Genome Size Variation: Consequences and Evolution

Ilia Leitch and Martin Lysak

Genome size variation: consequences and evolution

(i) How genome size varies across plants

(ii) What are the consequence of this variation

(iii) How did such variation evolve

C-value paradox

• **A. Boivin**, **R.** & **C. Vendrely** (Boivin et al. 1948, Venderely & Vendrely 1948) - systematic comparisons of DNA contents in different cattle tissues (thymus, liver), liver from pig and guinea pig:

– remarkable constancy DNA amount in different tissues

– approx. double of DNA content in sperm

– 'constancy in DNA amount is PROBABLY proportional to the number of genes'

• **Mirsky & Ris (1951)** discovered totally unexpected finding: 'an aquatic salamander has 70x as much DNA as is found in a cell of the domestic fowl'

• more studies confirmed the phenomenon >> **C-value paradox (C.A. Thomas, 1971)**

- simple organisms have more DNA than complex ones

- organisms have more DNA than would be predicted from gene number

- some morphologically similar groups have divergent DNA contents

Non-coding DNA was not known in that time.

Today C-value paradox is replaced by **C-value enigma (T.R. Gregory, 2001)**

The Origin, Evolution and Proposed Stabilization of the Terms 'Genome Size' and 'C-Value' to Describe Nuclear DNA Contents

JOHANN GREILHUBER[1,*], JAROSLAV DOLEŽEL[2], MARTIN A. LYSÁK[3] and MICHAEL D. BENNETT[3]

- **Holoploid genome** – the whole chromosome set with chromosome number **n** (irrespective of polyploidy, aneuploidy etc.)
- **Monoploid genome** – one chromosome set of an organism and its DNA having the chromosome base number **x**
- **Genome size** – covering term for the amount of DNA in both holoploid and monoploid genomes

Sometimes terminology matters…

• **C-value** – DNA content of a holoploid genome with chromosome number **n**

• **1C-value** – DNA content of one non-replicated holoploid genome with chromosome number n (= the half of a holoploid non-reduced genome with the chromosome number 2n); cf. 2C-value, 4C-value,…

• **C_x-value** – DNA content of a monoploid genome with chromosome base number **x**

• **Diploids:** 1C-value = 1C_x-value

• **Polyploids:** example 2C-value of allohexaploid wheat (*Triticum aestivum*; 2n=6x=42) is 34.6 →→ 1C-value: 17.3 pg; 1C_x-value: 5.8 pg (34.6 : 6)

Remember ! 1 pg = 980 Mbp

'C' means Constant (Swift H. 1950**.** *Proc. Natl. Acad. Sci. USA* 36: 643-654.)

Plant DNA C-values database

www.kew.org/genomesize/homepage.html

5150 species

Land plants

4427 angiosperms

207 gymnosperms

87 pteridophytes

176 bryophytes

Algae

91 Chlorophyta

44 Phaeophyta

118 Rhodophyta

Kew

PLANTS PEOPLE POSSIBILITIES

C-values in angiosperms range nearly 2000-fold

Genlisea margaretae

1C = 0.065 pg

Greilhuber *et al*. 2006. Plant Biology 8: 770-777

Utricularia gibba

1C = 0.990 pg

Greilhuber *et al*. 2006. Plant Biology 8: 770-777

Trillium ludovicianum

1C = 11.15 pg

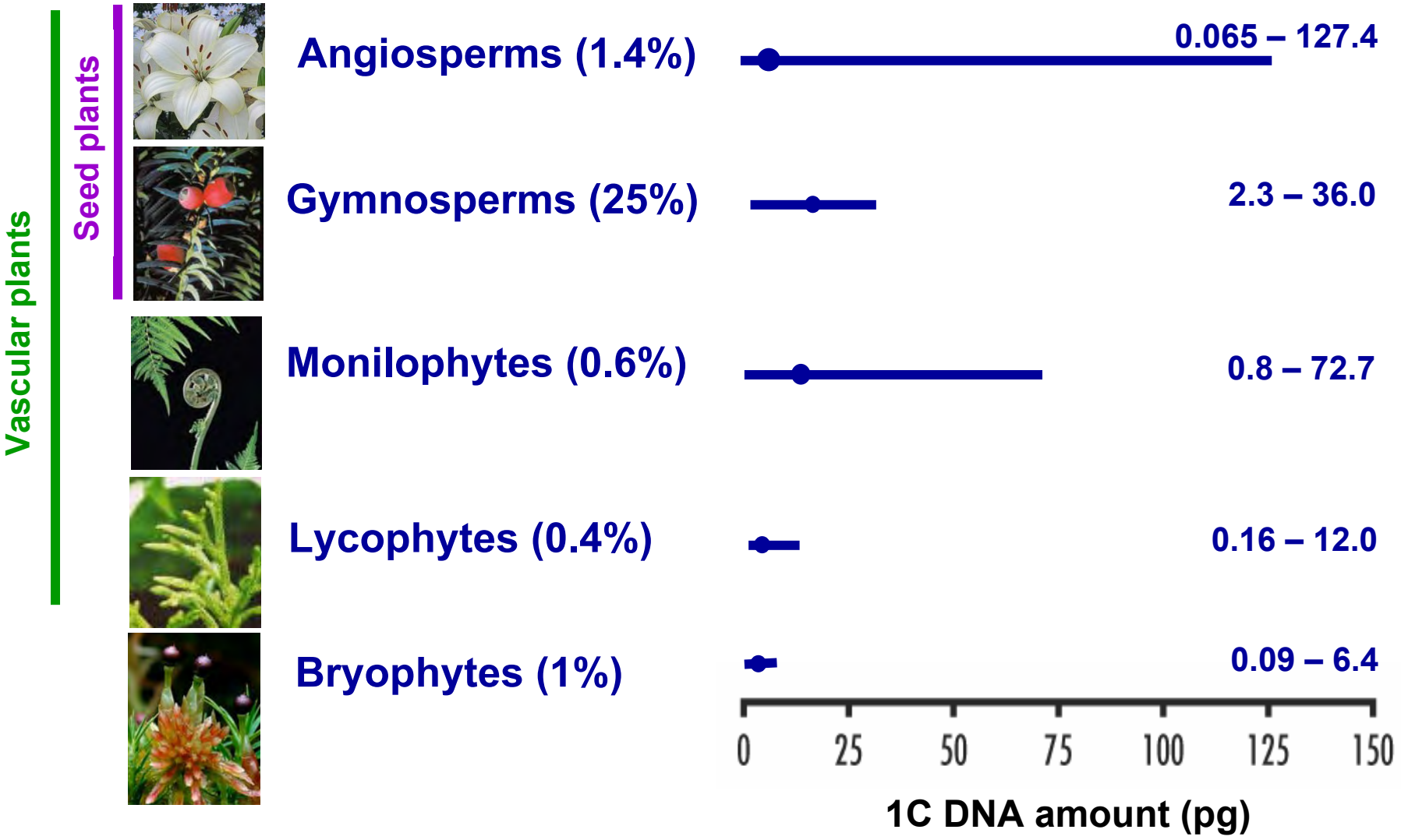
Gill. Taxon 22: 1331-1338

Fritillaria assyriaca

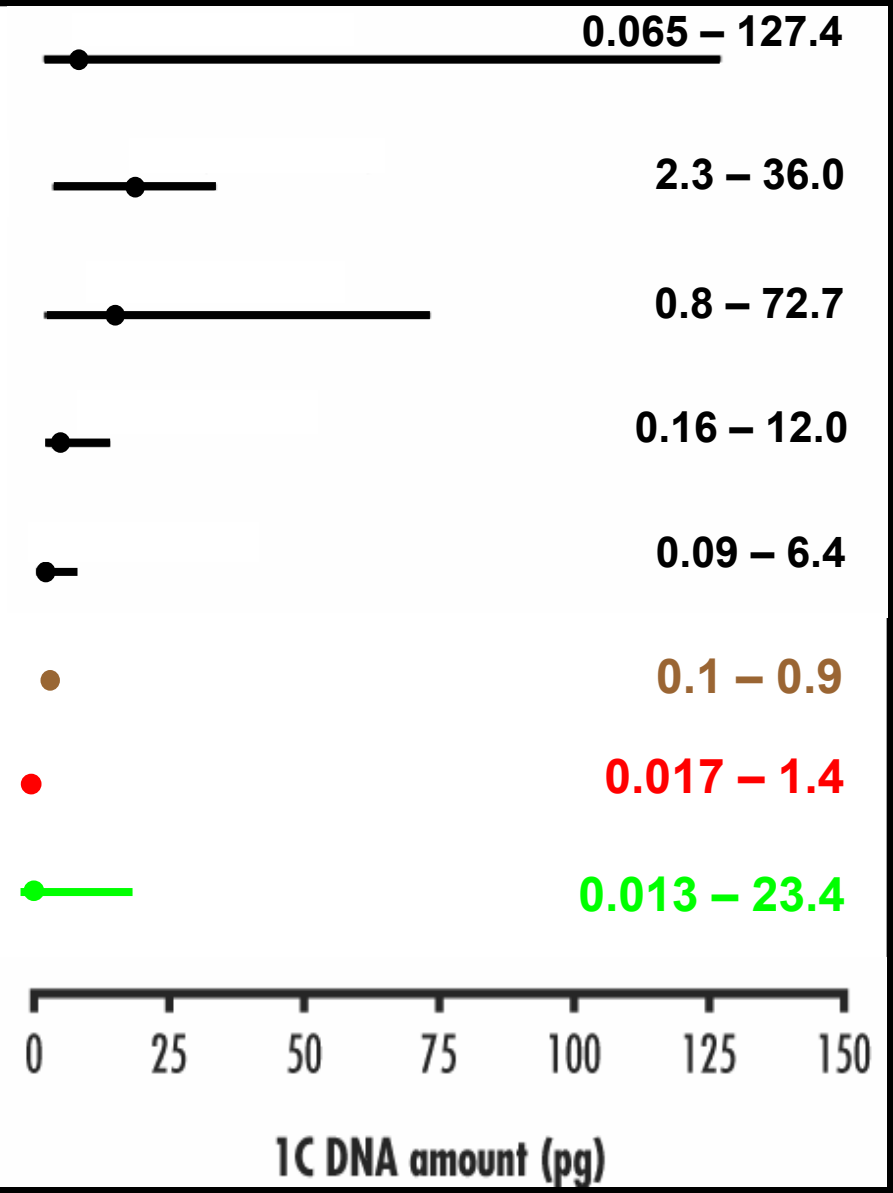
1C = 127.4 pg

Bennett. 1972. Proc. Roy. Soc. Lond. B 181: 109-135.

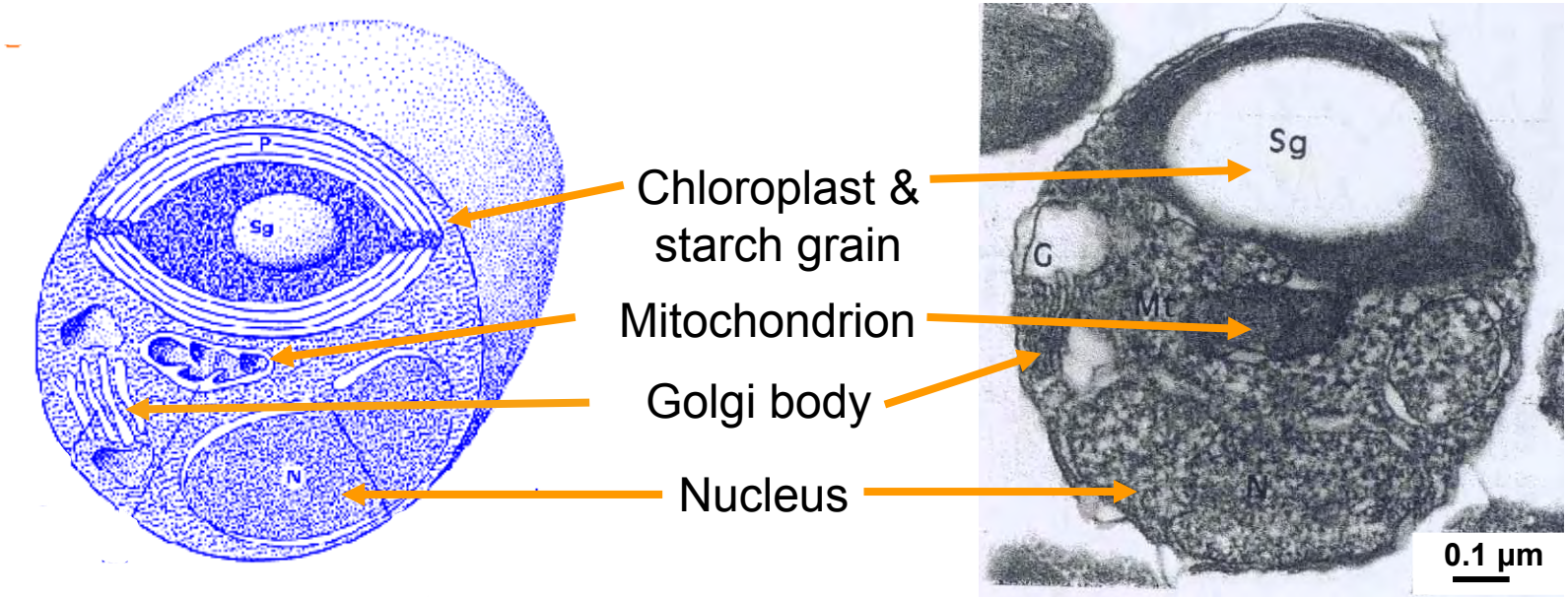
Range of DNA amounts in land plants



Range of DNA amounts in algae



Smallest, free living plant
Ostreococcus tauri (Prasinophyta)



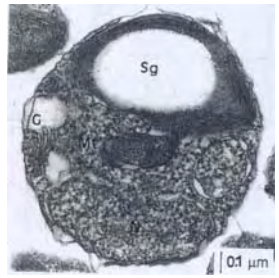
1C = 0.013 pg (12.5 Mb)

(smallest chromosome of *A. thaliana* has 17 Mb)

Chrétiennot-Dinet *et al.* (1995)
Phycologia 34: 285-292.

Unicellular green alga *Ostreococcus tauri* (Prasinophyceae): the world's smallest free-living eukaryote

Derelle et al. (2006) **Genome analysis of the smallest free-living eukaryote** *Ostreococcus tauri* **unveils many unique features**. *PNAS* 103



The green lineage is reportedly 1,500 million years old, evolving shortly after the endosymbiosis event that gave rise to early photosynthetic eukaryotes.

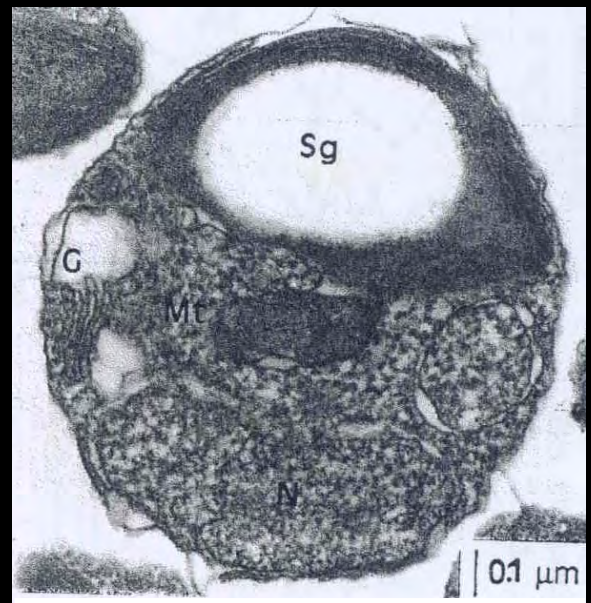
Overall, the 12.56-Mb nuclear genome has an extremely high gene density, in part because of extensive reduction of intergenic regions and other forms of compaction such as gene fusion.

However, the genome is structurally complex. It exhibits previously unobserved levels of heterogeneity for a eukaryote. Two chromosomes differ structurally from the other eighteen. Both have a significantly biased G+C content, and, remarkably, they contain the majority of transposable elements.

Genome size variation

Fritillaria assyriaca
1C = 127.4 pg

Genlisea margaretae
1C = 0.065 pg



Ostreococcus tauri
1C = 0.01 pg

C-value enigma

Gregory TR. 2001. Coincidence, co-evolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews* **76:** 65-101.

Mechanisms of DNA amount increase in plants

- (Paleo)polyploidy
- Amplification of transposable elements, TEs (especially retrotransposons)

Bennetzen & Kellogg (1997)

Do Plants Have a One-Way Ticket to Genomic Obesity?

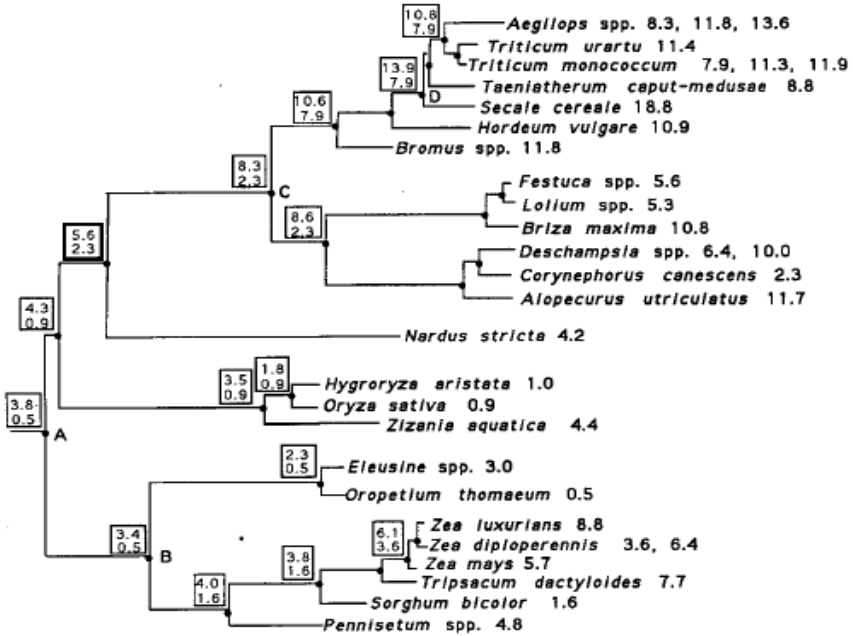


Figure 1. Phylogeny of Some Diploid Grasses for Which Genome Size Is Known.

We do agree, however, that a mechanism that leads to rapid increases in genome size does exist. Thus, unless evidence for a comprehensive mechanism for removing interspersed repetitive DNAs is found, and/or strong selective pressures for reducing genome size can be determined, we must conclude that plants may indeed have a one way ticket to larger genome sizes.

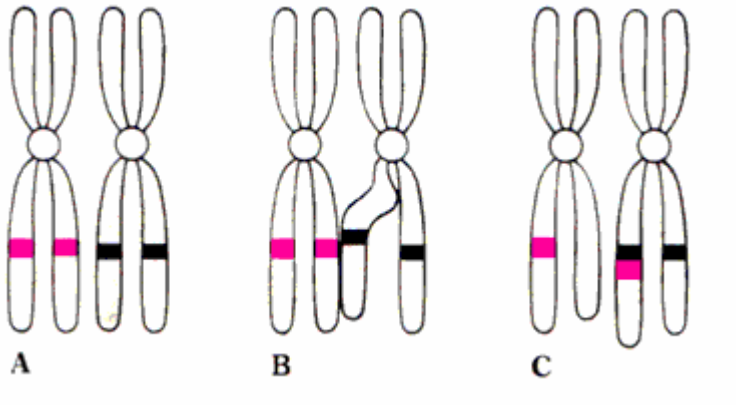
Approx. linear relationship between genome size and the amount of TEs

Mechanisms of DNA amount decrease in plants

Until recenty, only mechanisms responsible for DNA amount increases were known („one-way ticket to genomic obesity“ - Bennetzen and Kellogg, 1997)

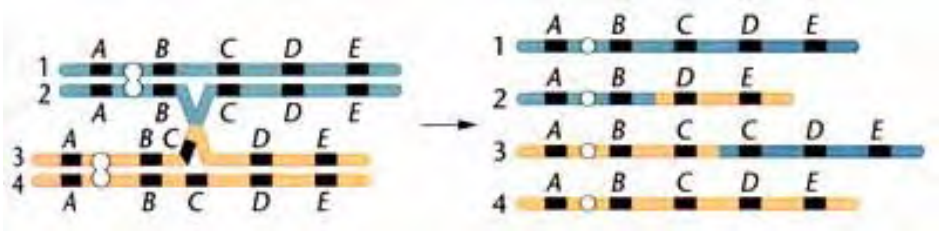
- **recombinational mechanisms**
- **repair of double-stranded breaks (DSBs)**
- **chromosome rearrangements (loss of centromeres, NORs)**

DNA loss (and DNA gain) through unequal homologous recombination (unequal crossing over)



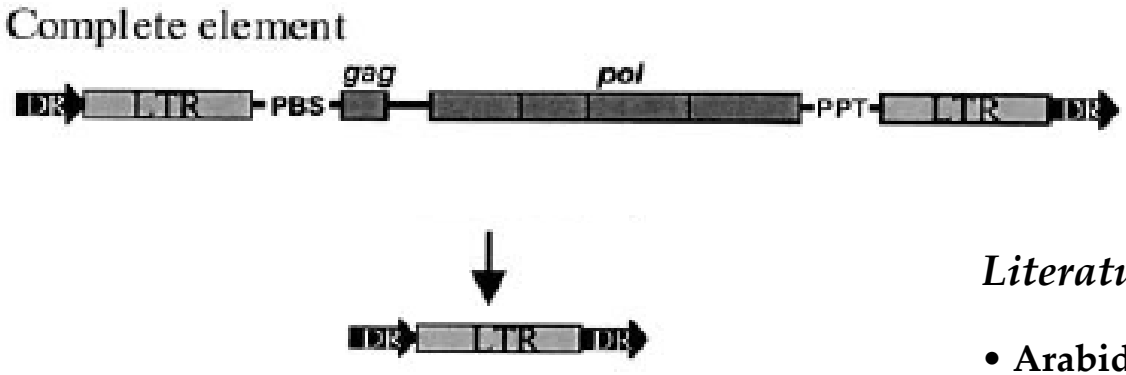
Sometimes during meiosis two chromatids from homologous chromosomes **(A)** are misaligned during a crossocer event **(B)** as a result, one chromatid gained a duplicated region and the another lost a deleted region **(C)**. The duplication as well as the deletion are inherited by resulting gametes.

…and another example



Several studies showed that plant genomes can be downsized through structural modifications of LTR retrotransposons

Structure of a complete element, with a direct repeat (DR) of flanking target-site DNA, two long terminal repeats (LTRs), a primer-binding site (PBS), and polypurine tract (PPT) needed for element replication and encoded gene products (*gag, pol*)



Devos *et al.* 2002

Literature

- **Arabidopsis**

Devos et al. (2002) Genome Res 12: 1075-1079.

- ***Hordeum*** **(barley)**

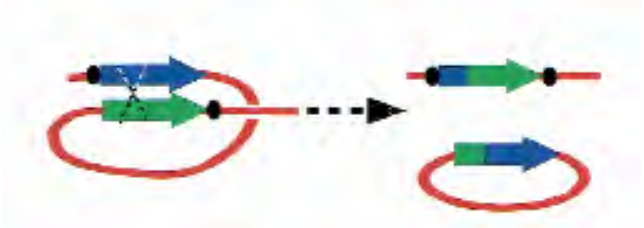
Shirasu et al. (2000), Genome Res 10: 908-915.
Vicient et al. (1999), Plant Cell 11: 1769–1784.
Kalendar et al. (2000), Proc Natl Acad Sci USA 97: 6603-6607.

- ***Oryza sativa*** **(rice)**

Vitte & Panaud (2003), Mol Biol Evol 20: 528–540.

Intra-chromosomal (intra-element) recombination between LTRs of LTR retrotransposons

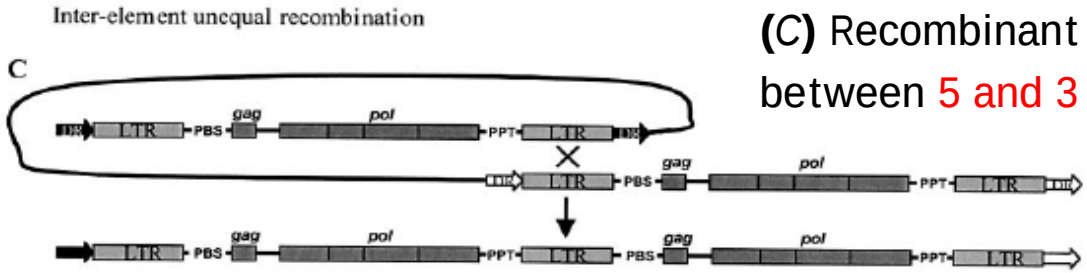
An excess of LTRs (solo LTRs) was found across the genus *Hordeum* (Kalendar *et al*. 2000, Shirasu *et al*. 2000)



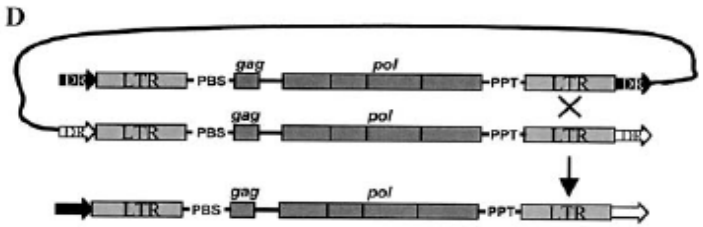
Model for recombination between two LTRs, resulting in a single recombinant LTR in the genome and a closed circle, bearing an LTR and the internal domain, which is then lost.

Unequal intrastrand recombination between LTR retrotransposons (illegitimate recombination)

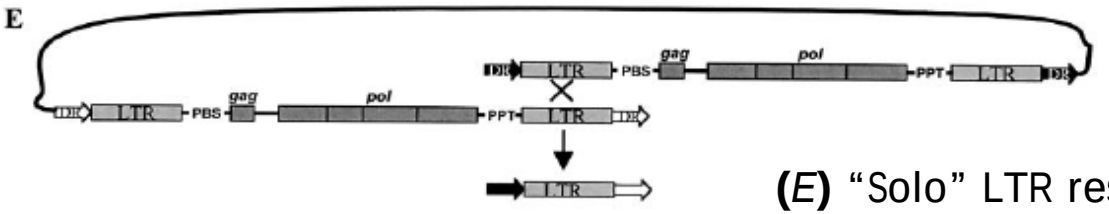
Arabidopsis: Devos *et al.* (2002)



(*C***)** Recombinant element resulting from recombination between 5 and 3 LTRs of two adjacent elements.



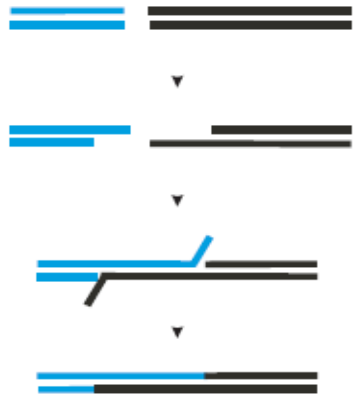
*(D***)** "Complete" element resulting from unequal recombination between two 5 LTRs, two 3 LTRs, or the internal regions of two elements.



(*E***)** "Solo" LTR resulting from recombination between 3 and 5 LTRs of two elements.

Double-strand break repair and genome size

Background



Puchta (2005) Double-strand breaks (DSBs) have to be eliminated before genomes can be replicated. Therefore, the repair of DSBs is critical for the survival of all organisms. Generally, DSBs can be repaired via two different pathways:

- homologous recombination (HR)
- non-homologous end-joining (NHEJ; also known as illegitimate recombination)

Literature

Kirik et al. (2000) *EMBO Journal* 19: 5562-5566.
Orel et al. (2003) *Plant Journal* 35: 604-612.
Orel and Puchta (2003) *Plant Mol Biology* 51: 523-531.
Puchta (2005) *J Exp Bot* 56: 1-14.

Double-strand break repair and genome size

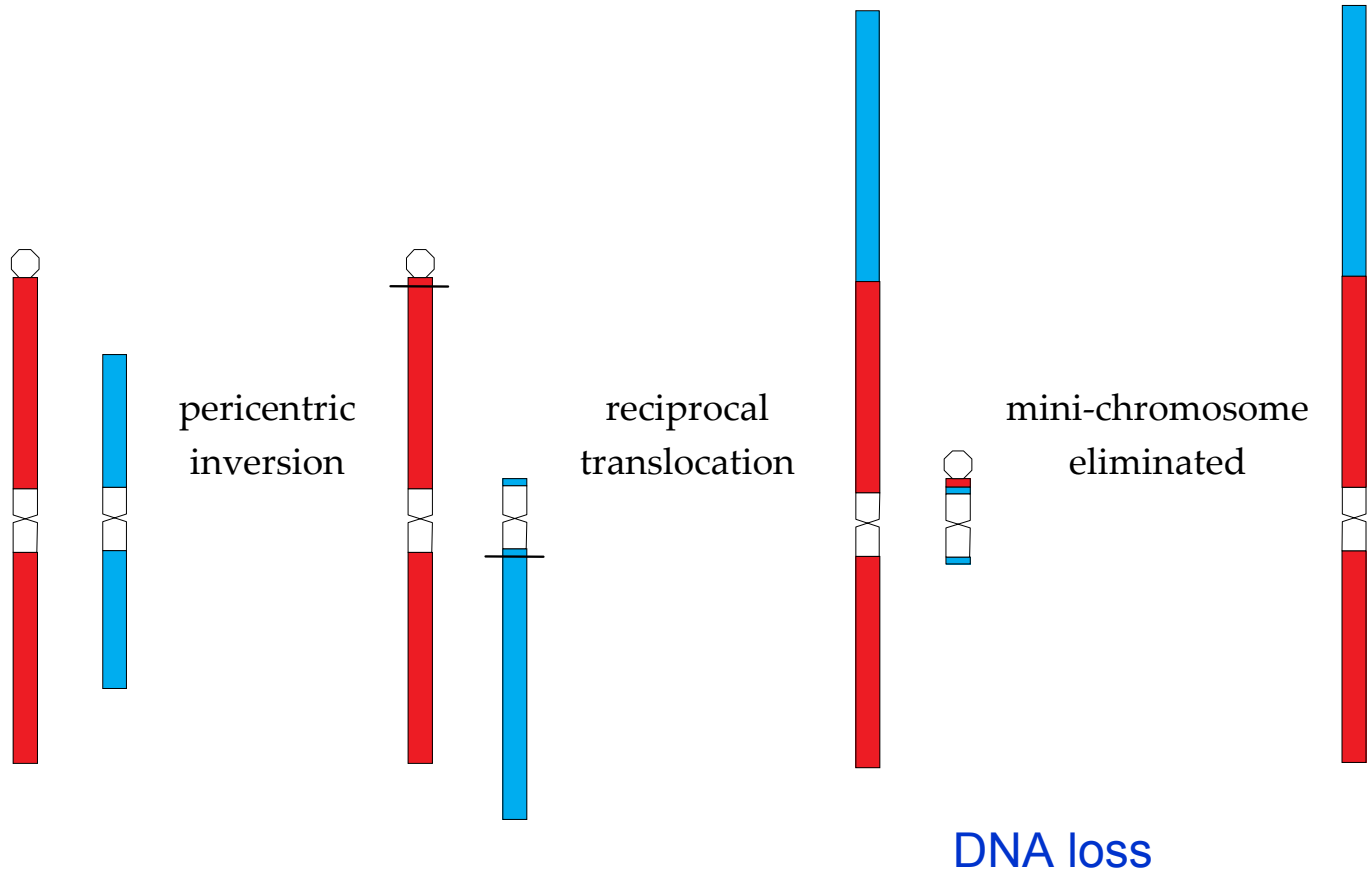
Kirik et al. (2000) A comparison of deletion formation in somatic cells of tobacco and Arabidopsis, two plant species with an over 20-fold difference in genome size.

Surprisingly large differences in DSB repair were found:
- the overall length of the deletions was about one-third shorter in tobacco than in Arabidopsis. Thus, there is an inverse correlation between genome size and the medium length of deletions could be detected.

Reasons?

Orel and Puchta (2003) During DSB repair the size of a deletion depends on the processing of DNA ends. If broken ends are not religated directly the processing of such ends might result in the loss of DNA at the break site. Depending on the efficiency of DNA degradation more or less information will be lost. Indications were found that plasmid DNA is degraded to a lesser extent in tobacco than Arabidopsis.

Chromosome number reduction via pericentric inversion–reciprocal translocation events

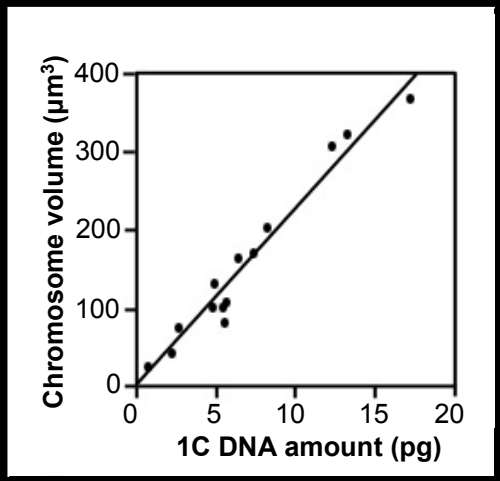


Is there genome size equilibrium ?

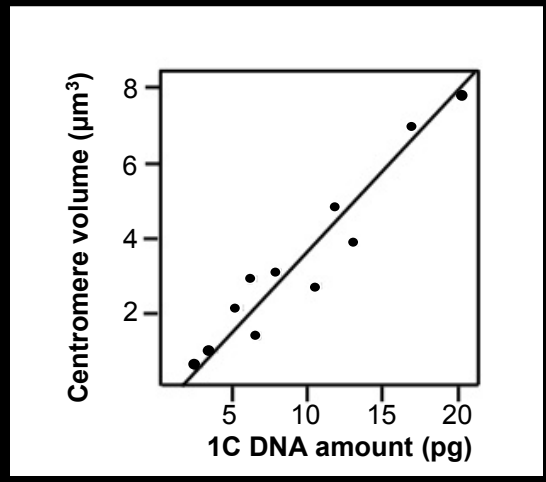
Genome size
INCREASE

Genome size
DECREASE

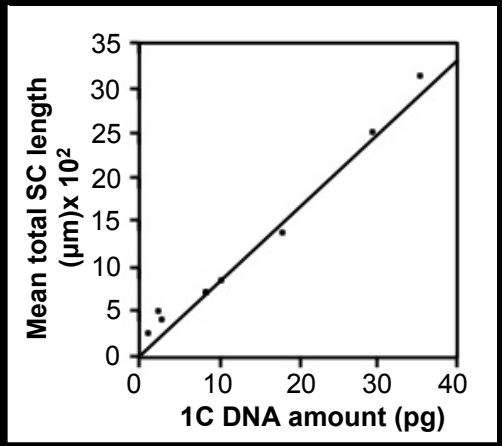
Variation of genome size: Consequences at nuclear level



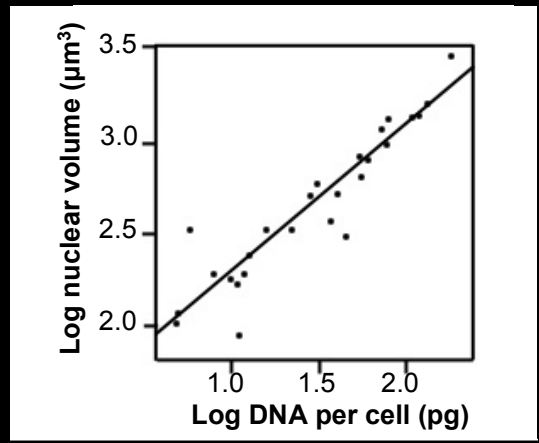
Bennett *et al.* 1983.
J. Cell Sci. **63:** 173-179.



Bennett *et al.* 1981.
J. Cell Sci. **47:** 91-115.

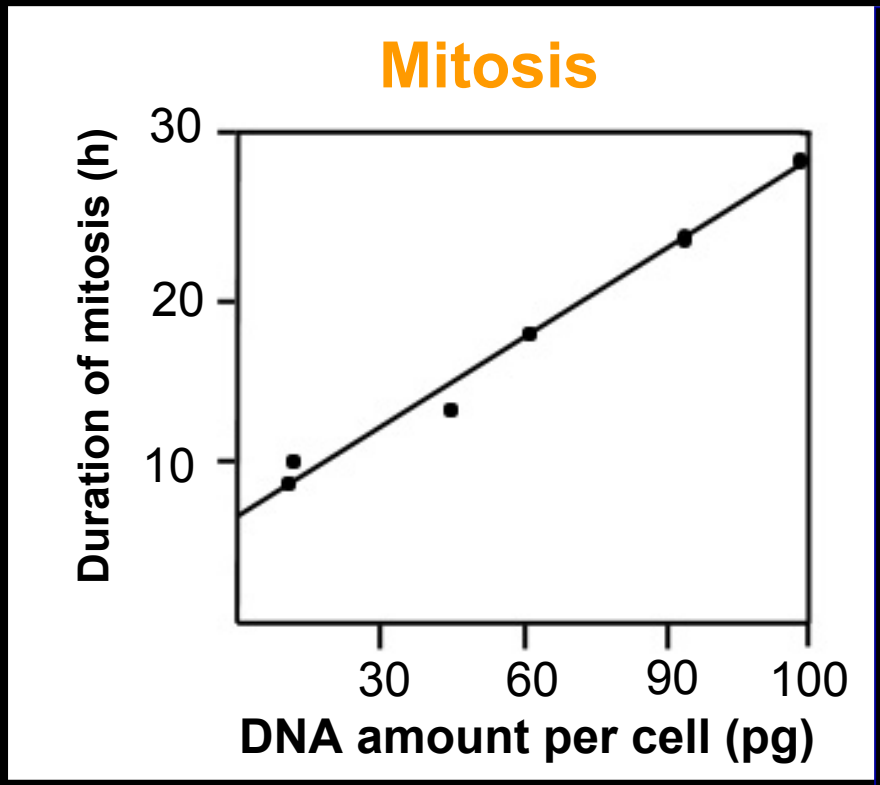


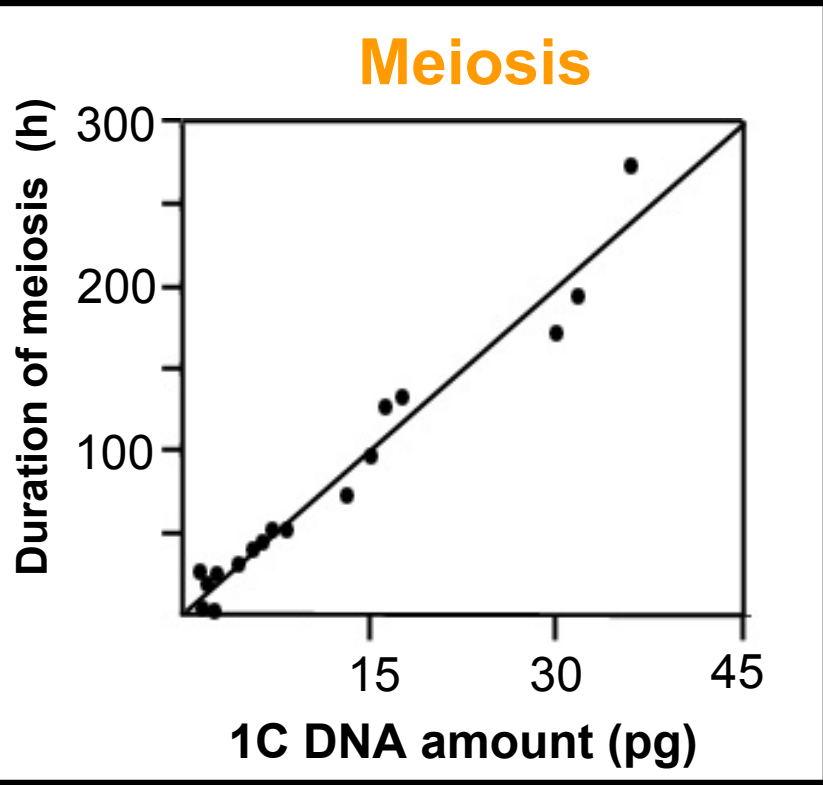
Anderson *et al*. 1985.
Exp. Cell Res. **156:** 367-378.



Baetcke *et al*. 1967.
Proc. Natl. Acad. Sci. **USA 58:** 533-540.

Variation of genome size: Consequences of timing

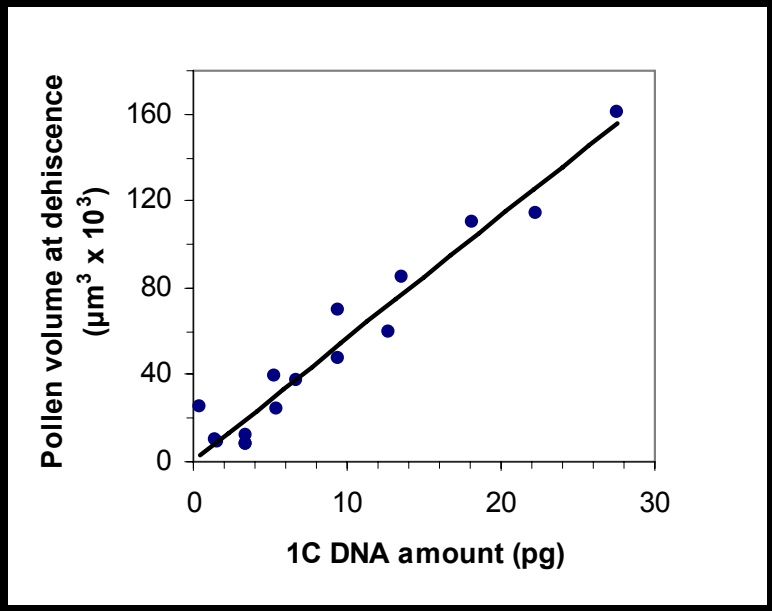


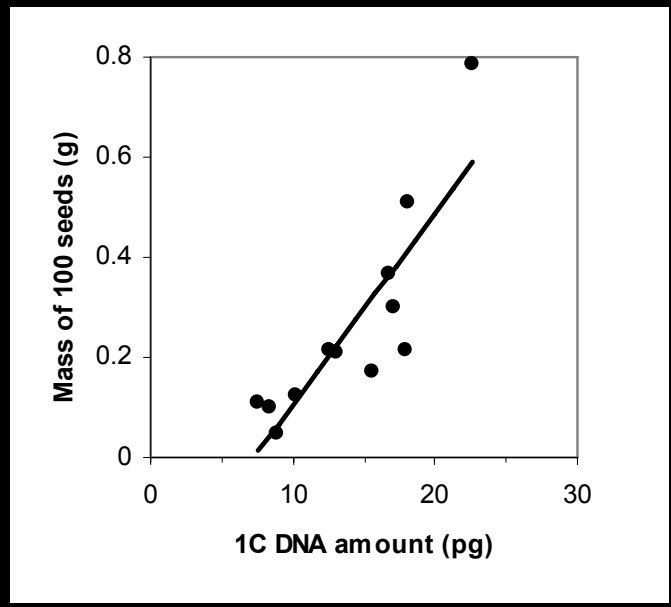


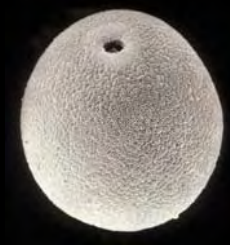
Van't Hof & Sparrow AH. 1963.
Proc. Natl. Acad. Sci. USA **49:** 897-902.

Bennett MD. 1977.
Phil. Trans. Roy. Soc. B **277:** 201-277.

Variation of genome size: Consequences at cell and tissue level







Relationship between pollen volume and DNA amount in 16 grass species.

Bennett *et al*. 1972

Relationship between seed weight and DNA amount in 12 *Allium* species.

Bennett *et al*. 1972

Consequences of variation in DNA amount

Whole plant level

a) **Life cycle options**

b) Life strategy options

c) Ecology options

d) Coping with environmental change

Consequences of variation in DNA amount

Whole plant level

a) Life cycle options

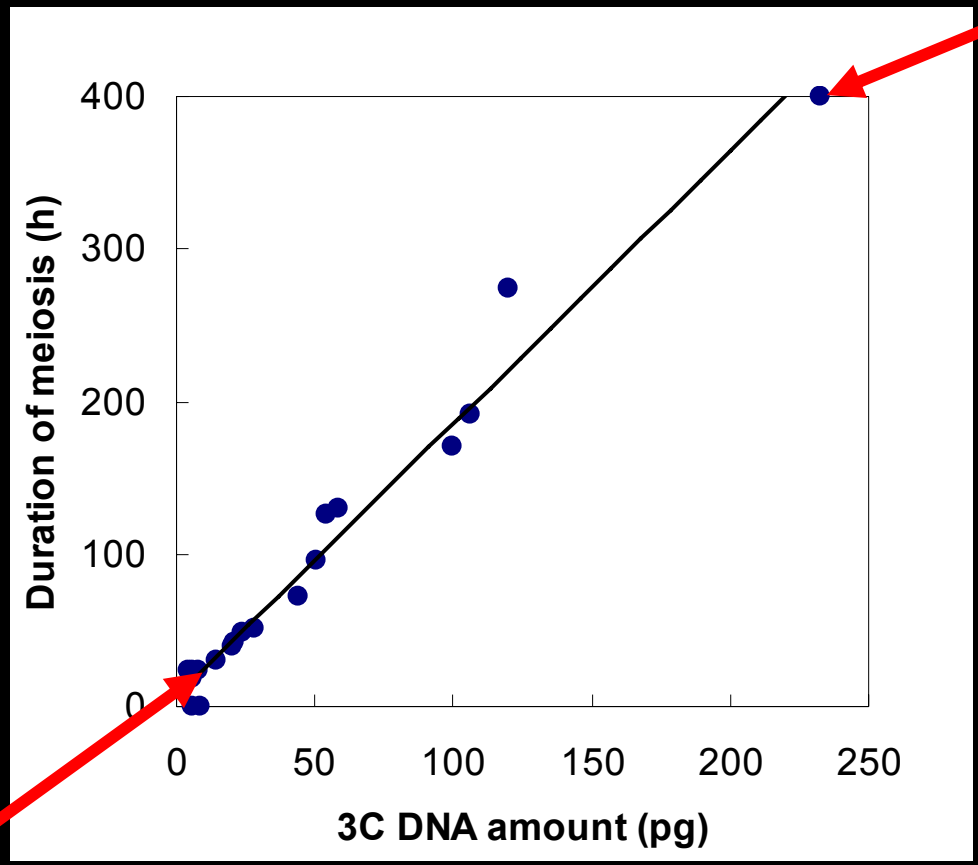
Bennett MD. 1972.

Nuclear DNA content and minimum generation time in herbaceous plants.

Proceedings of the Royal Society of London Series B-Biological Sciences **181:** 109-135.

Consequences: life cycle options

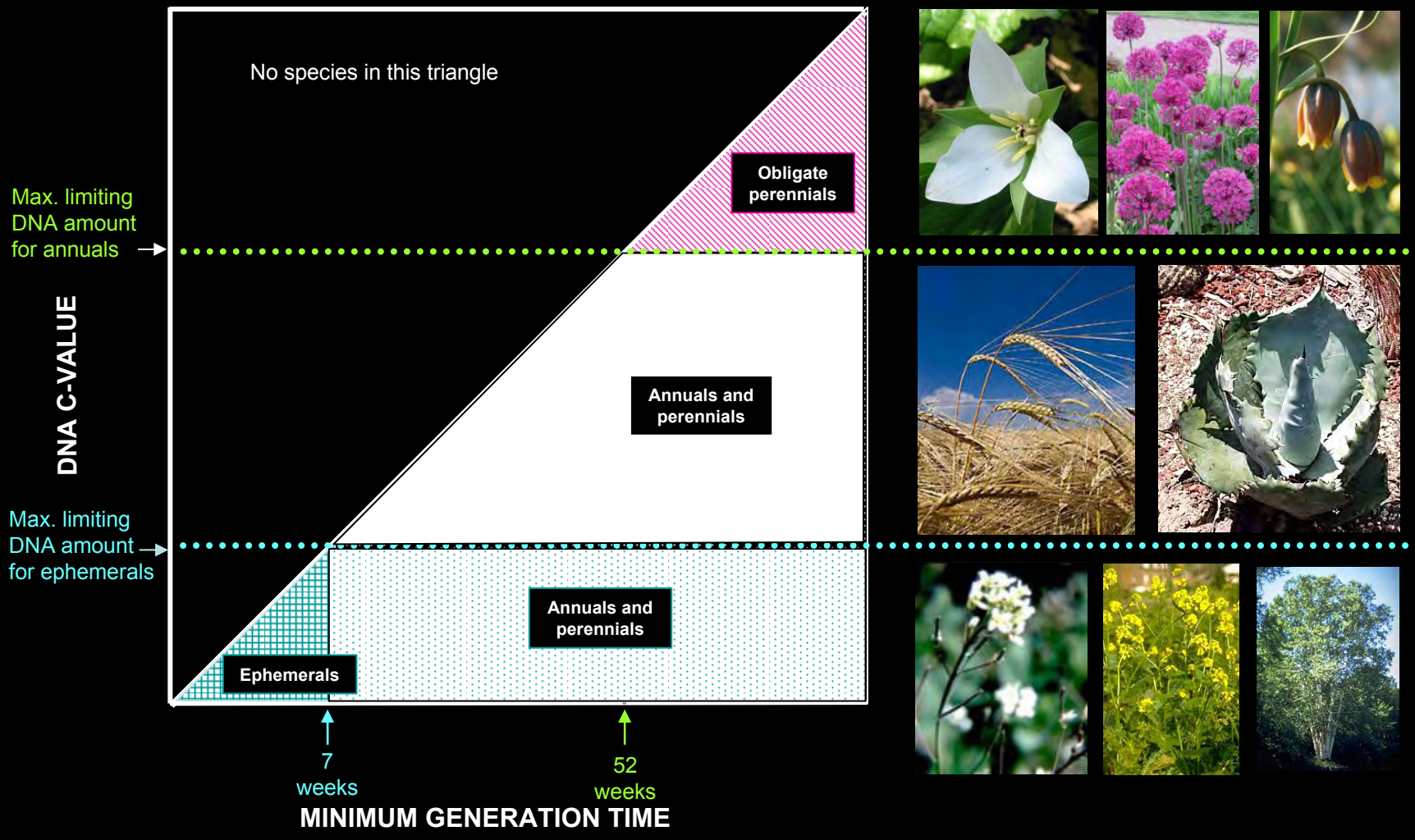
Arabidopsis thaliana
1C = 0.16 pg



Fritillaria meleagris
1C = 70.7 pg

Bennett MD. 1977.
Phil. Trans. Roy. Soc. B **277:** 201-277.

Consequences: life cycle options



Bennett MD. 1972. *Proc. Roy. Soc. Lond. B* **181**: 109-135

Consequences of variation in DNA amount

Life cycle options:

Conclusions

- DNA amount can impose limits on the type of life cycle a species can display
- Species with small genomes may be ephemerals, annuals or perennials
- Species with large genomes are restricted to being obligate perennials

Consequences of variation in DNA amount

Whole plant level

a) Life cycle options

b) Life strategy options

c) Ecology options

d) Coping with environmental change

Consequences of variation in DNA amount

Whole plant level

b) Life strategy options:

Potential to become a weed

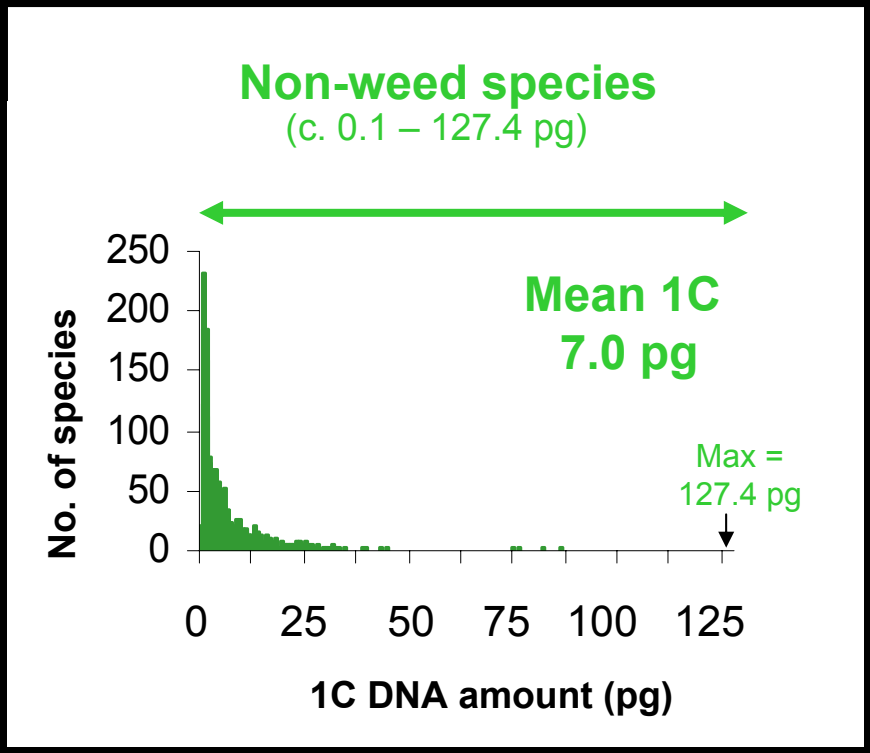
Bennett, Leitch & Hanson. 1998.
DNA amounts in two samples of angiosperm weeds.
Annals of Botany **82:** 121-134.

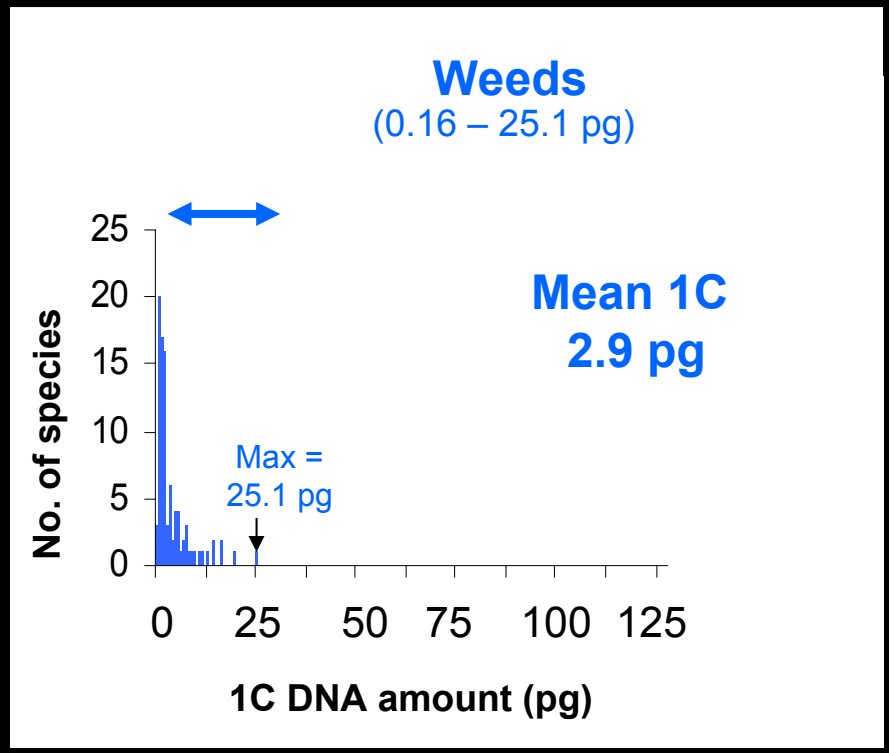
Consequences: option to be a weed

Method

DNA amounts for 156 angiosperms recognised as weeds compared with 2685 non-weed species

Consequences: option to be a weed





Bennett, Leitch & Hanson. 1998.
DNA amounts in two samples of angiosperm weeds.
Annals of Botany **82:** 121-134.

Success of an invasive weed

- Rapid establishment or completion of reproductive development
- Short generation time
- Rapid production of many small seeds

Consequences of variation in DNA amount

Whole plant level

a) Life cycle options

b) Life style options

c) Ecology options

d) Coping with environmental change

Genome size and latitude

Pop.	Several *Picea sitchensis*	Miksche 1967, 1971
Sp.	Tropical vs. temperate grasses	Avdulov 1931
Sp.	329 tropical vs. 527 temperate plants	Levin and Funderburg 1979
Sp.	17 Poaceae and 15 Fabaceae crops	Bennett 1976
Pop.	24 *Berberis* in Patagonia	Bottini *et al*. 2000

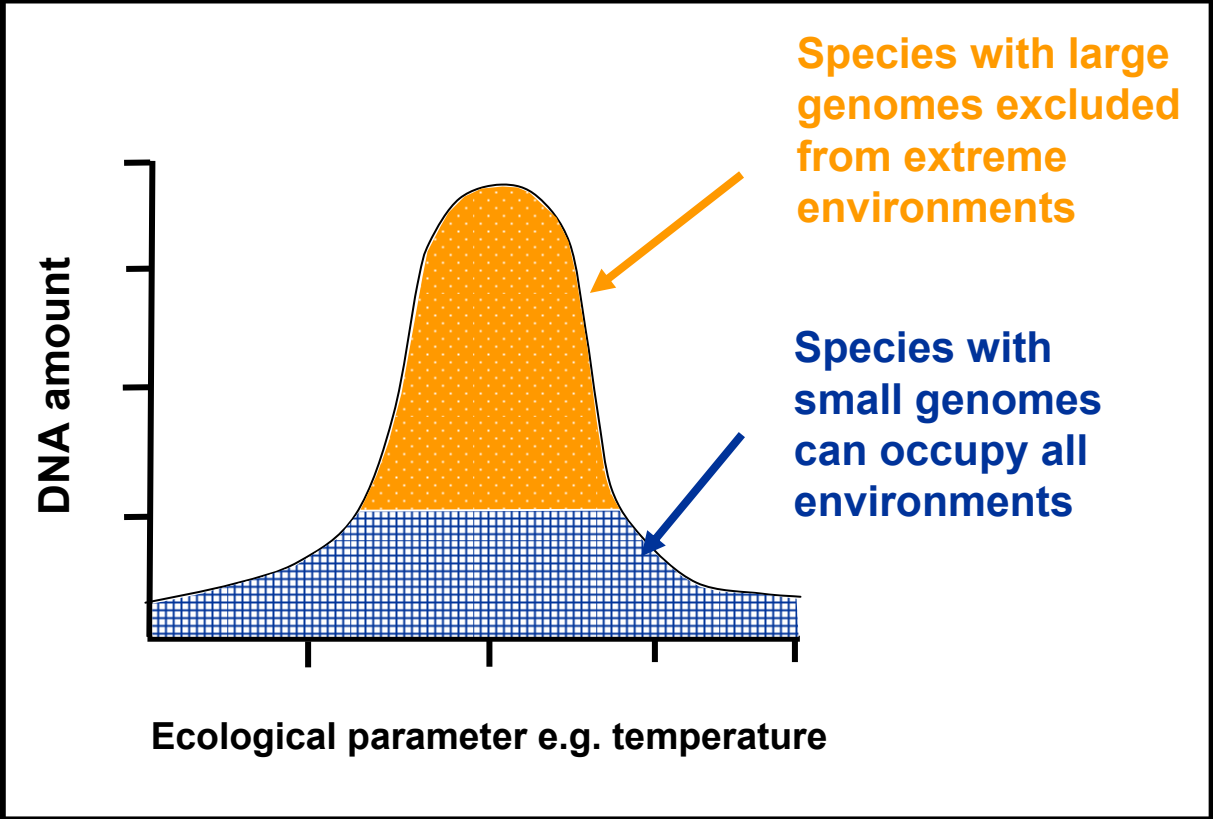
+ correlation

Consequences: ecology options

401 species in the state of California

Knight & Ackerly. 2002.
Variation in nuclear DNA content across environmental gradients: a quantile regression analysis.
Ecology Letters **5:** 66-76.

Consequences: ecology options



Knight & Ackerly. 2002.
Ecology Letters **5:** 66-76.

Consequences: ecology options

Summary

- The relationship between genome size and environmental factors is not uniform but appears to be stronger for species with large genomes
- Species with large genomes are excluded from extreme environments

Consequences of variation in DNA amount

Whole plant level

a) Life cycle options

b) Life style options

c) Ecology options

d) Coping with environmental change

(i) Pollution

(ii) Threat of extinction

Consequences: Coping with environmental change

Pollution: The question

What effect does genome size have on the survival of plants in lead polluted soils?

B. Vilhar, T. Vidic and J. Greilhuber

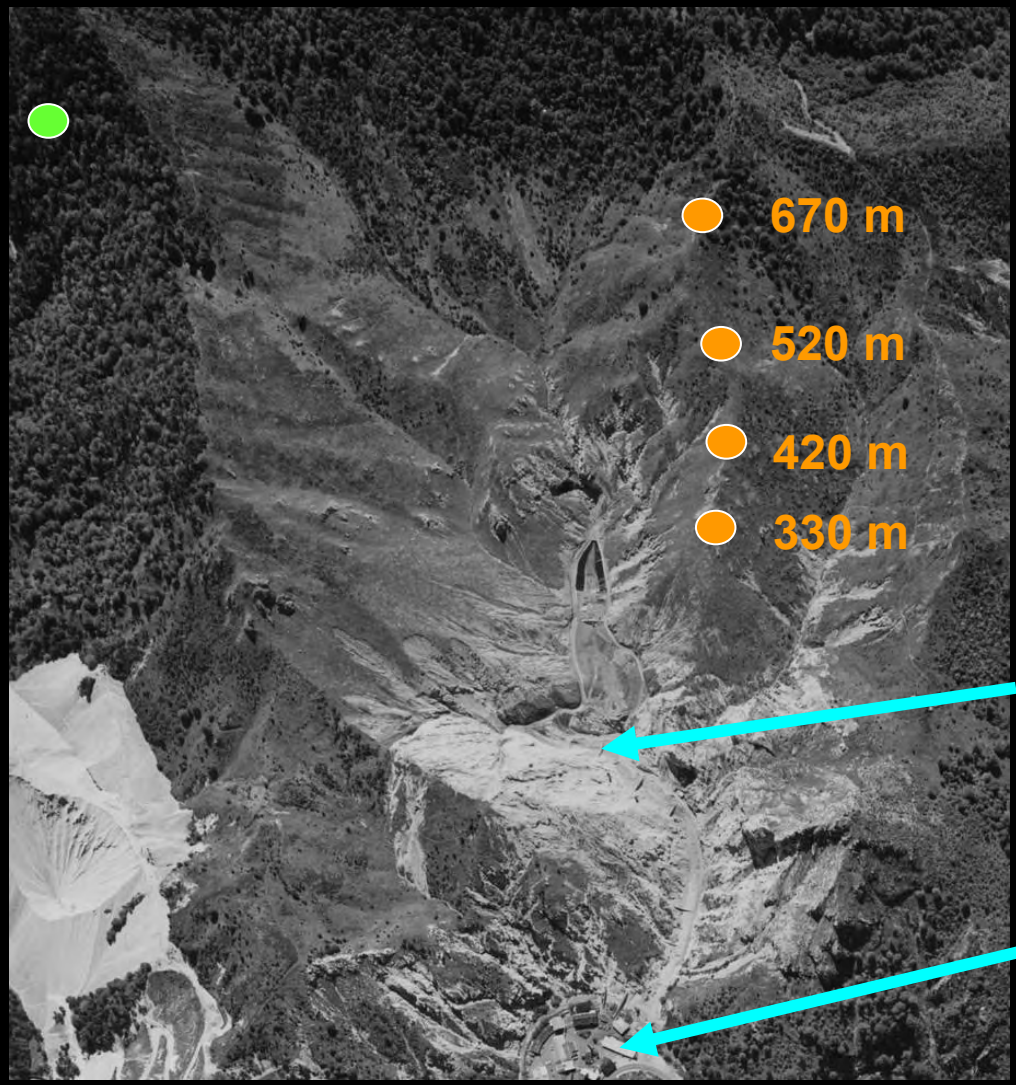
Consequences: Genome size and pollution

Dolina smrti in Slovenia

Consequences: Genome size and pollution

Reference plot

Conc. lead in soil = 0.1%



Conc. of lead in soil

0.7%

1.5%

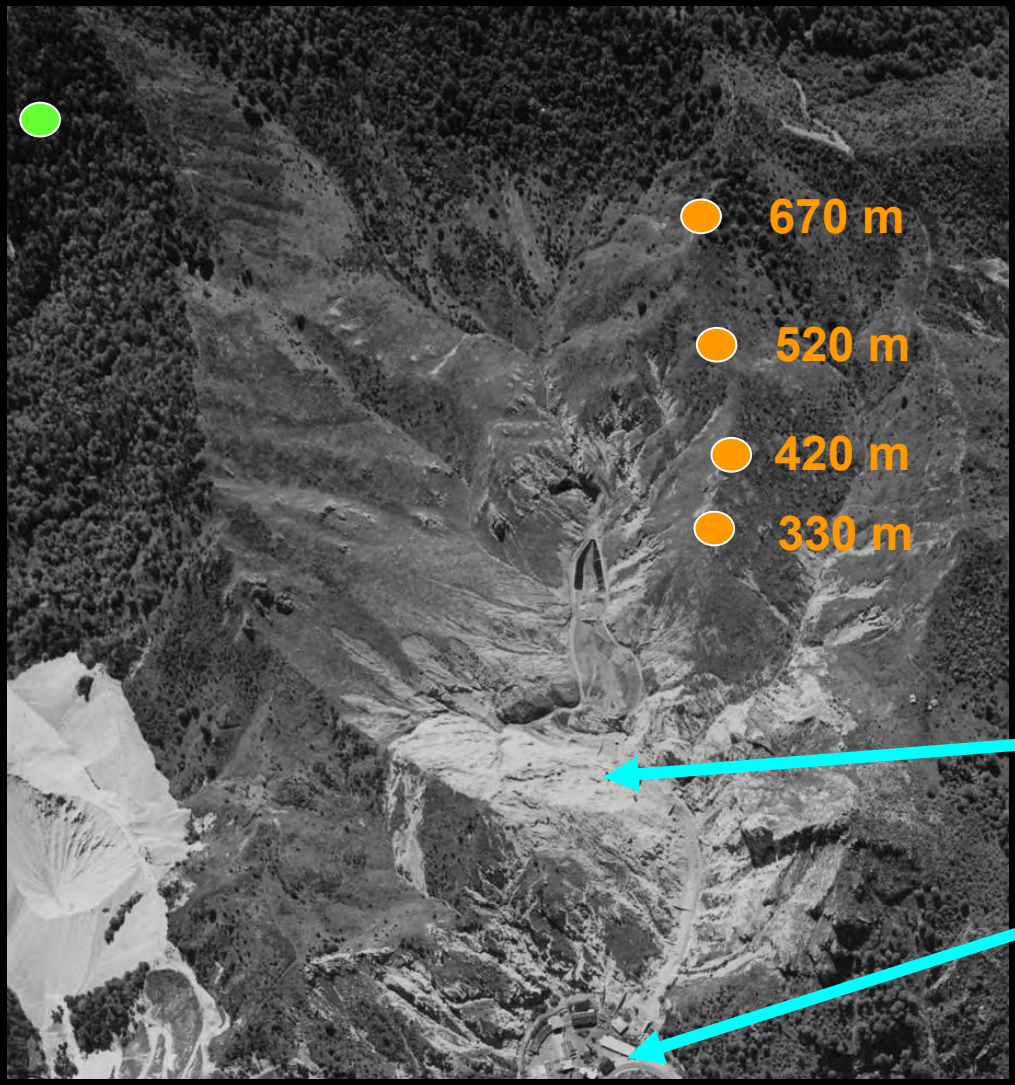
2.0%

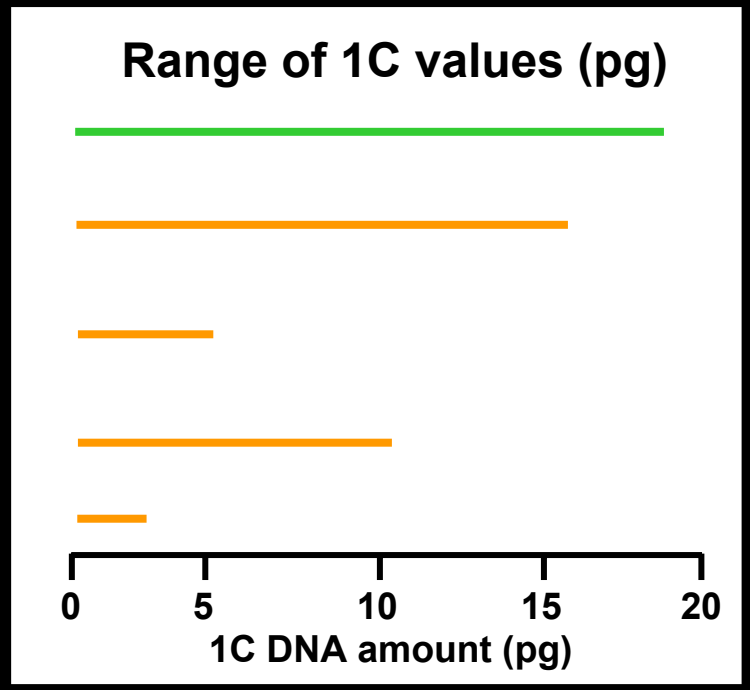
3.3%

Smelter chimney

Lead smelter

Consequences: Genome size and pollution



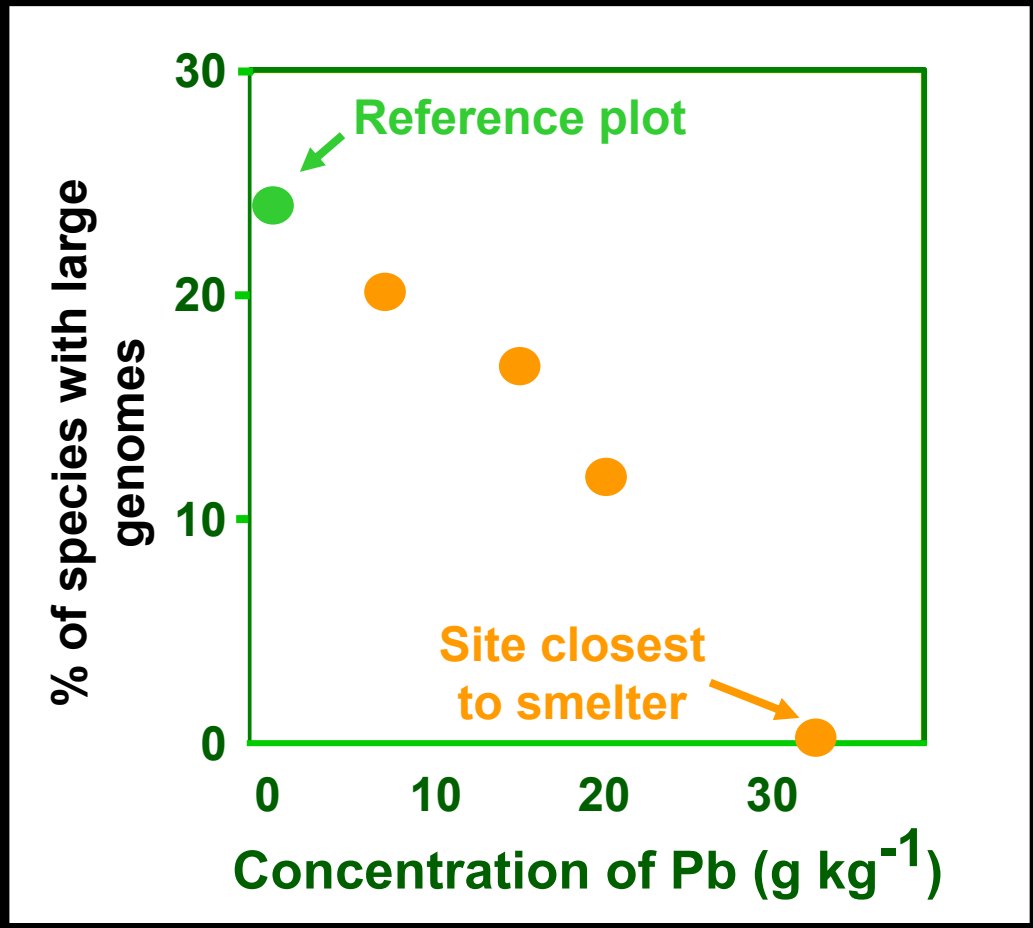


Smelter chimney

Lead smelter

Consequences: Genome size and pollution

Percentage of species with large genomes in individual plots



Consequences: Genome size and pollution

Conclusions

Species with large genomes are at selective disadvantage in extreme environmental conditions induced by pollution.

Consequences of variation in DNA amount

Whole plant level

a) Life cycle options

b) Life style options

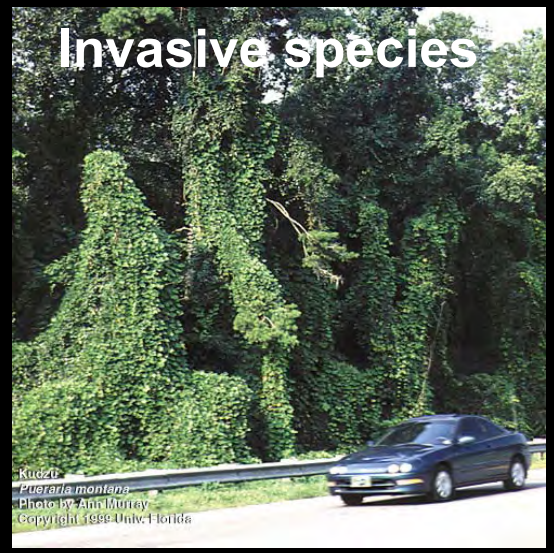
c) Ecology options

d) Coping with environmental change

(i) Pollution

(ii) Threat of extinction

Consequences:
Genome size and threat of extinction



Is genome size important?

Vinogradov AE. 2003.
Selfish DNA is maladaptive: evidence from the plant Red List.
Trends in Genetics **19:** 609-614.

Consequences: Genome size and threat of extinction

Data and analysis

→ 3036 species →

→ Global concern = 305

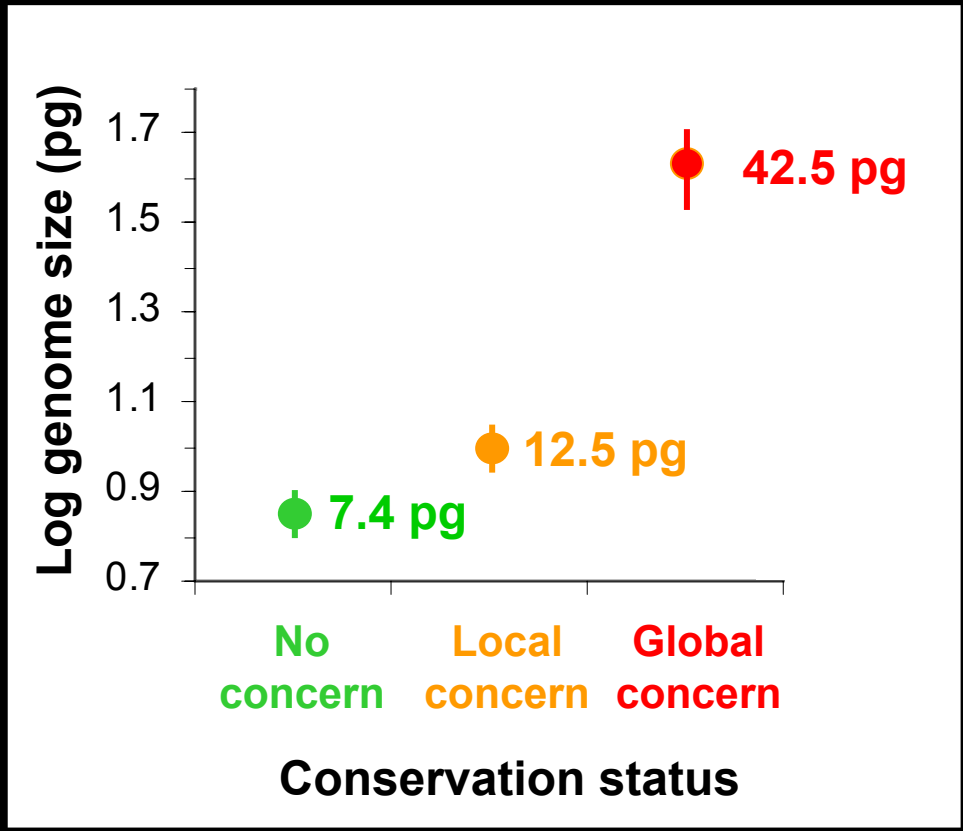
→ Local concern = 1329

→ No concern = 1402

Vinogradov AE. 2003.
Selfish DNA is maladaptive: evidence from the plant Red List.
Trends in Genetics **19:** 609-614.

Consequences: Genome size and threat of extinction

Results



Vinogradov AE. 2003.
Trends in Genetics **19:** 609-614.

Consequences: Genome size and threat of extinction

Analysis within families

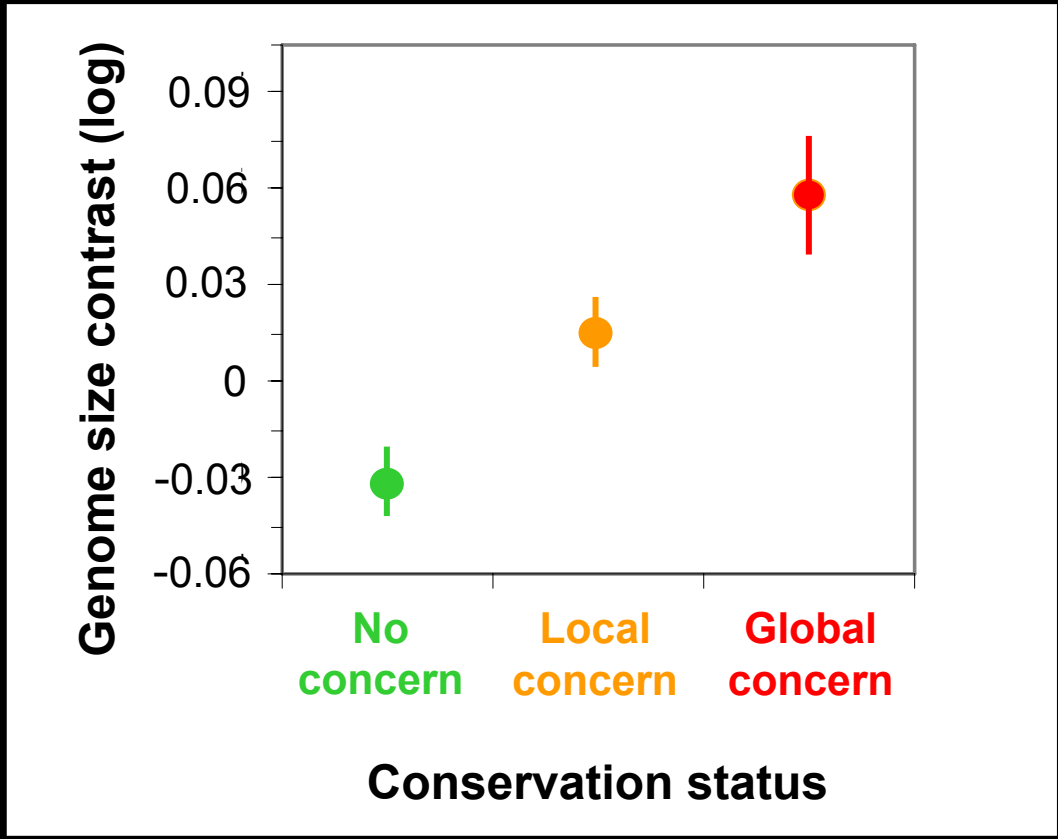
$$\text{Genome size contrast} = \frac{\text{Genome size for species}}{\text{Mean genome size of family}}$$

$$\text{Genome size contrast} = \frac{32.4}{7.7} = 4.2$$

$$\text{Genome size contrast} = \frac{1.2}{4.3} = 0.3$$

Consequences:
Genome size and threat of extinction

Results



Vinogradov AE. 2003.
Trends in Genetics **19:** 609-614.

Consequences:
Genome size and threat of extinction

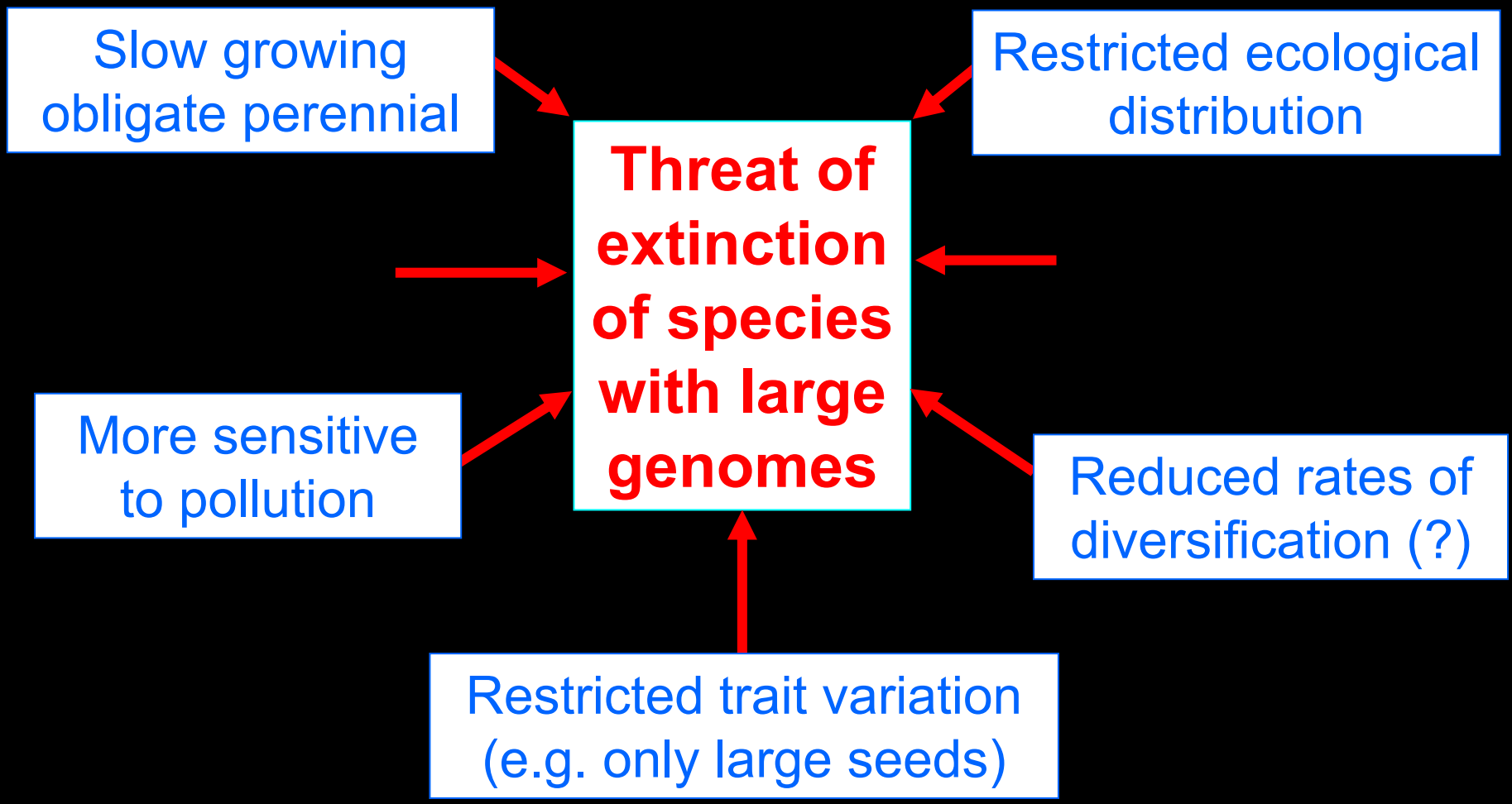
Conclusions

Species with large genomes are at greater risk of extinction than those with small genomes.

- Independent of life cycle type (at least partially)

- Independent of polyploidy

Consequences: Genome size and threat of extinction



DNA amount variation and consequences

Summary

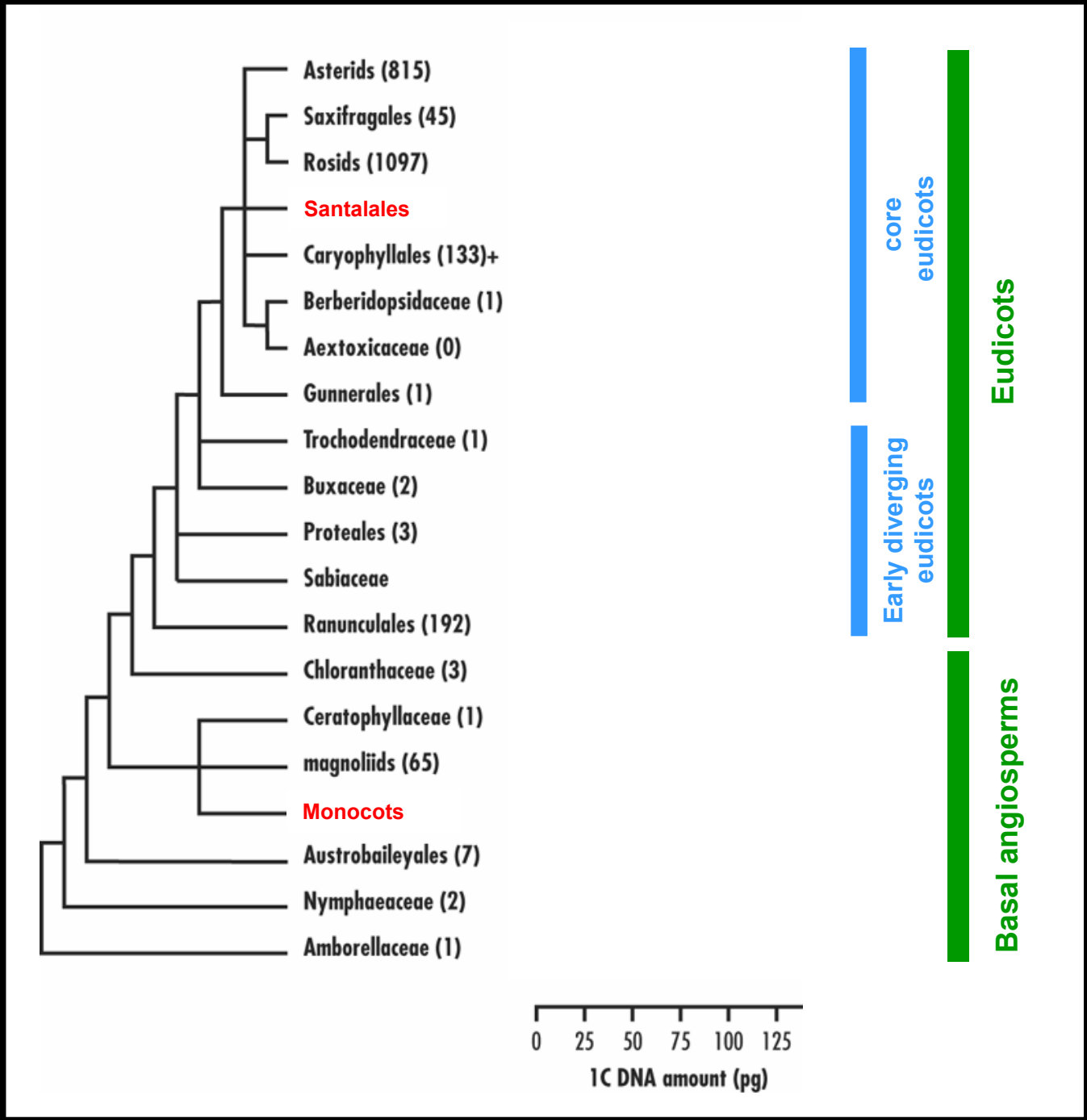
- Huge variation in DNA amount in plants
- Consequences of this variation visible at:
  - Cellular level
  - Tissue level
  - Whole organism level
- Possession of large genomes appear to impose constraints which operate at:
  - Functional level
  - Ecological level
  - Evolutionary level

Variation of genome size

1. Consequences of genome size variation in plants

2. **Evolution of genome size variation**

Phylogenetic tree of angiosperms and C-value data



Modified from:
Leitch, Chase, Bennett MD. 1998. *Annals of Botany* **82:** 85-94.

Angiosperms with 'very large' genomes (≥ 35 pg)

MONOCOTS

- **Liliales**
  - Liliaceae
  - Melanthiaceae

- **Asparagales**
  - Alliaceae
  - Alstroemeriaceae
  - Orchidaceae
  - Xanthorrhoeaceae

- **Commelinids**
  - Commelinaceae

CORE EUDICOTS

- **Santalales**
  - Santalaceae

Viscum

Large scale analysis of genome size evolution in angiosperms

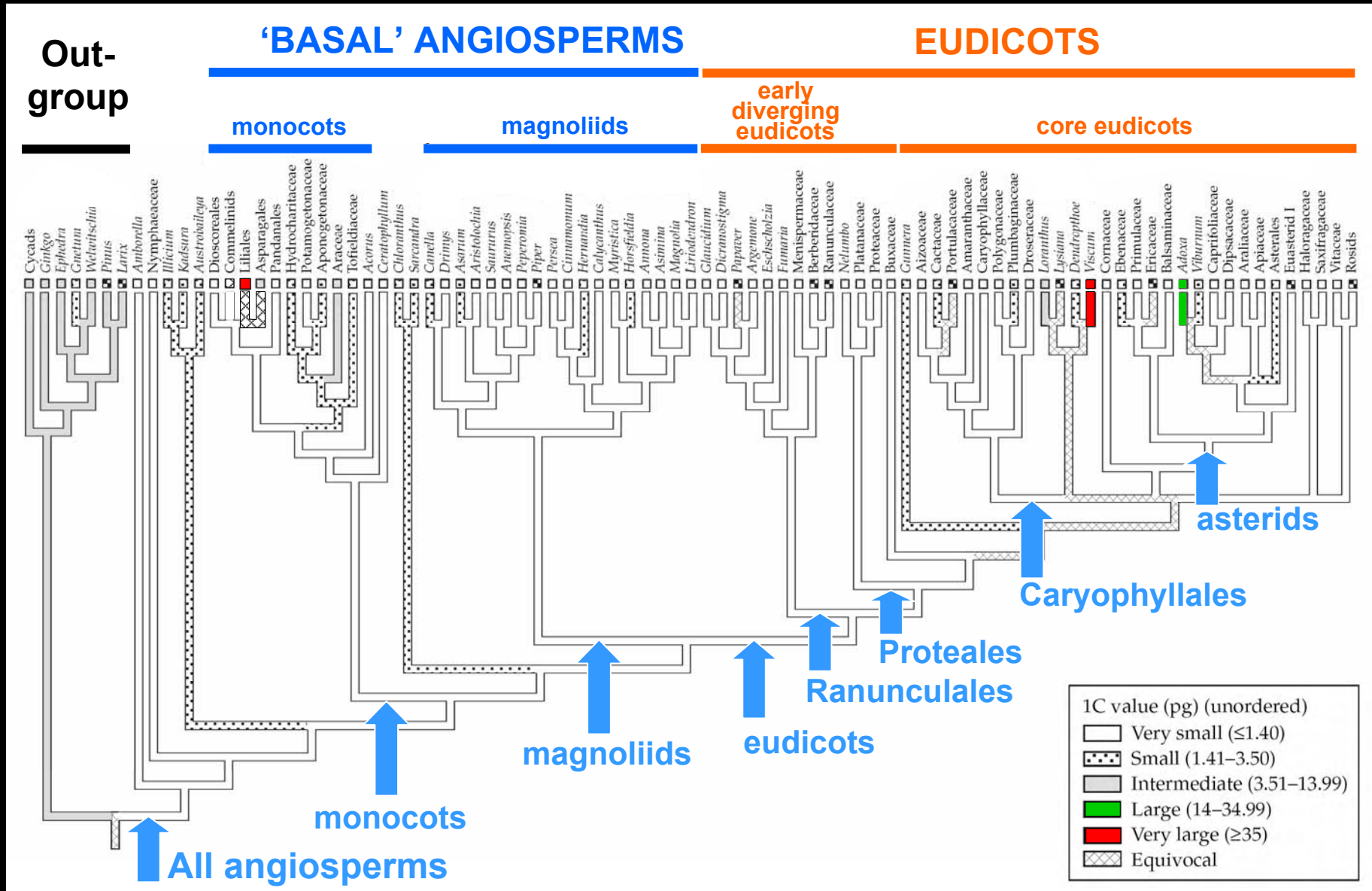
Data: Genome sizes for 4,119 species

Method: The 'All most parsimonious states' resolving option of MacClade

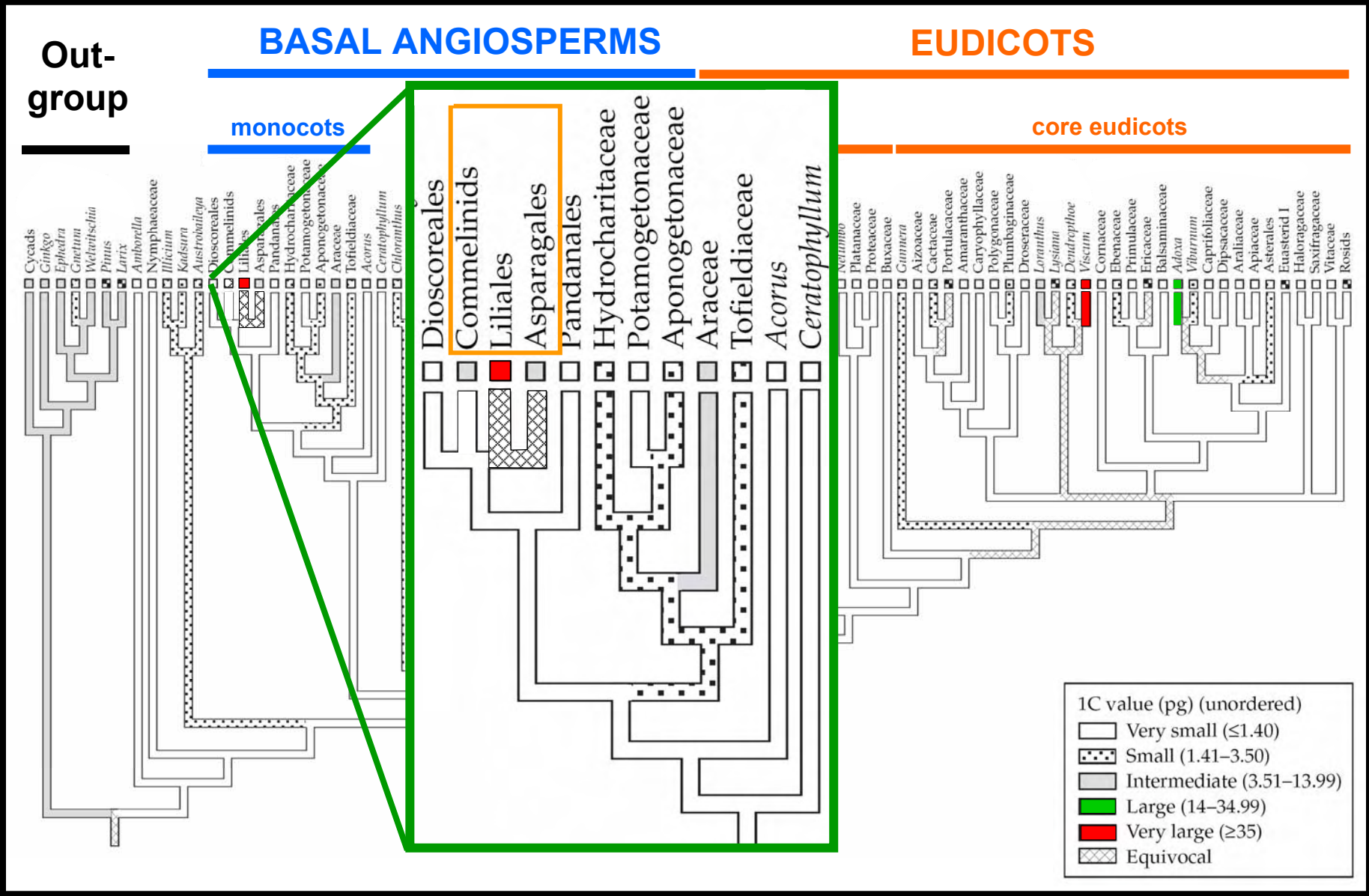
C-value range	Description
≤ 1.4 pg	Very small
≤ 3.5 pg	Small
3.51 – 13.99 pg	Intermediate
≥ 14 pg	Large
≥ 35 pg	Very large

Soltis, Soltis, Bennett, Leitch. 2003. *Am J Bot* **90:** 1596-1603.

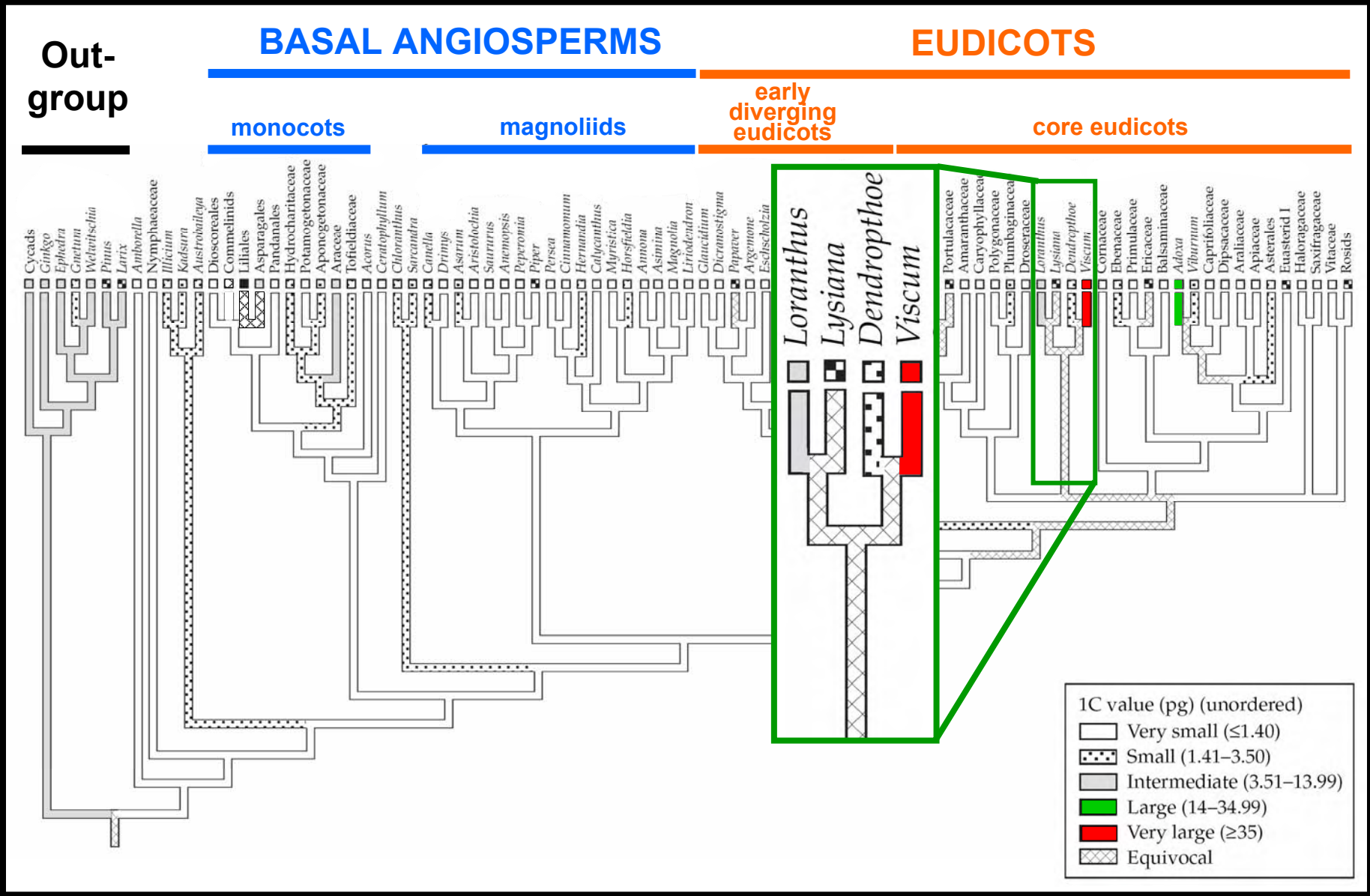
Large scale analysis of genome size evolution in angiosperms



Reconstruction of C-value diversification across angiosperms



Reconstruction of C-value diversification across angiosperms



Acknowledgements

Jodrell Laboratory,Royal Botanic Gardens, Kew, UK

Mike Bennett

Mike Fay

Lynda Hanson

Masaryk University, Brno,Czech Republic

Martin Lysak

California Polytechnic State University, USA

Charley Knight

Arjun Pendharkar

Biological Sciences, Yale University, USA

Jeremy Beaulieu

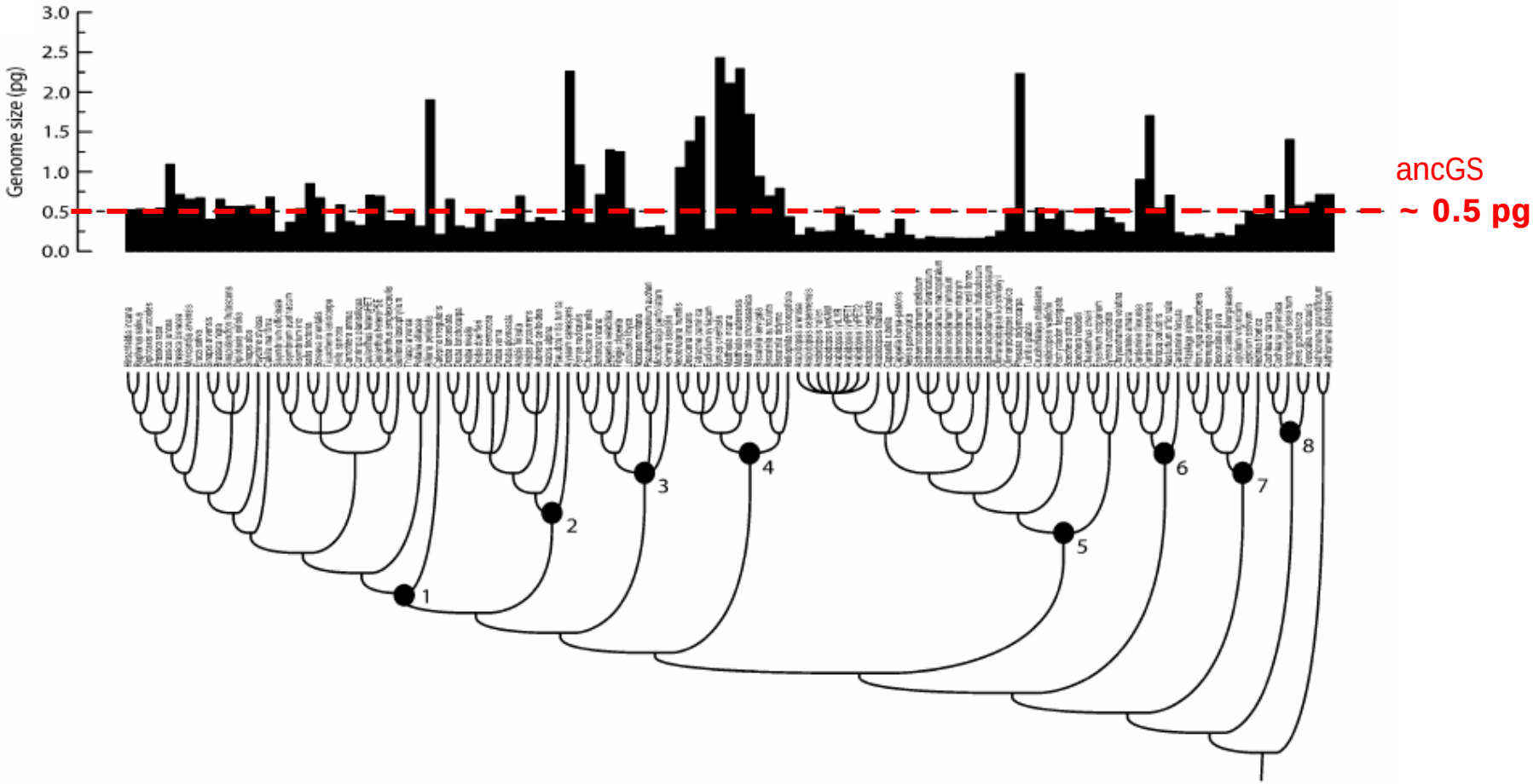
The Dynamic Ups and Downs of Genome Size Evolution in *Brassicaceae*

Martin Lysak, Marcus Koch,† Jeremy Beaulieu,‡ Armin Meister,§ and Ilia Leitch‖*

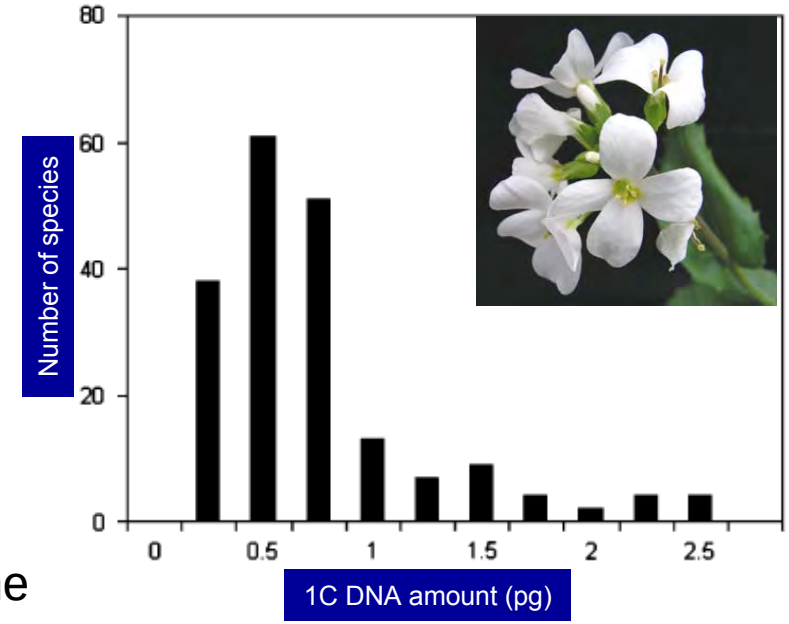
*Department of Functional Genomics and Proteomics, Institute of Experimental Biology, Masaryk University, Brno, Czech Republic; †Heidelberg Institute of Plant Sciences, Biodiversity and Plant Systematics, University of Heidelberg, Heidelberg, Germany; ‡Department of Ecology and Evolutionary Biology, Yale University; §Leibniz-Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany; and ‖Jodrell Laboratory, Royal Botanic Gardens, Richmond, Surrey, Kew, United Kingdom

Crucifers (*Brassicaceae*, *Cruciferae*) are a large family comprising some 338 genera and c. 3,700 species. The family includes important crops as well as several model species in various fields of plant research. This paper reports new genome size (GS) data for more than 100 cruciferous species in addition to previously published C values (the DNA amount in the unreplicated gametic nuclei) to give a data set comprising 185 *Brassicaceae* taxa, including all but 1 of the 25 tribes currently recognized. Evolution of GS was analyzed within a phylogenetic framework based on gene trees built from five data sets (*mat*K, *chs*, *adh*, *trn*LF, and ITS). Despite the 16.2-fold variation across the family, most *Brassicaceae* species are characterized by very small genomes with a mean 1C value of 0.63 pg. The ancestral genome size (ancGS) for *Brassicaceae* was reconstructed as $^{anc}1C = 0.50$ pg. Approximately 50% of crucifer taxa analyzed showed a decrease in GS compared with the ancGS. The remaining species showed an increase in GS although this was generally moderate, with significant increases in C value found only in the tribes Anchonieae and Physarieae. Using statistical approaches to analyze GS, evolutionary gains or losses in GS were seen to have accumulated disproportionately faster within longer branches. However, we also found that GS has not changed substantially through time and most likely evolves passively (i.e., a tempo that cannot be distinguished between neutral evolution and weak forms of selection). The data reveal an apparent paradox between the narrow range of small GSs over long evolutionary time periods despite evidence of dynamic genomic processes that have the potential to lead to genome obesity (e.g., transposable element amplification and polyploidy). To resolve this, it is suggested that mechanisms to suppress amplification and to eliminate amplified DNA must be active in *Brassicaceae* although their control and mode of operation are still poorly understood.

The distribution of 1C values for 120 crucifer taxa superimposed on the supertree phylogeny built from five data sets (*mat*K, *chs*, *adh*, *trn*LF, and ITS)

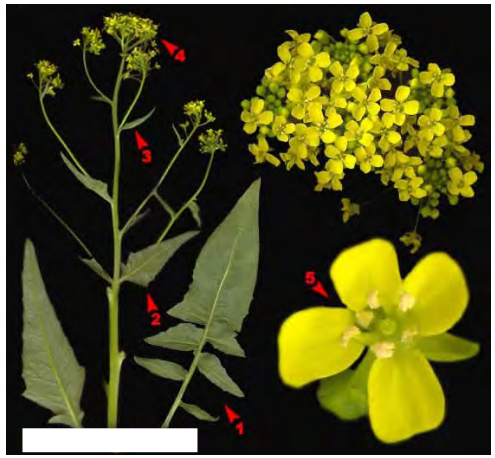


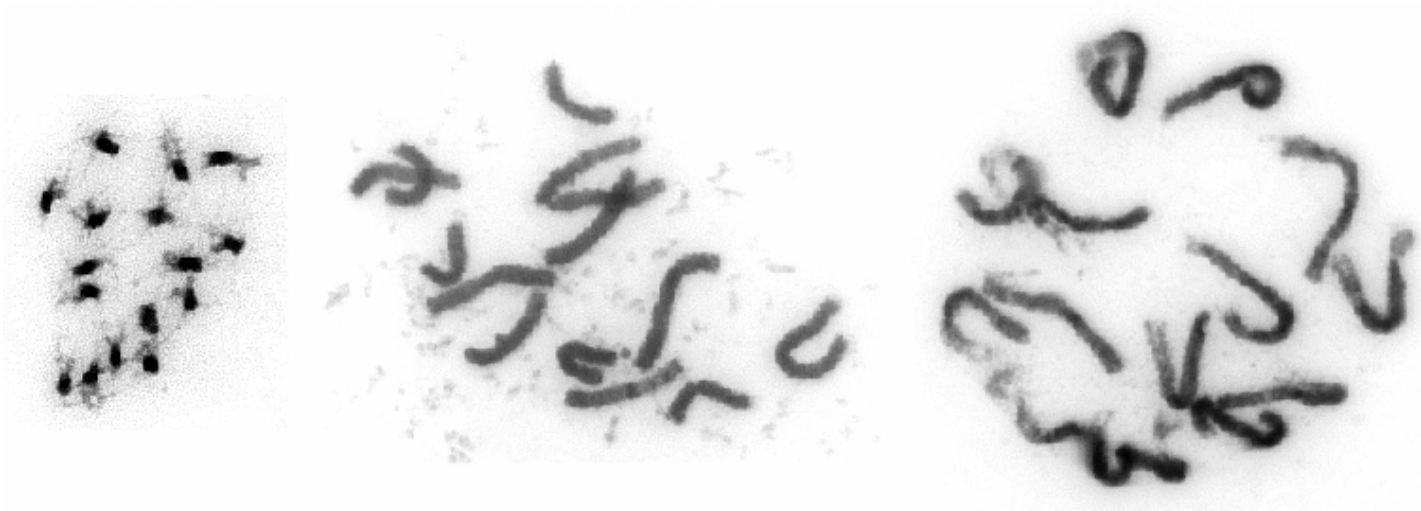
Basic facts on GS evolution in *Brassicaceae*



- the 16.2-fold GS variation across the family
- mean 1C = 0.63 pg
- crucifers have very small (≤ 1.4 pg) or small (≤ 3.5 pg) genomes (Leitch et al. 1998)
- ancGS 1C = 0.54 pg
- GS has not changed substantially through time
- GS evolves most likely passively (i.e., a tempo that cannot be distinguished between neutral evolution and weak forms of selection)
- an apparent paradox of GS evolution: the narrow range of small GSs vs. genomic processes leading to genome obesity (e.g., transposable element amplification and polyploidy)
- poorly understood mechanisms suppressing amplification and/or eliminating amplified DNA must be active

Paradox of GS evolution in *Brassicaceae*:
large genomes in (some) species with a low chromosome number





Mode of genome evolution in some crucifer lineages