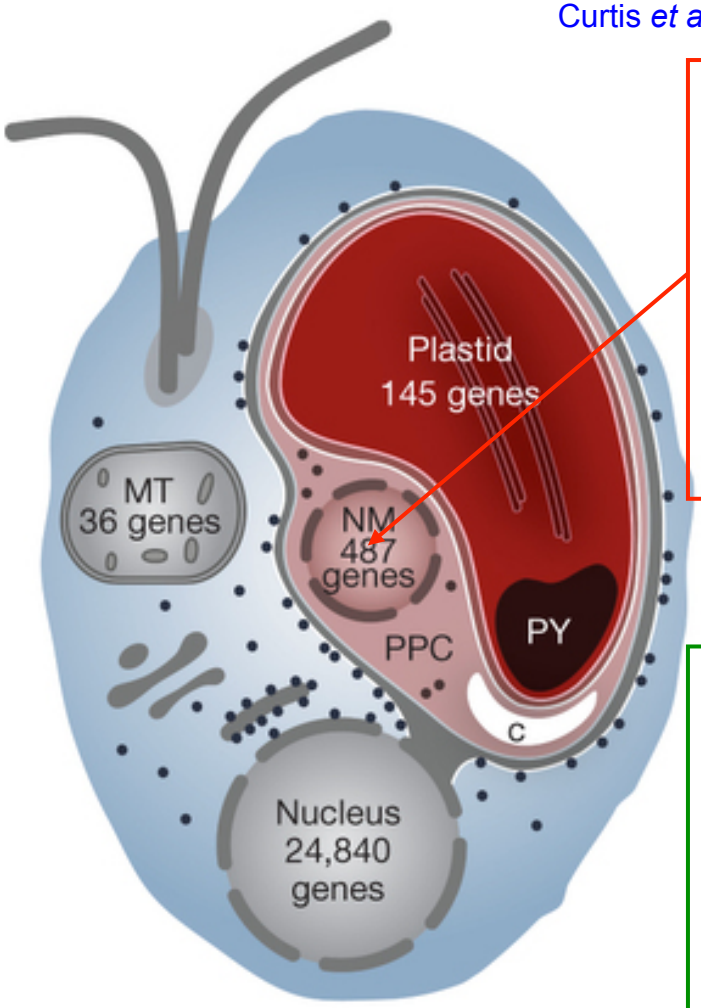
Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs

Curtis *et al.* (2012) *Nature* **492**: 59-65

Nucleomorph:
3 chromosomes:
196, 181, 174 kb

inverted repeats at chromosome ends
(rDNA, ubiquitin-conjugating enzyme gene)

Douglas *et al.*, 2001 *Nature* **410**: 1091-1096

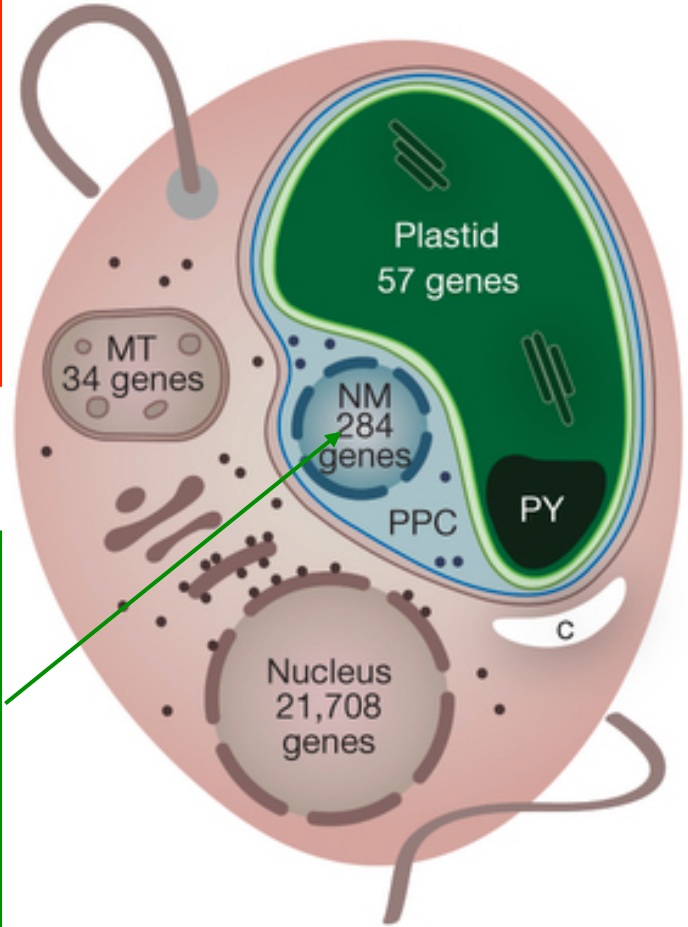


Guillardia theta

Nucleomorph:
3 chromosomes:
141, 134, 98 kb

inverted repeats at chromosome ends
(rDNA, DnaK pseudogenes)

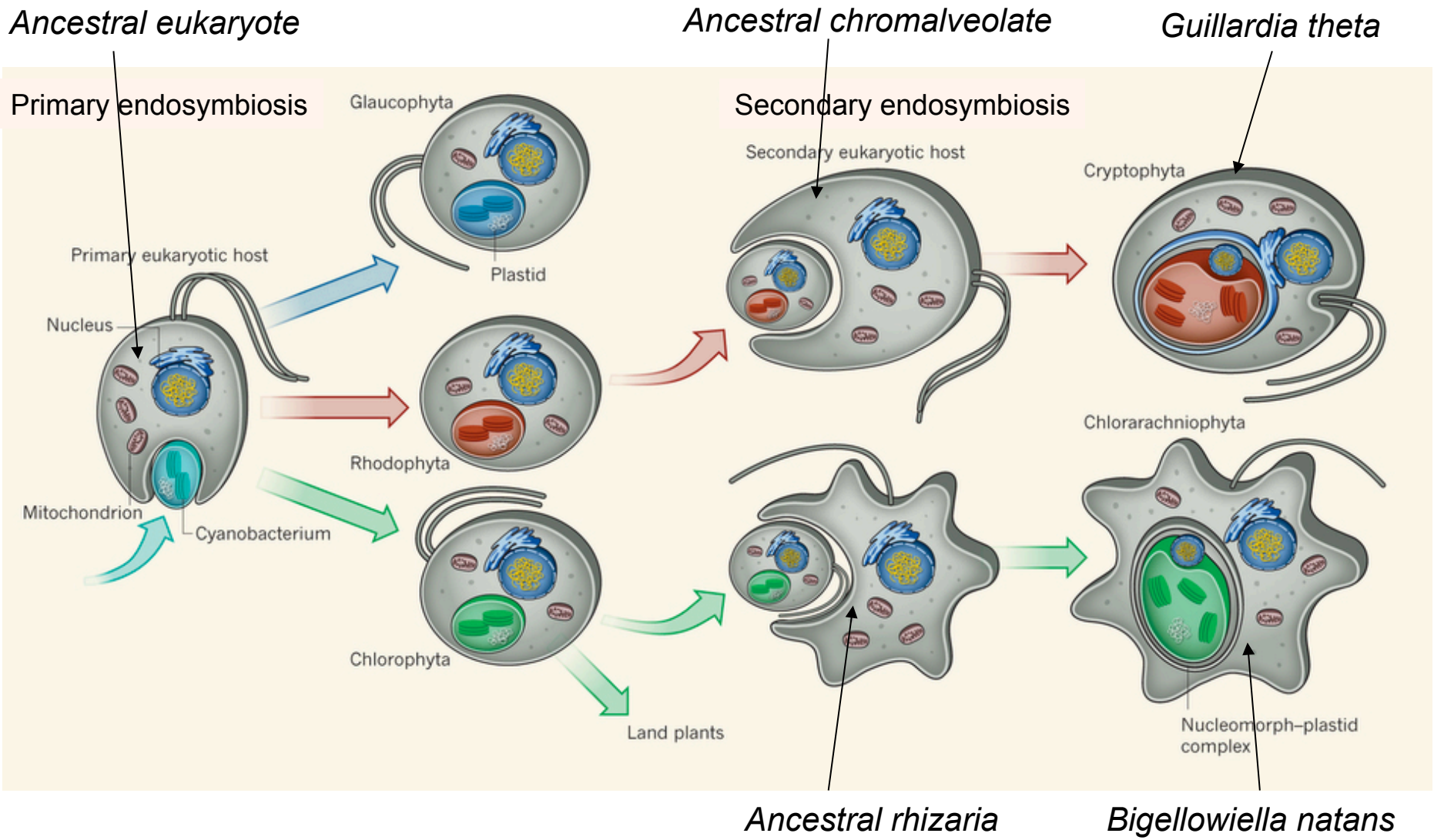
Gilson *et al.*, 2006 *PNAS* **103**: 9566-9571



Bigelowiella natans

Genome size (Mb)	87.2	94.7
Split CDS (%)	*80*	*86*
Mean exons /gene	6.4	8.8
Mean intron size (nuc.)	110	184

Gould, 2012, *Nature* **492**: 46-48



The world of eukaryotes from genomics

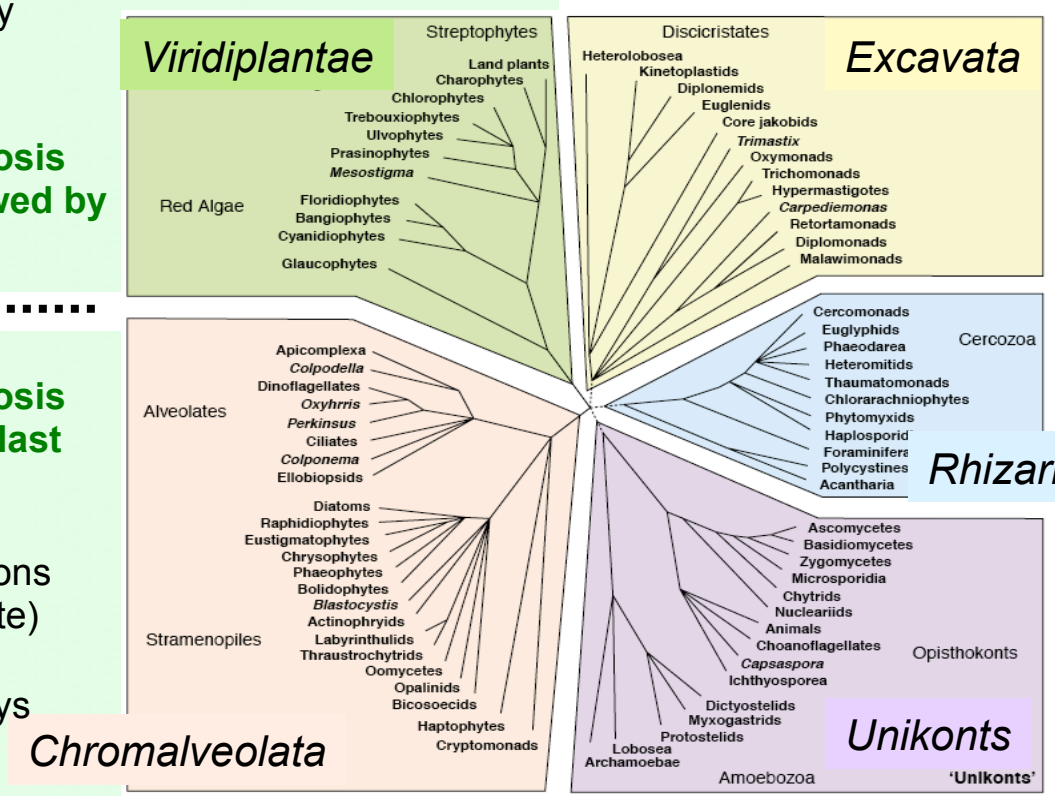
From small, compact genomes to massively duplicated genomes

Primary endosymbiosis (chloroplasts) followed by chloroplast loss

Primary endosymbiosis followed by chloroplast loss

Genome reconstructions (with RNA intermediate)

Recent endosymbiosys (red algae)



Massive RNA edition
Regressive evolution

Recent endosymbiosis (green algae)

What else ?

From small, compact genomes to extremely large genomes

Gene duplication, genome duplication, intron gain and loss, horizontal gene acquisition, frequent secondary loss (regressive evolution)

Range of genome size variation

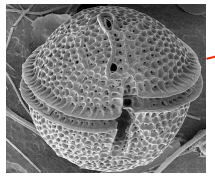
Fritillaria assyriaca
100 000 Mb

Zea mais
3 000 Mb

1000 X

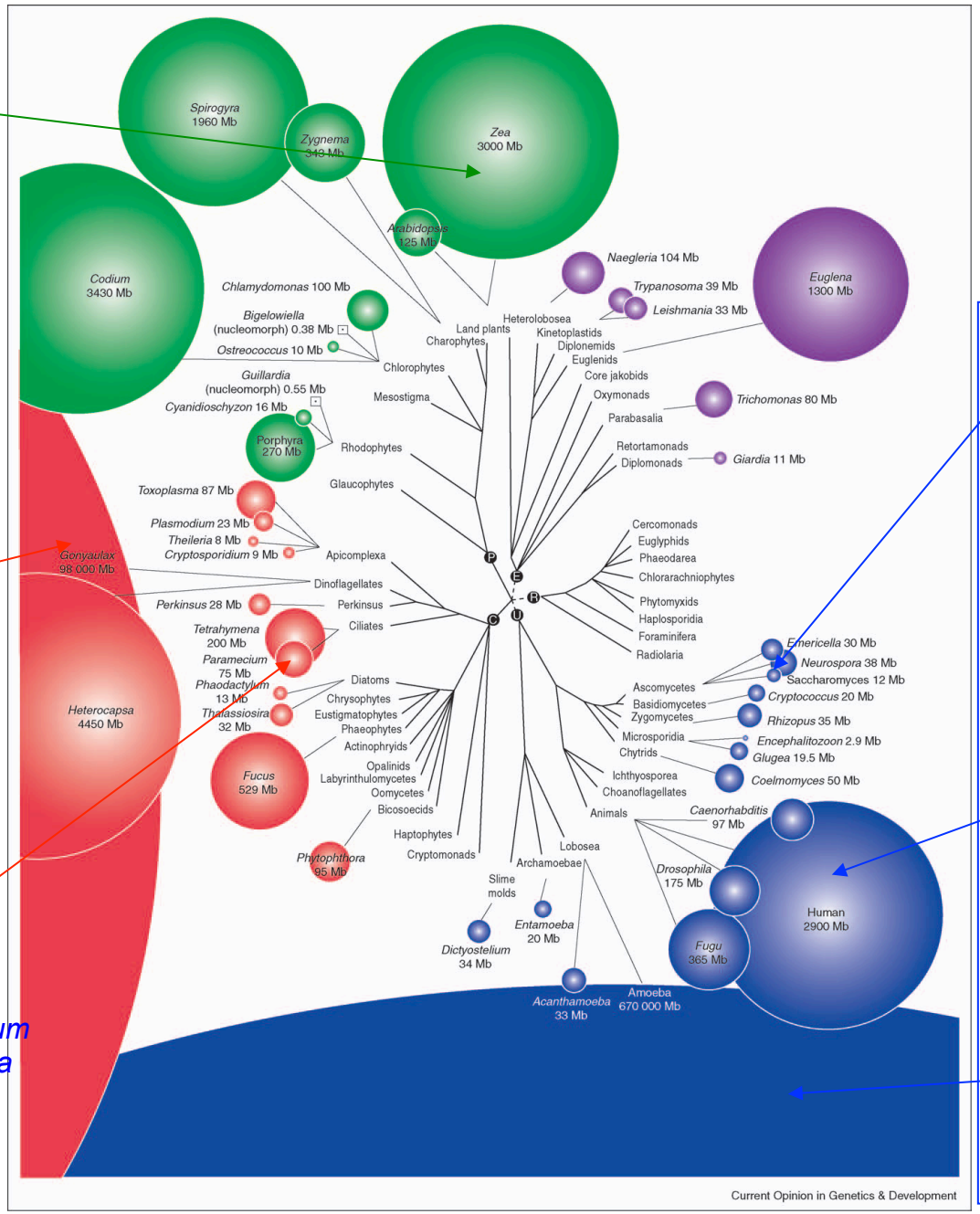
Vitis vinifera
487 Mb

1000 X

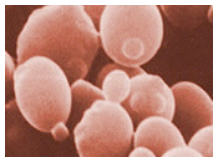


Gonyaulax grindleyi
98 000 Mb

Paramecium tetraurelia
100 Mb

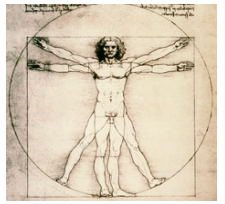


55 000 X

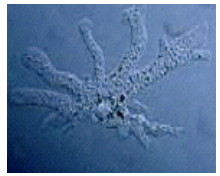


Saccharomyces cerevisiae 12 Mb

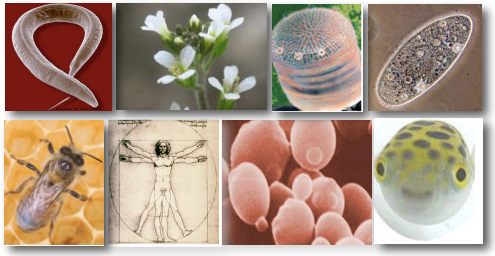
Necturus lewisi
100 000 Mb



Homo sapiens
2 900 Mb



Amoeba dubia
670 000 Mb

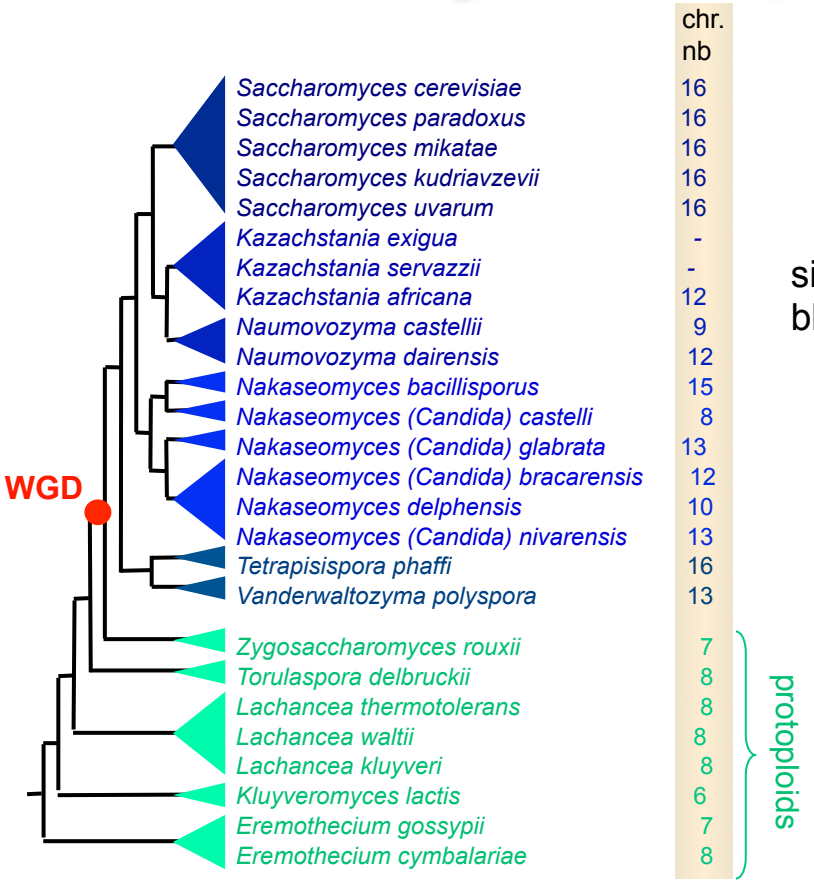


Le monde des eucaryotes, vu par la génomique comparative

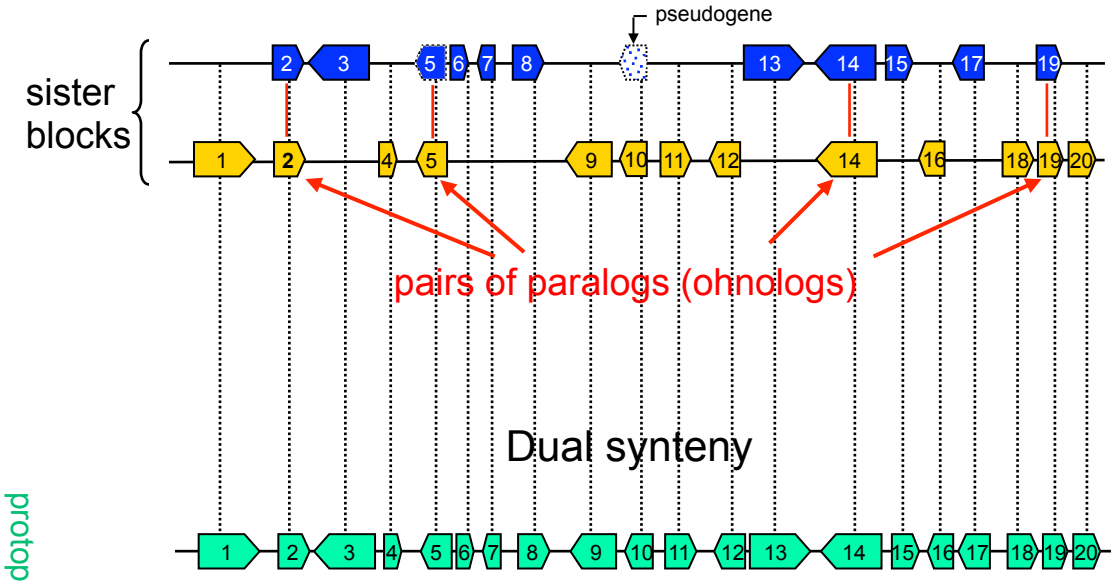
- Les mécanismes moléculaires de l'évolution des génomes eucaryotes

La formation *de novo* de gènes, *frameshifts* et ARNs

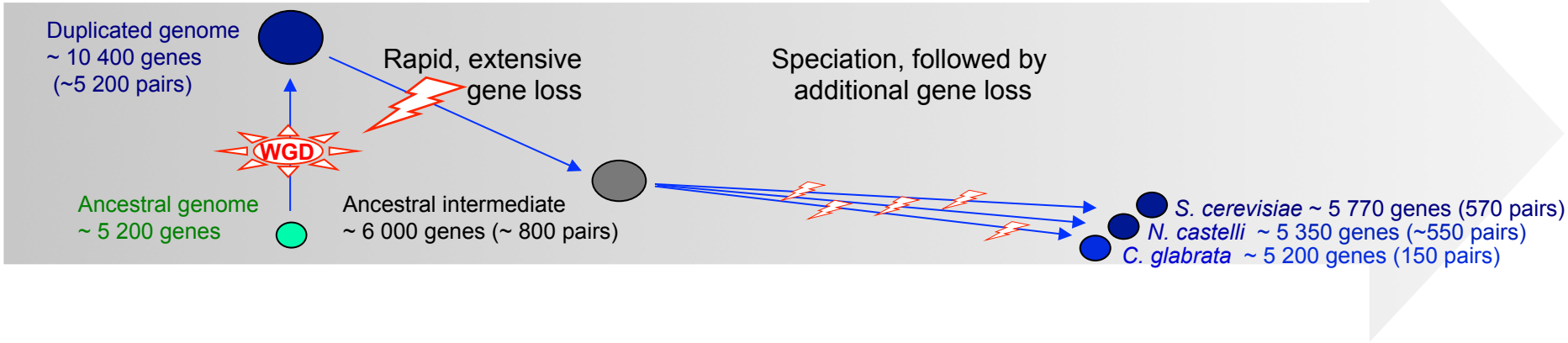
Whole-genome duplication in *Saccharomycetaceae*



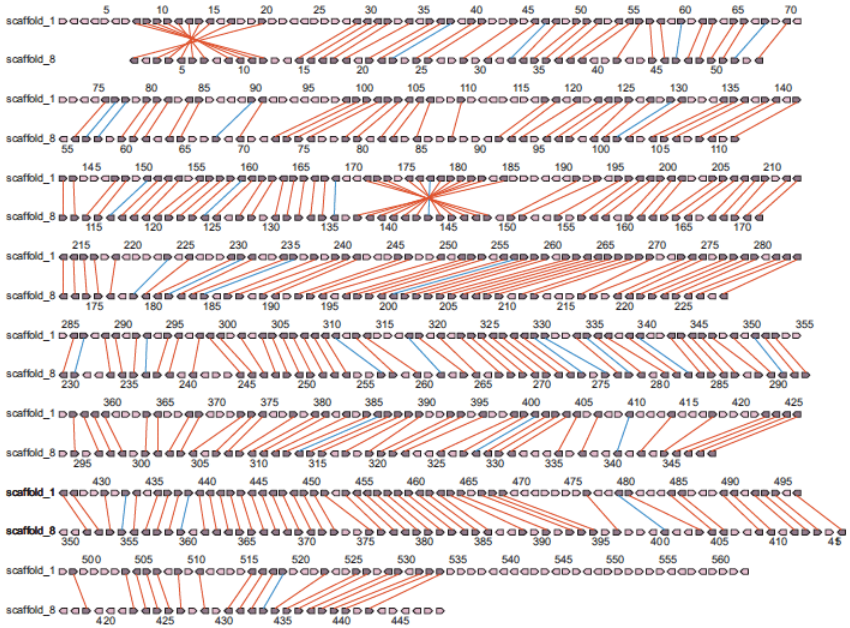
Wolfe and Shields, 1997, *Nature* **387**, 708–713

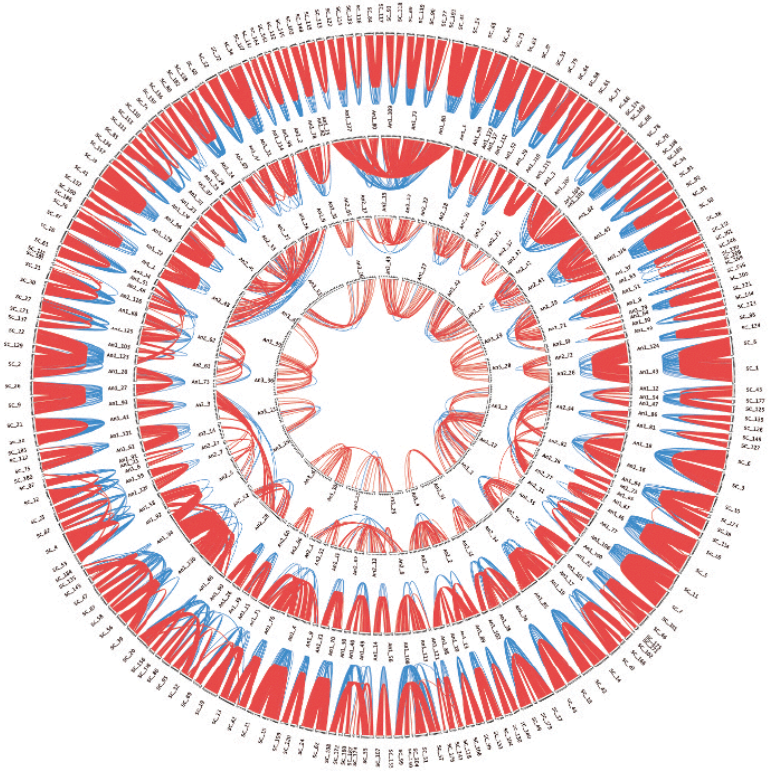


Kellis *et al.*, 2004, *Nature* **428**: 617-624
Dietrich *et al.*, 2004, Science **304**: 304-307

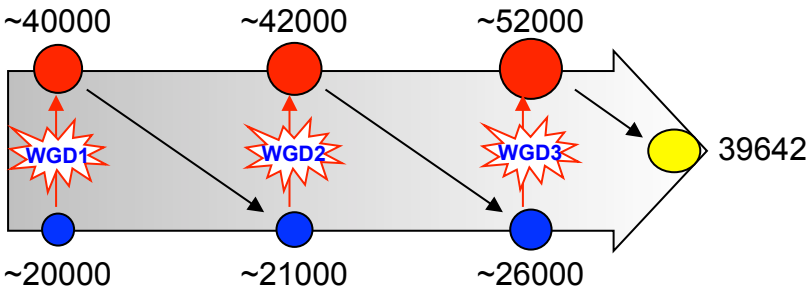


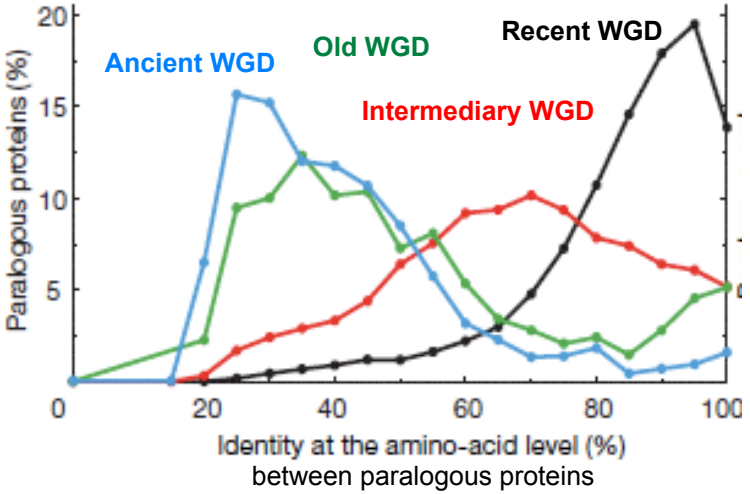
Several succesive genome duplications in *Paramecium*





Comparison of two scaffolds originating from a common ancestor after a recent WGD in P*aramecium. tetraurelia*

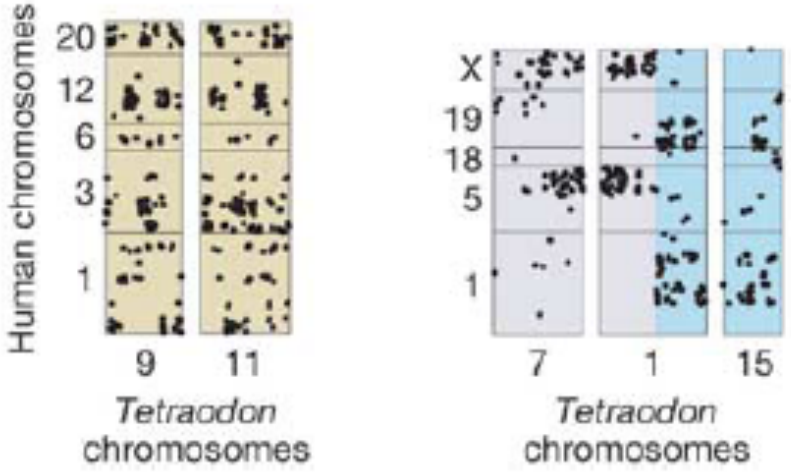




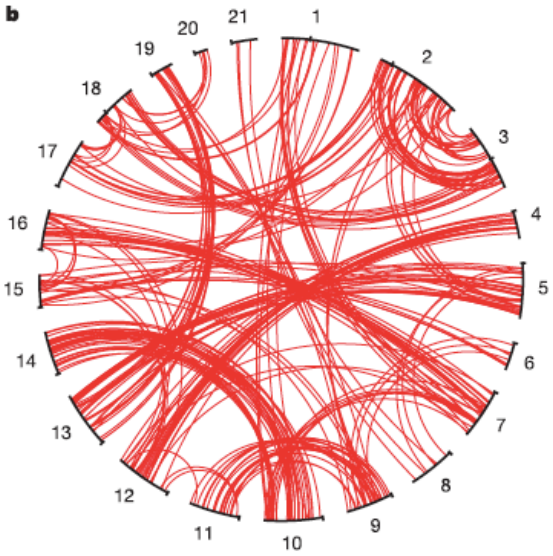
Whole-genome duplication at the origin of *Actinopterygii* fishes

Jaillon *et al.* 2004 *Nature* **431**: 946-957

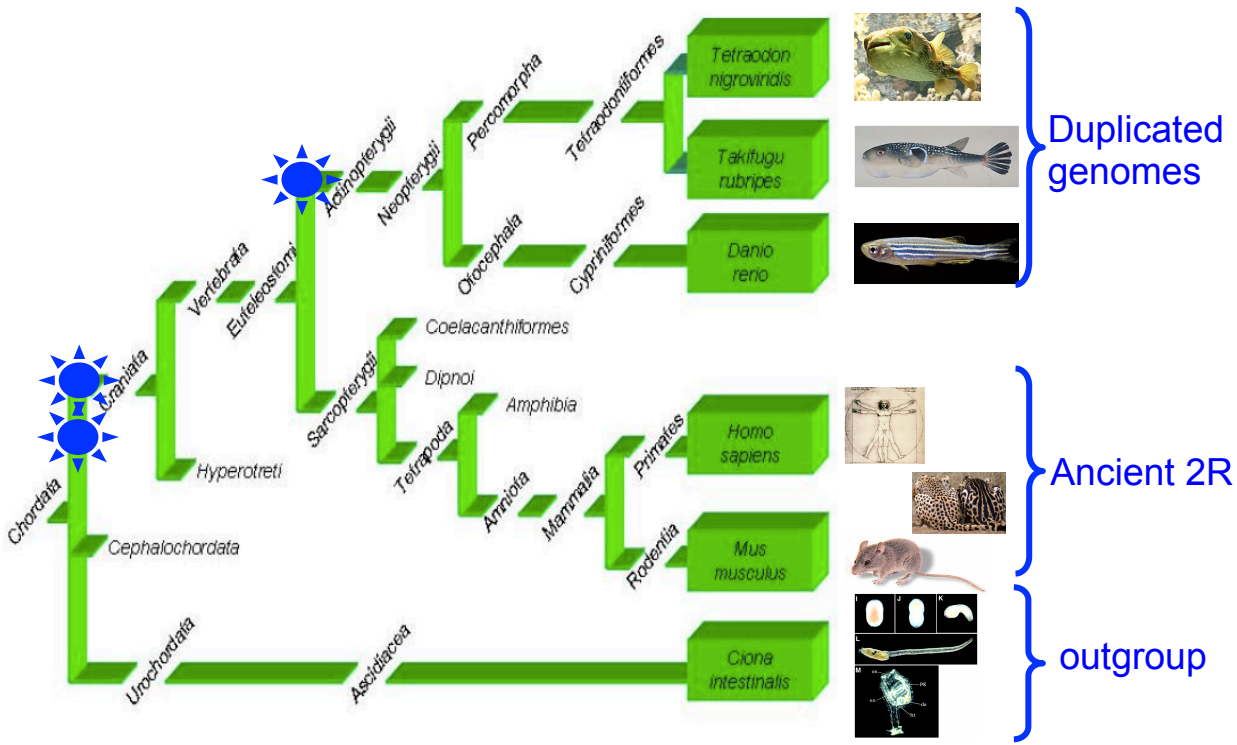
Tetraodon negroviridis



Dual synteny (Human-*Tetraodon*)



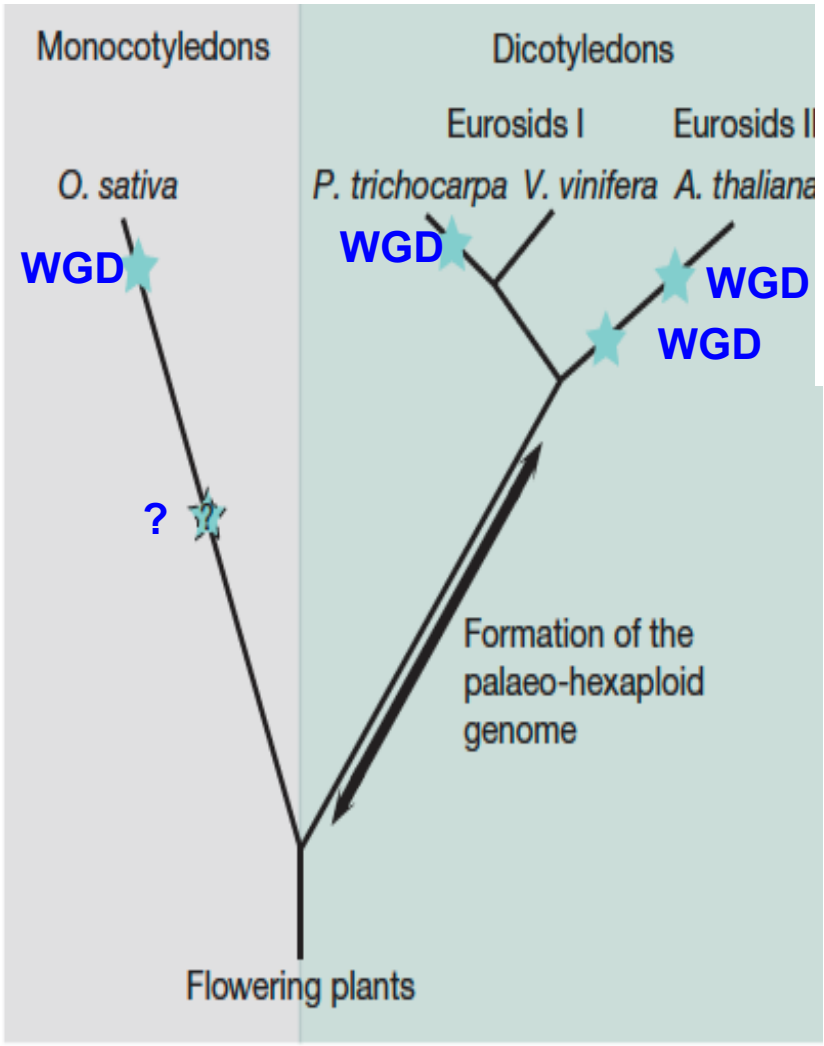
Sister blocks of synteny

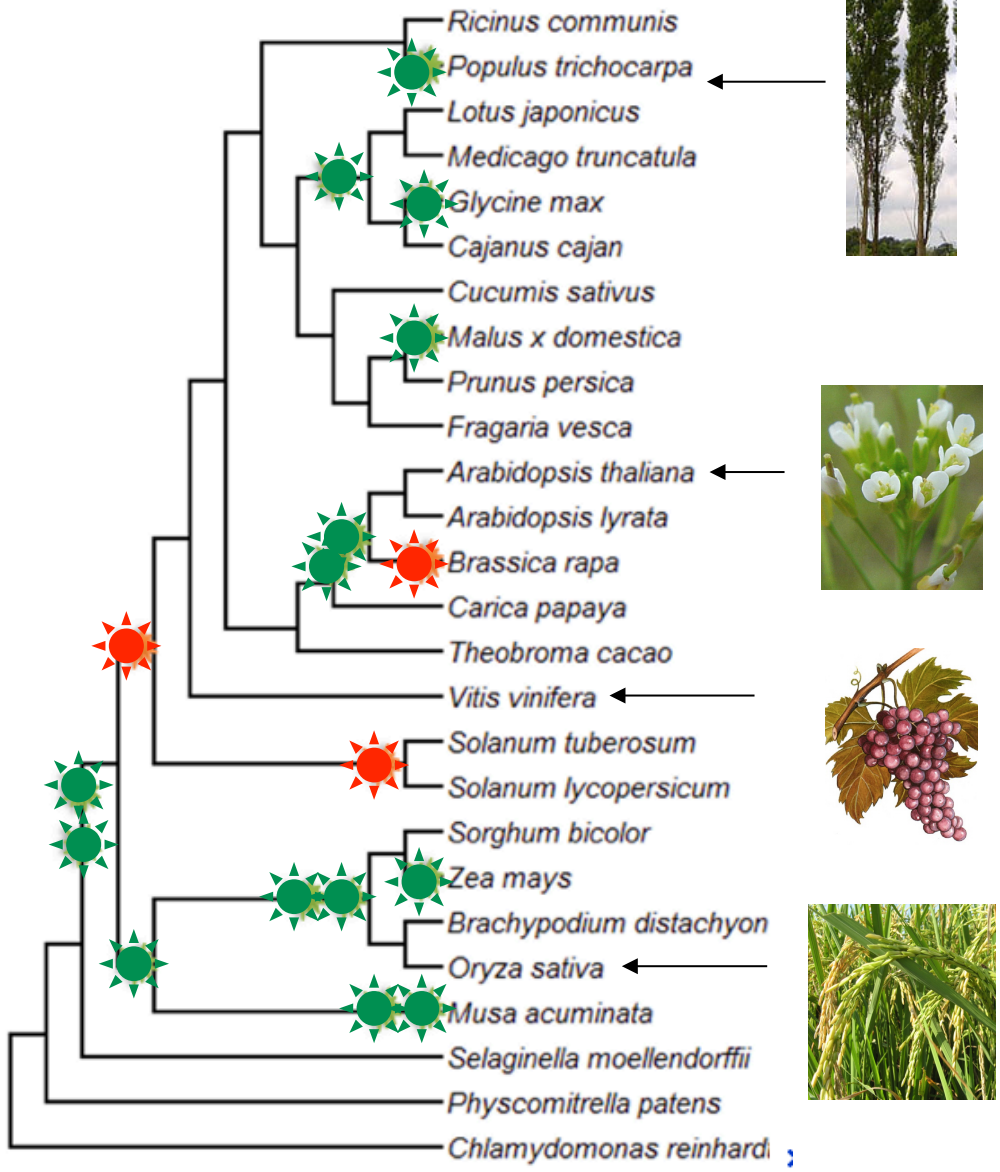


Multiple genome duplications and triplication in *Streptophyta*

The grapevine genome sequence suggests ances hexaploidization in major angiosperm phyla

The French–Italian Public Consortium for Grapevine Genome Characterization*





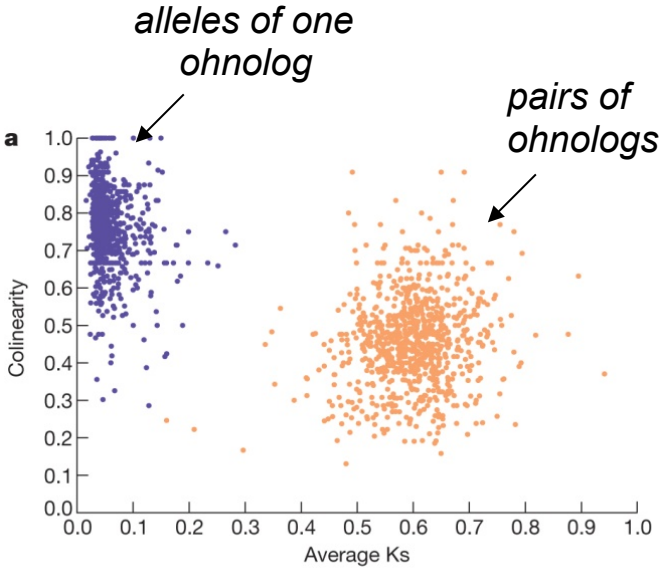
Atypical genomes in metazoa

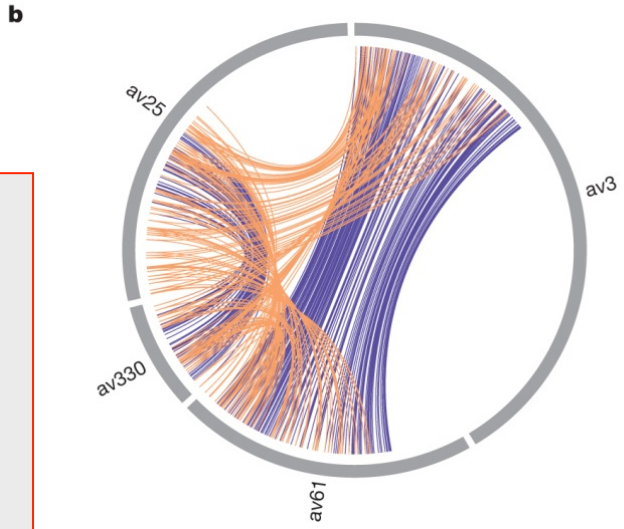
Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*

Nature (2013) ***503****: 453-457*

Jean-François Flot[1,2,3,4,5,6], Boris Hespeels[1,2], Xiang Li[1,2], Benjamin Noel[3], Irina Arkhipova[7], Etienne G. J. Danchin[8,9,10], Andreas Hejnol[11], Bernard Henrissat[12], Romain Koszul[13], Jean-Marc Aury[3], Valérie Barbe[3], Roxane-Marie Barthélémy[14], Jens Bast[15], Georgii A. Bazykin[16,17], Olivier Chabrol[14], Arnaud Couloux[3], Martine Da Rocha[8,9,10], Corinne Da Silva[3], Eugene Gladyshev[7], Philippe Gouret[14], Oskar Hallatschek[6,18], Bette Hecox-Lea[7,19], Karine Labadie[3], Benjamin Lejeune[1,2], Oliver Piskurek[20], Julie Poulain[3], Fernando Rodriguez[7], Joseph F. Ryan[11], Olga A. Vakhrusheva[16,17], Eric Wajnberg[8,9,10], Bénédicte Wirth[14], Irina Yushenova[7], Manolis Kellis[21], Alexey S. Kondrashov[16,22], David B. Mark Welch[7], Pierre Pontarotti[14], Jean Weissenbach[3,4,5], Patrick Wincker[3,4,5], Olivier Jaillon[3,4,5,21]* & Karine Van Doninck[1,2]*

Total genome size 244 Mb
(including 26 Mb homozygous 2x)
49,300 protein-coding genes





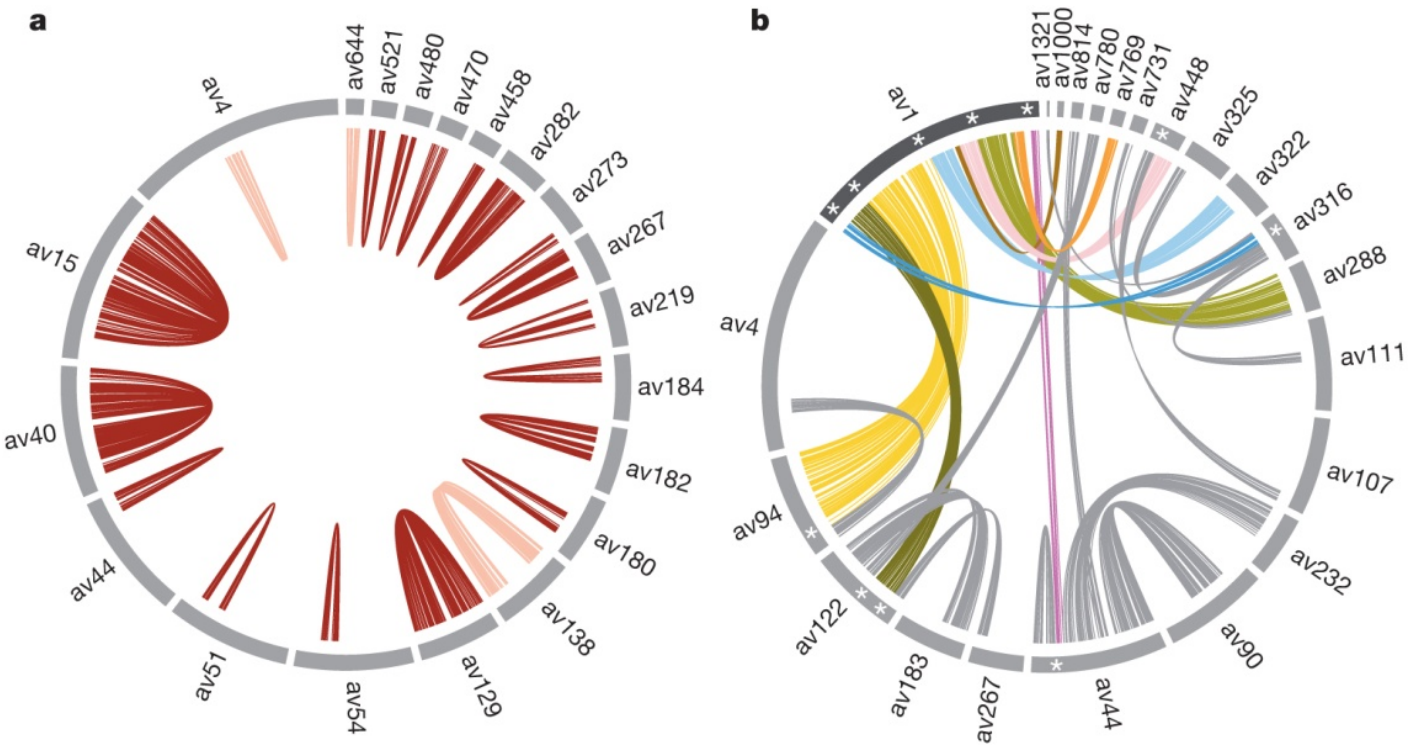
Example of a genomic quartet of 4 scaffolds

Numerous homologous blocks (collinear regions) forming two groups:
- pairs of ***ohnologs*** *(mean 74 % identity) corresponding to an ancient genome duplicaton*
-pairs of ***alleles*** *(mean 96 % identity) for each member of the pair of ohnologs.*

Their coexistence in the same genome forms ***quartets*** *that, altogether, cover*
40 % of the entire genome (---> a ***locally tetraploid*** *genome)*

Atypical genomes in metazoa

The genomic structure is incompatible with conventional meiosis

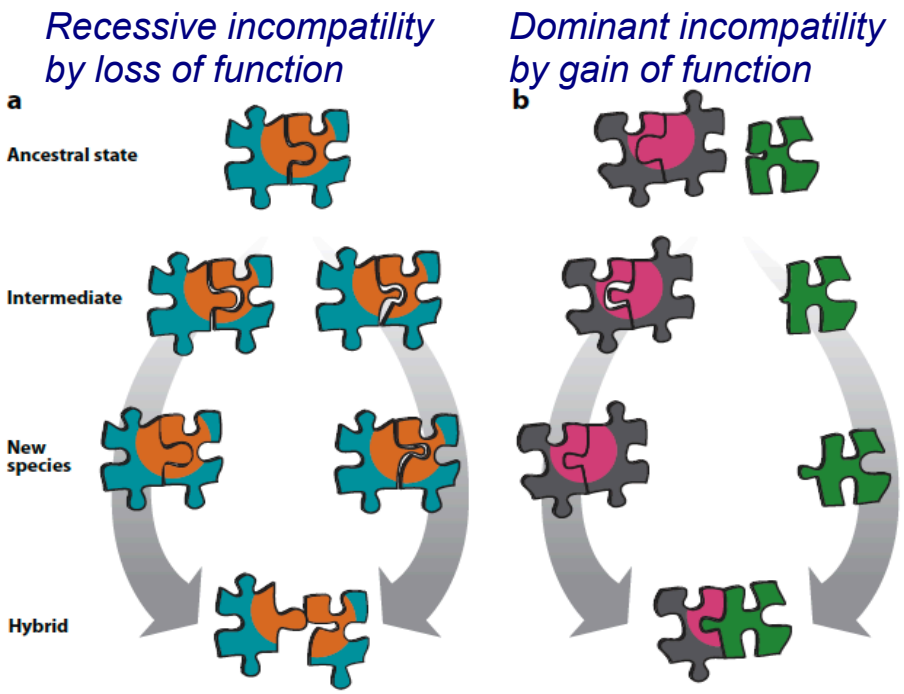


*Allelic pairs are found on the **same** chromosomes*

*Colinear regions do not extend to **entire** chromosome scale*

*Multiple traces of **gene conversion** between gene copies reassort alleles without meiosis*
*Multiple traces of **horizontal gene acquisition** of non-metazoan origin throughout the genome (8%)*

Bateson-Dobzhansky-Müller incompatibility



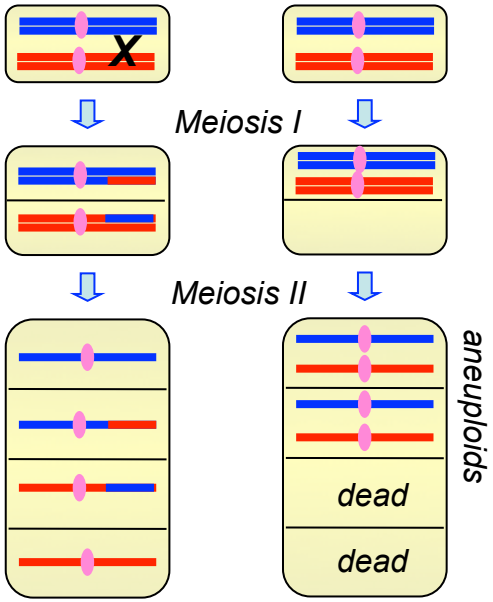
However: very few examples of DM gene pairs identified

Drosophila:	*Lbr - Hmr: heterochromatin turnover*
	Zbr: pericentric satellite DNA
	OdsH - Ovd: heterochromatin-mediated sex ratio distortion
	Nup96 - Nup 160: nucleopore proteins
Xiphophorus:	*Xmrk2 deregulated expression*
Mus:	*Prdm9 meiotic recombination hotspot*
Oryza:	*S5: proteolytic enzyme in cell wall*
	SaM -SaF: SUMO E3 ligase

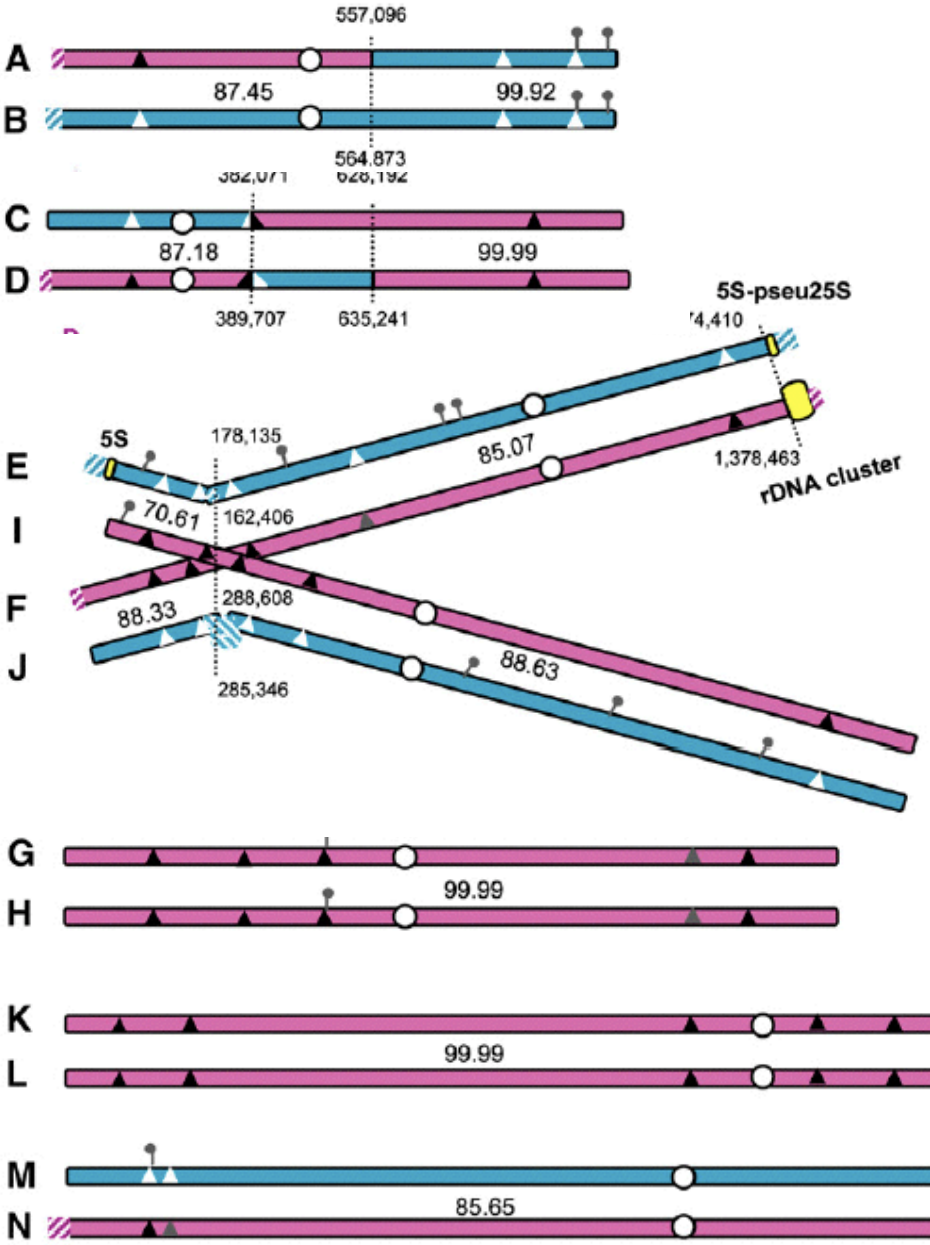
Saccharomyces: nucleo-mitochondrial interactions
AEP2 (nuc) - Oli1 (mt): lack of Oli1 translation
MRS1 (nuc) - Cox1 (mt): lack of Cox1 intron splicing

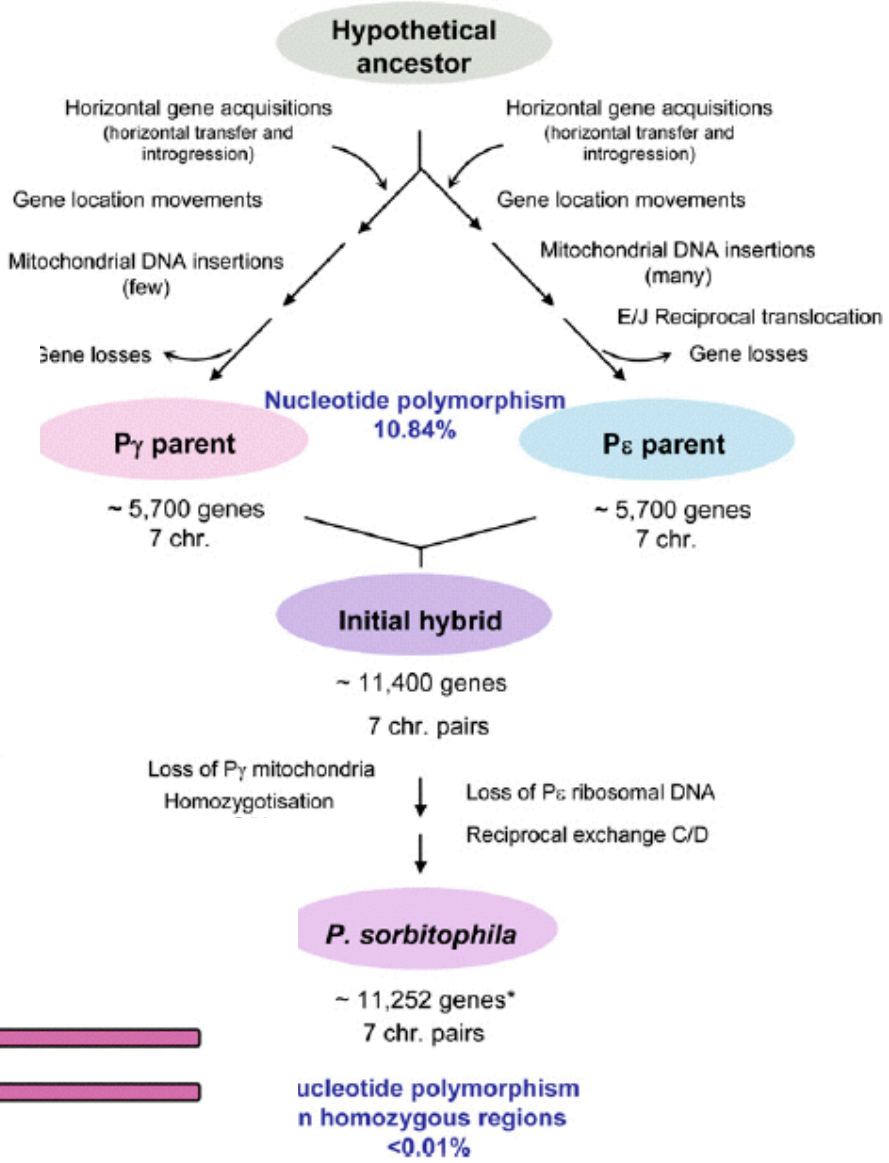
Hybrid meiotic sterility

Sequence divergence --> mismatch repair system --> no COV --> aneuploidy



Pichia sorbitophila: a recent hybrid in the process of resolution



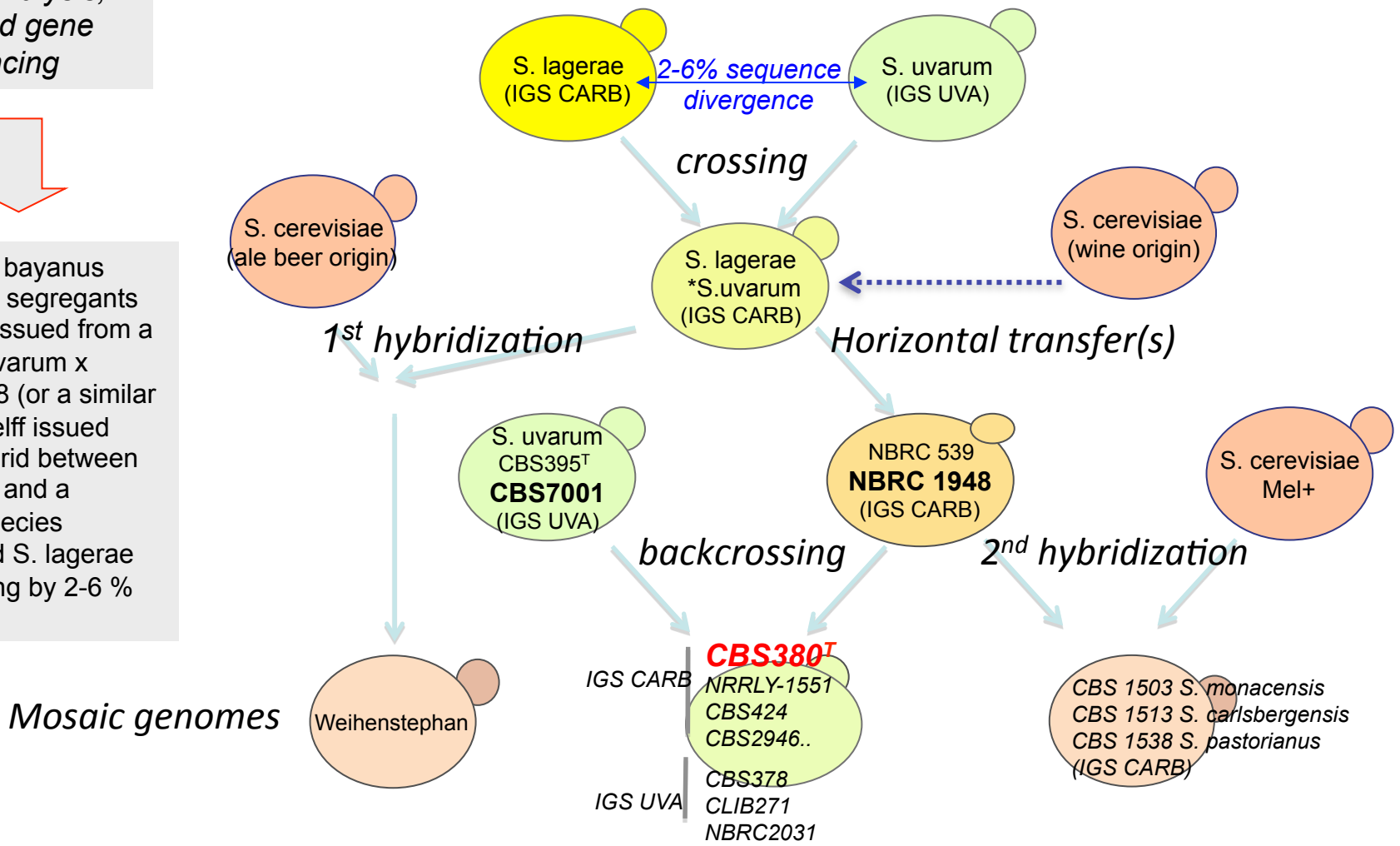


Even type-strains may have a complex hybrid ancestry

*Nguyen et al., 2011. PLoS ONE **6**: e25821*

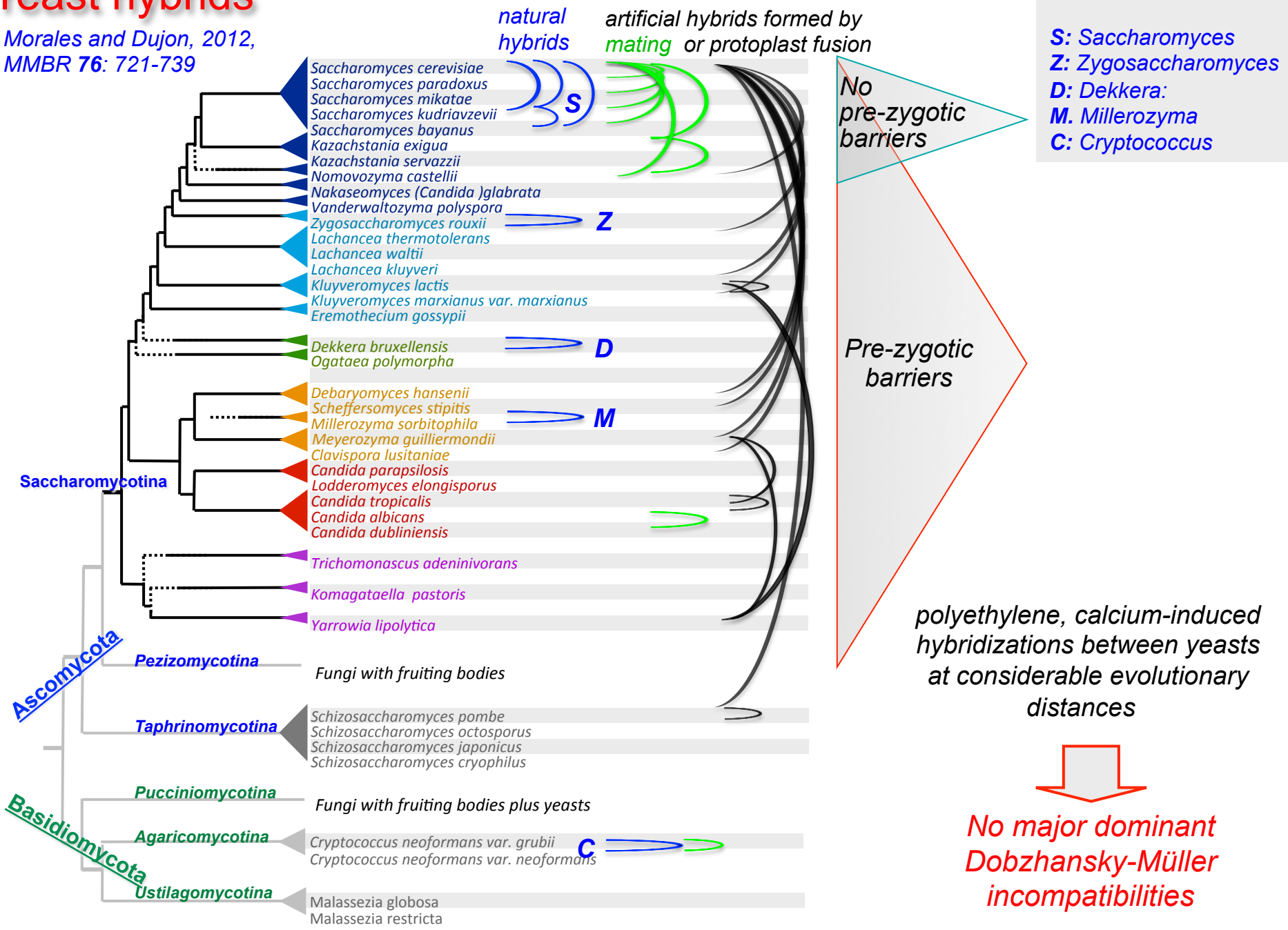
AFLP analysis, selected gene sequencing

Current S. bayanus strains are segregants of tetrads issued from a cross S. uvarum x NBRC1948 (or a similar strain), itselff issued from a hybrid between S. uvarum and a missing species designated S. lagerae and differing by 2-6 % SNPs.



Yeast hybrids

*Morales and Dujon, 2012, MMBR **76**: 721-739*



Natural hybrids

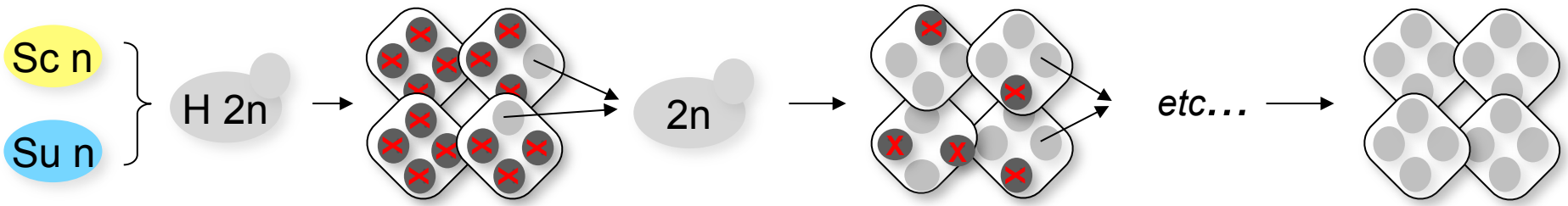
***S:** Saccharomyces*
***Z:** Zygosaccharomyces*
***D:** Dekkera:*
***M.** Millerozyma*
***C:** Cryptococcus*

polyethylene, calcium-induced hybridizations between yeasts at considerable evolutionary distances

No major dominant Dobzhansky-Müller incompatibilities

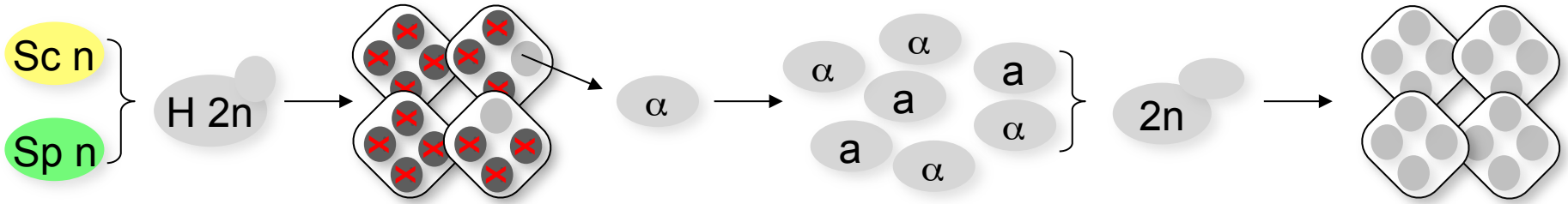
Spontaneous processes rapidly restore fertlity

1- Gradual recovery of fertility during successive meiotic cycles



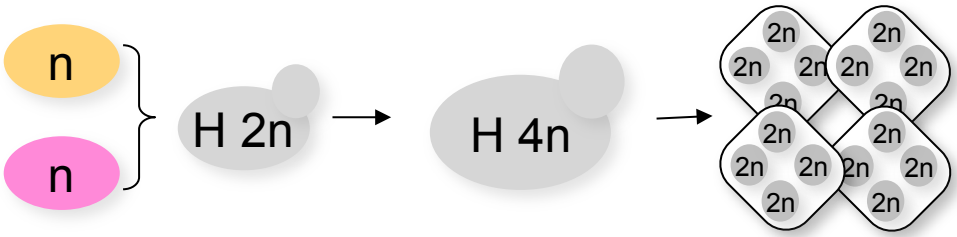
Antunovics et al., (2005) FEMS Yeast Res. ***5****: 1141-1150*

2- Recovery of fertility after mating-type switching



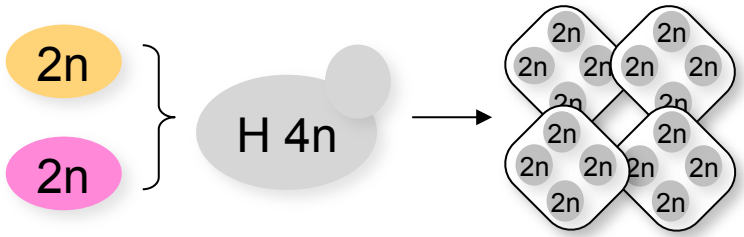
Greig et al (2002) Science **298**: 1773-1775

3- Recovery of fertility by endoreplication



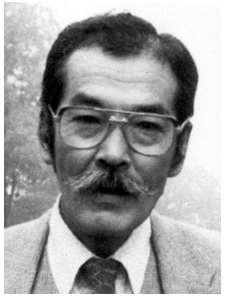
Sebastini et al (2002) Res. Mic. **153**: 53-58

4- Fertility of allotetraploids

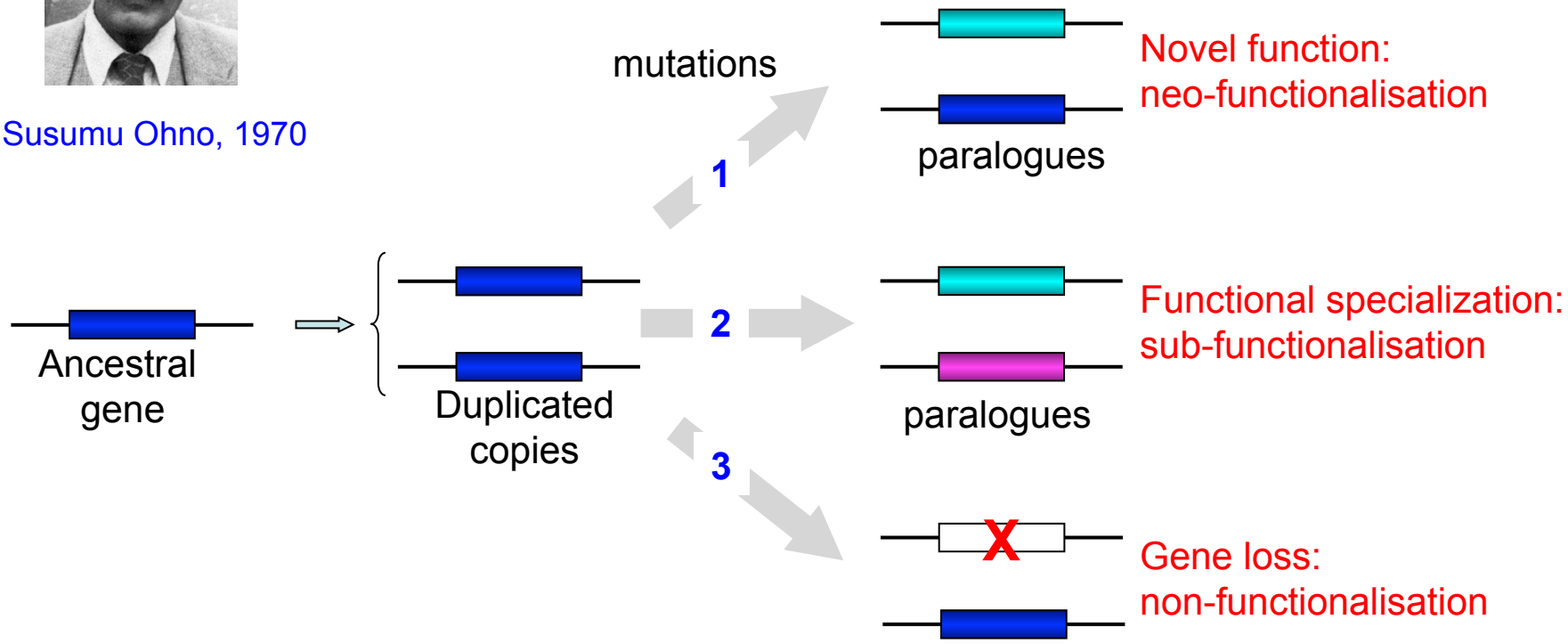


Greig et al (2002) Proc. Roy. Soc. **269**: 1167-1171

Consequences of gene duplications



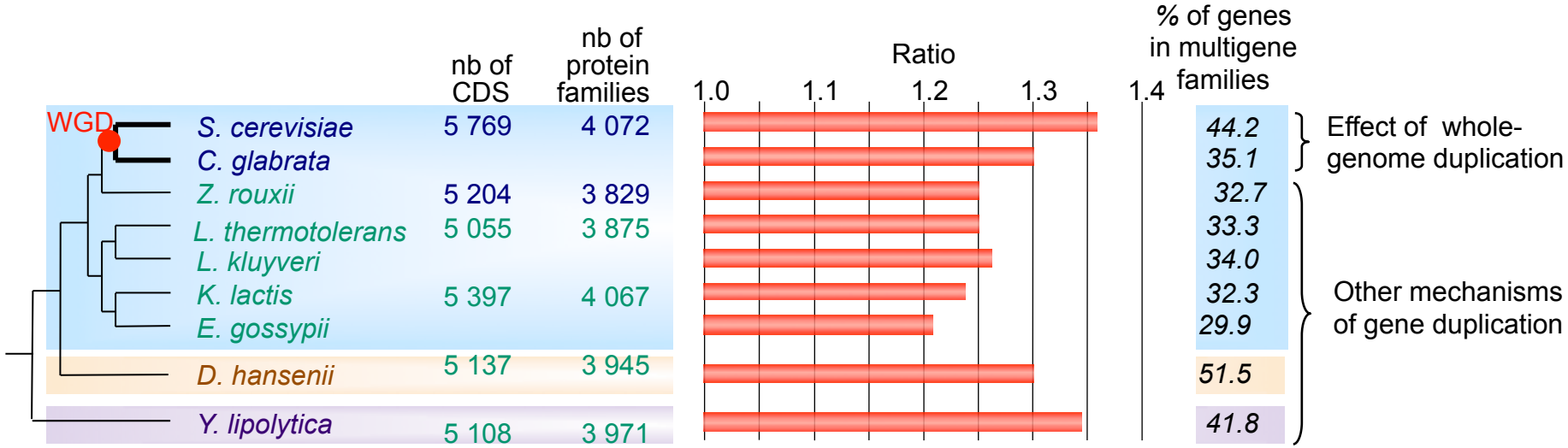
Susumu Ohno, 1970



All genomes show high number of paralogous gene copies resulting from the combination of whole-genome duplications, segmental duplications, tandem gene duplications

A genome is only a snapshot in time of continuous gene duplications and losses.

Duplications and genome redundancy



Single gene duplications:

in tandem: few arrays in most yeast genomes, more arrays in some species

ectopic: no direct indication that this mechanism exists

through RNA intermediate:
phenomemon is experimentally demonstrated, but extant retrogenes are very difficult to identify in yeast genomes

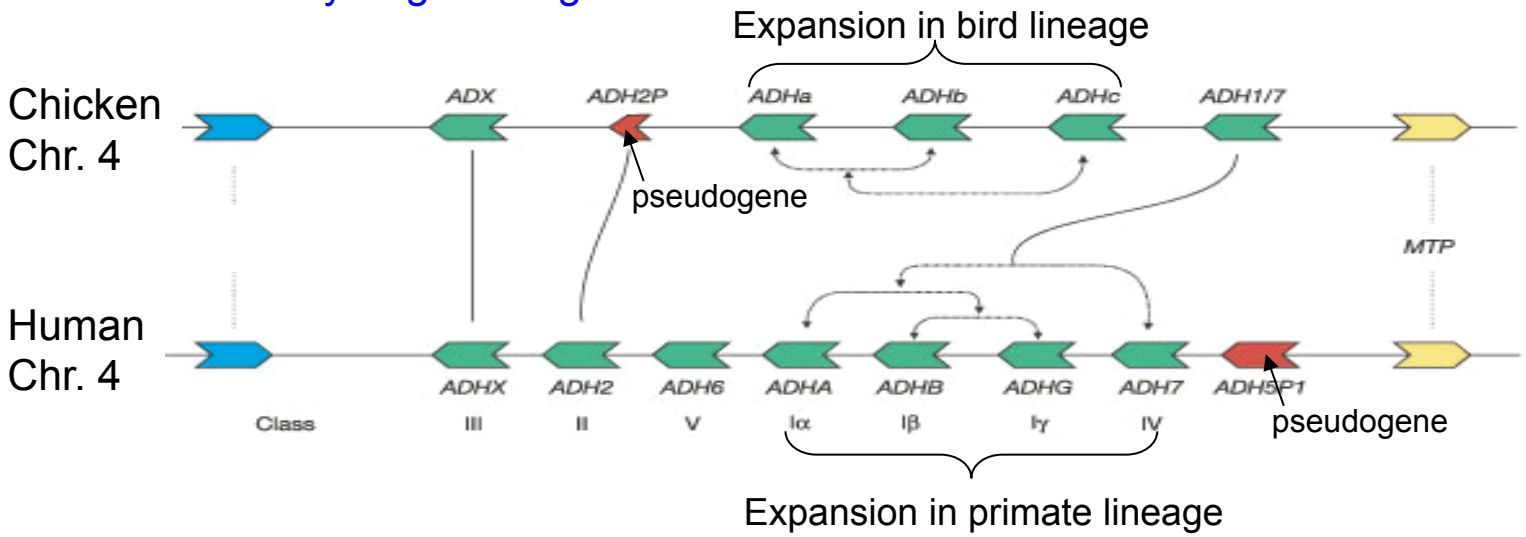
many dispersed paralogs owing to very ancient evolutionary lineages

Segmental duplications:

phenomemon is experimentally demonstrated, but examples are very rare in yeast genomes, except in subtelomeric regions

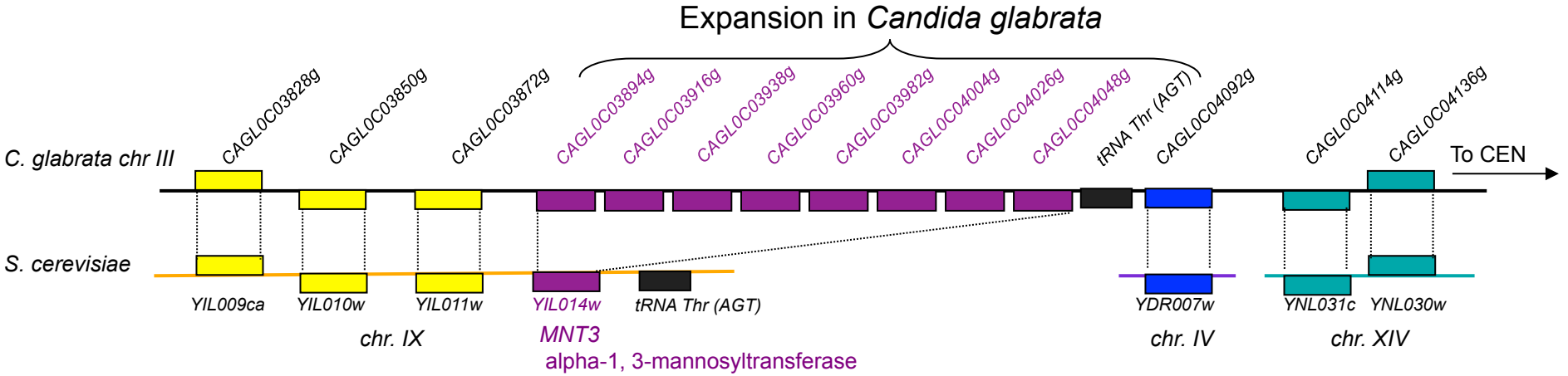
Formation and evolution of tandem gene arrays

Example 1: alcohol dehydrogenase gene cluster in vertebrates



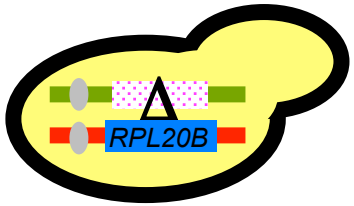
Example 2: alpha 1-3 mannosyl transferase in yeasts

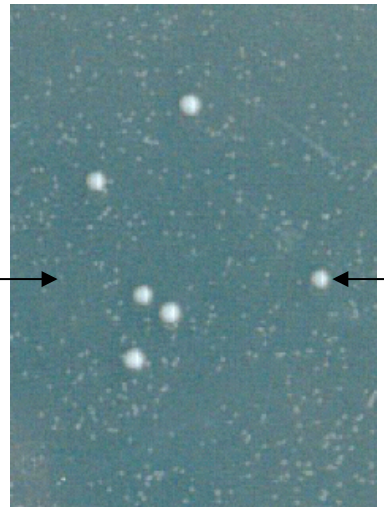
Elevated sequence divergence between gene copies suggests **subfunctionalization** and **adaptive evolution**

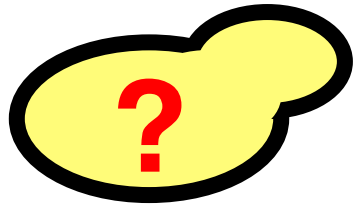


Segmental duplications in *Saccharomyces cerevisiae*

Koszul *et al*. 2004 *EMBO J.* **23**: 234-243

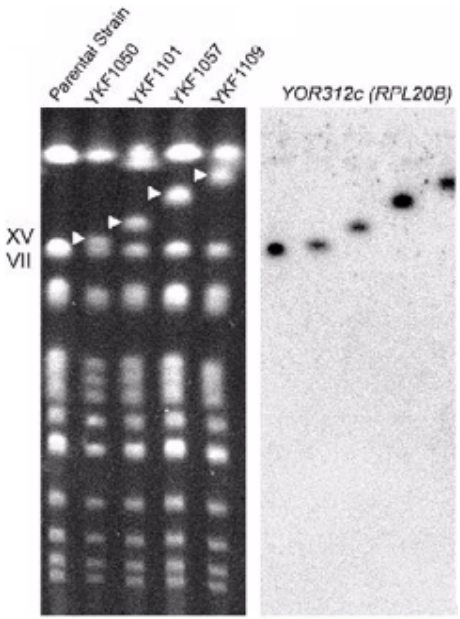




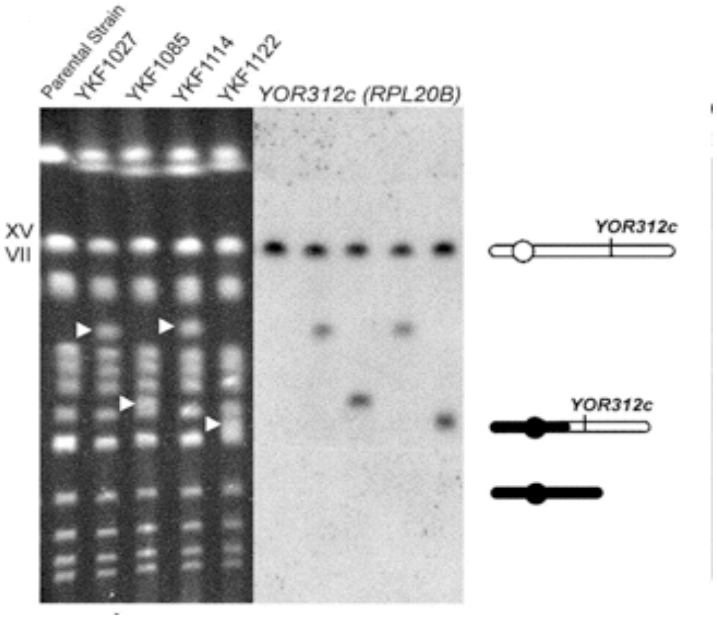


micro-colonies

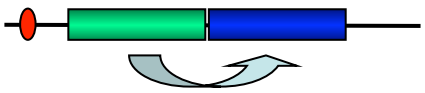
normal size colonies



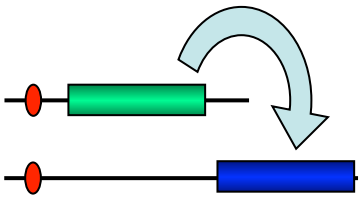
Pulsed field gel electrophoresis



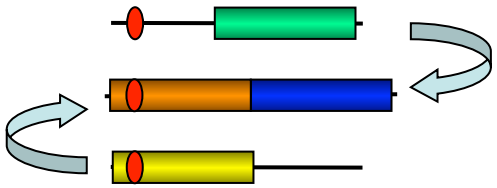
Pulsed field gel electrophoresis



Intrachromosomal duplications



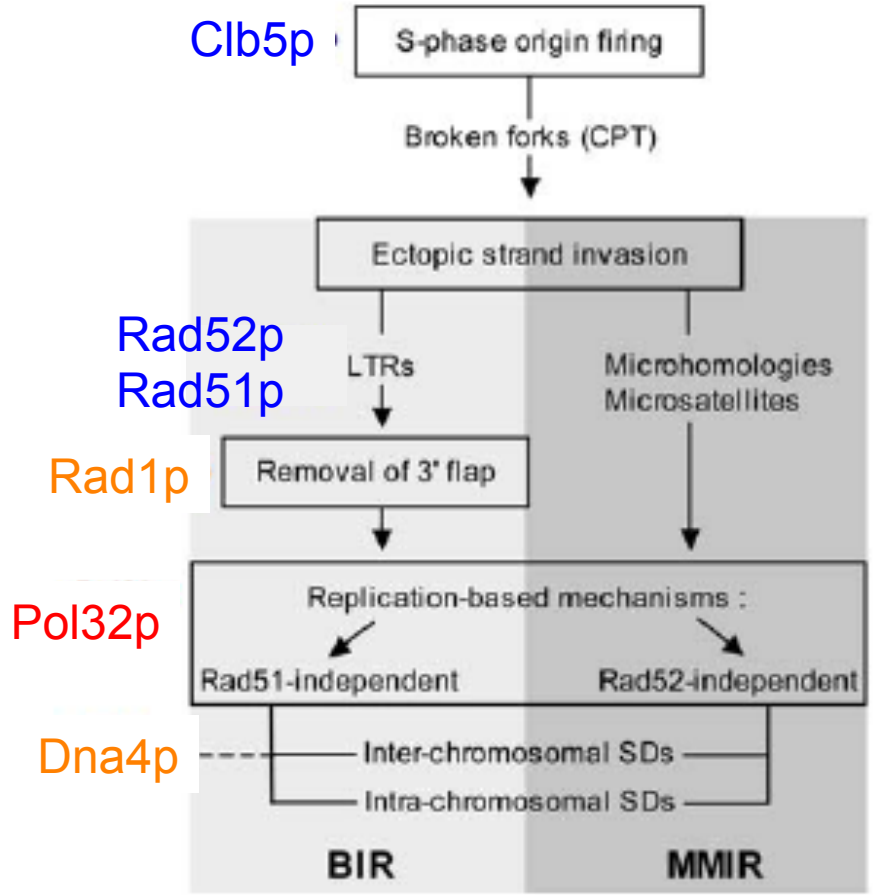
Interchromosomal duplications

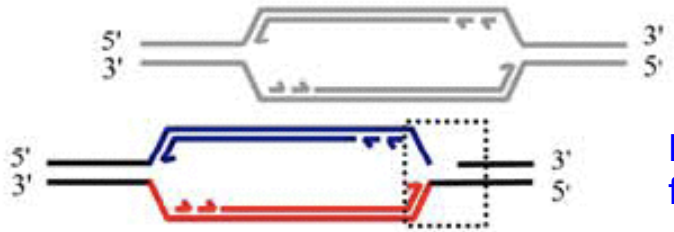


Neo-chromosome

Molecular mechanism of formation of segmental duplications

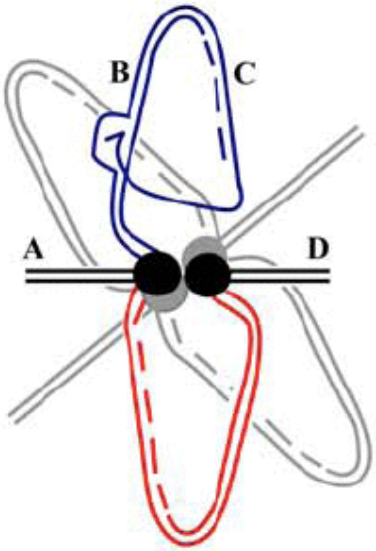
Payen *et al.* 2008 *PLoS Genetics* **4**: e1000175





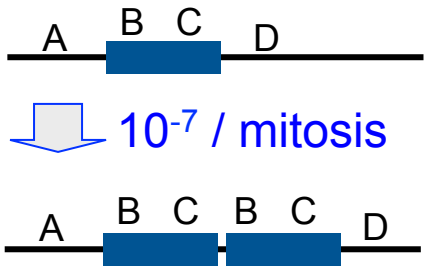
Broken fork

Abnormal reinitiation



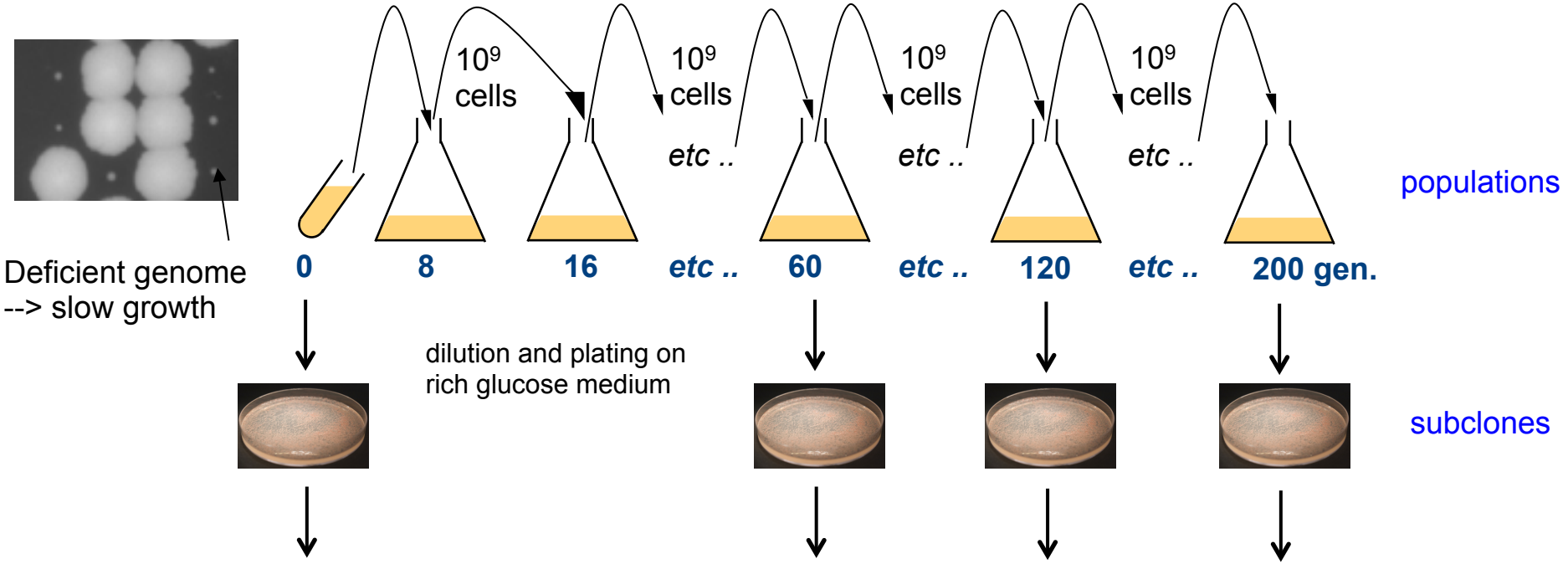
Effect of gene mutation on the formation of segmental duplications.

Red: abolished > essential protein
Orange: reduced > protein involved
Bleu: invreased > control protein



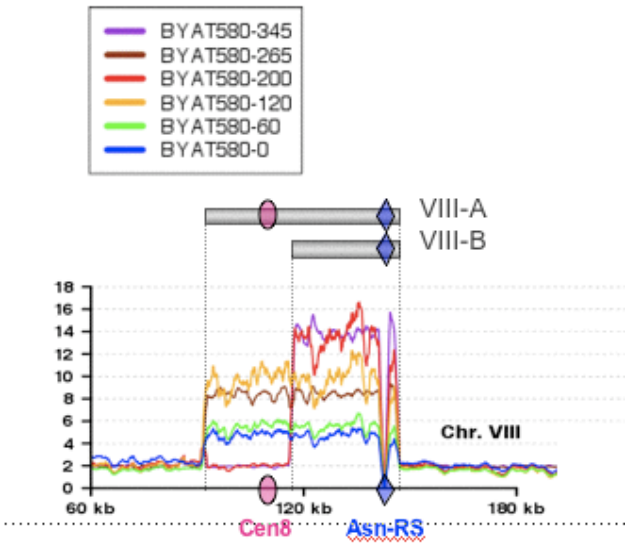
Experimental evolution by serial transfer of populations

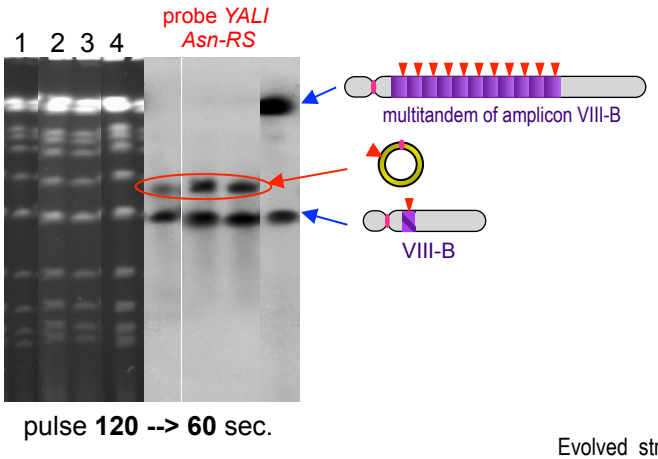
- two-liter batch cultures in rich glucose medium at 30°C with aeration
- serial transfers with population bottlenecks of 10^9 cells

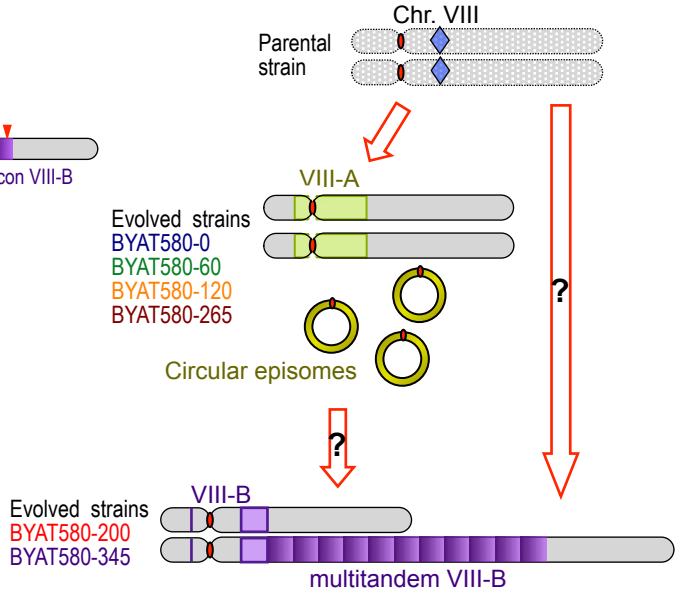


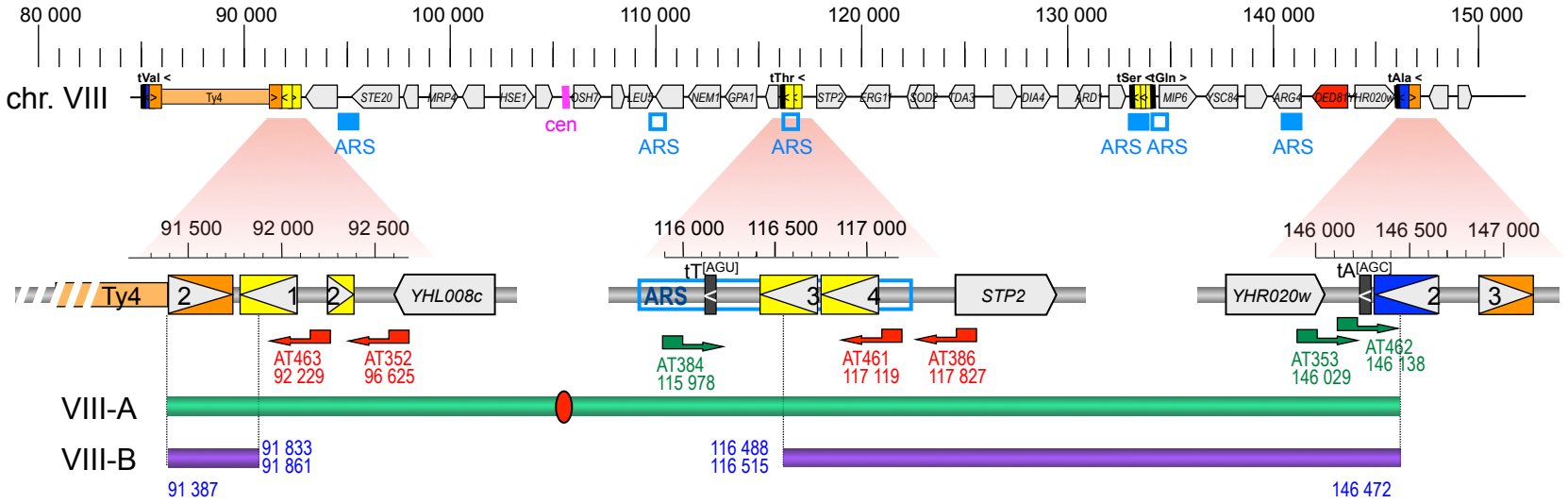
Fitness measurement (growth rate)

Whole genome deep sequencing (Solexa HiSeq2000)


BYAT580-345
BYAT580-265
BYAT580-200
BYAT580-120
BYAT580-60
BYAT580-0
VIII-A
VIII-B
18
16
14
12
10
8
6
4
2
0
Chr. VIII
60 kb
120 kb
180 kb
Cen8
Asn-RS

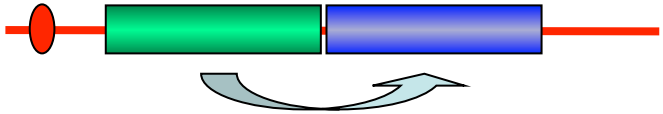

1 2 3 4
probe YALI Asn-RS
multitandem of amplicon VIII-B
VIII-B
pulse 120 --> 60 sec.


Chr. VIII
Parental strain
VIII-A
Evolved strains
BYAT580-0
BYAT580-60
BYAT580-120
BYAT580-265
Circular episomes
?
?
VIII-B
Evolved strains
BYAT580-200
BYAT580-345
multitandem VIII-B

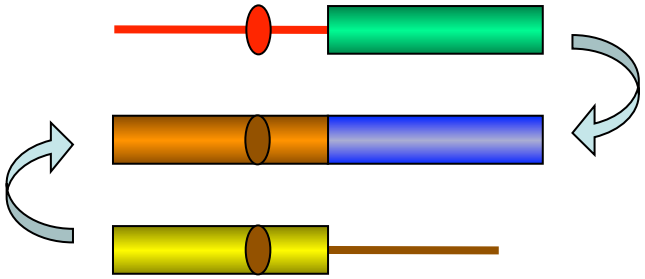

80 000
90 000
100 000
110 000
120 000
130 000
140 000
150 000
chr. VIII
tVal
Ty4
STE20
MRP4
HSE1
cen
DSH7
LEU3
NEM1
GPA1
tThr
STP2
ERG1
SOD2
DIA3
DIA4
ARD1
tSer
tGln
MIP6
YSC84
ARG4
DED81
HR020w
tAla
ARS
ARS
ARS
ARS ARS
ARS
91 500
92 000
92 500
116 000
116 500
117 000
146 000
146 500
147 000
Ty4
YHL008c
tT[AGU]
ARS
STP2
tA[AGC]
YHR020w
AT463
92 229
AT352
96 625
AT384
115 978
AT461
117 119
AT386
117 827
AT353
146 029
AT462
146 138
VIII-A
VIII-B
91 833
91 861
91 387
116 488
116 515
146 472

Types of chromosomal dynamics

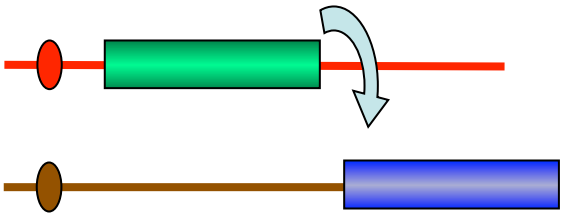
(a) Intrachromosomal



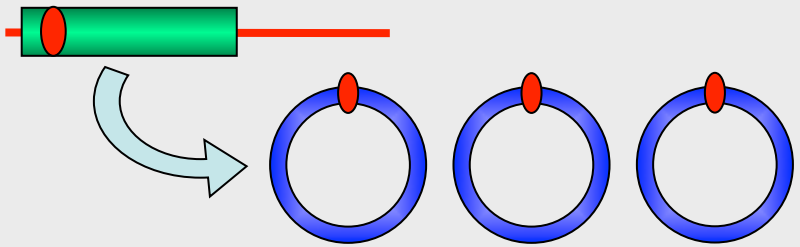
(d) Neo-chromosome



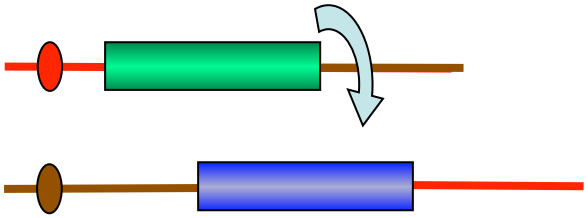
(b) Interchromosomal (subtelomeric addition)



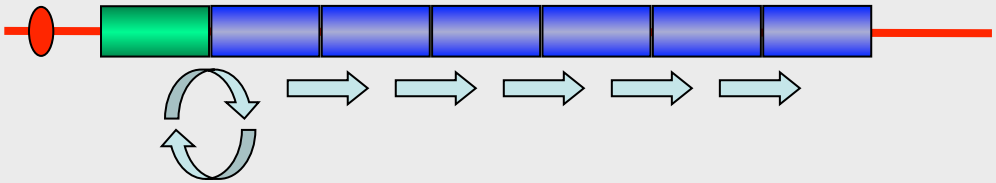
(e) Episome



(c) Interchromosomal (reciprocal translocation)

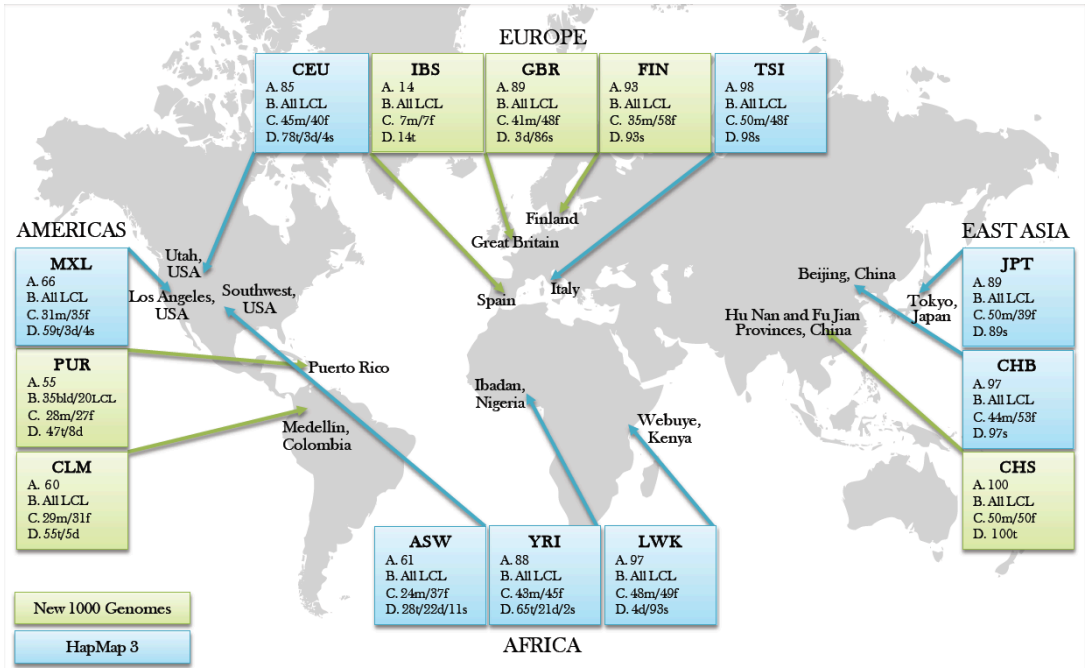


(f) Multitandems



Polymorphic variation of the human genome

The 1000 Genomes Project Project Consortium (2012) *Nature* **491**: 56-65



1 092 adults from 14 populations

Sequencing method

genome-wide low coverage
high coverage of coding exons
a few parents-child trios

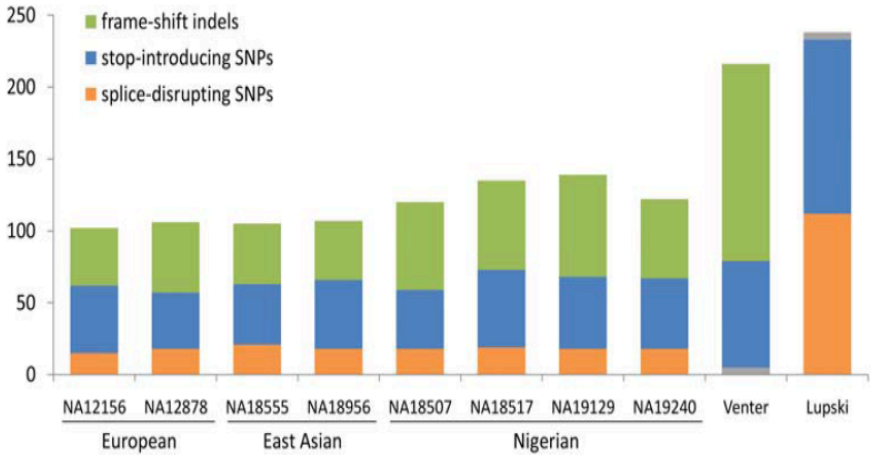
Results

38 millions of SNPs
1.4 millions of indels
14 000 large délétions

Compred to the reference genome, each individual genome carries:

3 millions SNPs (~ 40 % homozygote)
400 000 indels (~ 40 % homozygote)
Tens of deletions and duplications of average size 8 kb, totalling **~ 400 kb** (0,01 % of genome)

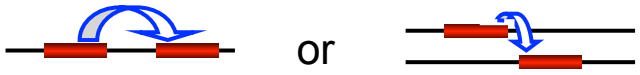
In total > **100 loss of fonction** per genome (including **30** homozygote)



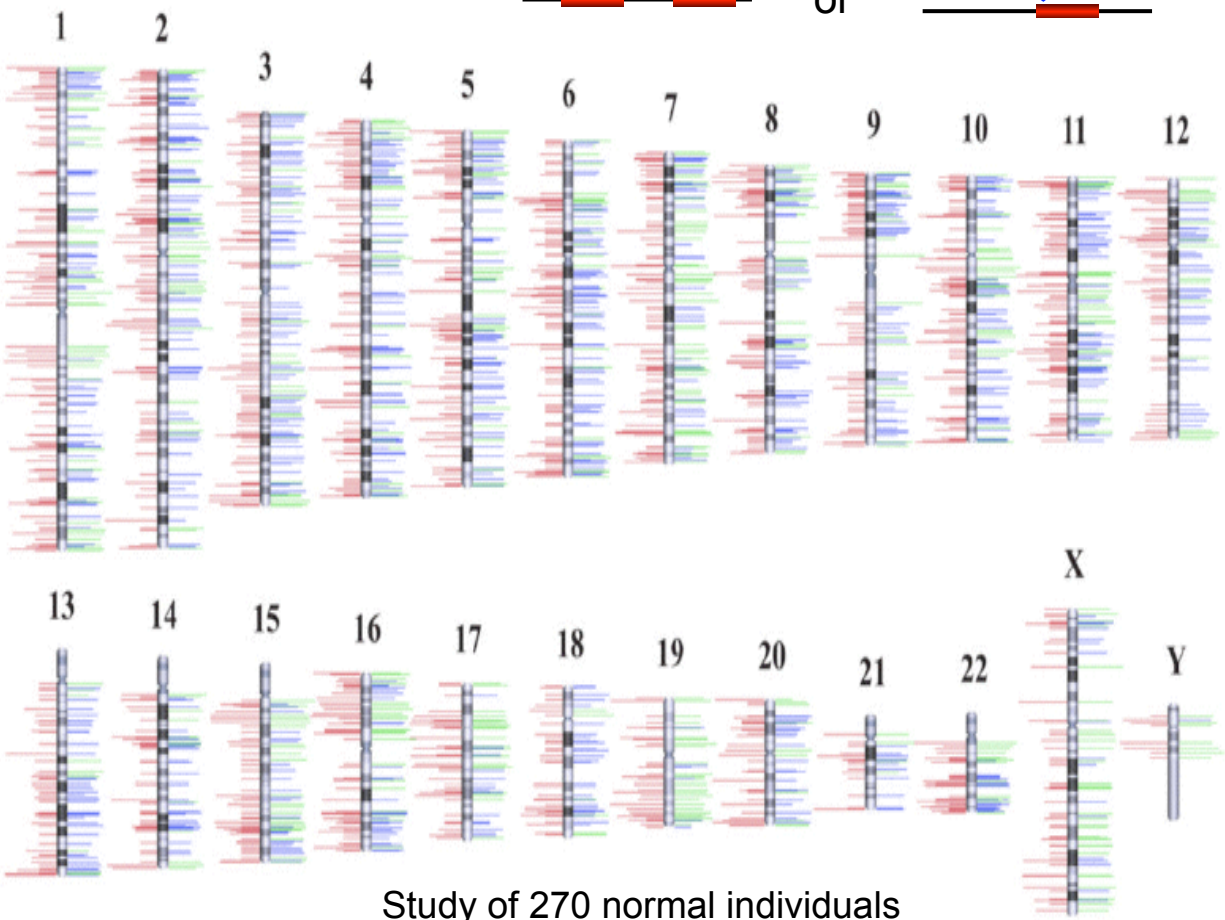
MacArthur and Tyler-Smith (2010) *Human Mol. Genetics* **19**: R125-130

Duplications and deletions in normal individuals

Common sites of segmental duplications



or



Number of cases: 1, 10, 100

Size (kb): <10, 100, 1000

Study of 270 normal individuals
(Europa, Africa, Asia)

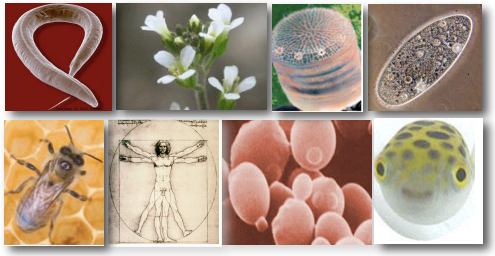
- **1 447** sites identified (**~ 12 %** of genome)
- each individual carry duplications covering **~ 5%** of genome

Gene birth **Gene loss**

From parent-child trios: *ca.* 550-750 kb missing at each generation
(Conrad *et al.*, 2006 *Nature Genetics* **38**, 75-81; McCarroll *et al.*, 2006 *Nature Genetics* **38**, 86- 92; Hinds *et al.*, 2006 *Nature Genetics* **38**, 82- 85)

One new gene created on average for every 21 newborn babies.

60 new genes appeared in the human genome since its separation from chimpanzee. Most new genes are highly expressed in cerebral cortex and testicles.
(Wu *et al.*, 2011, *PLoS Genetics* **7**: e1002379)



Le monde des eucaryotes, vu par la génomique comparative

Les mécanismes moléculaires de l'évolution des génomes eucaryotes

- La formation *de novo* de gènes, *frameshifts* et ARNs

Comment naissent les gènes ?

- Réassemblage d'éléments préexistants (au niveau ADN ou au niveau ARN)
- Création *de novo* de séquences codantes ou d'ARN non codants

Mécanismes disponibles au niveau de l'ADN

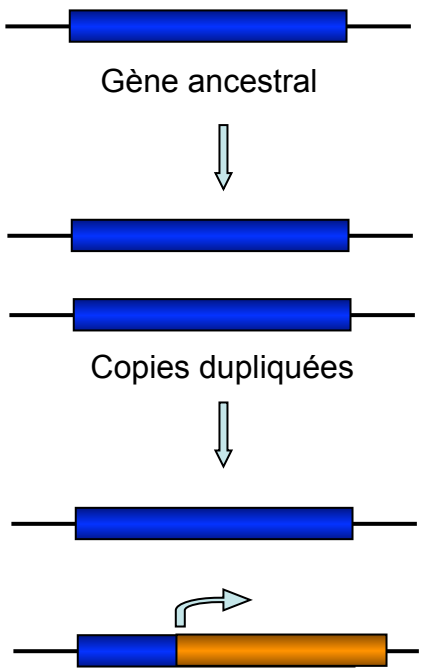
duplications - divergence de séquence (tandems, segments, génome entier)
fusions / fissions (au niveau des jonctions de segements chromosomiques)
frameshifts
création de nouveaux promoteurs et de nouvelles jonctions intron-exons (divergence de séquence)
transcription-traduction antisens
existence de protogènes

Mécanismes disponibles au niveau de l'ARN

exon shuffling
rétrogènes
exonisation de séquences non-codantes
capture de promoteurs existants et création d'exon 5' non codants

Formation de nouveaux gènes par *frameshifts*

Frameshifts



Chez la bactérie *Flavobacterium sp.* (portant un plasmide K172), une **nouvelle fonction enzymatique capable de dégrader le nylon** est apparue spontanément au début des années 1980.

Il s'agit d'une **protéine de 392 acides-aminés** montrant une **activité hydrolase** sur les oligomères linéaires d'acide 6-amino hexanoîque (nylon).

Après examen de la séquence, on trouve que cet enzyme est produit par la **traduction hors de phase** d'une séquence préexistante codant une protéine de 427 acides-aminés **riche en arginine**.

(Ohno 1984 *PNAS* **81**: 2421-2425).

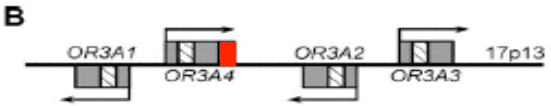
Formation de nouveaux gènes par *frameshifts* (génome humain)

Okamura *et al.*, (2006), *Genomics* **88**: 690-697

Examen systématique de nouvelles séquences capables de coder des protéines dans le génome humain par *frameshifts* de gènes ancestraux. Examen de 23 052 séquences de mRNA (RefSeq après curation).

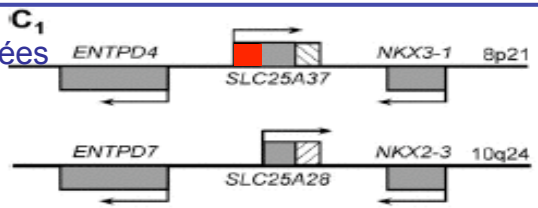
Tandems

Serotonin-receptor genes

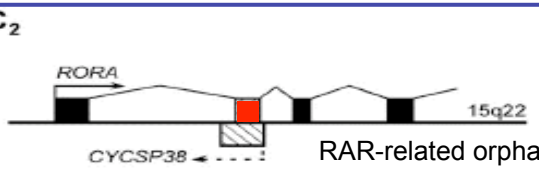


Olfactory-receptor genes

Duplications dispersées

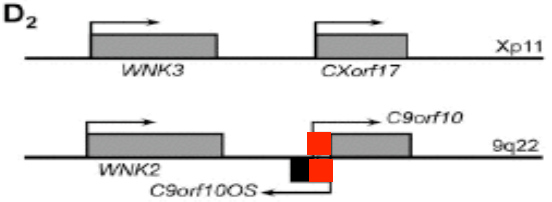


Mitochondrial transporter genes

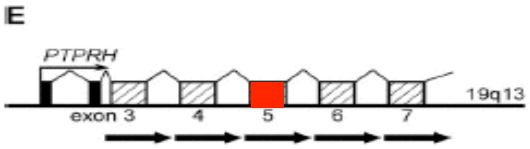


RAR-related orphan receptor gene with antisense piece of cytochrome c gene

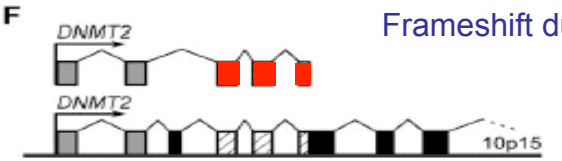
Gènes chevauchants



Frameshifts internes



repeats of a 267 bp motif



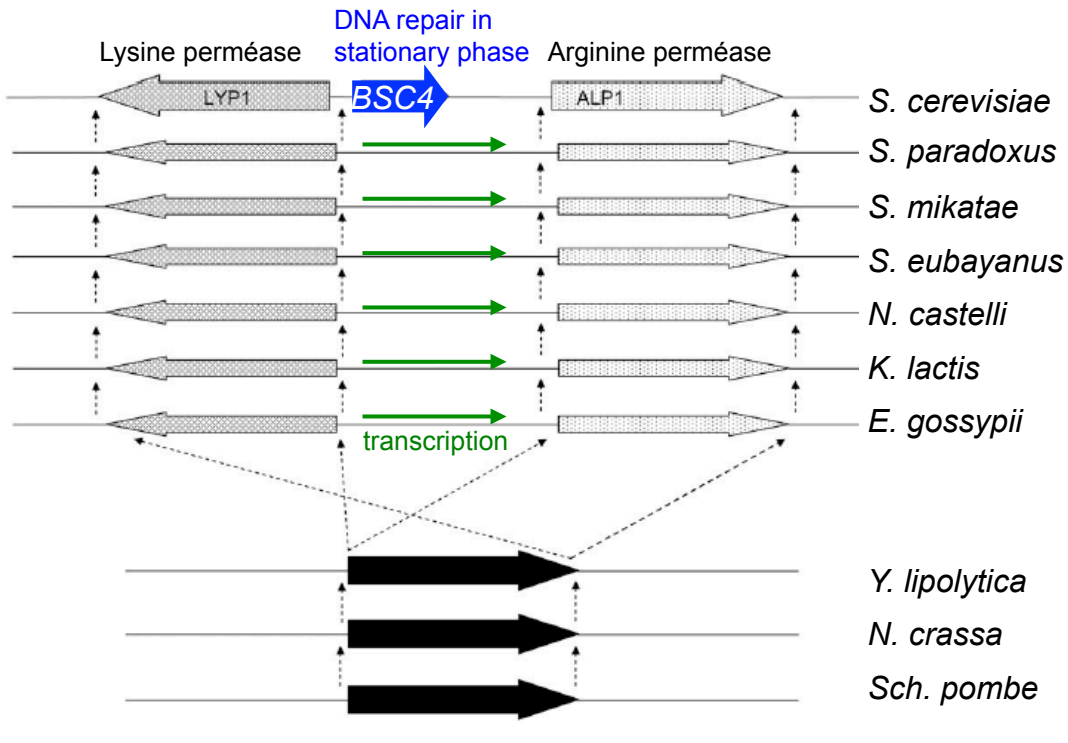
Frameshift du à épissage alternatif

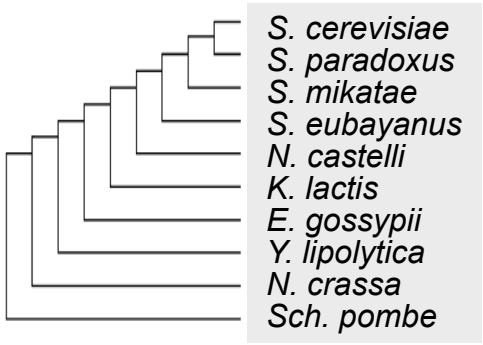
➢ 470 *frameshifts* (17 - 397 codons, moyenne 56 codons) parmi paires de paralogues examinées

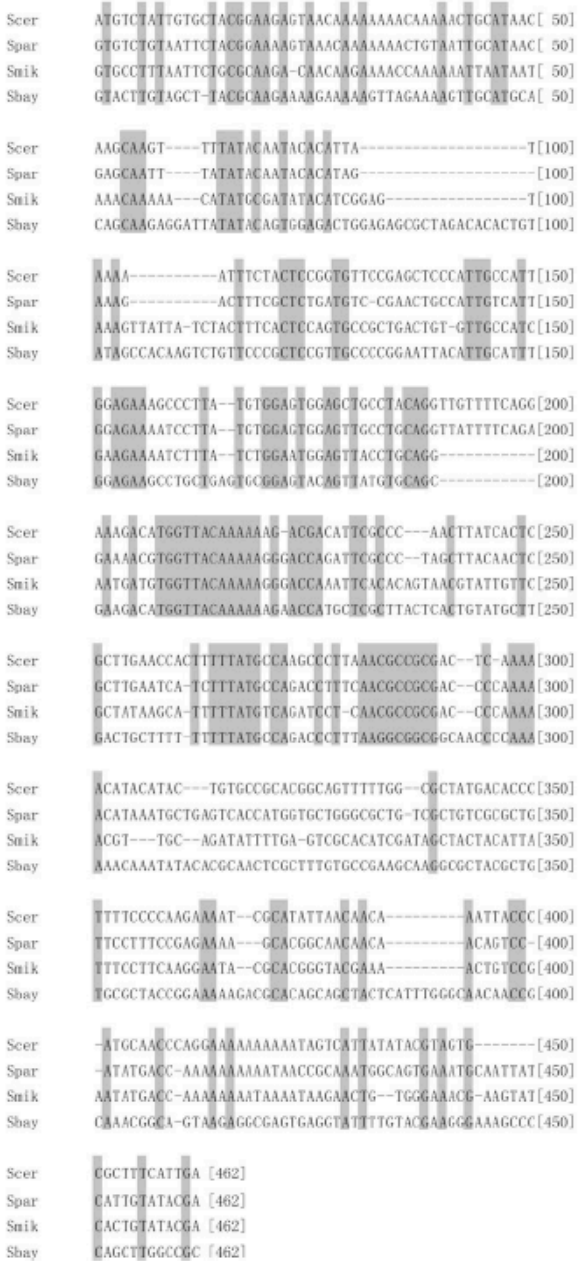
Autres données, revue de Raes and Van de Peer 2005 *Tends in Genetics* **21**: 428-431

Création *de novo* de gène dans une zone transcrite

Cai *et al.*, 2008 *Genetics* **179**: 487-496

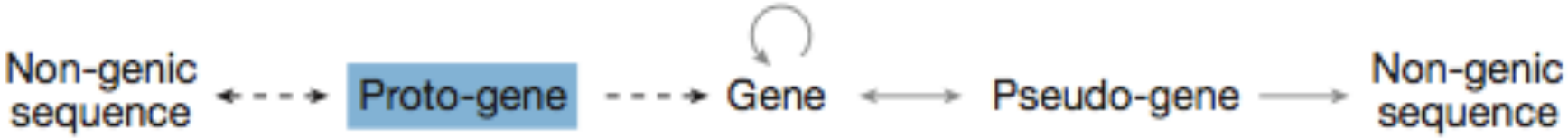


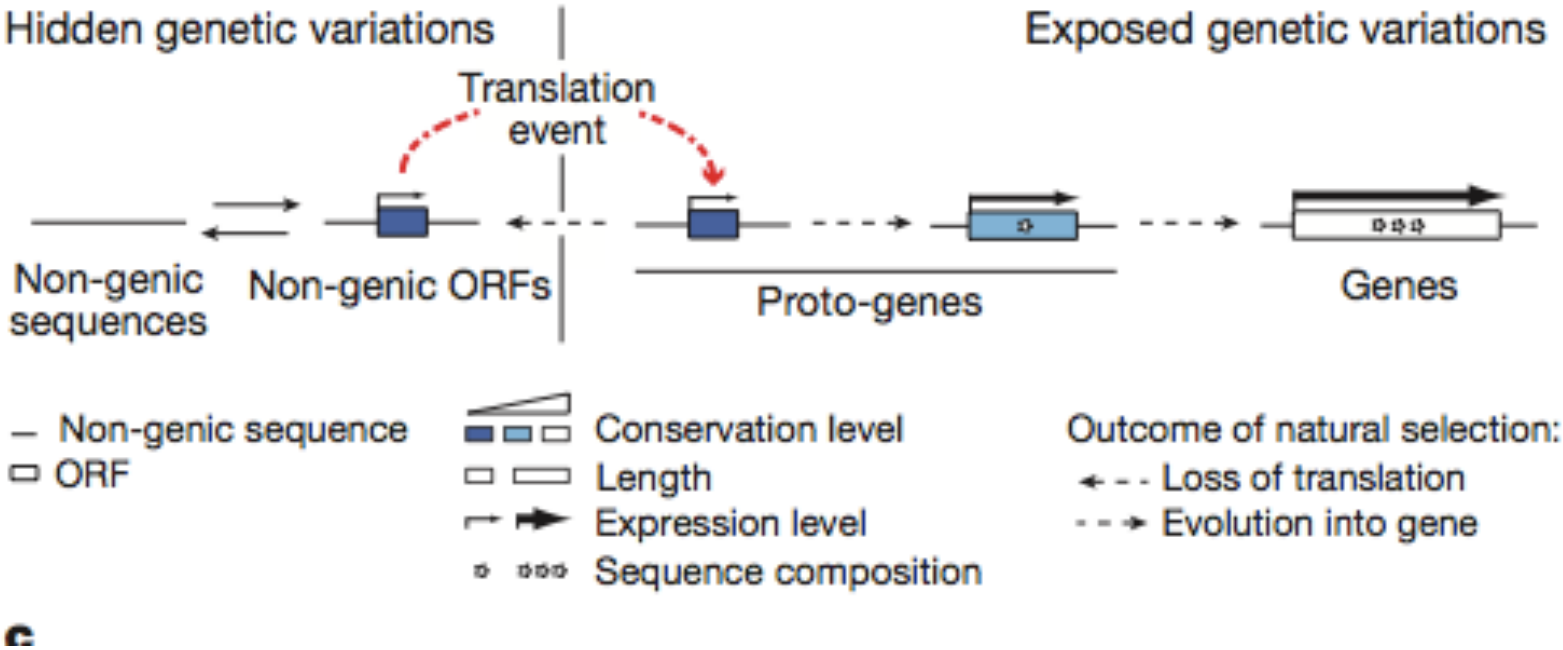




Notion de protogènes

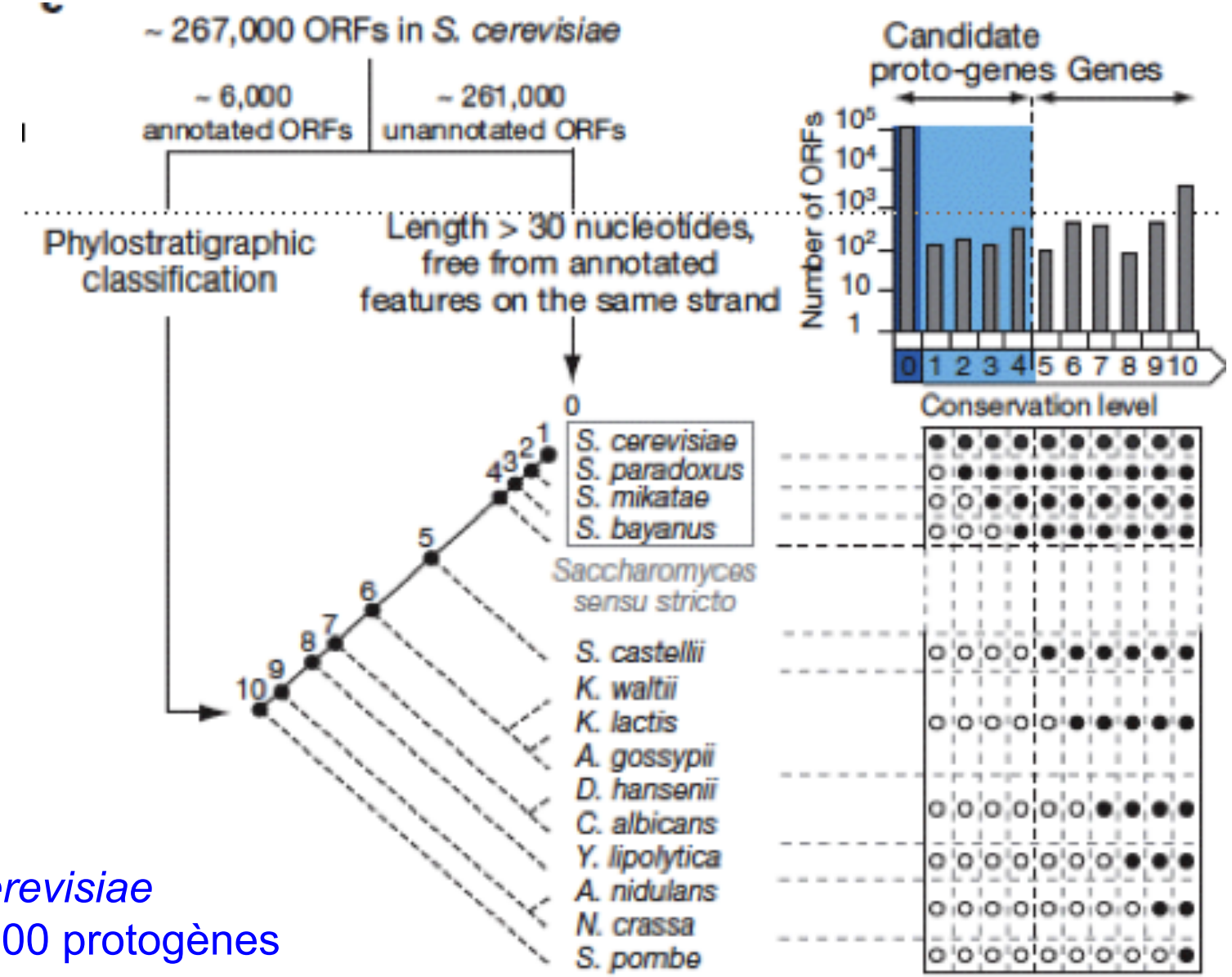
Carvunis *et al.* 2012 *Nature* **487**: 370-374





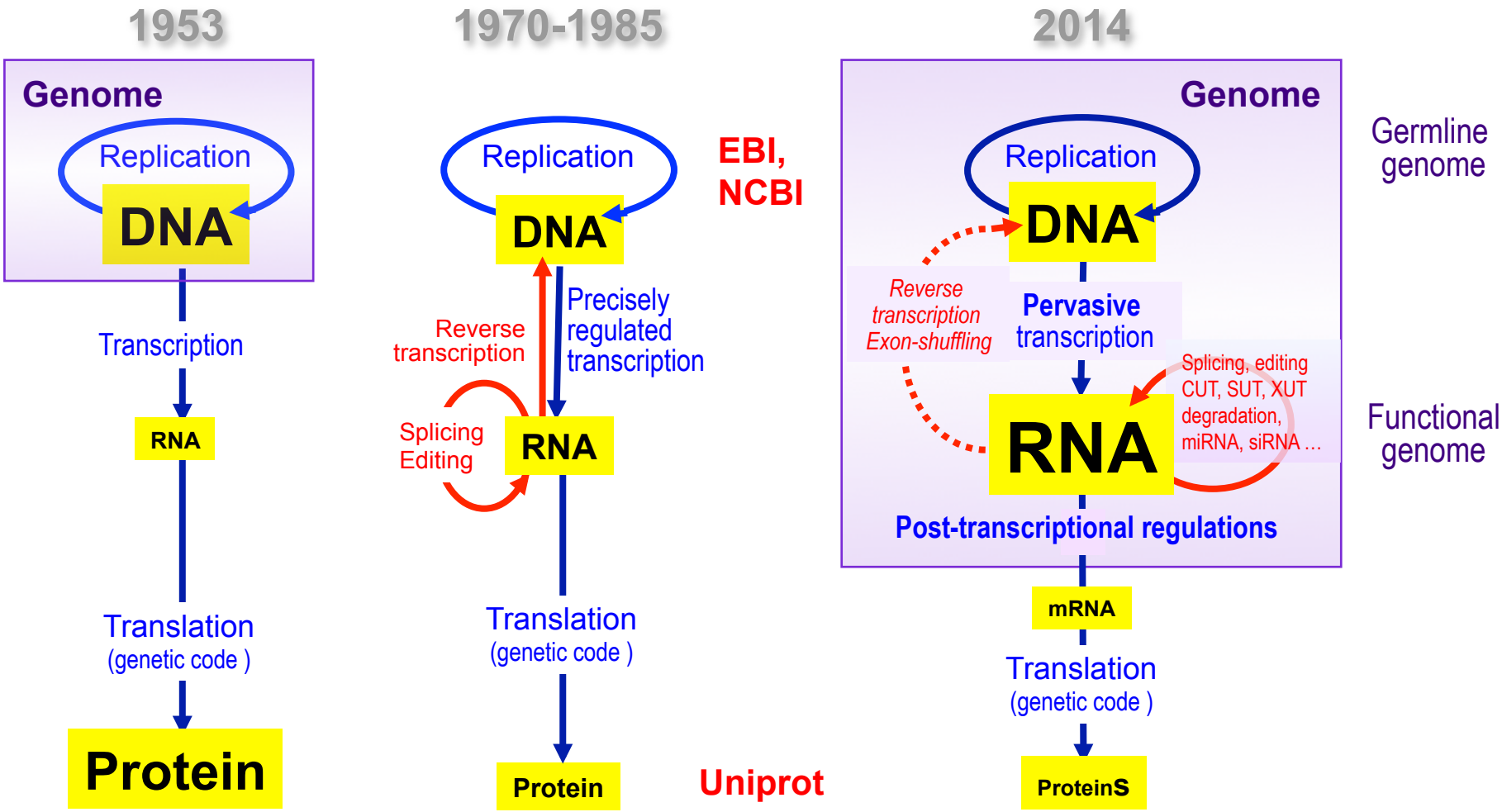
Notion de protogènes

Carvunis *et al.* 2012 *Nature* **487**: 370-374



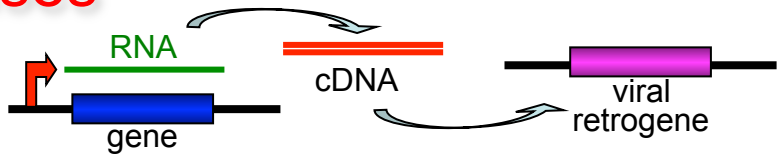
S. cerevisiae
~ 1 900 protogènes

The central dogma of molecular biology, revisited



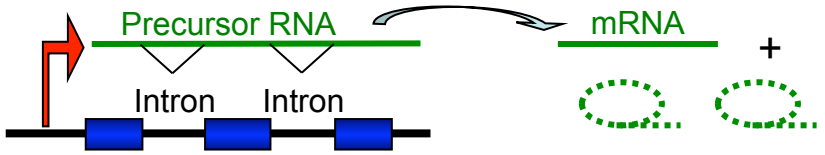
RNA surprises

1970 reverse transcription
Chicken virus



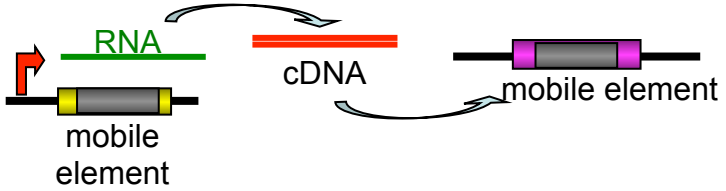
1977 introns
Mammalian virus, rabbit globin gene

1983 RNA catalysis
Tetrahymena nuclear intron, *E. coli* RNAse P

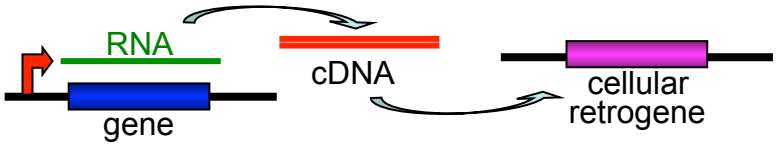


1983 RNA editing
Trypanosoma

1985 retrotransposons
Saccharomyces cerevisiae

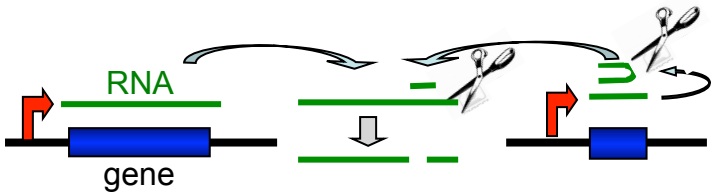


1993 cellular retrogenes
Drosophila melanogaster

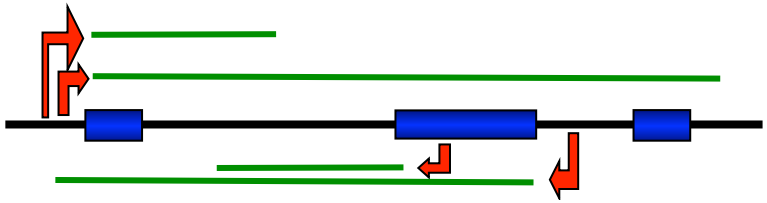


2000 RNA interference
(1980) petunias, fungi, *Caenorhabditis elegans*

2002 micro RNAs
Caenorhabditis elegans



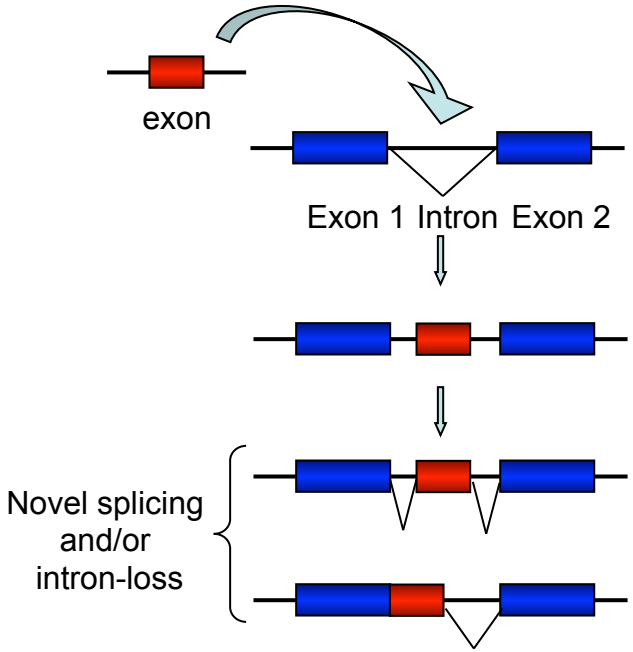
2012 pervasive transcription
S. cerevisiae, human



RNA-mediated events

Walter Gilbert

Exon shuffling



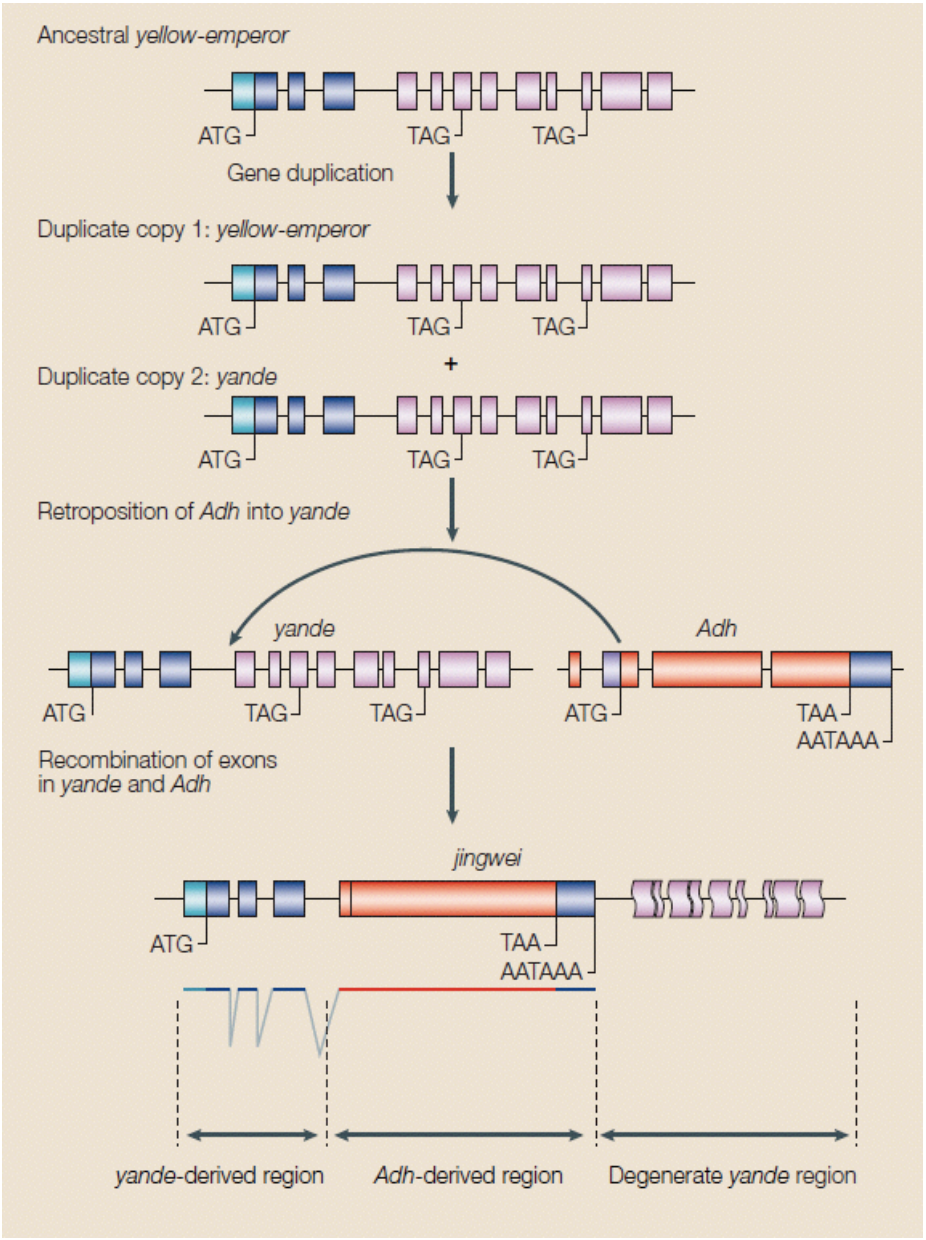
~ 19 % of human exons

Retrogenes

~ 1 % of human exons

Exonization of mobile elements

~ 4 % of human exons

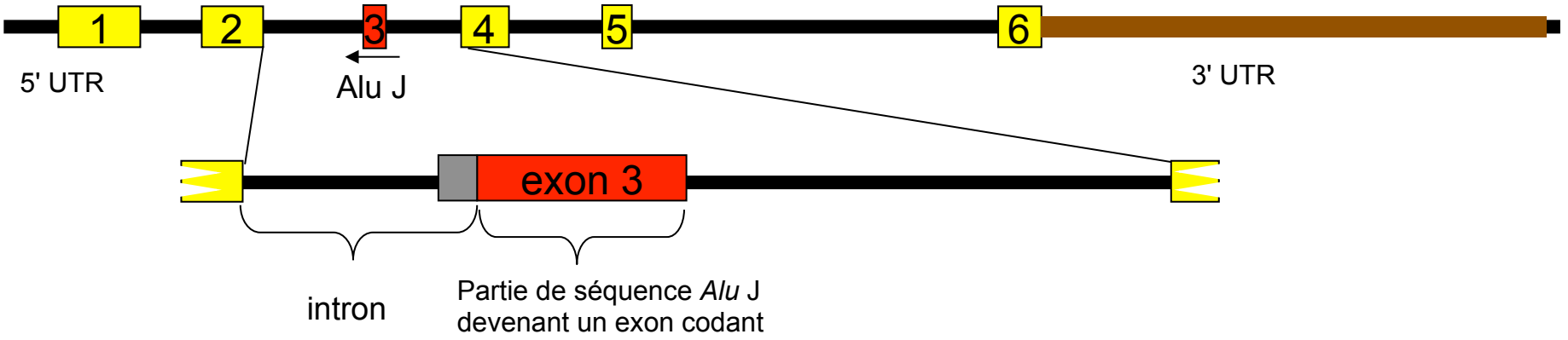


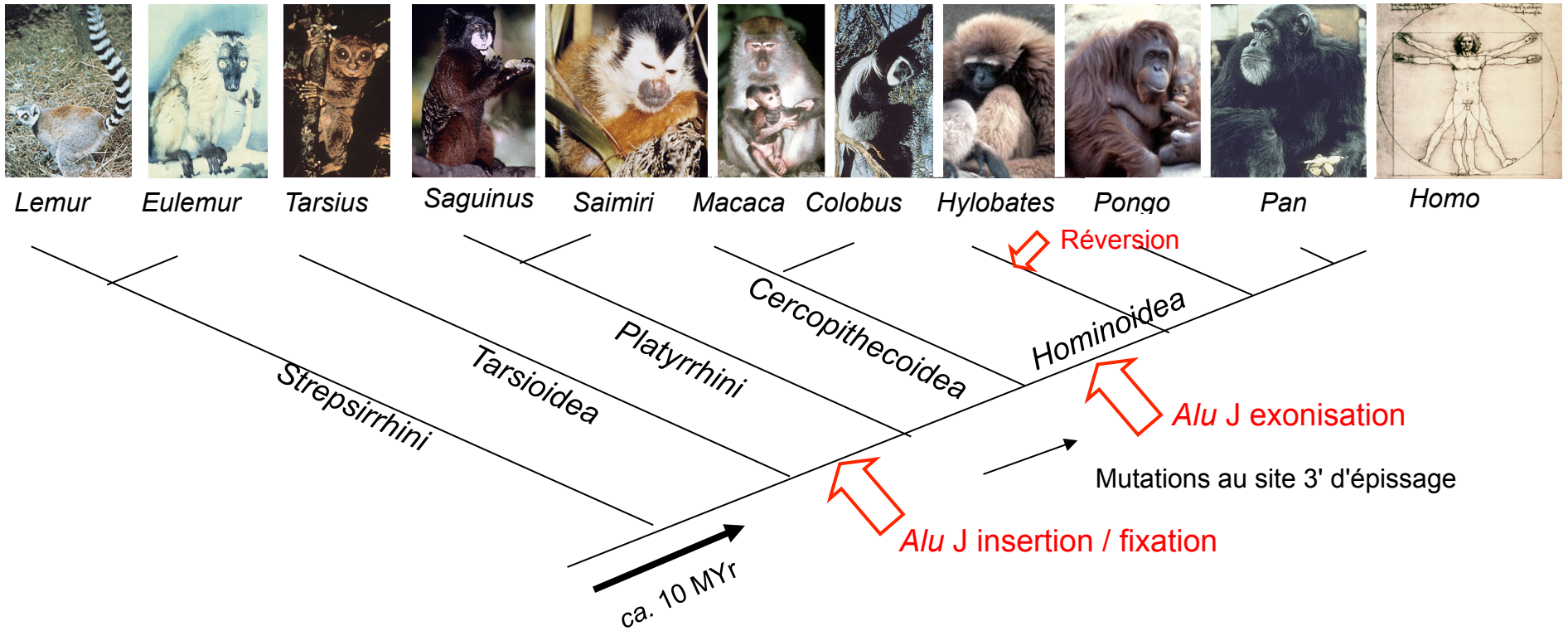
Long and Langley (1993) *Science* **260**: 91-95
Long *et al.* (2003) *Nature Genetics Reviews* **4**: 865- 875

Exonisation d'éléments mobiles

Gène humain *RPE2-1* Ribulose-5-phosphate-3-épimerase

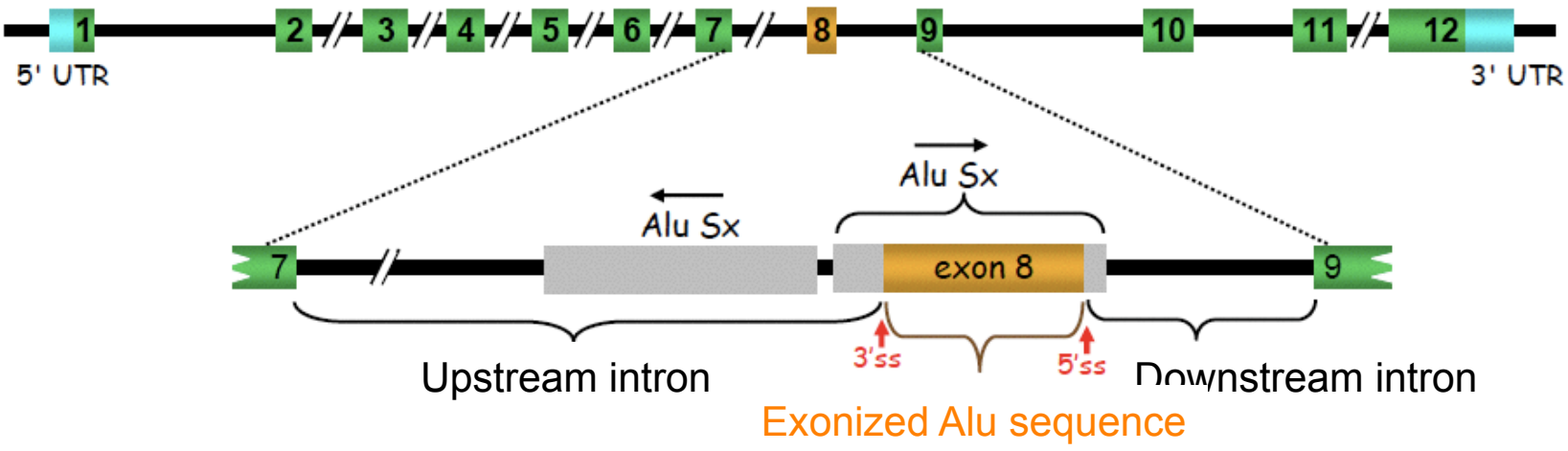
Krull *et al.*, (2005) *Mol. Biol. Evol.* **22,** 1702-1711



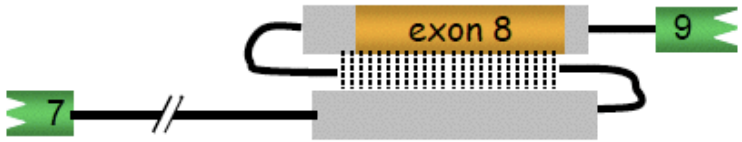


Exonization of mobile elements

Nuclear prelamin A Recognition Factor (Möller-Krull *et al.*, 2008 *J. Mol. Biol.*))



A to I RNA editing generates the 3' SS and eliminates an in-frame stop codon



Exonization of mobile elements