The largest genome of them all?
(Plant genome size diversity : focus on Melanthiaceae)

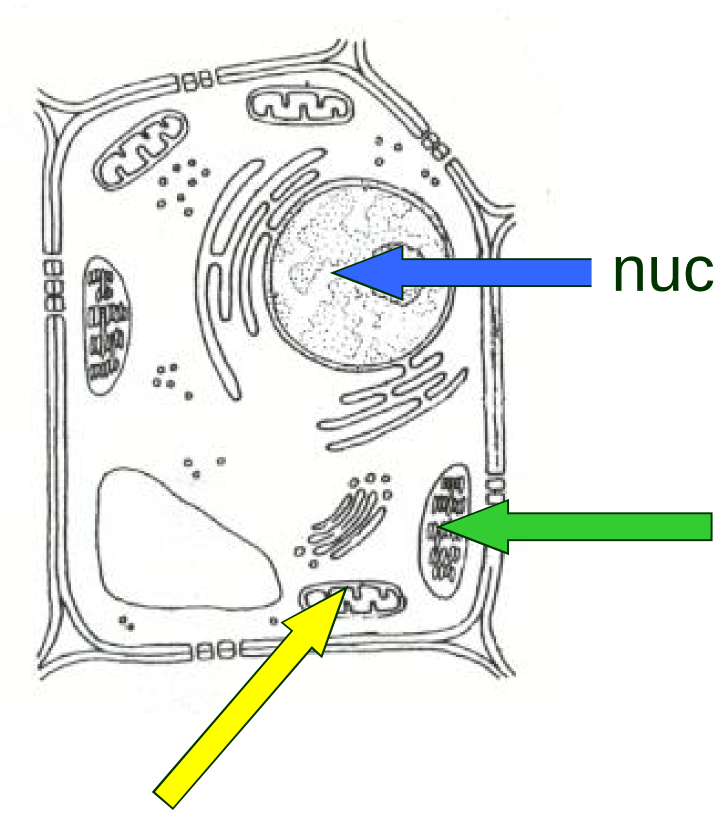
Michael F. Fay, Jaume Pellicer and Ilia J. Leitch, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB
Wendy Zomlefer, University of Georgia, Athens, GA, USA

Talk Outline

- Genome size: concepts and nomenclature
- How do we estimate genome size?
- What do we know?
- Why this family?
- Results for Melanthiaceae
- Future prospects

Pseudotrillium rivale

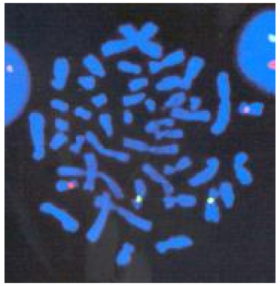
Where is the DNA?



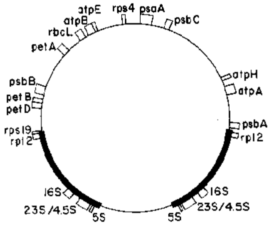
nucleus

chloroplast

mitochondrion
maternal only

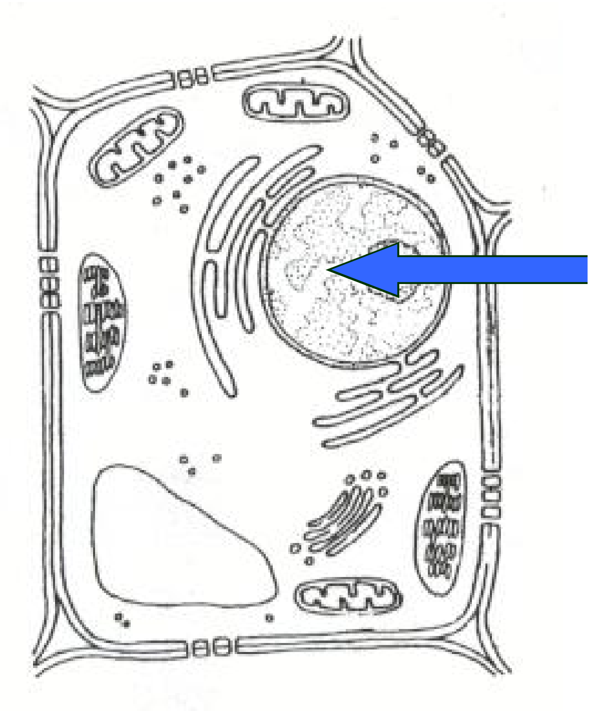


maternal
& paternal

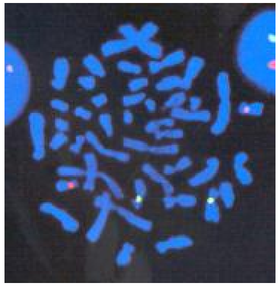


maternal only

This talk relates to nuclear genome size



nucleus



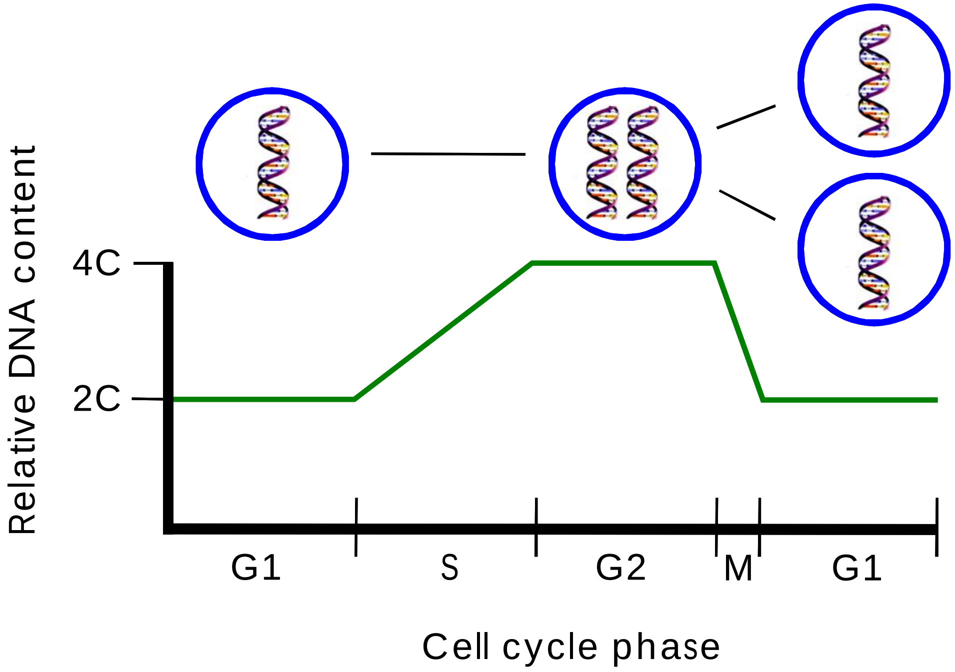
maternal
& paternal

Genome size: terms and nomenclature

- C-value (Swift 1950): CONSTANT, the amount of nuclear DNA characteristic of a particular genotype

C-TERMINOLOGIES:

  - 1C: DNA content of the unreplicated haploid chromosome complement
  - 1Cx: DNA amount of an unreplicated monoploid genome with chromosome number x



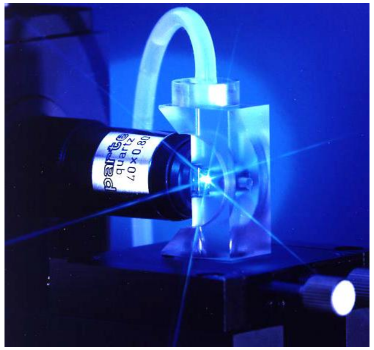
Pfosser et al. (2007)

How do we estimate genome size?

- Biochemical methods
- Flow cytometry ⬅
- Microdensitometry
- Scanning microdensitometry
- Chemical extraction
- Feulgen cytophotometry
- Feulgen densitometry
- Feulgen microdensitometry ⬅
- Feulgen microspectrophotometry
- Feulgen scanning densitometry
- Feulgen image scanning

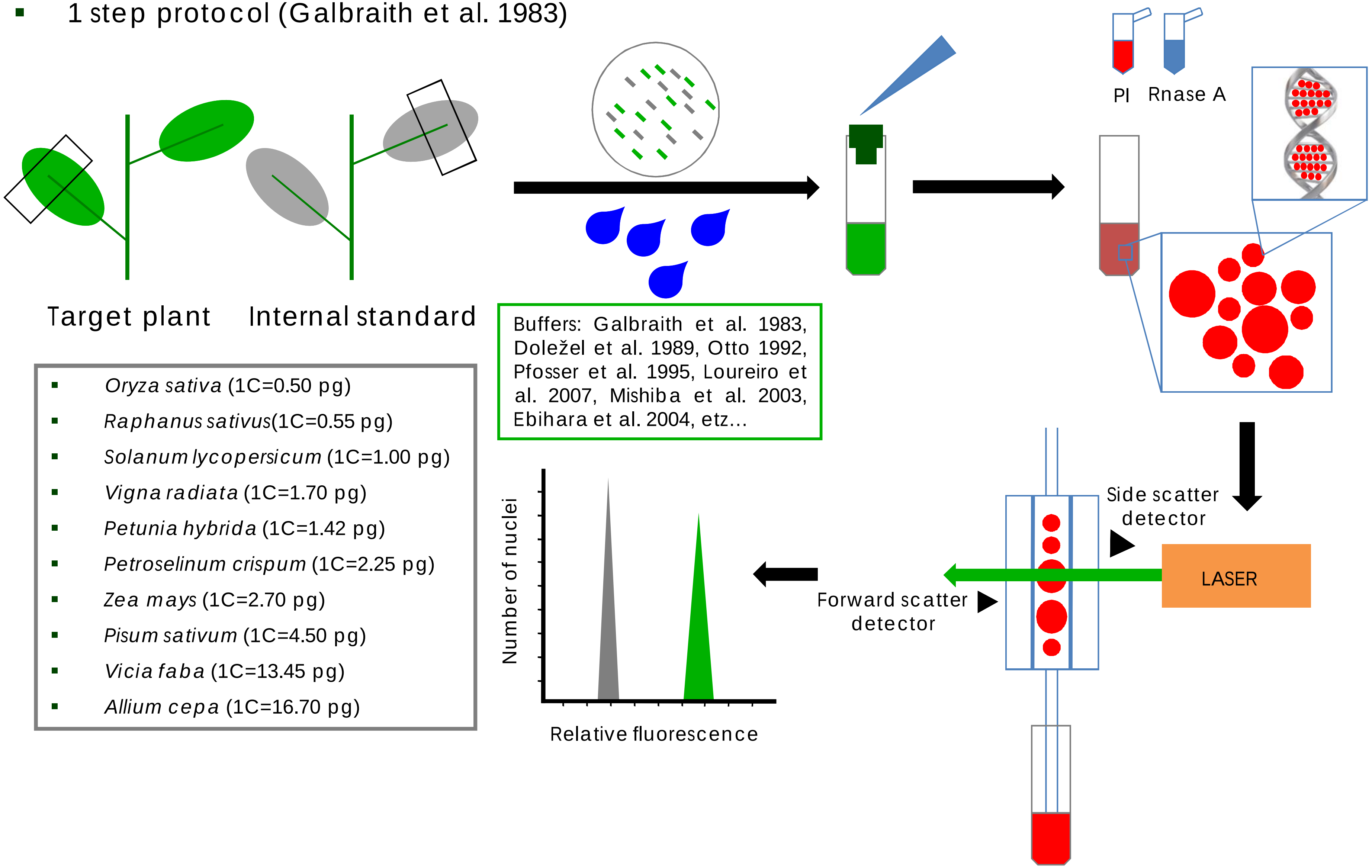
WHY flow cytometry ?

- Convenient and easy
- Fast
- Reliable
- Analysis of representative numbers of nuclei
- Non-destructive (small amount of materials)
- Facilitates the analysis of subpopulations with different DNA contents
- High precision and repeatability



How do we estimate genome size: flow cytometry

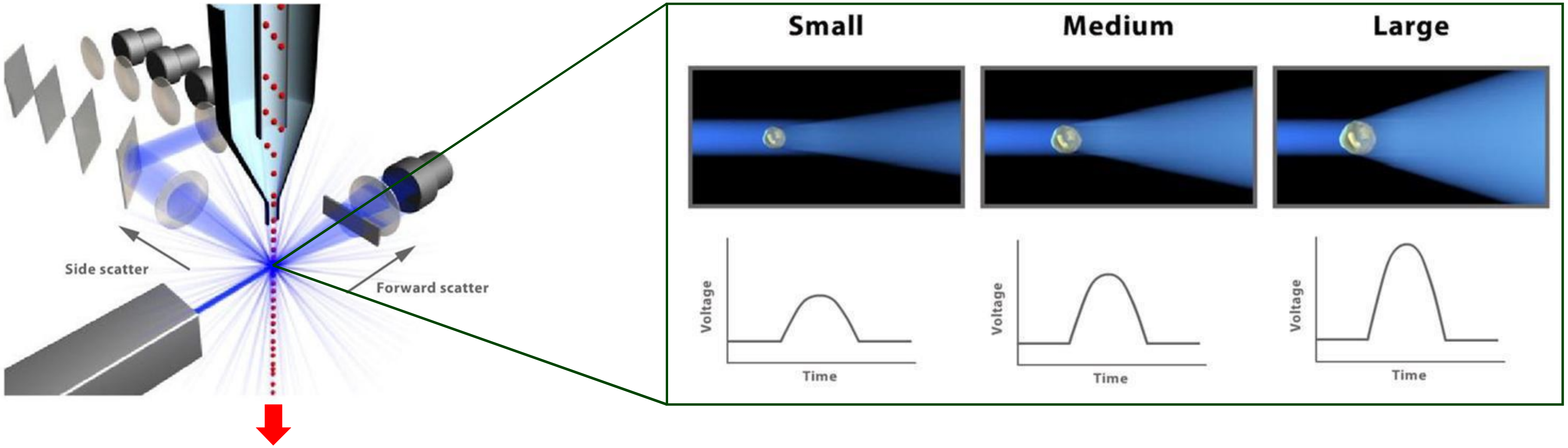
- 1 step protocol (Galbraith et al. 1983)



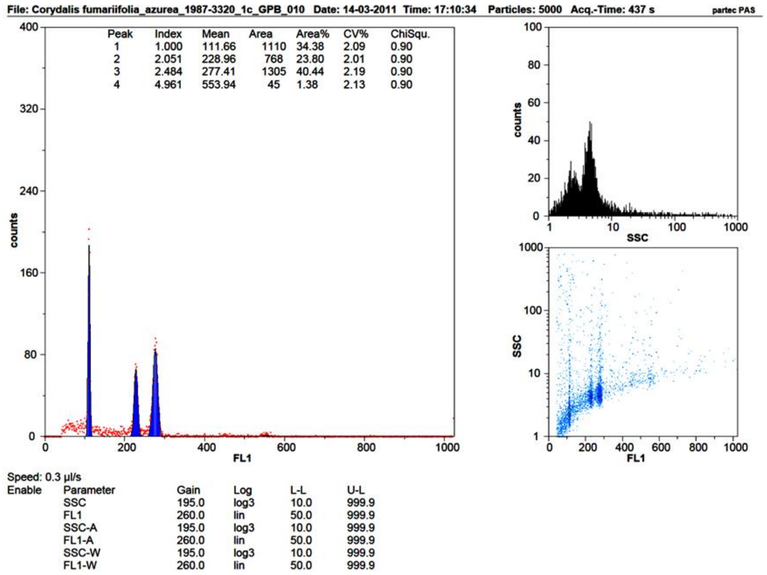
- *Oryza sativa* (1C=0.50 pg)
- *Raphanus sativus*(1C=0.55 pg)
- *Solanum lycopersicum* (1C=1.00 pg)
- *Vigna radiata* (1C=1.70 pg)
- *Petunia hybrida* (1C=1.42 pg)
- *Petroselinum crispum* (1C=2.25 pg)
- *Zea mays* (1C=2.70 pg)
- *Pisum sativum* (1C=4.50 pg)
- *Vicia faba* (1C=13.45 pg)
- *Allium cepa* (1C=16.70 pg)

Buffers: Galbraith et al. 1983, Doležel et al. 1989, Otto 1992, Pfosser et al. 1995, Loureiro et al. 2007, Mishiba et al. 2003, Ebihara et al. 2004, etz...

How do we estimate genome size: flow cytometry



- Nuclei produce a unique light scatter which in turn generates a unique voltage pulse
- Two light detectors:
    - FSC: particle size
    - SSC: granularity and complexity



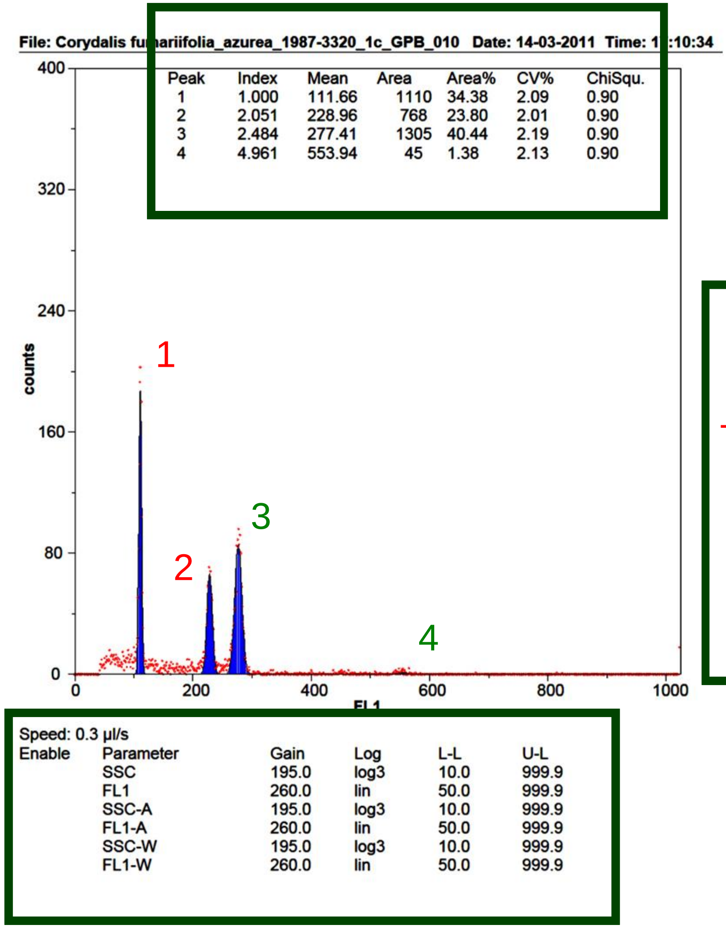
```
File: Corydalis fumariifolia_azurea_1987-3320_1c_GPB_010   Date: 14-03-2011   Time: 17:10:34   Particles: 5000   Acq.-Time: 437 s        partec PAS

Peak   Index   Mean     Area   Area%   CV%   ChiSqu.
 1     1.000   111.66   1110   34.38   2.09   0.90
 2     2.051   228.96    768   23.80   2.01   0.90
 3     2.484   277.41   1305   40.44   2.19   0.90
 4     4.961   553.94     45    1.38   2.13   0.90
```

```
Speed: 0.3 µl/s
Enable   Parameter        Gain    Log    L-L    U-L
         SSC              195.0   log3   10.0   999.9
         FL1              260.0   lin    50.0   999.9
         SSC-A            195.0   log3   10.0   999.9
         FL1-A            260.0   lin    50.0   999.9
         SSC-W            195.0   log3   10.0   999.9
         FL1-W            260.0   lin    50.0   999.9
```

How do we estimate genome size: flow cytometry

- The output: flow histogram



File: Corydalis fu ariifolia_azurea_1987-3320_1c_GPB_010 Date: 14-03-2011 Time: 1 :10:34

```
Peak    Index    Mean      Area   Area%   CV%    ChiSqu.
 1      1.000    111.66    1110   34.38   2.09   0.90
 2      2.051    228.96     768   23.80   2.01   0.90
 3      2.484    277.41    1305   40.44   2.19   0.90
 4      4.961    553.94      45    1.38   2.13   0.90
```

```
Speed: 0.3 µl/s
Enable   Parameter        Gain    Log    L-L    U-L
         SSC              195.0   log3   10.0   999.9
         FL1              260.0   lin    50.0   999.9
         SSC-A            195.0   log3   10.0   999.9
         FL1-A            260.0   lin    50.0   999.9
         SSC-W            195.0   log3   10.0   999.9
         FL1-W            260.0   lin    50.0   999.9
```

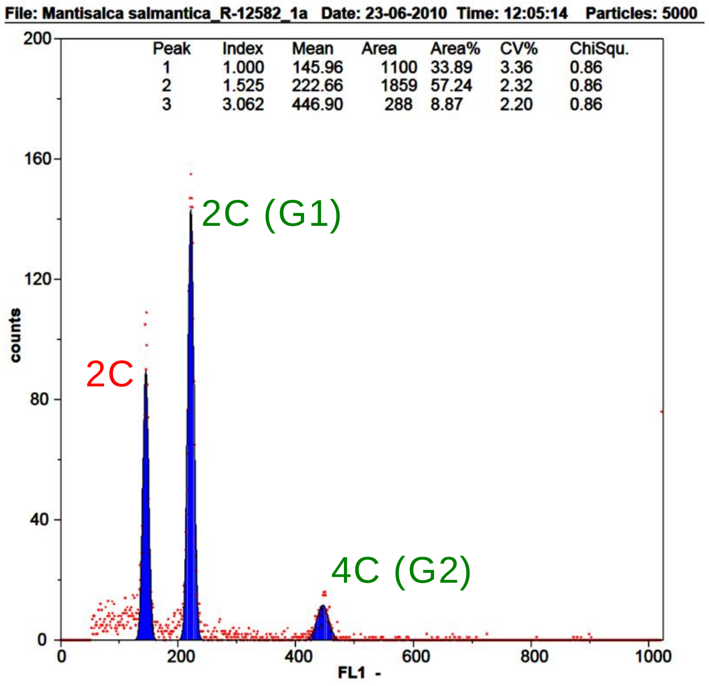
- Target species:
  - Peak 1 (G1): 2C
  - Peak 2 (G2): 4C
- Internal standard species:
  - Peak 3 (G1): 2C
  - Peak 4 (G2): 4C

$$\text{Target 2C value (pg/Mbp)} = \text{standard 2C} \times \frac{\text{Target G1 mean Peak (1) position}}{\text{standard G1 mean peak (3) position}}$$

1 pg=978 Mbp (Doležel et al. 2003)

(1 pg≈1 billion bp)

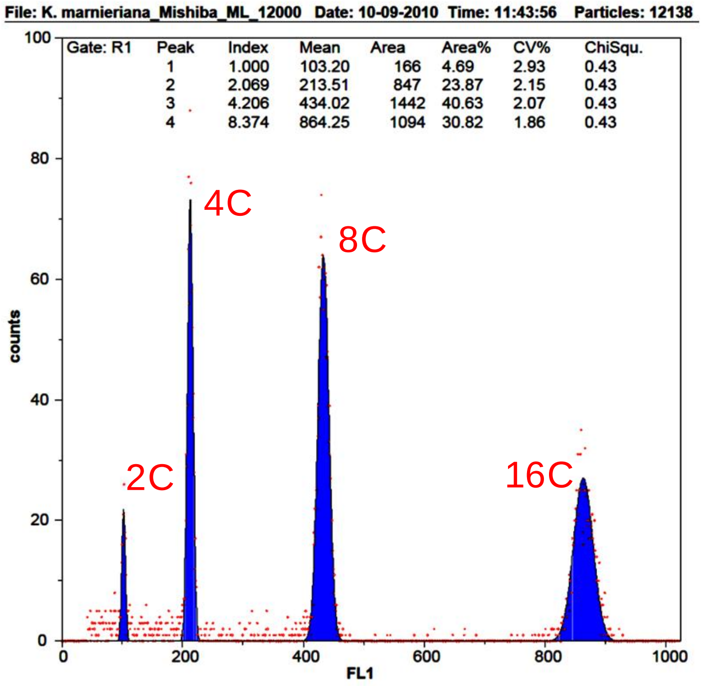
How do we estimate genome size: flow cytometry



```
File: Mantisalca salmantica_R-12582_1a   Date: 23-06-2010   Time: 12:05:14   Particles: 5000
Peak   Index   Mean     Area   Area%   CV%    ChiSqu.
 1     1.000   145.96   1100   33.89   3.36   0.86
 2     1.525   222.66   1859   57.24   2.32   0.86
 3     3.062   446.90    288   8.87    2.20   0.86
```

Target species

Internal standard species



```
File: K. marnieriana_Mishiba_ML_12000   Date: 10-09-2010   Time: 11:43:56   Particles: 12138
Gate: R1   Peak   Index   Mean     Area   Area%   CV%    ChiSqu.
            1     1.000   103.20    166   4.69    2.93   0.43
            2     2.069   213.51    847   23.87   2.15   0.43
            3     4.206   434.02   1442   40.63   2.07   0.43
            4     8.374   864.25   1094   30.82   1.86   0.43
```

- Enables the analysis of subpopulations with different DNA contents (endopolyploidy)

What do we know?

- Genome size estimations provide relevant data for:
  - Choice of genetic model species for genome sequencing
  - Construction of genetic libraries
  - Studies dealing with genetic diversity (AFLP, RFLP, SSR, RAPD)
  - Large scale hypotheses about how plant genomes change in the frameworks of ecology, conservation and evolution

Plant DNA C-values
Database (release 5.0;
Bennett & Leitch, 2010)

Release 1.0 (Loureiro et al. 2008)

A genome size database
in the Asteraceae
Release 1.0 (Garnatje et al. 2011)

What do we know: GS diversity

Genome size in angiosperms varies ca. 2400-fold

Genlisea margaretae, Lentibulariaceae

1C=0.0648 pg; Greilhuber et al. (2006)

Paris japonica, Melanthiaceae

1C=152.23 pg; Pellicer et al. (2010)

Genome size variation in animals

REPTILES

5-fold

Mionecton skink

Chalcides mionecton

1C = 1.1 pg

Spur-thighed tortoise

Testudo graeca

1C = 5.4 pg

MAMMALS

5-fold

Bent winged bat

Miniopterus schreibersi

1C = 1.73 pg

Red viscacha rat

Tympanoctomys barrerae

1C = 8.40 pg

BIRDS

2.4-fold

Black chinned hummingbird

Archilochus alexandri

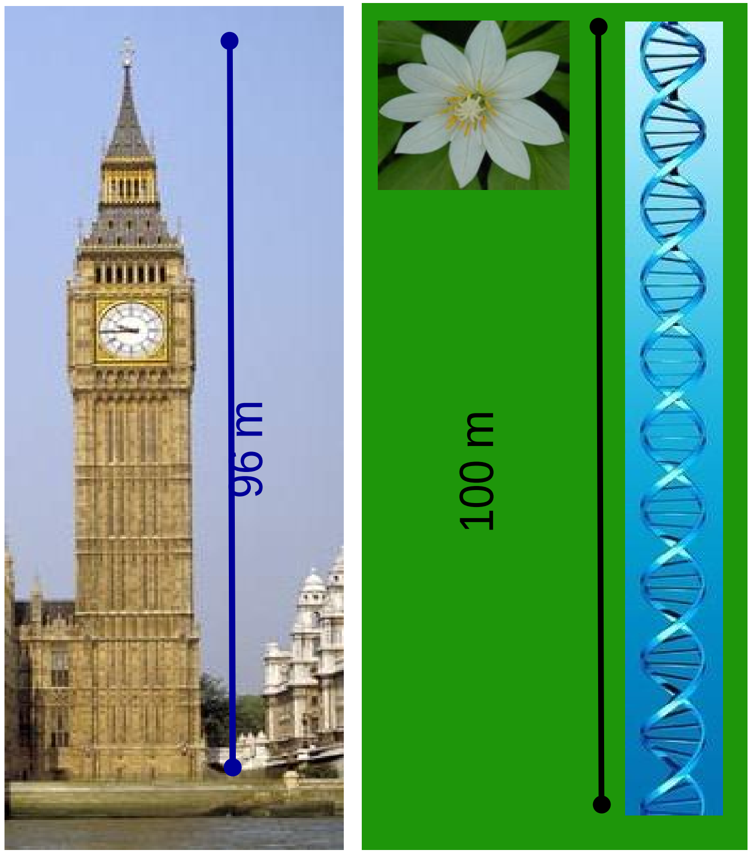
1C = 0.91 pg

Ostrich

Struthio camelus

1C = 2.16 pg

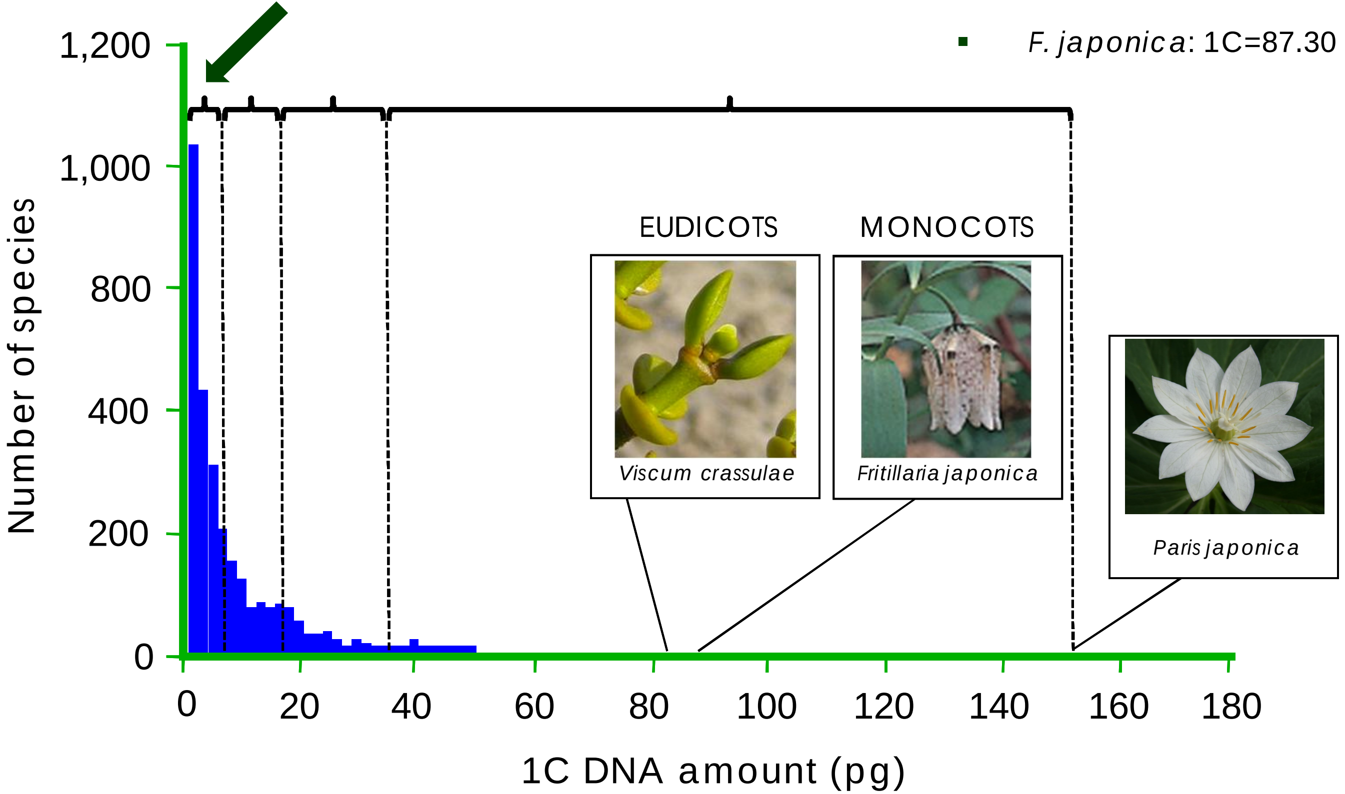
What does that much DNA look like?



What do we know?

But… the vast majority of angiosperms have small or very small genomes

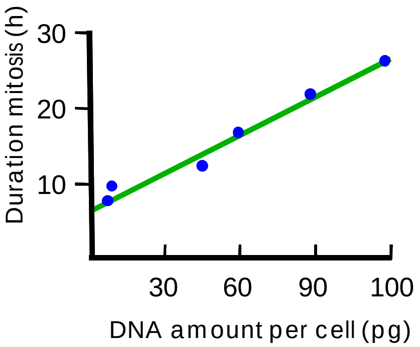
- *V. crassulae*: 1C=82.00 pg (Zonneveld et al. 2010)
- *F. japonica*: 1C=87.30 pg (Ambrožová et al. 2011)



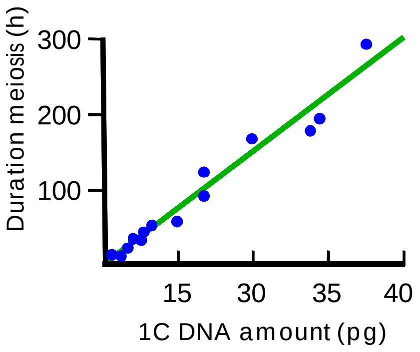
From Garcia et al. (2010) including data from Plant C-values Database (Bennett and Leitch 2010)

Consequences of genome size variation

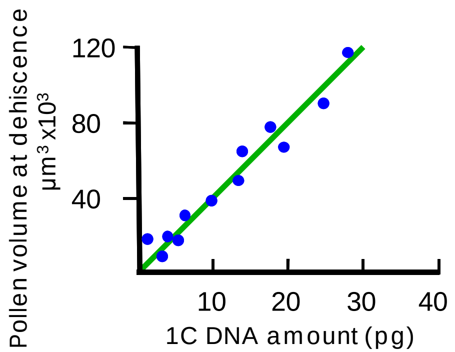
- GS has been correlated with different biological and ecological parameters:



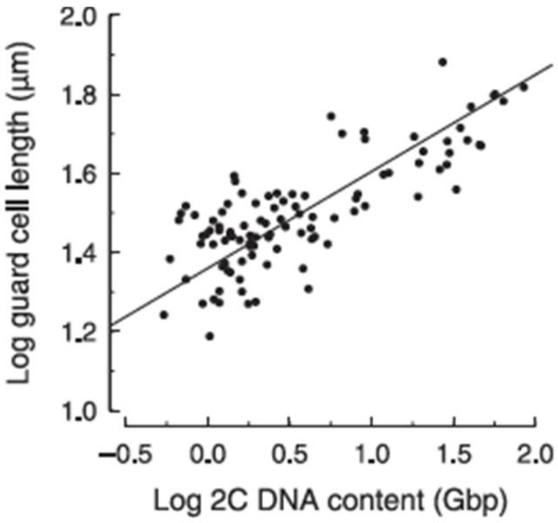
Van’t Hof & Sparrow (1963)



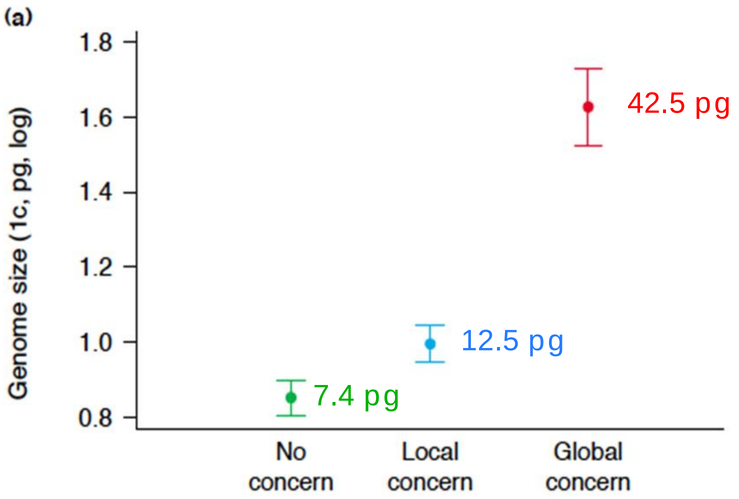
Bennett (1977)



Bennett et al. (1972)



Beaulieu et al. (2008)



Vinogradov (2003)

Consequences of genome size variation

Species with large genomes at greater risk of extinction

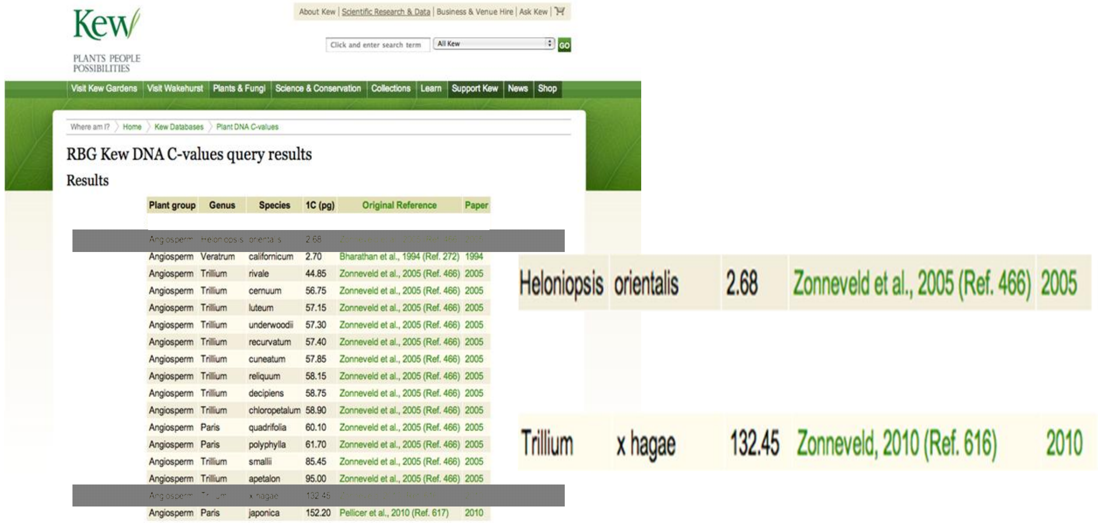
Vinogradov AE. 2003.
Trends Genet 19**:** 609-614.

What don’t we know?

- What mechanisms are involved in these size shifts? (some ideas...)
- How do some plants and animals keep their genomes so small (lean)?
- Is there any real advantage to having an enormous (obese) genome?
- Have plants with the largest genomes lost their ability to identify and eliminate “foreign” DNA?
- What are the biosynthetic costs of a genome so large?
- Etc.

Why do we work on this family?

MELANTHIACEAE

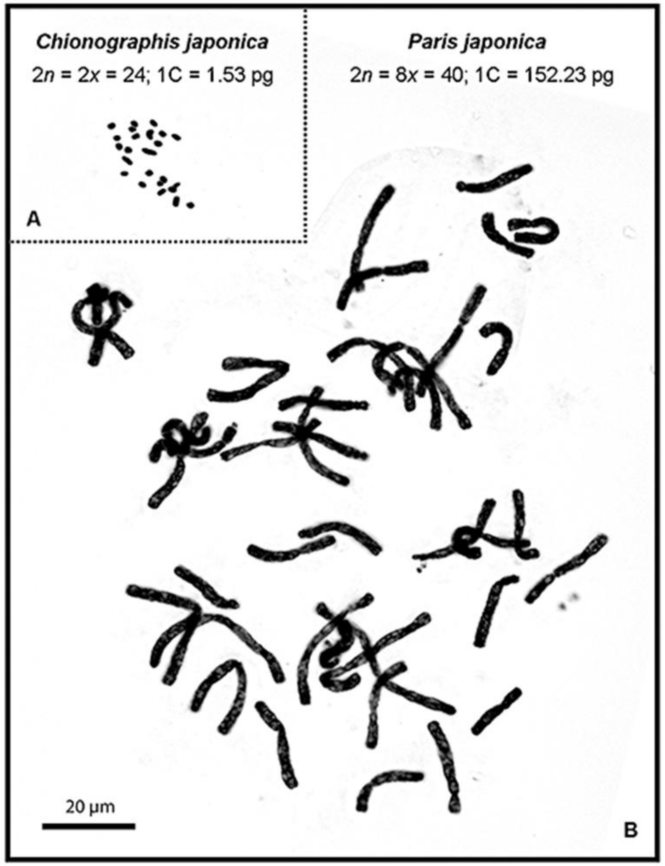


RBG Kew DNA C-values query results

Results

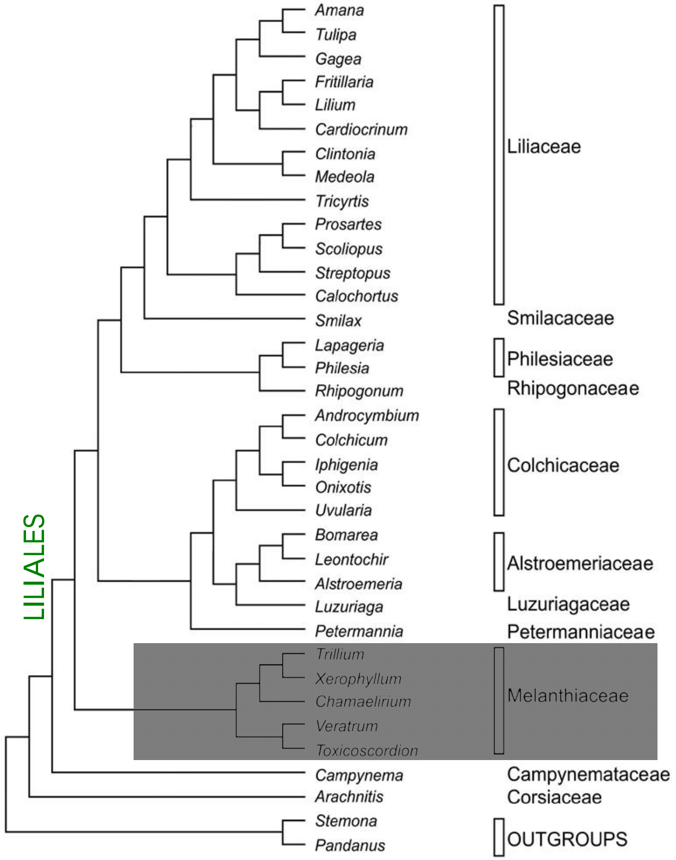
Plant group	Genus	Species	1C (pg)	Original Reference	Paper
Angiosperm	Heloniopsis	orientalis	2.68	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Veratrum	californicum	2.70	Bharathan et al., 1994 (Ref. 272)	1994
Angiosperm	Trillium	rivale	44.85	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	cernuum	56.75	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	luteum	57.15	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	underwoodii	57.30	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	recurvatum	57.40	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	cuneatum	57.85	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	reliquum	58.15	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	decipiens	58.75	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	chloropetalum	58.90	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Paris	quadrifolia	60.10	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Paris	polyphylla	61.70	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	smallii	85.45	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	apetalon	95.00	Zonneveld et al., 2005 (Ref. 466)	2005
Angiosperm	Trillium	x hagae	132.45	Zonneveld, 2010 (Ref. 616)	2010
Angiosperm	Paris	japonica	152.20	Pellicer et al., 2010 (Ref. 617)	2010

Heloniopsis	orientalis	2.68	Zonneveld et al., 2005 (Ref. 466)	2005
Trillium	x hagae	132.45	Zonneveld, 2010 (Ref. 616)	2010



- Few published GS estimations available
- Evidence of a wide range of GS variation
- Variable chromosome number and size
- Phylogenetically well understood

Melanthiaceae



Fay et al. (2006)

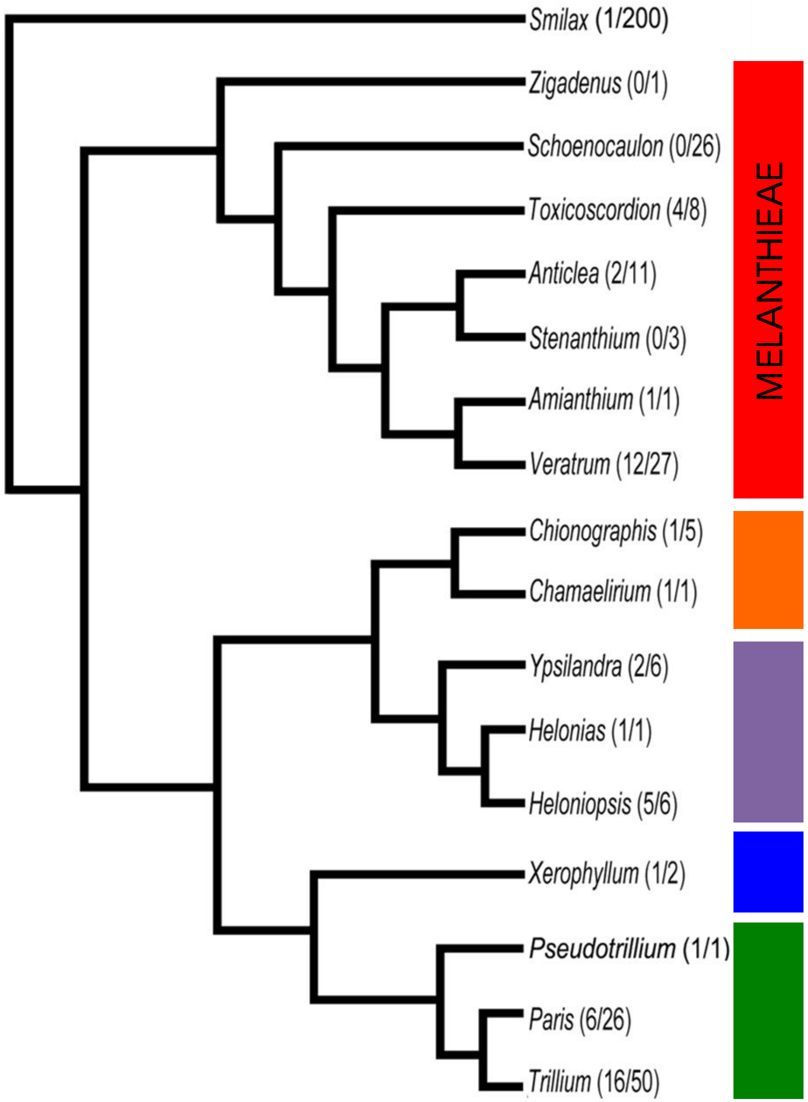
- Member of Liliales
- c. 160 species and 16 genera
- Woodland and/or alpine perennial herbs
- Main distribution: from temperate to Arctic zones of the Northern Hemisphere

Heloniopsis kawanoi

C-values in Melanthiaceae

Genus	Min. (pg)	Max. (pg)	Mean (pg)	No. species with C-values	Representation (%)
Amianthium	-	-	3.38	1	100
Anticlea	1.58	4.29	1.58	2	18
Chamaelirium	-	-	0.98	1	100
Chionographis	2.99	3.05	3.02	1	20
Helonias	3.19	3.23	3.21	1	100
Heloniopsis	2.50	4.91	3.63	5	84
Paris	42.25	152.23	66.47	7	27
Pseudotrillium	-	-	27.51	1	100
Toxicoscordion	4.83	5.85	5.01	4	50
Trillium	47.51	68.64	68.21	16	32
Veratrum	2.07	4.31	3.02	12	44
Xerophyllum	-	-	2.96	1	50
Ypsilandra	4.82	5.50	5.16	2	33

C-values in Melanthiaceae



Not sampled:

- *Zigadenus*
- *Schoenocaulon*
- *Stenanthium*

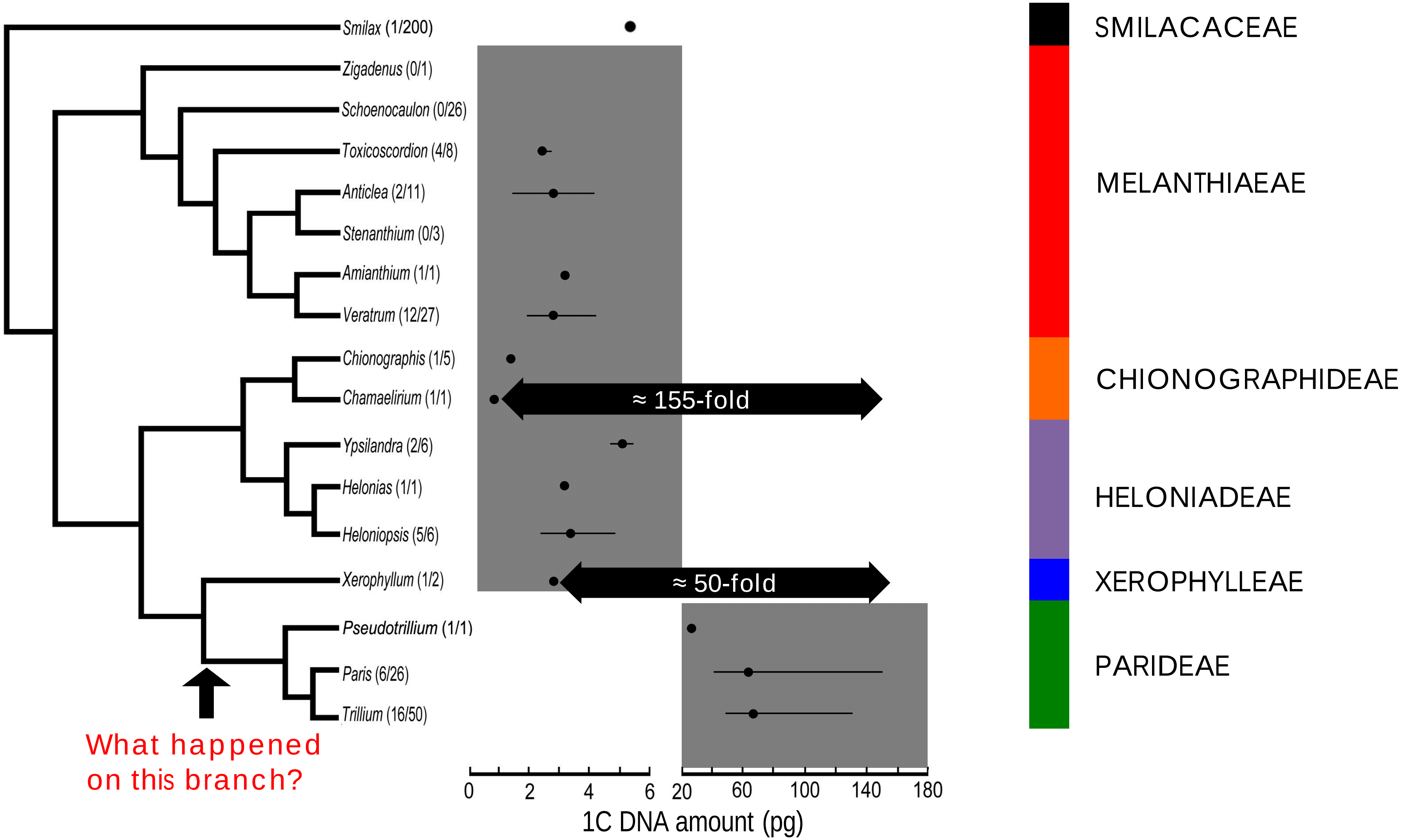
Z. glaberrimus

S. texanum

S. gramineum

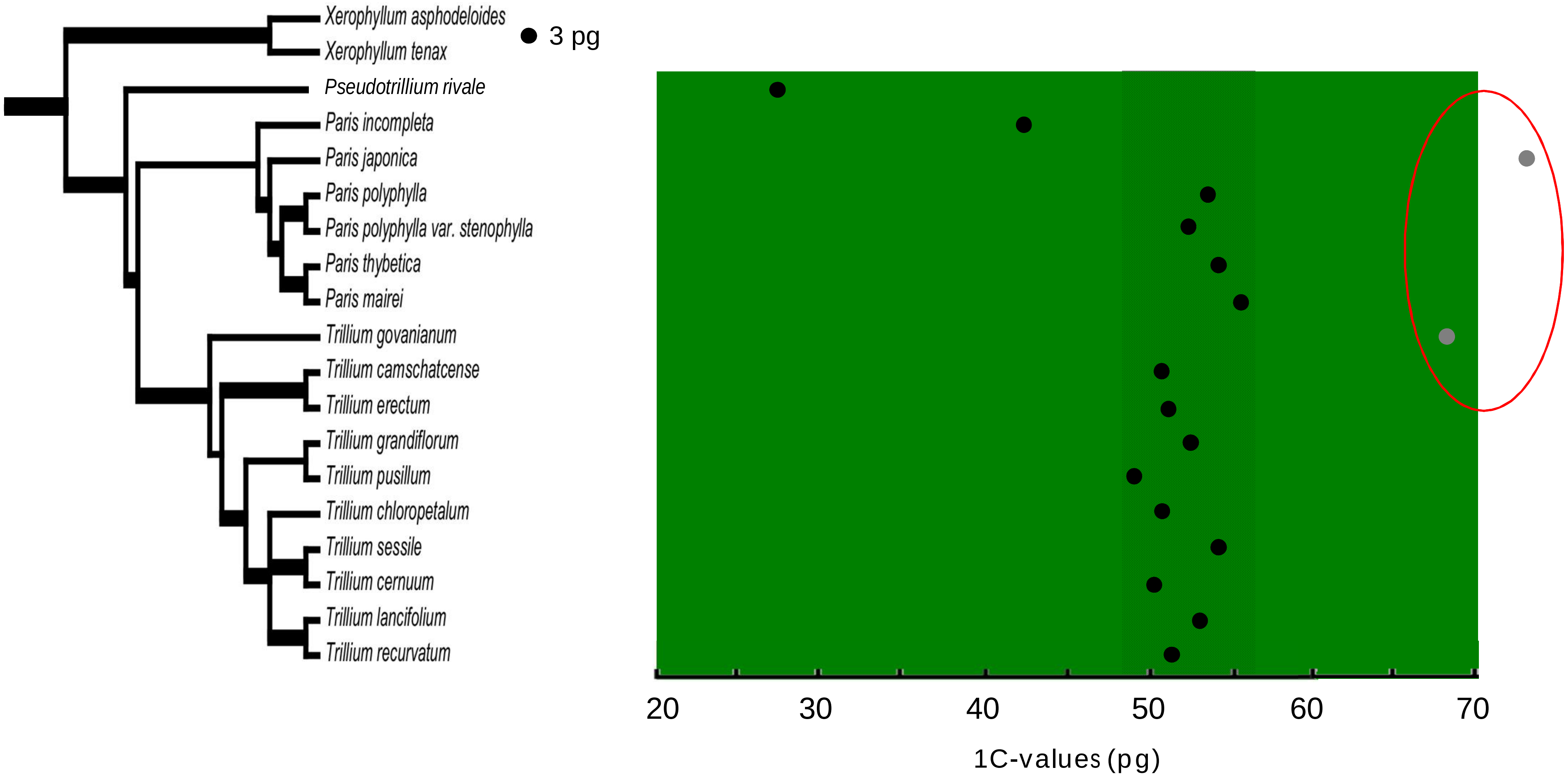
GS variation in Melanthiaceae

- Melanthiaceae have been revealed as one of the most diverse in terms of GS

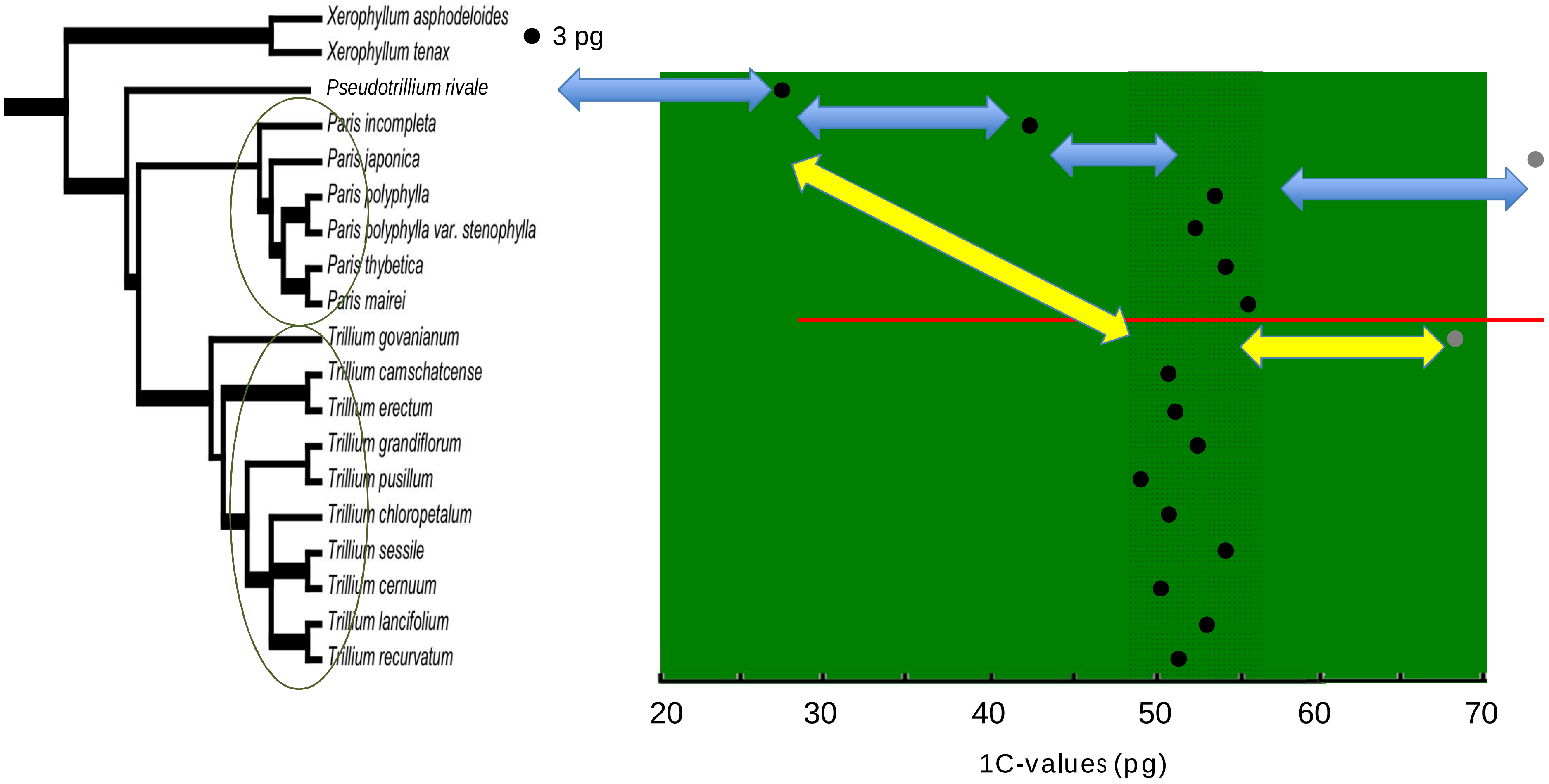


What happened on this branch?

GS variation in Parideae



GS variation in Parideae: jumps in size



Future prospects...

- To complete family sampling: C-values and chromosome data
- To analyse the mode of evolution of genome size in a phylogenetic context and infer ancestral character states
- To study the main mechanisms inducing genome size increase in plants:
  - Polyploidy
  - Repetitive DNA sequences (TEs, Satellite DNA etc.)

Acknowledgements

- Dr. Ralf Kynast (RBG Kew)
- Dr. Oriane Hidalgo (Ohio University)
- Dr. Lola Lledó, Laszlo Csiba (RBG Kew)
- Kit Strange, Katie Price, Richard Kernick, Carlos Magdalena (RBG Kew)

Kew
ROYAL BOTANIC GARDENS