Unraveling the life dynamics of Sirevirus LTR retrotransposons: major players in the organization and evolution of maize and other angiosperm genomes

Alexandros Bousios
Sireviruses – What was known

- One of three Copia genera
- Plant-specific and young genus
- Putative retroviral properties
- Few publications on Sireviruses
- Scarce reference on the Sirevirus origin of some elements
- Difficulty in assigning LTR-RTNs into genera, research at this classification level is missing

Havecker et al. 2004 – Genome Biology
The unique genome structure of Sireviruses

Consensus:

primer binding site

integrase signal

multiple PPT signature

inverted repeat

Bousios et al. 2010 – BMC Genomics
Novel type of repeated motif with modular organization

- Exceptionally high copy number
- Dyad symmetry
- Core CGG-CCG signature
- Form CpG islands
- Upstream of a conserved promoter in the typical regulatory locus of LTR-RTNs
- Surprisingly, also upstream of the ENV

Bousios et al. 2010 – BMC Genomics
at the sequence level...
Sireviruses – What we know now

- Ancient genus with vastly divergent members
- Short highly conserved motifs in key non-coding domains that are critical for the life cycle of LTR-RTNs
- Astonishing conservation – hosts diverged 140-150 mya
- Only LTR-RTN genus with such an intriguing genome structure
The MASiVE algorithm

- Genomic sequence
  - LTRharvest
  - Vmatch
  - LTRharvest/PPT signature overlap
    - 23777 elements
    - LTRharvest: 74179
      PPT signatures: 43405
    - PPT500
    - PBS500
    - gene-wise
    - Pfam RT domain
      - -6093
      - -3727
    - RT
    - Vmatch
    - Sirevirus oligomers
      - -2051
    - PPT20/PBS20
    - LTR500
      - -1053
      - -234
    - Sirevirus LTR-RTNs
      - 10619 elements

Darzentas et al. 2010 - Bioinformatics
Sirevirus infiltration patterns in maize

Schnable et al. 2009 – Science

Baucom et al. 2009 – Plos Genetics

Bousios et al. 2011 – The Plant Journal
Sirevirus phylogenetic diversity in maize

Phylogenetic tree of maize Sireviruses based on the RT gene

Table 1. Properties of the Sirevirus families identified in the maize genome

<table>
<thead>
<tr>
<th>Family</th>
<th>FL</th>
<th>solo</th>
<th>frag</th>
<th>FL/solo</th>
<th>FL/frag</th>
<th>solo/frag</th>
<th>avg age (my)</th>
<th>FL</th>
<th>LTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opie</td>
<td>5310 +1790</td>
<td>2028</td>
<td>9826</td>
<td>2.6</td>
<td>0.5</td>
<td>0.2</td>
<td>0.90</td>
<td>9117</td>
<td>1254</td>
</tr>
<tr>
<td>Ri</td>
<td>4865 +772</td>
<td>2421</td>
<td>11377</td>
<td>2.0</td>
<td>0.4</td>
<td>0.2</td>
<td>0.94</td>
<td>9519</td>
<td>1271</td>
</tr>
<tr>
<td>Jienv</td>
<td>175 +175</td>
<td>103</td>
<td>469</td>
<td>1.7</td>
<td>0.4</td>
<td>0.2</td>
<td>0.71</td>
<td>12123</td>
<td>1534</td>
</tr>
<tr>
<td>Giepun</td>
<td>143</td>
<td>180</td>
<td>698</td>
<td>0.8</td>
<td>0.2</td>
<td>0.3</td>
<td>0.76</td>
<td>12666</td>
<td>1469</td>
</tr>
<tr>
<td>Hopie</td>
<td>74 +34</td>
<td>149</td>
<td>478</td>
<td>0.5</td>
<td>0.2</td>
<td>0.3</td>
<td>1.03</td>
<td>11696</td>
<td>1675</td>
</tr>
<tr>
<td>Djiap</td>
<td>15 +6</td>
<td>38</td>
<td>28</td>
<td>0.4</td>
<td>0.5</td>
<td>1.4</td>
<td>2.0</td>
<td>10783</td>
<td>1525</td>
</tr>
<tr>
<td>Other</td>
<td>37</td>
<td>19</td>
<td>59</td>
<td>1.9</td>
<td>0.6</td>
<td>0.3</td>
<td>1.59</td>
<td>10449</td>
<td>1279</td>
</tr>
<tr>
<td>Total</td>
<td>10619</td>
<td>4938</td>
<td>22935</td>
<td>2.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.92</td>
<td>9424</td>
<td>1273</td>
</tr>
</tbody>
</table>

Bousios et al. 2011 – The Plant Journal
Temporal activity of Sireviruses in the maize genome

Bousios et al. 2011 – The Plant Journal
Sireviruses integration biases in the maize genome

within and near genes...

for a palindromic sequence motif...

Bousios et al. 2011 – The Plant Journal
Sirevirus LTRs are methylation and recombination hotspots

Epigenetic modification profiles of Sirevirus genomes

Bousios et al. 2011 – The Plant Journal

- LTR indels length (bp) distribution
- Indels topology on the Sirevirus LTRs
The turbulent life of Sireviruses in the maize genome

- 21% of the genome – 90% of the Copia complement
- Intense amplification period the past 600,000 years
- Plethora of families with different life and genome characteristics
- Colonize gene-rich areas, mediating gene diversification and the formation of gene islands
- Sirevirus LTRs may be the epicenter of interactions with the host control system
MASiVEdb: the Sirevirus Plant Retrotransposon Database

http://bat.infspire.org/databases/masivedb/
Sireviruses dynamics in other angiosperm genomes

**Graphs and Tables**

- **Bar Chart**: Comparison of the number of Sireviruses across various plant species, with a focus on Zea mays and Glycine max.
- **Line Graph**: Time of integration (mya) vs. % of Sireviruses for different plant species, including Glycine max, Lotus japonicus, Sorghum bicolor, Theobroma cocoa, and Zea mays.

**Legend**
- Total SVs
- with ENV

**Plant Species**
- Osatj
- Tcoc
- Vvin
- Bdis
- Osati
- Fves
- Atha
- Zea mays
- Glycine max
Thank you for your attention…