PLANT OF THE DAY

- *Paris japonica* (native of Japan)
- Family – Melanthaceae
- Largest Eukaryotic genome – 150 Gbp
- DNA from a single cell stretched out end-to-end would be taller than 300 feet (91 m)

PLANT OF THE DAY

- *Paris japonica* (native of Japan)
- Largest plant genome known to science – 150 Gbp

- Oak – circa 0.9 Gb

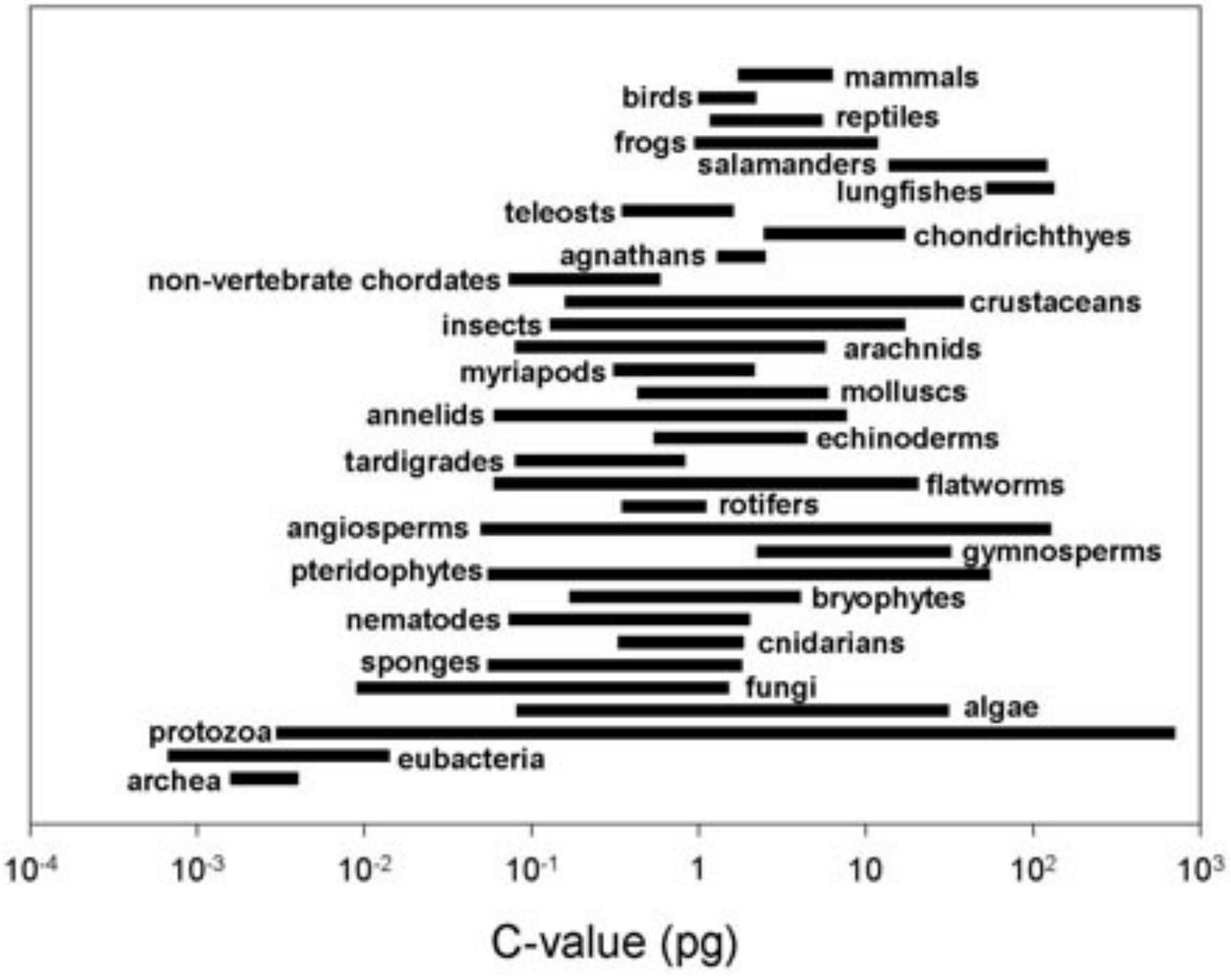
Big Questions

Why do genome sizes vary wildly among organisms with similar levels of cellular and developmental complexity (the C-Value paradox)?

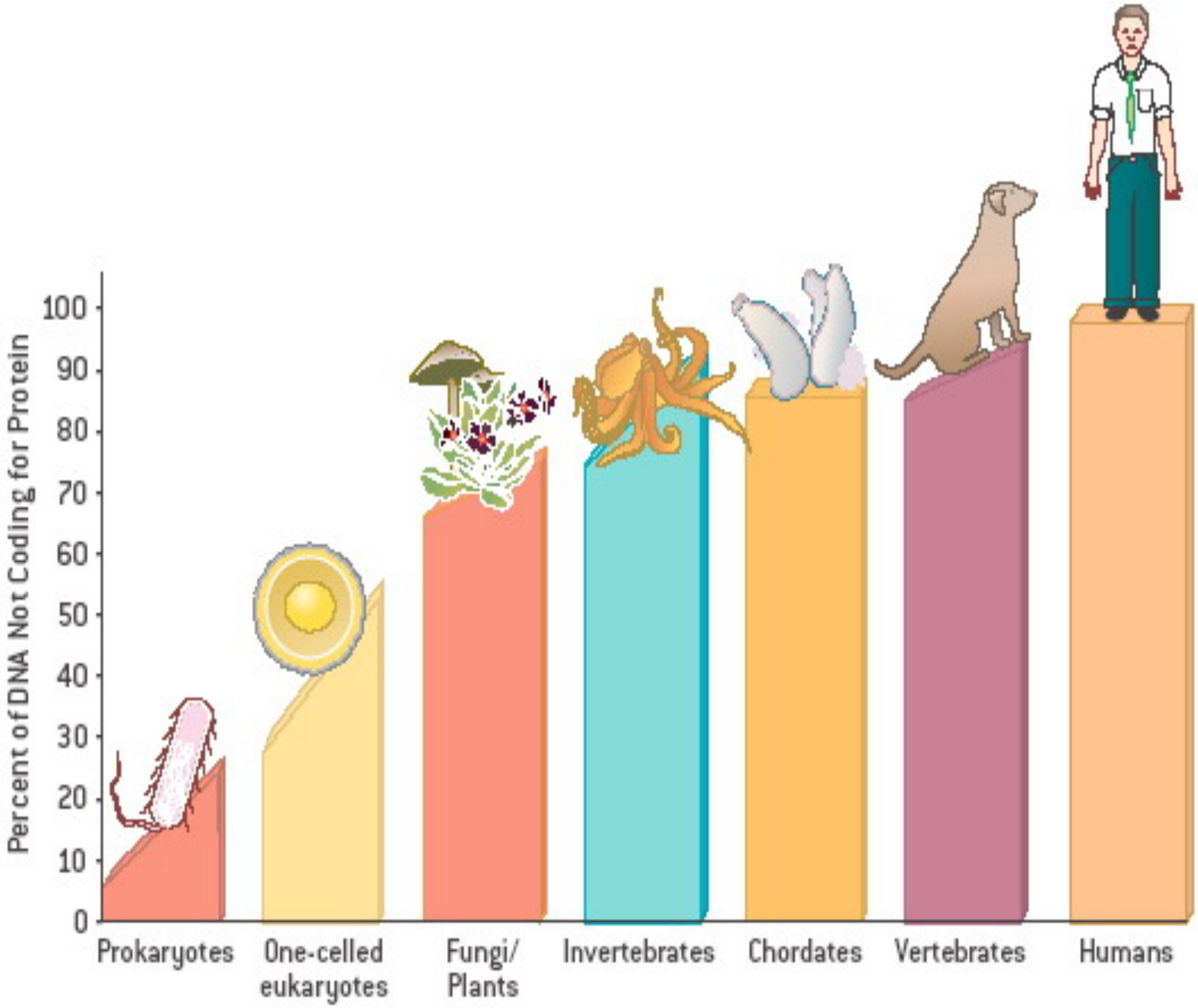
What is the function and evolutionary role of repetitive elements?

C-Value Paradox

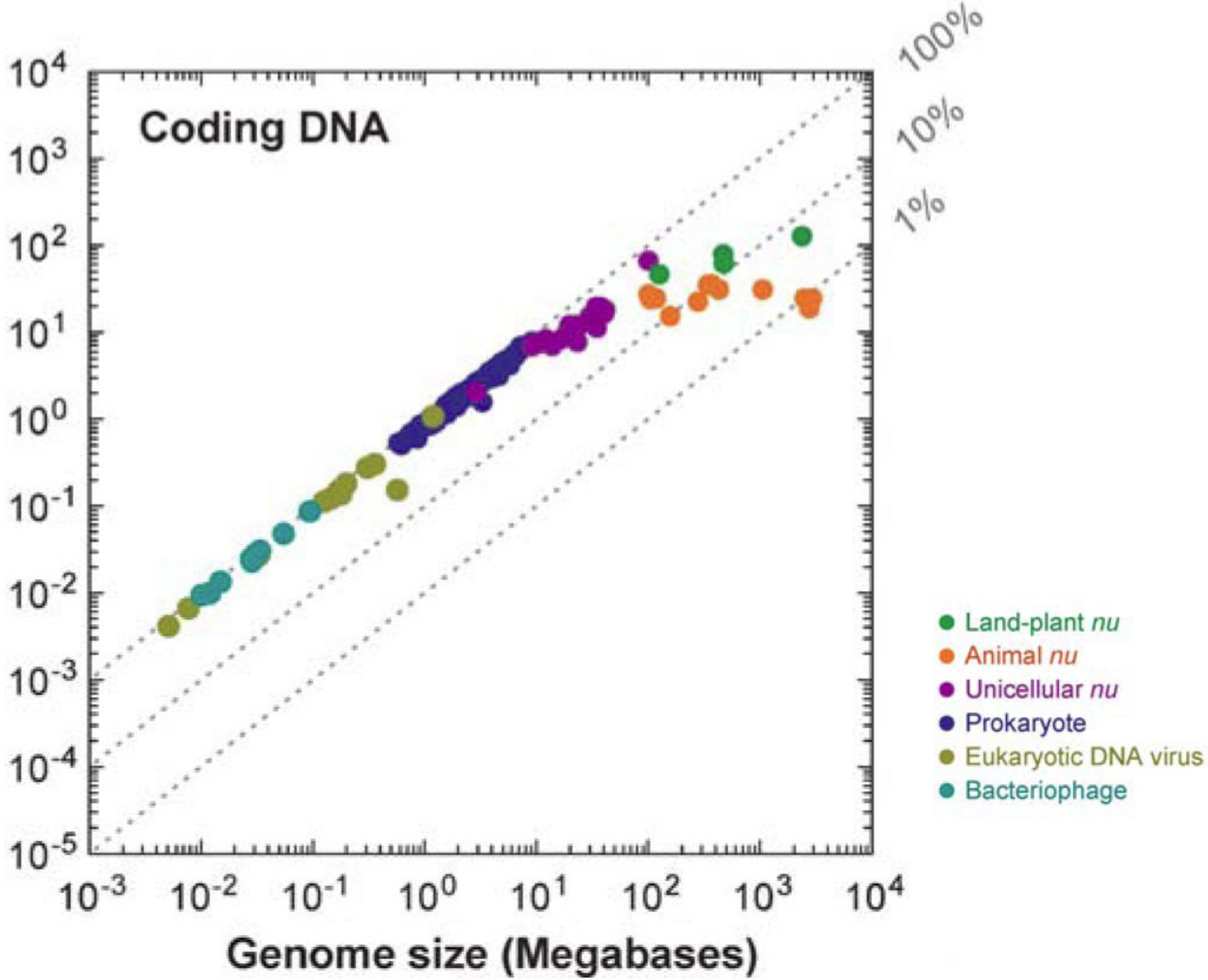
Why do genome sizes vary wildly among organisms with similar levels of cellular and developmental complexity?



Percent of DNA non-coding



NONPROTEIN-CODING SEQUENCES make up only a small fraction of the DNA of prokaryotes. Among eukaryotes, as their complexity increases, generally so, too, does the proportion of their DNA that does not code for protein. The noncoding sequences have been considered junk, but perhaps it actually helps to explain organisms' complexity.


Coding DNA
100%
10%
1%
10^4
10^3
10^2
10^1
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-3
10^-2
10^-1
10^0
10^1
10^2
10^3
10^4
Genome size (Megabases)
● Land-plant *nu*
● Animal *nu*
● Unicellular *nu*
● Prokaryote
● Eukaryotic DNA virus
● Bacteriophage

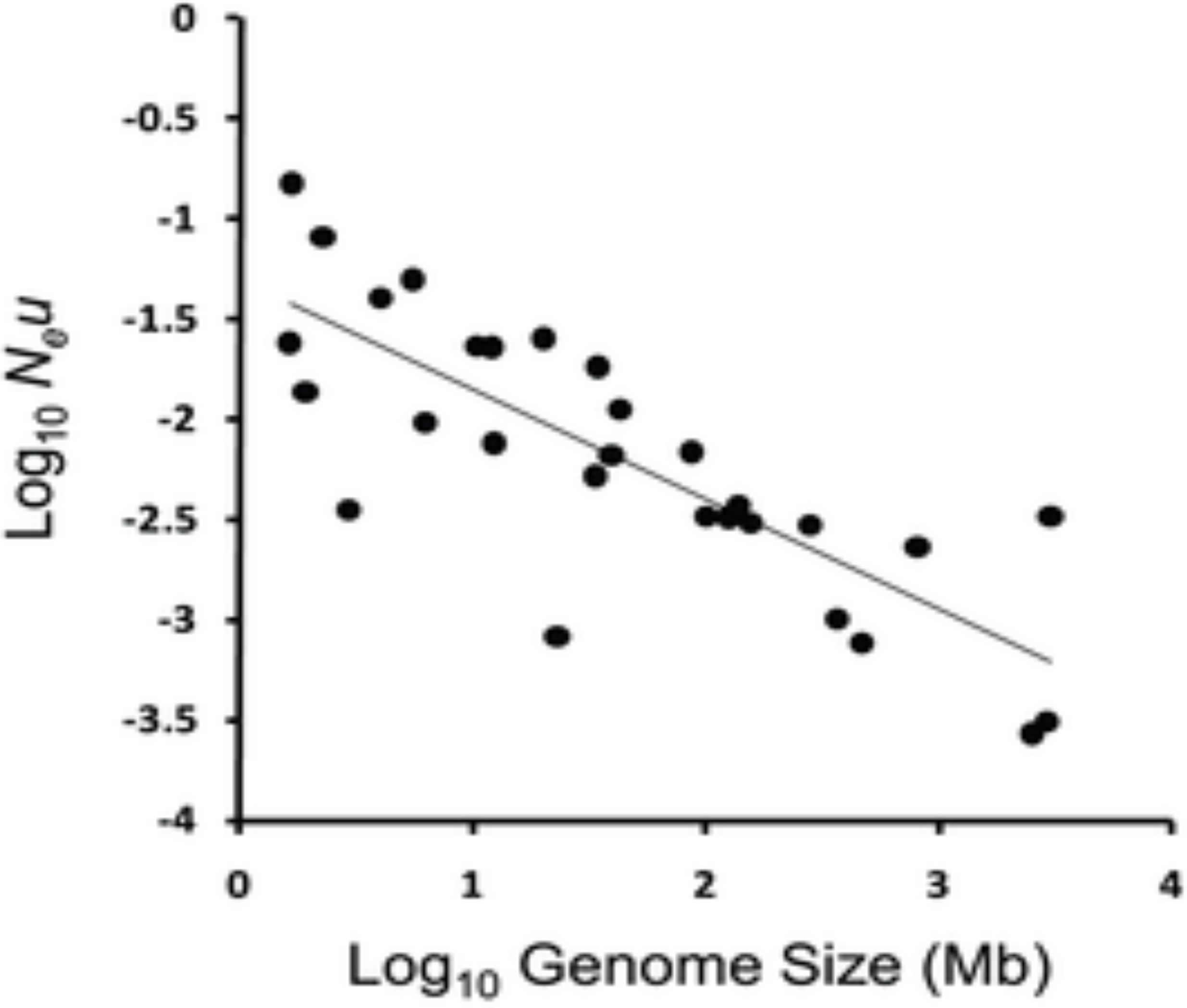
C-Value Paradox

Why does the percent of non-coding DNA vary wildly among organisms with similar levels of cellular and developmental complexity?

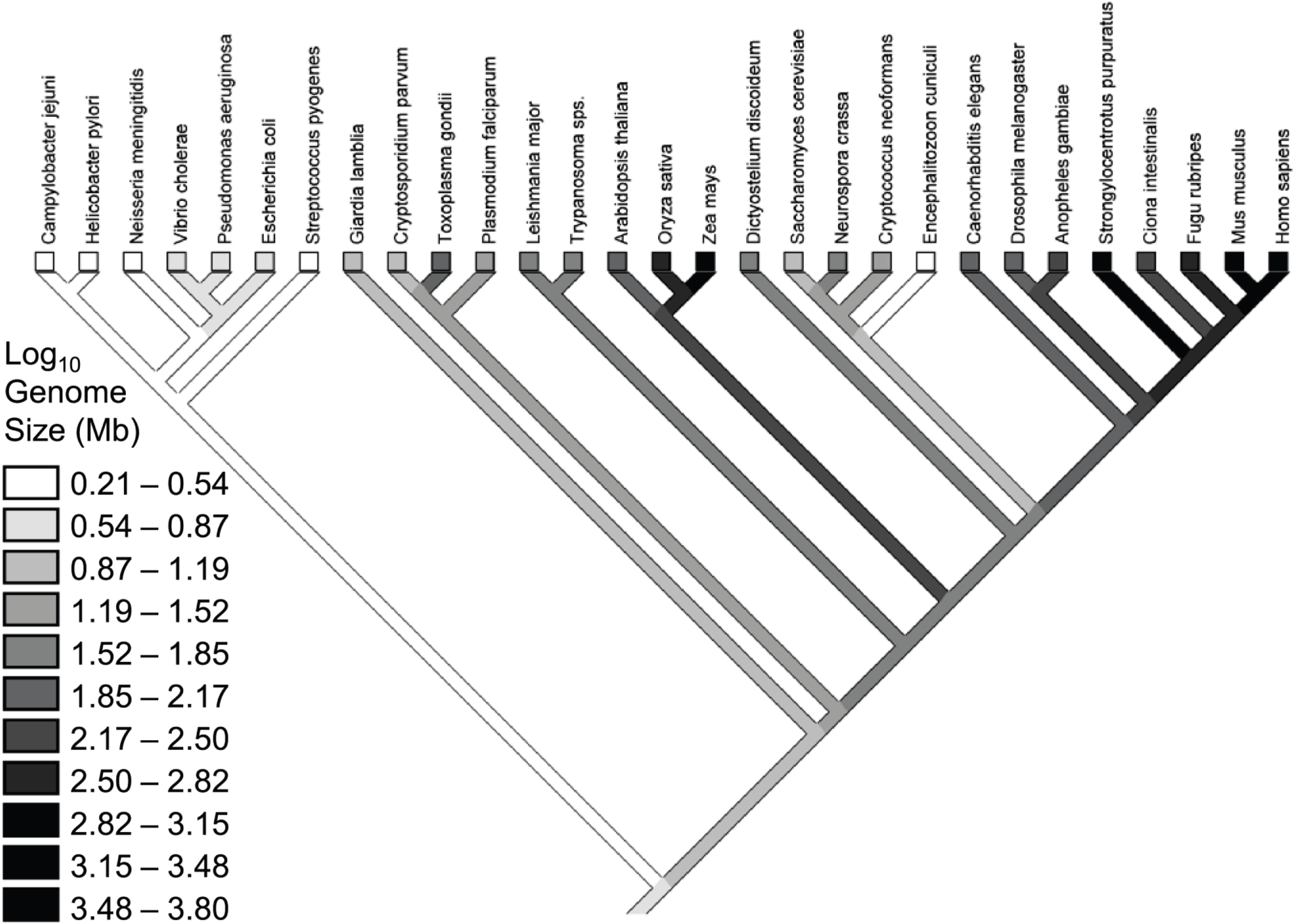
Hypotheses:

1)Selfish DNA – most non-coding DNA consists for selfish elements capable of proliferating until the cost to host fitness becomes prohibitive.
2)Bulk DNA – genome size has a direct effect on nuclear volume, cell size, and cell division rate, all of which influence important life history features.
3)Metabolic cost of DNA limits genome size.
4)Interspecific variation in mutational tendency to delete excess DNA.
5)Population size and mutational hazard of excess DNA, especially gain of function mutations.
6)All of the above

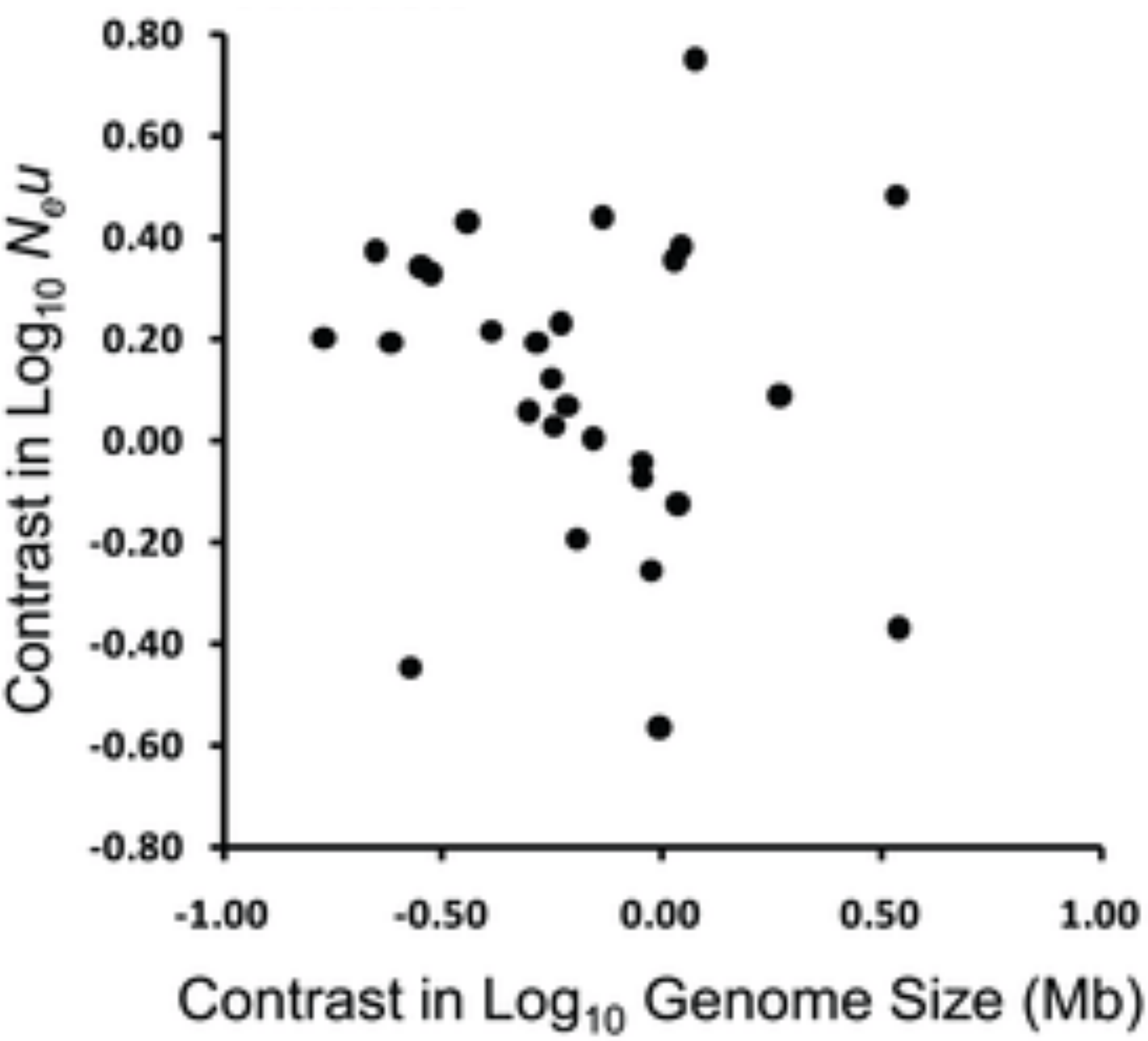
Large genome size, small Ne



Mike Lynch


Campylobacter jejuni
Helicobacter pylori
Neisseria meningitidis
Vibrio cholerae
Pseudomonas aeruginosa
Escherichia coli
Streptococcus pyogenes
Giardia lamblia
Cryptosporidium parvum
Toxoplasma gondii
Plasmodium falciparum
Leishmania major
Trypanosoma sps.
Arabidopsis thaliana
Oryza sativa
Zea mays
Dictyostelium discoideum
Saccharomyces cerevisiae
Neurospora crassa
Cryptococcus neoformans
Encephalitozoon cuniculi
Caenorhabditis elegans
Drosophila melanogaster
Anopheles gambiae
Strongylocentrotus purpuratus
Ciona intestinalis
Fugu rubripes
Mus musculus
Homo sapiens
Log10 Genome Size (Mb)
0.21 – 0.54
0.54 – 0.87
0.87 – 1.19
1.19 – 1.52
1.52 – 1.85
1.85 – 2.17
2.17 – 2.50
2.50 – 2.82
2.82 – 3.15
3.15 – 3.48
3.48 – 3.80

Phylogenetically independent contrasts



Ken Whitney

Plant Nuclear Genome Size Variation

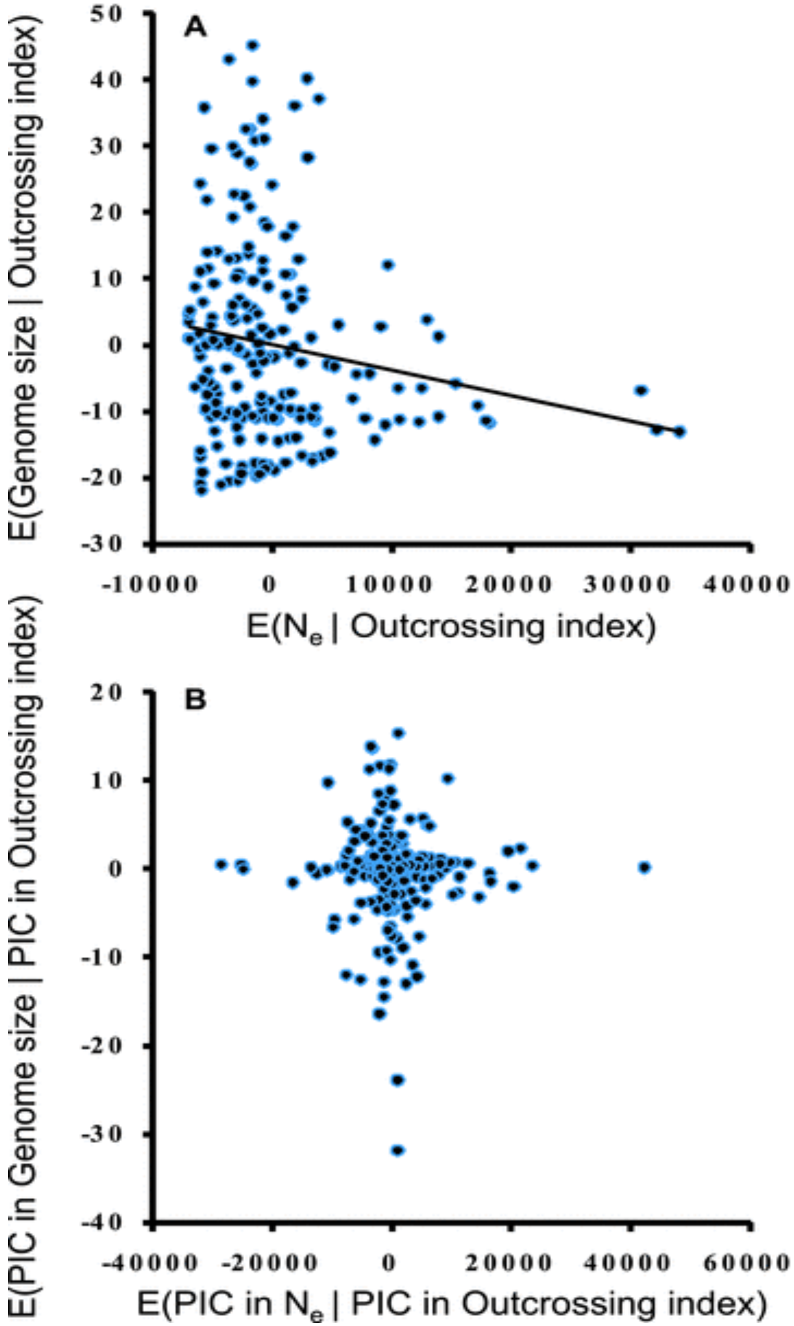
7058 species surveyed and databased
~ 2300X difference

63 mbp

150,000 mbp

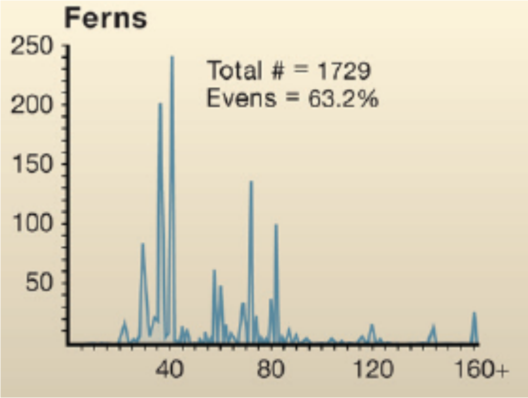
http://data.kew.org/cvalues/

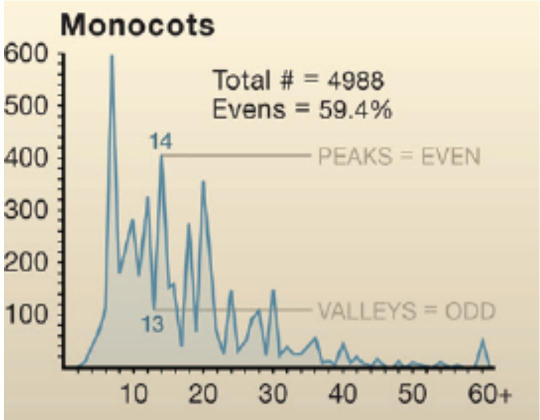
A ROLE FOR NONADAPTIVE PROCESSES IN PLANT GENOME SIZE EVOLUTION?

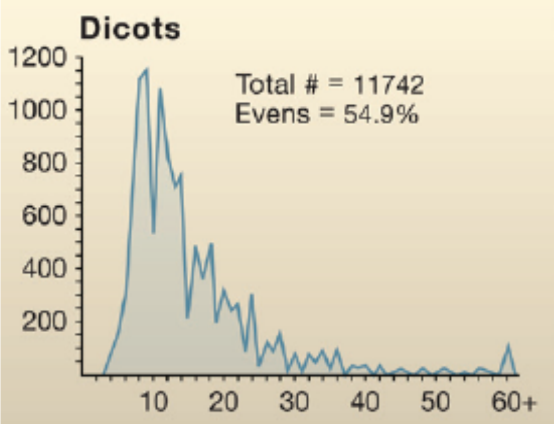


Evolution
Volume 64, Issue 7, pages 2097-2109, 9 FEB 2010

Chromosome Number Variation







Chromosome numbers vary *n* = 2 to *n* = ~680

Euploid variation – polyploidy

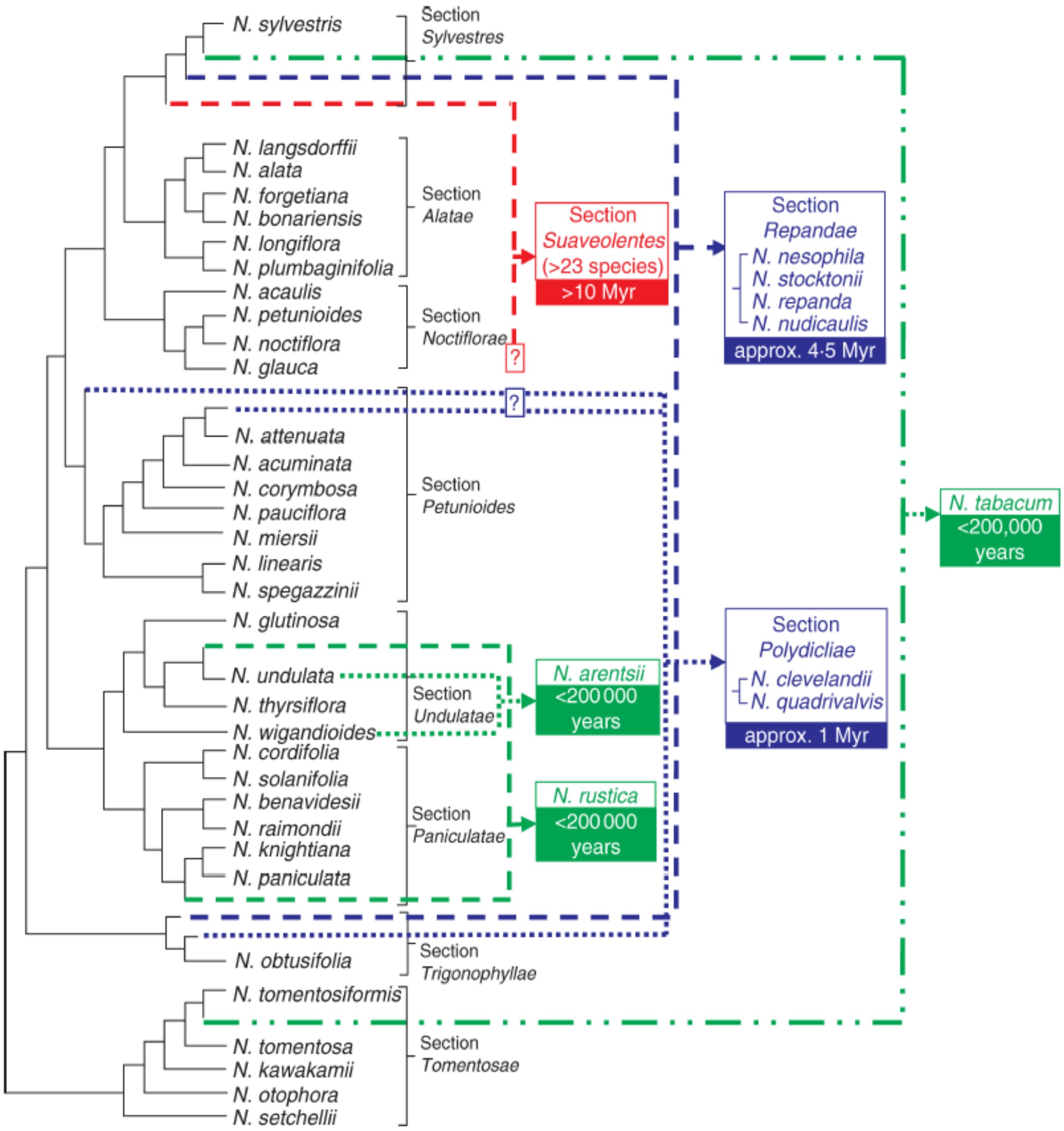
~35% of vascular plants are neopolyploids

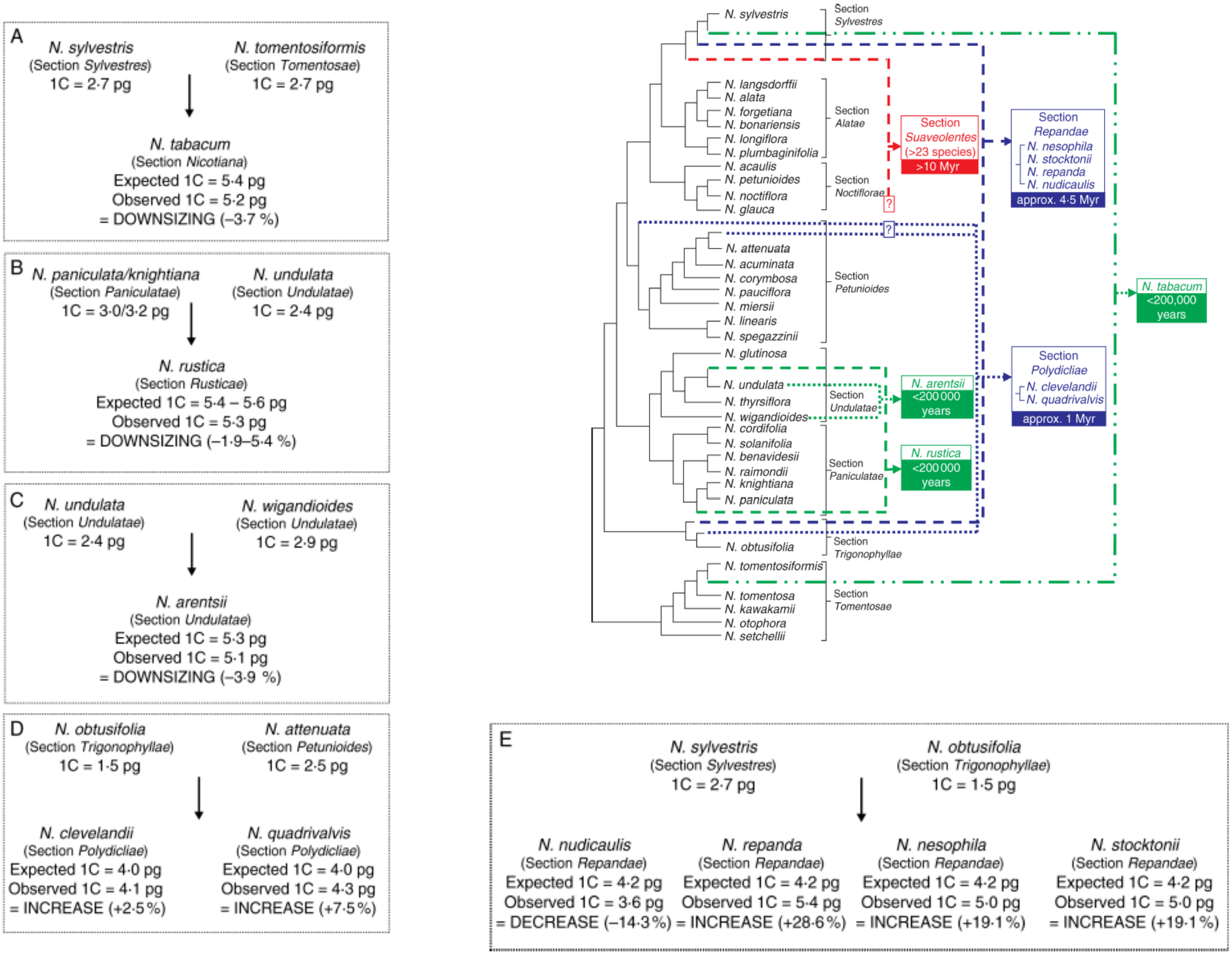
All are paleopolyploids

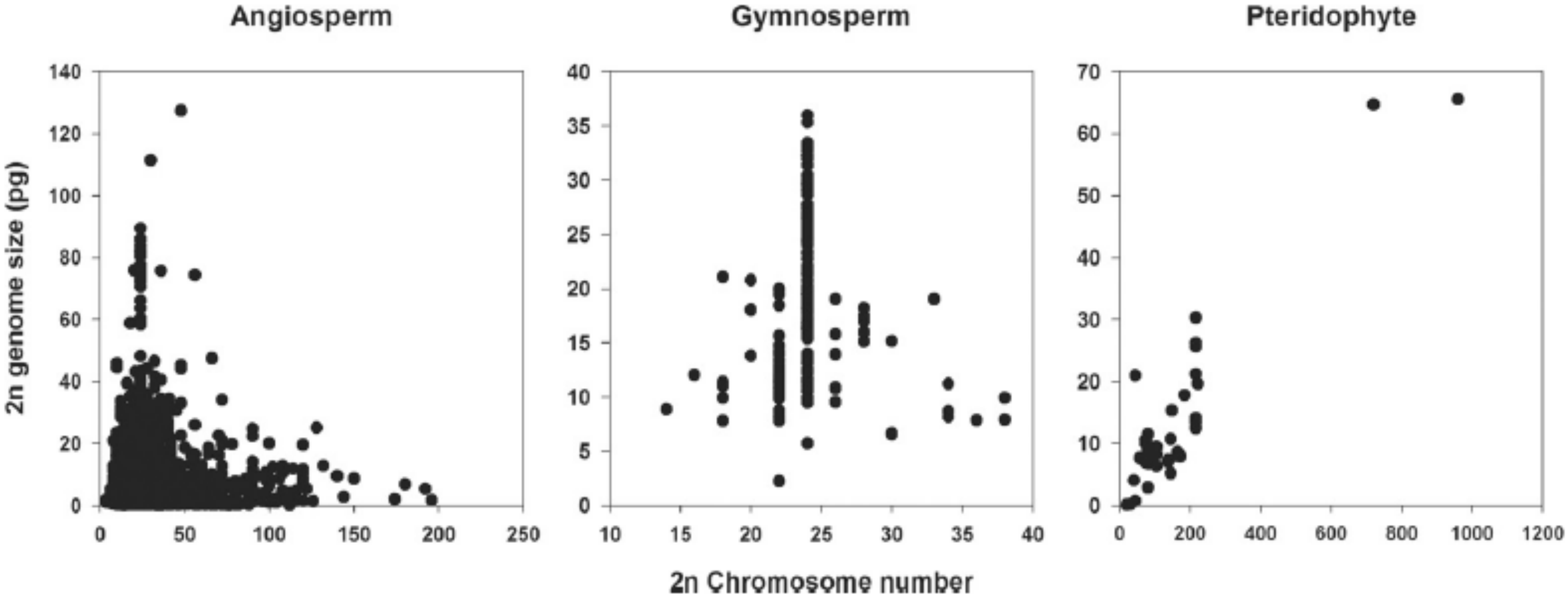
Aneuploid variation – gain or less of one or more chromosomes

Post polyploid genome size change is variable

- -Additive sum of parents
- -Increase
- -Decrease




A
N. sylvestris
(Section Sylvestres)
1C = 2·7 pg
N. tomentosiformis
(Section Tomentosae)
1C = 2·7 pg
N. tabacum
(Section Nicotiana)
Expected 1C = 5·4 pg
Observed 1C = 5·2 pg
= DOWNSIZING (–3·7 %)
B
N. paniculata/knightiana
(Section Paniculatae)
1C = 3·0/3·2 pg
N. undulata
(Section Undulatae)
1C = 2·4 pg
N. rustica
(Section Rusticae)
Expected 1C = 5·4 – 5·6 pg
Observed 1C = 5·3 pg
= DOWNSIZING (–1·9–5·4 %)
C
N. undulata
(Section Undulatae)
1C = 2·4 pg
N. wigandioides
(Section Undulatae)
1C = 2·9 pg
N. arentsii
(Section Undulatae)
Expected 1C = 5·3 pg
Observed 1C = 5·1 pg
= DOWNSIZING (–3·9 %)
D
N. obtusifolia
(Section Trigonophyllae)
1C = 1·5 pg
N. attenuata
(Section Petunioides)
1C = 2·5 pg
N. clevelandii
(Section Polydicliae)
Expected 1C = 4·0 pg
Observed 1C = 4·1 pg
= INCREASE (+2·5 %)
N. quadrivalvis
(Section Polydicliae)
Expected 1C = 4·0 pg
Observed 1C = 4·3 pg
= INCREASE (+7·5 %)
N. sylvestris
Section Sylvestres
N. langsdorffii
N. alata
N. forgetiana
N. bonariensis
N. longiflora
N. plumbaginifolia
Section Alatae
N. acaulis
N. petunioides
N. noctiflora
N. glauca
Section Noctiflorae
Section Suaveolentes (>23 species)
>10 Myr
?
Section Repandae
N. nesophila
N. stocktonii
N. repanda
N. nudicaulis
approx. 4·5 Myr
?
N. attenuata
N. acuminata
N. corymbosa
N. pauciflora
N. miersii
N. linearis
N. spegazzinii
Section Petunioides
N. tabacum
<200,000 years
N. glutinosa
N. undulata
N. thyrsiflora
N. wigandioides
Section Undulatae
N. arentsii
<200 000 years
Section Polydicliae
N. clevelandii
N. quadrivalvis
approx. 1 Myr
N. cordifolia
N. solanifolia
N. benavidesii
N. raimondii
N. knightiana
N. paniculata
Section Paniculatae
N. rustica
<200 000 years
N. obtusifolia
Section Trigonophyllae
N. tomentosiformis
N. tomentosa
N. kawakamii
N. otophora
N. setchellii
Section Tomentosae
E
N. sylvestris
(Section Sylvestres)
1C = 2·7 pg
N. obtusifolia
(Section Trigonophyllae)
1C = 1·5 pg
N. nudicaulis
(Section Repandae)
Expected 1C = 4·2 pg
Observed 1C = 3·6 pg
= DECREASE (–14·3 %)
N. repanda
(Section Repandae)
Expected 1C = 4·2 pg
Observed 1C = 5·4 pg
= INCREASE (+28·6 %)
N. nesophila
(Section Repandae)
Expected 1C = 4·2 pg
Observed 1C = 5·0 pg
= INCREASE (+19·1 %)
N. stocktonii
(Section Repandae)
Expected 1C = 4·2 pg
Observed 1C = 5·0 pg
= INCREASE (+19·1 %)



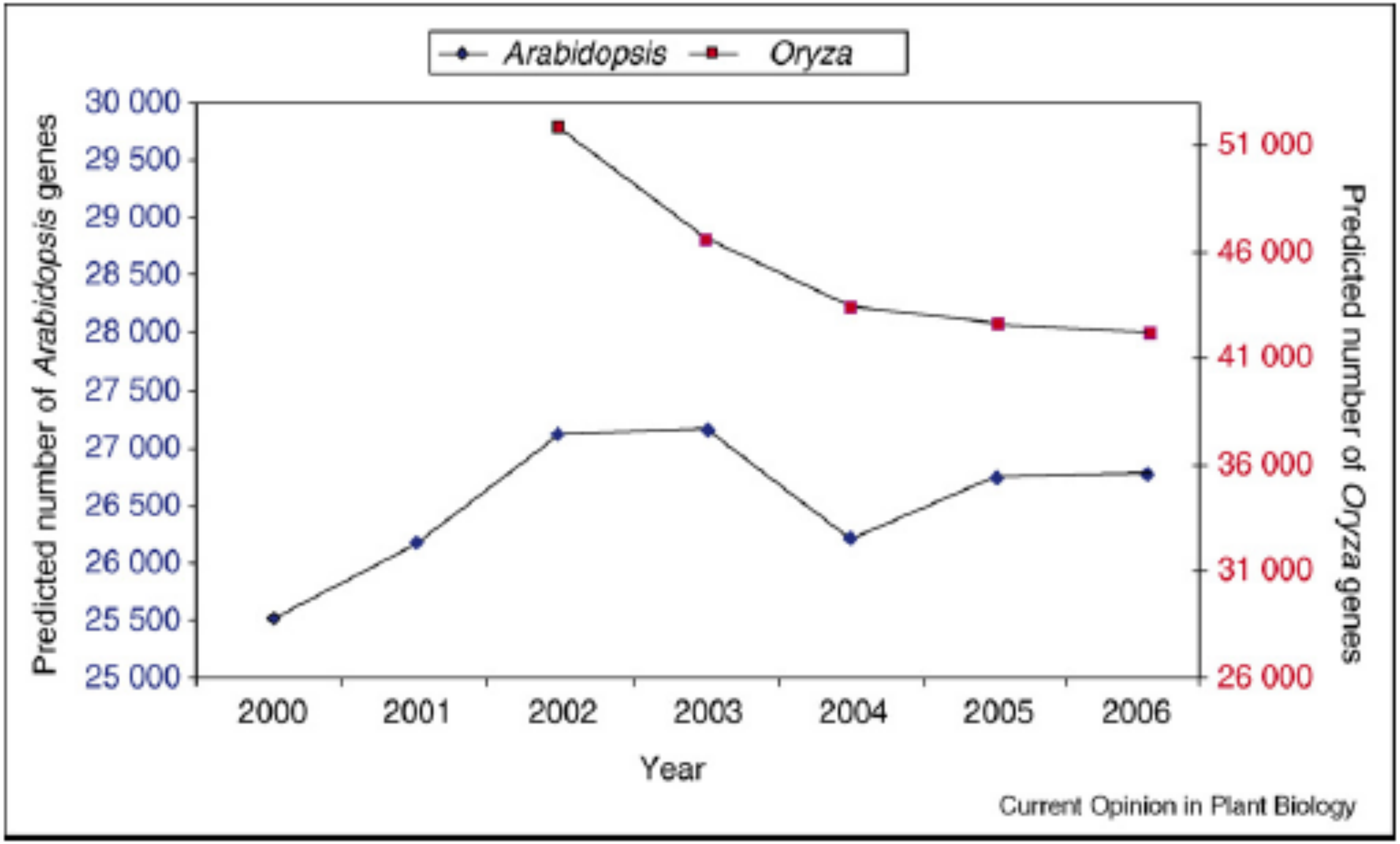
Correlation of Chromosome Number and Genome Size

Angiosperm r = -0.023

Gymnosperm r = 0.106

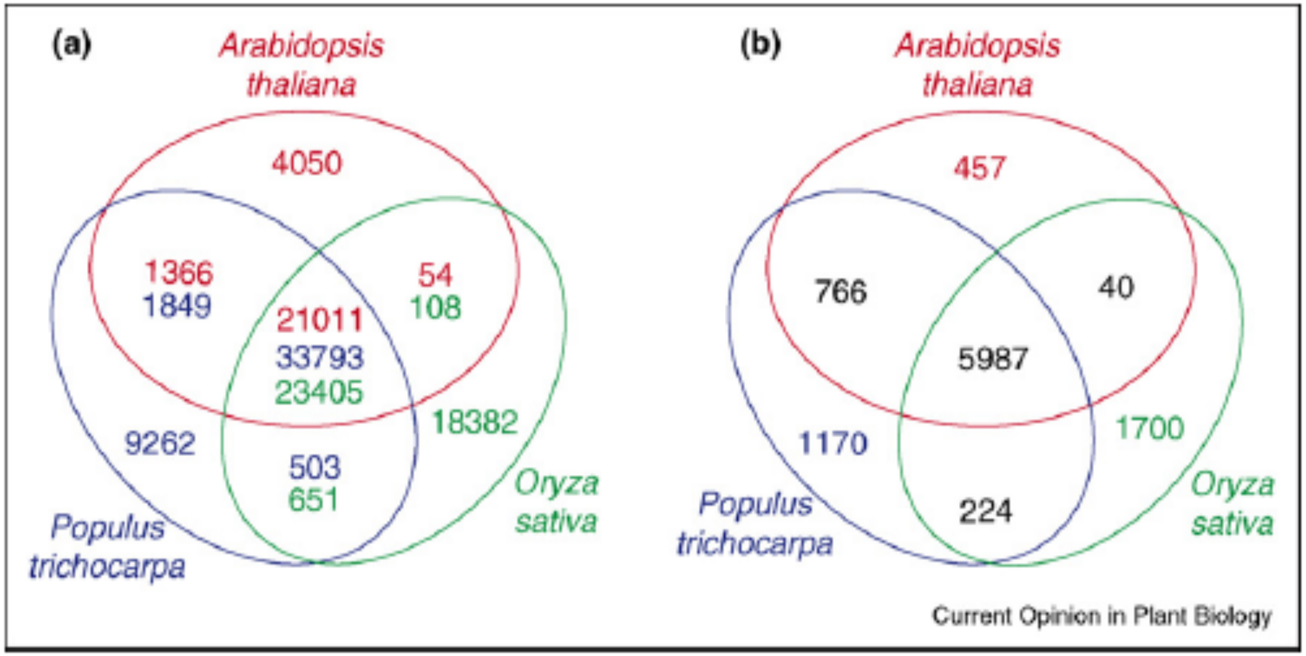
Pteridophytes r = 0.913

Variation in the number of nuclear genes



Current Opinion in Plant Biology

Plant Nuclear Gene Overlap



Current Opinion in Plant Biology

~90% of genes have homologs in other genomes

Does not appear to be large differences – most genomes around 40,000

Not a substantial contributor to variation in genome size

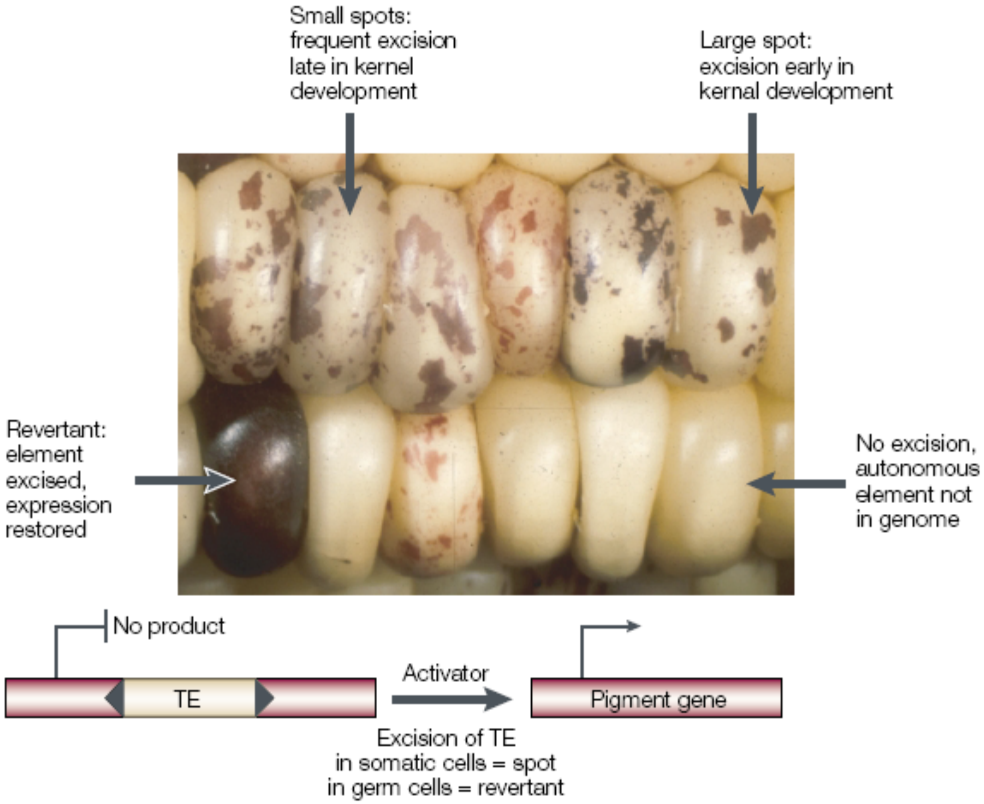
Transposable Elements (TEs)

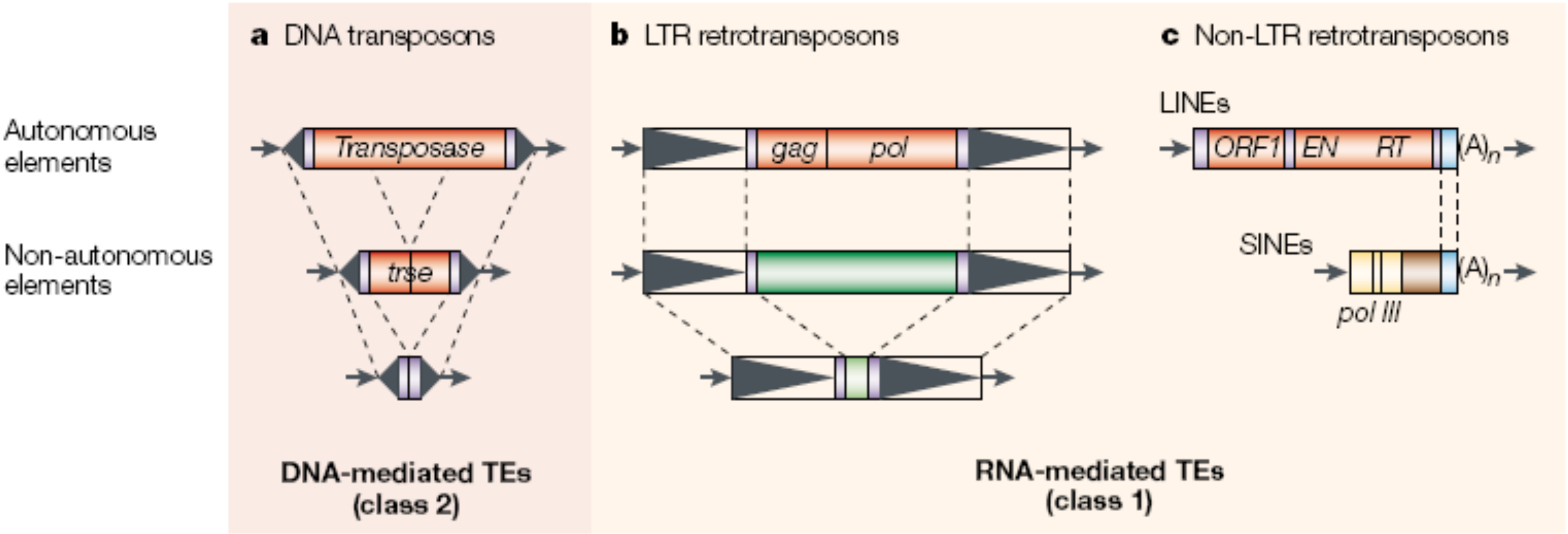
50-80% of plant genomes are TEs

Discovered by Barbara McClintock by studying unstable corn kernel phenotypes

Fragments of DNA that can insert into new chromosomal locations

Often duplicate themselves during the process of moving around





Class 1 TEs use RNA intermediates to move around and undergo duplicative transposition

Class 2 TEs are excised during transposition and may undergo “cut and paste” transposition with no duplication or “gap repair” where the gap is filled with a copy of the transposon

Autonomous elements contain necessary genes for transposition

Non-autonomous elements rely on products of other elements for transposition

MITES: Miniature Inverted Repeat Transposable Elements

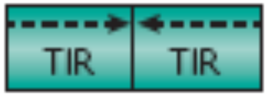
Class 2 elements found in or near genes

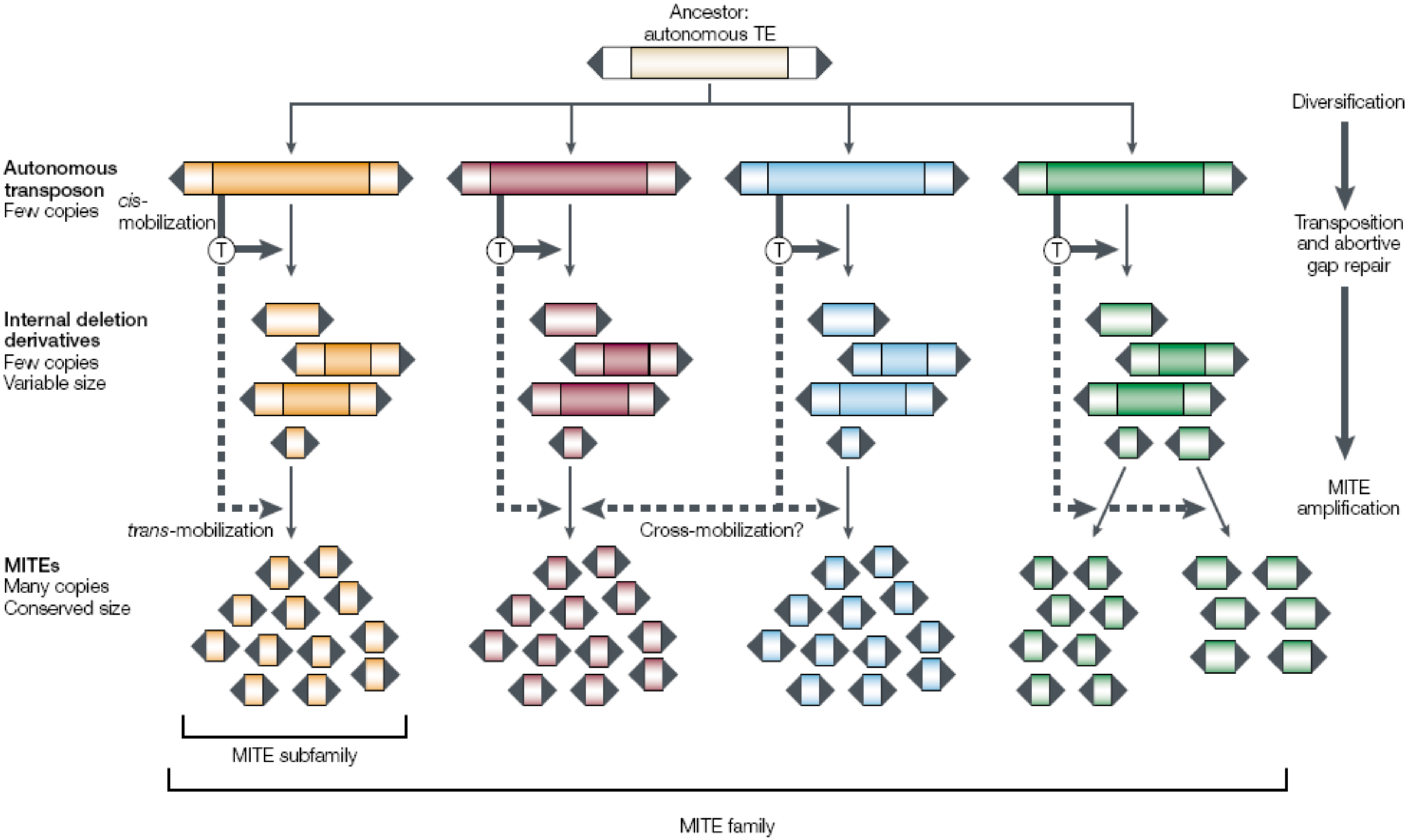
A few dozen to few hundred base pairs in length

Two inverted repeats

Non-autonomous – activated by other autonomous TEs

6% of *Arabidopsis* and 12% of rice genomes are composed of MITES




Ancestor:
autonomous TE
Diversification
Autonomous transposon
Few copies
cis-mobilization
T
Transposition and abortive gap repair
Internal deletion derivatives
Few copies
Variable size
MITE amplification
trans-mobilization
Cross-mobilization?
MITEs
Many copies
Conserved size
MITE subfamily
MITE family

LTRs: Long Terminal Repeat Retrotransposons

Class 1 elements found between genes

Autonomous – self activating

Duplicative transposition

Single largest component of plant genomes

50-70% of maize genome is LTR

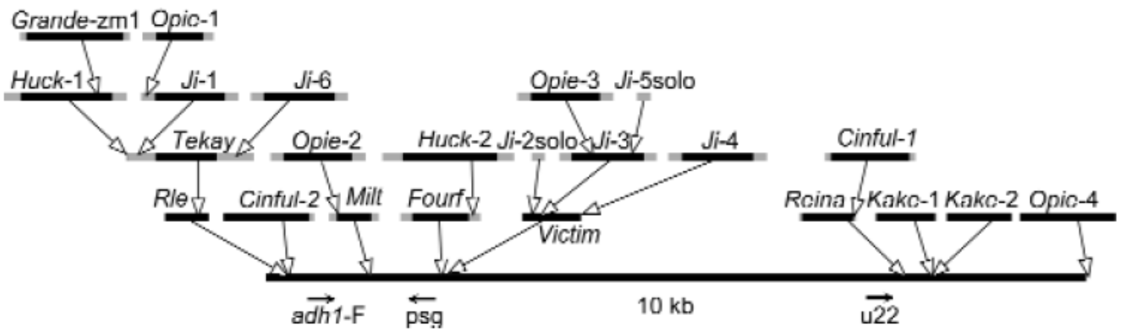
LTR Driven Genome Expansion in Maize

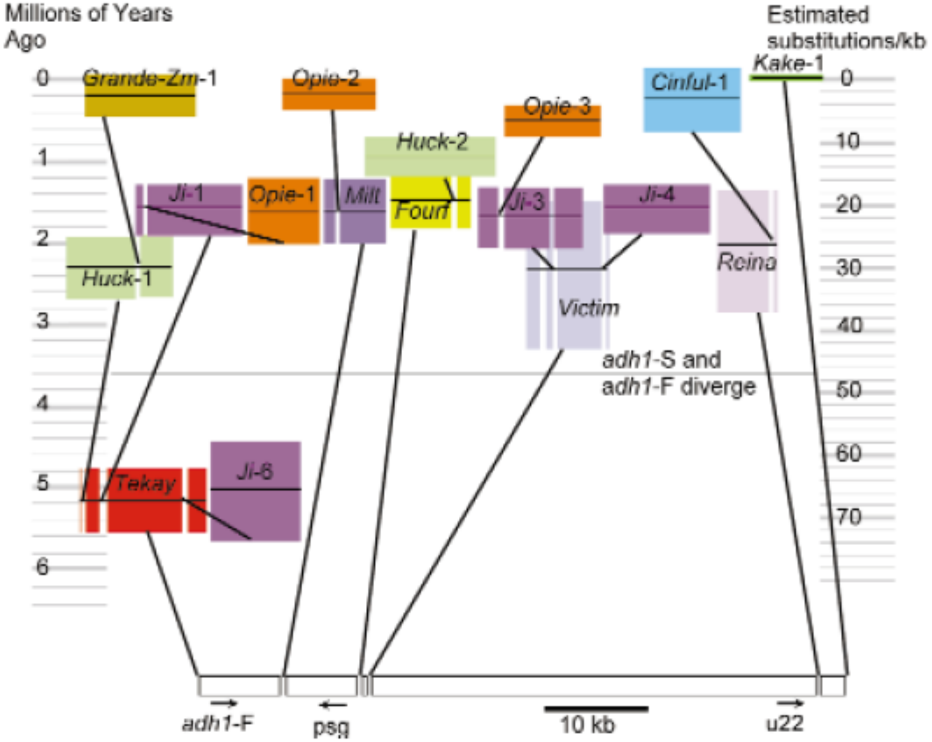
240 KB contig containing *adh1*
~60% of sequence was
retrotransposon

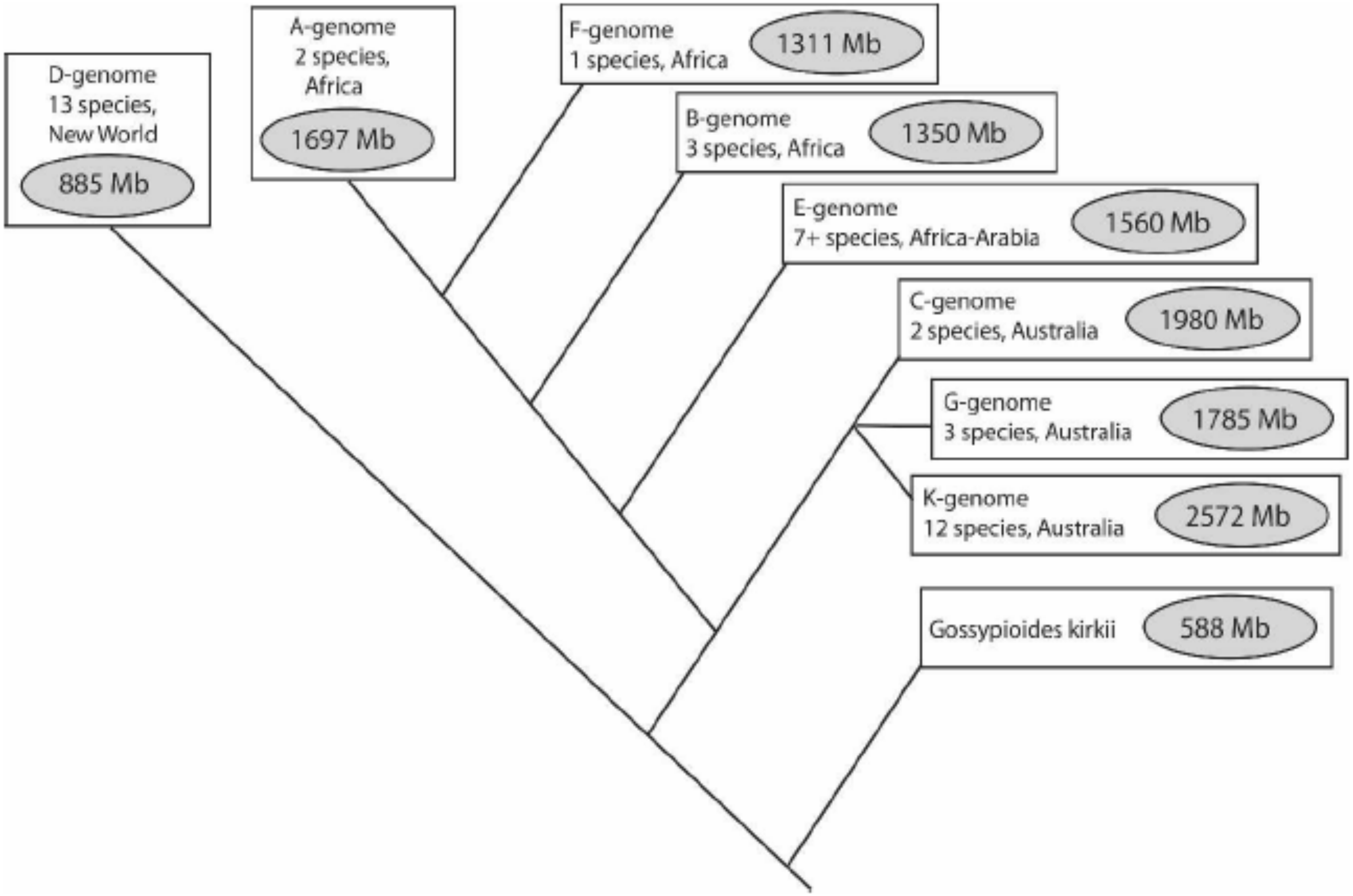
Copy numbers 10 – 1000

Comparison to syntetic region in
Sorghum showed no evidence of
transposons

Dating suggest doubling of maize
genome in less than 6 million years






D-genome
13 species,
New World
885 Mb
A-genome
2 species,
Africa
1697 Mb
F-genome
1 species, Africa
1311 Mb
B-genome
3 species, Africa
1350 Mb
E-genome
7+ species, Africa-Arabia
1560 Mb
C-genome
2 species, Australia
1980 Mb
G-genome
3 species, Australia
1785 Mb
K-genome
12 species, Australia
2572 Mb
Gossypioides kirkii
588 Mb

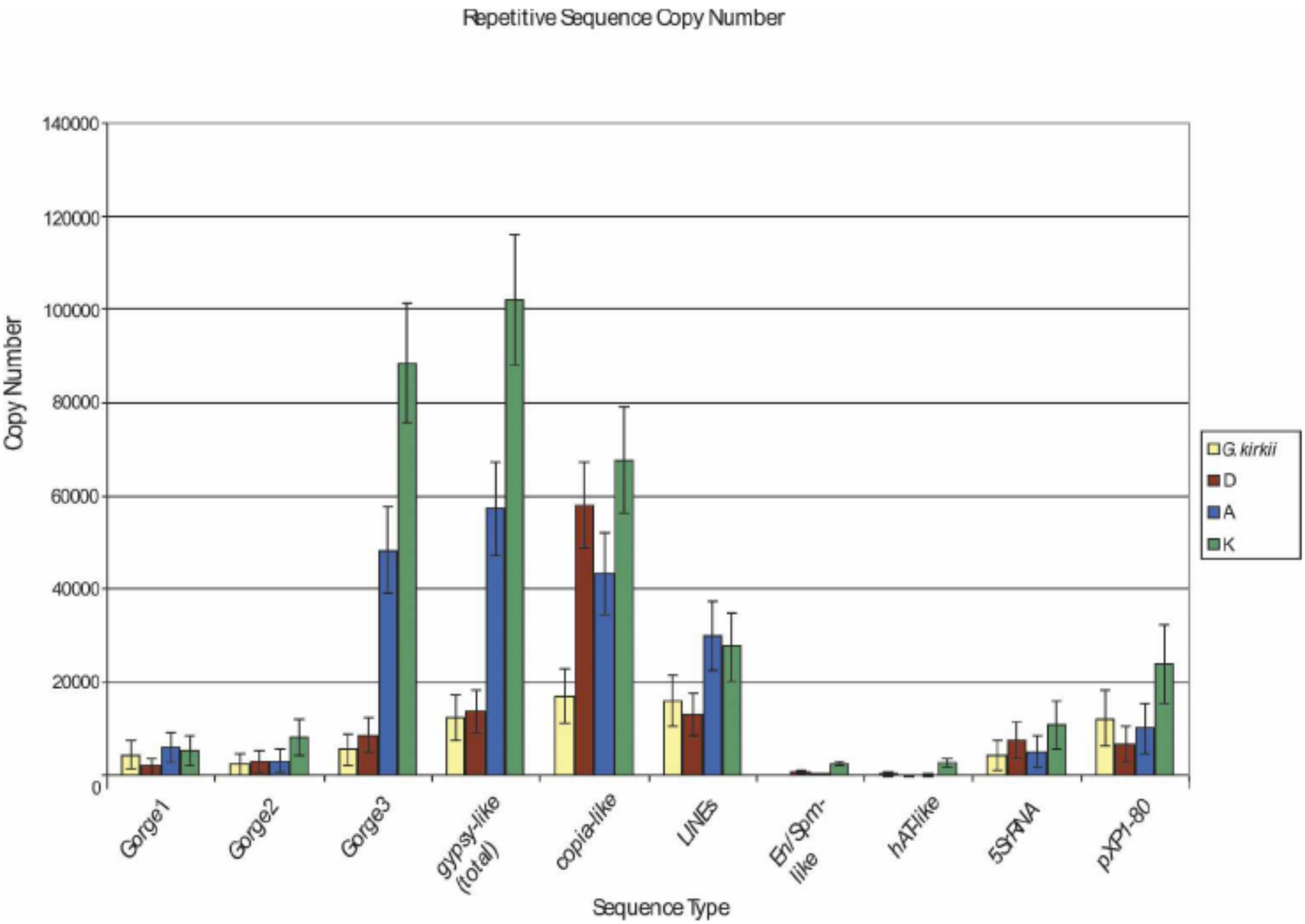

Repetitive Sequence Copy Number
Copy Number
140000
120000
100000
80000
60000
40000
20000
0
Gorge1
Gorge2
Gorge3
gypsy-like (total)
copia-like
LINEs
En/Spm-like
hAT-like
5SrRNA
pXP1-80
Sequence Type
G. kirkii
D
A
K

Table 2. Repetitive element copy number and density estimates

	G. kirkii Outgroup 588 Mb	*G. raimondii* D genome 880 Mb	*G. herbaceum* A genome 1667 Mb	*G. exiguum* K genome 2460 Mb
Tandem repeats				
5SrRNA	4279 ± 3227	7675 ± 3826	5073 ± 3379	10,794 ± 5082
pXP1–80	12,264 ± 6098	6573 ± 3956	10,101 ± 5392	23,795 ± 8528
Class II transposons				
En/Spm-like	120 ± 138	835 ± 326	343 ± 216	2514 ± 602
	~0.2%	~0.9%	~0.2%	~1.0%
hAT-like	305 ± 352	81 ± 163	263 ± 304	2615 ± 986
	~0.2%	<0.1%	<0.1%	~0.4%
Class II Total	3.5 Mb	12 Mb	5 Mb	42 Mb
	<0.1%	1.0%	<0.1%	~1.4
Class I retrotransposons				
copia-like	17,006 ± 5765	57,956 ± 9300	43,181 ± 8774	67,700 ± 11,324
	9.7%–19.7%	28%–38.7%	10.7%–16.1%	11.7%–16.5%
LINE	16,006 ± 5597	13,011 ± 4502	30,000 ± 7335	27,563 ± 7271
	5.1%–10.6%	2.8%–5.7%	4.0%–6.5%	2.4%–4.1%
GORGE1 gypsy-like	4502 ± 2992	1971 ± 1762	5909 ± 3273	5319 ± 3205
	2.4%–11.9%	0.2%–3.9%	1.5%–5.2%	0.8%–3.2%
GORGE2 gypsy-like	2500 ± 2233	3154 ± 2227	3181 ± 2403	8221 ± 3983
	0.4%–7.5%	1.0%–5.7%	0.4%–3.2%	1.6%–4.7%
GORGE3 gypsy-like	5502 ± 3305	8674 ± 3683	48,181 ± 9257	88,492 ± 12,904
	3.5%–13.9%	5.3%–13.0%	22.0%–32.6%	28.8%–38.6%
Class I Total	255 Mb	465 Mb	865 Mb	1400 Mb
	42%	53%	52%	58%

Rate of Transposable Element Insertion and Fitness Effects

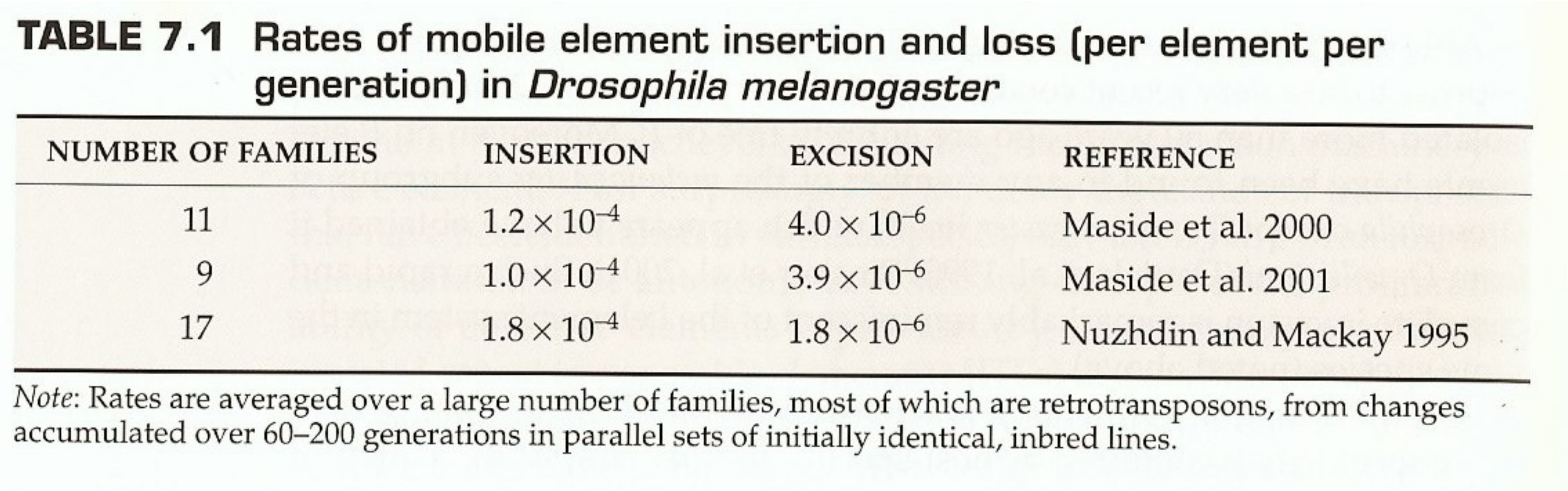


TABLE 7.1 Rates of mobile element insertion and loss (per element per generation) in *Drosophila melanogaster*

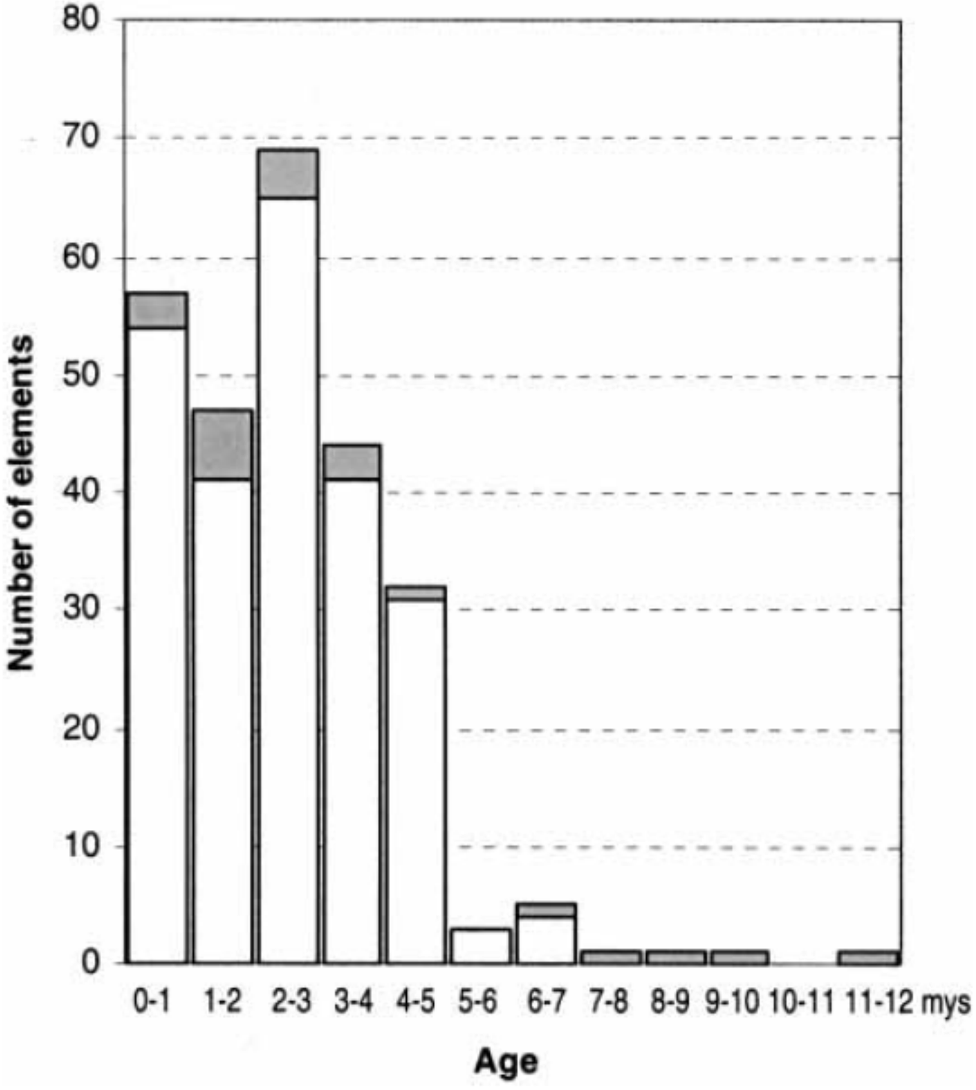
NUMBER OF FAMILIES	INSERTION	EXCISION	REFERENCE
11	1.2×10^{-4}	4.0×10^{-6}	Maside et al. 2000
9	1.0×10^{-4}	3.9×10^{-6}	Maside et al. 2001
17	1.8×10^{-4}	1.8×10^{-6}	Nuzhdin and Mackay 1995

Note: Rates are averaged over a large number of families, most of which are retrotransposons, from changes accumulated over 60–200 generations in parallel sets of initially identical, inbred lines.

Average reductions in fitness per insertion: 0.5 -1.5%

Are genome size and TE growth unchecked?

Ages of LTRs in Rice



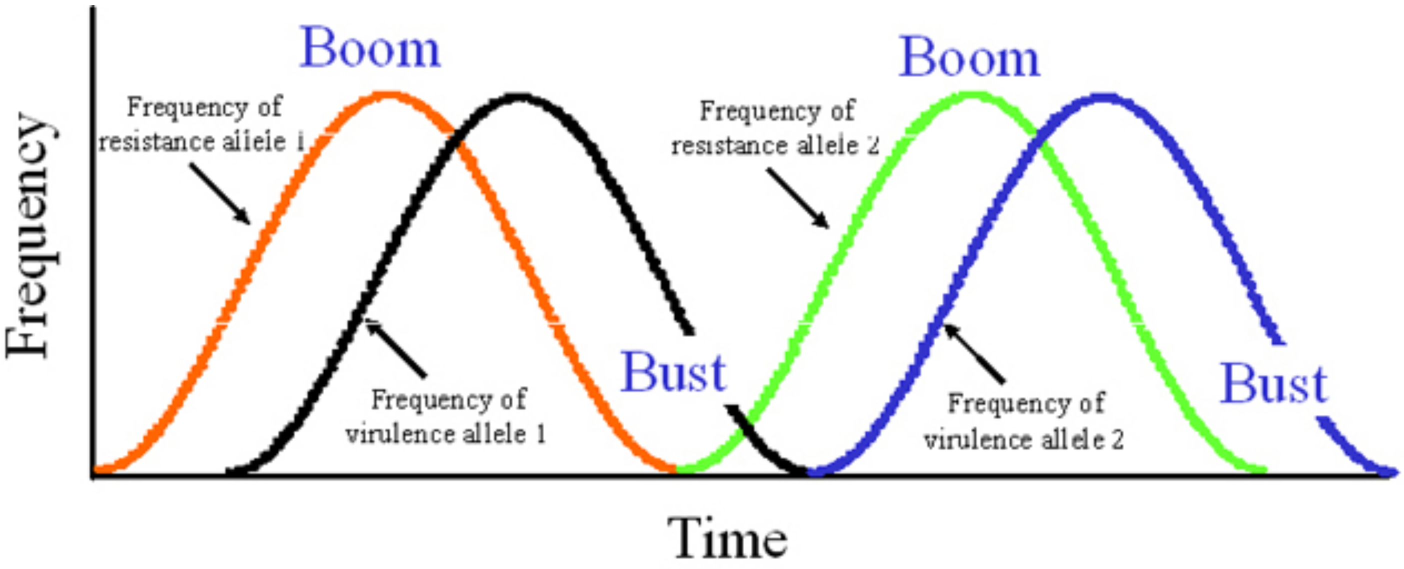
Bursts of LTR expansion

Hopi is currently active and accounts for 30% of rice genome

Half life of ~3 MYR

By examining the number of truncated LTRs, it appears that 61-78% of the DNA has been removed since insertion in the last 5 MYR

Boom & Bust Cycle Fueled by Hybridization & Stress

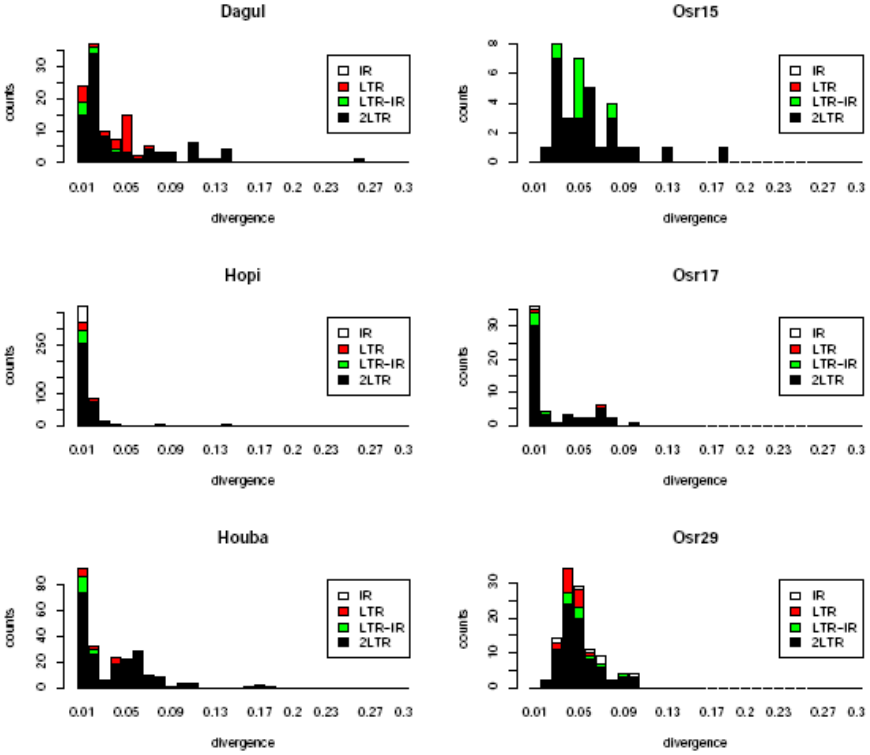


Data do not suggest stabilization

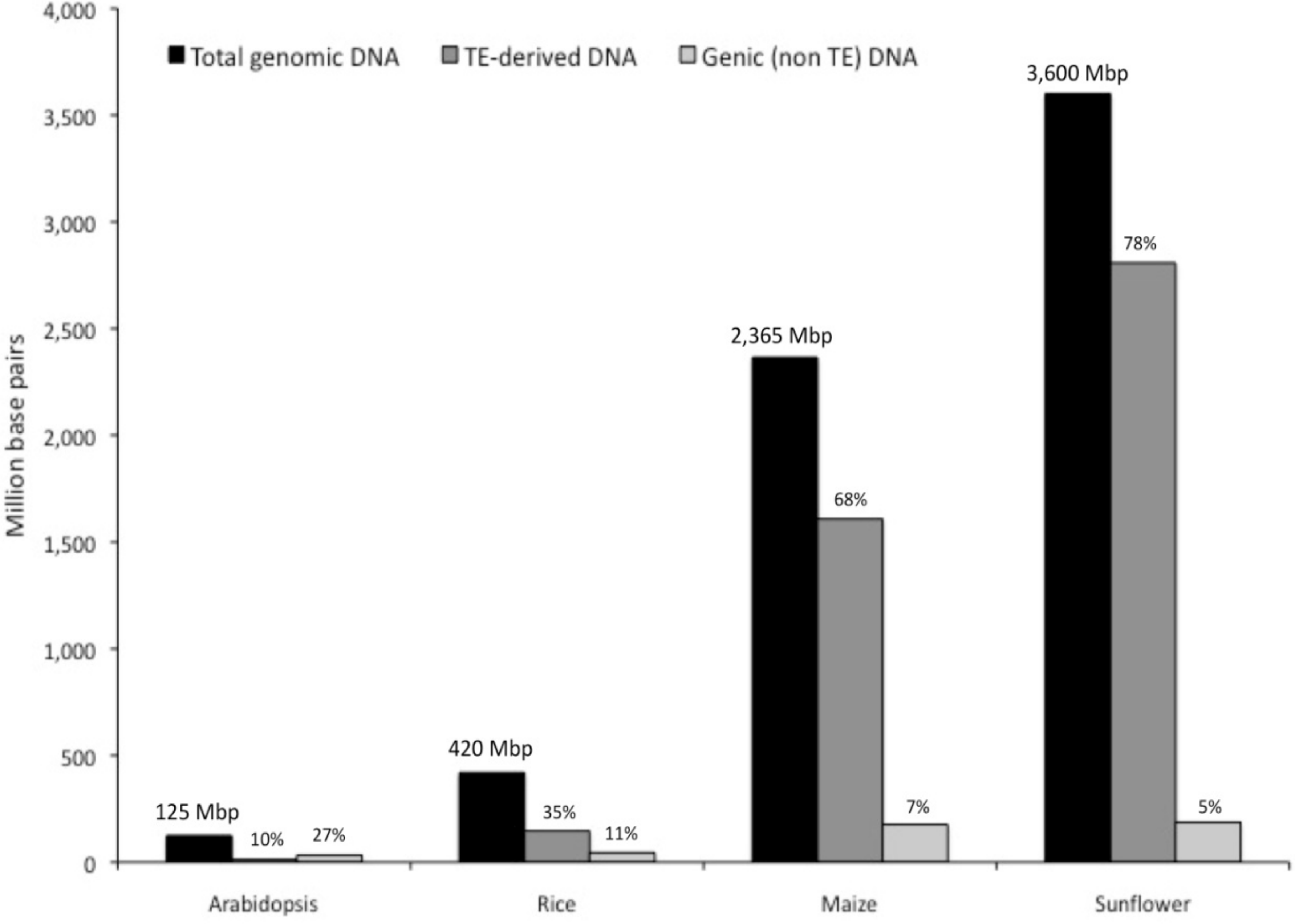
- -no old TEs
- -TEs demonstrate boom/bust patterns

TEs proliferate in naïve hosts (hybridization)

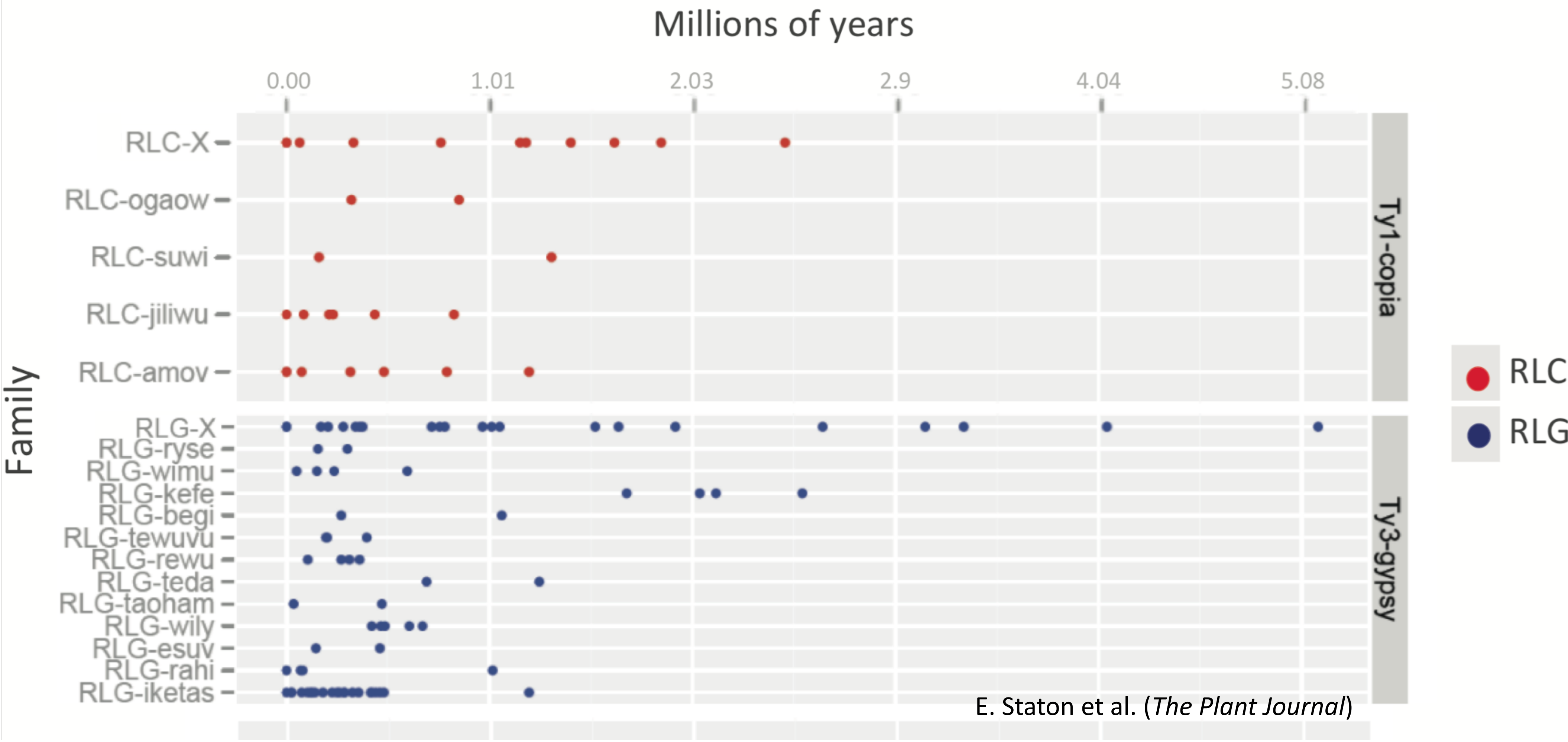
Stress overwhelms host ability to limit TEs



Large Repetitive Genomes Complicate Genome Assemblies



Rapid Turnover of LTR Retrotranspons Complicates Genome Assemblies



E. Staton et al. (*The Plant Journal*)

Unanswered Questions

Why is aneuploidy more common in flowering plants than ferns?

What is the role of TEs in adaptive evolution?

Is the proliferation of TEs following genomic stress adaptive or is it a maladaptive consequence of the breakdown of mechanisms that suppress TE amplification?

What is the metabolic cost of synthesizing DNA?