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.
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
Phylogenetically independent contrasts

Ken Whitney
Plant Nuclear Genome Size Variation

7058 species surveyed and databased
~ 2300X difference

63 mbp

150,000 mbp

Genlisea margaretae

[Link to data: http://data.kew.org/cvalues/]

A ROLE FOR NONADAPTIVE PROCESSES IN PLANT GENOME SIZE EVOLUTION?
Chromosome Number Variation

Chromosome numbers vary $n = 2$ to $n = \sim 680$

Euploid variation – polyploidy

$\sim 35\%$ of vascular plants are neopolyploids

All are paleopolyploids

Aneuploid variation – gain or loss of one or more chromosomes
Post polyploid genome size change is variable:
- Additive sum of parents
- Increase
- Decrease
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
Plant Nuclear Gene Overlap

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

<table>
<thead>
<tr>
<th>NUMBER OF FAMILIES</th>
<th>INSERTION</th>
<th>EXCISION</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>$1.2 \times 10^{-4}$</td>
<td>$4.0 \times 10^{-6}$</td>
<td>Maside et al. 2000</td>
</tr>
<tr>
<td>9</td>
<td>$1.0 \times 10^{-4}$</td>
<td>$3.9 \times 10^{-6}$</td>
<td>Maside et al. 2001</td>
</tr>
<tr>
<td>17</td>
<td>$1.8 \times 10^{-4}$</td>
<td>$1.8 \times 10^{-6}$</td>
<td>Nuzhdin and Mackay 1995</td>
</tr>
</tbody>
</table>

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

![Bar chart showing the proportion of different types of DNA in various species.](chart.png)
Rapid Turnover of LTR Retrotransposons Complicates Genome Assemblies

Age of LTR retrotransposon insertions

Millions of years

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?