Phylogenomics of the Brassicales Brassicaceae and Tribe Brassiceae

J. Chris Pires
University of Missouri
31 March 2014

@JChrisPires
Polyploid evolution and punctuated equilibrium: dynamic periods of change followed by stasis?

Are whole genome duplications (WGD) associated with morphological complexity and/or clade diversification?
Outline: Phylogenomics

I. Order Brassicales
II. Family Brassicaceae
III. Tribe Brassiceae
IV. Future Directions
Does polyploidy spur lineage diversification and complexity? Specifically, have nested polyploid events contributed to the diversification of lineages and anti-insectivore secondary metabolites (glucosinolates)? Revisit Ehrlich and Raven co-evolution/macroevolution ideas:

A. PHYLOGENOMICS – need trees to interpret WGD

B. TRANSCRIPTOMICS – need to find WGD

C. NETWORK BIOLOGY – measure WGD functional impact

On origin of novel traits and diversification
Phylogenomics using Illumina Sequencing

Genome Survey Sequencing (GSS)
1. Total DNA or mRNA
2. Multiplex Libraries
3. SE or PE reads
4. de novo Assembly (cp, mt, rDNA)
5. Annotation
6. Phylogenetics Analyses
($100/taxa; Steele et al. 2012 AJB)

<table>
<thead>
<tr>
<th>Species</th>
<th>Illumina Read Processing</th>
<th>Nuclear Genome</th>
<th>Mitochondrial Genome</th>
<th>Plastid Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw Reads</td>
<td>Cleaned Reads</td>
<td>Total Contigs</td>
<td>Total Sequence</td>
</tr>
<tr>
<td><strong>Myagrum</strong></td>
<td>14,299,919</td>
<td>13,621,738</td>
<td>22,569 Contigs</td>
<td>4,003,494 bp</td>
</tr>
<tr>
<td><strong>Caulanthus</strong></td>
<td>13,756,796</td>
<td>13,035,325</td>
<td>8,636 Contigs</td>
<td>2,452,489 bp</td>
</tr>
<tr>
<td><strong>Ionopsidium</strong></td>
<td>14,201,297</td>
<td>13,436,285</td>
<td>10,337bp - 100bp</td>
<td>2,569,346 bp</td>
</tr>
<tr>
<td><strong>Heliophila</strong></td>
<td>12,084,646</td>
<td>11,434,650</td>
<td>10,928bp - 100bp</td>
<td>1,470,345 bp</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>13,585,665</td>
<td>12,882,000</td>
<td>11,595 Contigs</td>
<td>2,623,919 bp</td>
</tr>
</tbody>
</table>
Brassicales: 18 families (4765 species)

- Phylogeny estimated with 1155 single copy nuclear genes
- Largely congruent with 78 plastid & 34 mitochondrial markers
- ML (RaxML), MP (PAUPRat), and Bayesian (BEAST)
- 100% Bootstrap support
Diversification Analyses show a “Lag-Time” for Speciation after WGD

I. Brassicaceae radiation:
- p-value = 0.0003430522
- bonf. p-value = 0.006861044
- “Mass-diversification” following At-α WGD occurred ~ 10 mya later

II. KT-Extinction radiation:
- p-value = 0.0019349845
- bonf. p-value = 0.038699690
- “Mass-diversification” following At-β WGD occurred minimum ~17 million years later

Thus, WGD alone does not lead to diversification

If lag, how may species radiations be associated with WGD?
Origin of Novel Pathways: WGDs

- Retained Whole Genome Duplicates (Neofunctionalized) At-β and At-α event – Novel Biosynthetic Pathways
  * Supported by Syntenic Blocks and Phylogenetic Analyses

* nearly ~50% GLS genes are tandem duplicates (Brassicaceae)

Hofberger et al., 2013 Genome Biol. Evol.
Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family
1. The Butterfly Plant Evolutionary Arms Race has a new fossil-calibrated phylogenomic chronogram

2. Origin of Novel Glucosinolate Pathways arose due to WGD, the $\text{At-}\beta$ WGD spurred arms race with indole GLS, and likely $\text{At-}\alpha$ WGD rearmed for some met-GLS

3. Specialist insect herbivores turn the ‘mustard oil bomb’ into a defensive ‘cyanide bomb’

4. Significant Adaptive Radiations on both sides
   A. Brassicales-feeding Pierid butterflies are 4X more species-rich
   B. Brassicales plants with Indolic GLS are 29X more species-rich
   C. Brassicaceae plants with Met GLS are 80X more species-rich
Outline: Phylogenomics

I. Order Brassicales
II. Family Brassicaceae
III. Tribe Brassiceae
IV. Future Directions
Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in family Brassicaceae

Collaborators: Chien-hsun Huang, Hong Ma, Ihsan Al-Shebaz, Marcus Koch, Klaus Mummenhoff and others

Current taxon sampling: 85 species
(across 33 tribes, 52 genera, 84 species)

Sequence data: 28 single copy nuclear genes from 6 genomes, 58 transcriptomes, and 27 PCR
Lineages I, II, III and tribes within the family Brassicaceae remain unresolved; taxon sampling shown in yellow
Outline: Phylogenomics

I. Order Brassicales
II. Family Brassicaceae
III. Tribe Brassiceae
IV. Future Directions
# Plant materials

- **Seed collection completed**
  - USDA, Madrid, Field trip

- **Plants were grown greenhouse**
  - Life Sciences Center MU

- **DNA and Herbarium vouchers**
  - Collections identified: Ihsan Al-Shebaz
Results

4-chloroplast marker phylogeny

1. Relationships among subtribe lineages in tribe Brassiceae more resolved

2. Recovered 7 previously identified clades and New African clade

3. Parallel evolution

Arias & Pires 2012 Taxon
**Tribe Brassiceae** has seven major lineages (or sub-tribes) identified using cp markers (Warwick 2005), but not well resolved.

<table>
<thead>
<tr>
<th>Genera</th>
<th>Oleracea</th>
<th>Nigra</th>
<th>Cakile</th>
<th>Vella</th>
<th>Zilla</th>
<th>Savignya</th>
<th>Crambe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Relationships among sub-tribes in Tribe Brassiceae are more resolved

Core Brassiceae

Oleracea (40 spp.)
Savignia (2 spp.)
Nigra (30 spp.)
Cakile (5 spp.)
Crambe (4 spp.)
Henophyton (15 spp.)
Zilla (5 spp.)
Vella (3 spp.)

Arias & Pires 2012 Taxon
Henophyton Lineage: New African clade

- Rare and Endemic genera from North Africa (Morocco and Algeria)
- DNA extracted from old herbarium specimens at MA

Henophyton
Trachystoma
Ammosperma
Pseuderucaria

Arias & Pires 2012 Taxon
Parallel evolution of morphological characters has led to polyphyletic genera

**Brassica**

**Erucastrum**

**Sinapis**

**Diplotaxis**
When did the tribe diversification occur?

The tribe origin & triplication event have been dated at

\[ \approx 28 \text{ to } 15.6 \text{ Mya (Beilstein & al., 2010)} \]

\[ \approx 23 \text{ to } 11 \text{ Mya (Couvreur & al., 2010)} \]
Tribe Brassiceae colonized North Africa about 24 MYA and moved clockwise through Europe.
Diversification dates agree with previous results; however, there is lag time between the whole genome triplication & diversification.
Tribe Brassiceae Conclusions


1. A new phylogeny of the tribe Brassiceae led to discovery of a New African clade (Henophyton)

2. Origin of tribe Brassiceae was inferred in the Saharo-Sindian region about 24 Ma. Ancestors colonized Europe through Western North Africa

3. Diversification shifts in the tribe do not coincide with the whole triplication event
Transcriptome Based Nuclear Tree Methods for Tribe Brassiceae

• RNA extracted from young leaf tissue
• Sequenced six species per lane on Illumina Hi-Seq with 2x100bp
• De novo transcriptome assembly and ortholog selection using the Agalma Pipeline (Dunn, et.al., 2013)
• Concatenated tree based on 1,663 loci with 72% occupancy estimated using RAxML (Stamatakis, 2014)
Outline: Phylogenomics

I. Order Brassicales
II. Family Brassicaceae
III. Tribe Brassicaceae
IV. Future Directions
How do genomes and networks evolve in *Brassica*?

- **Brassica**
  - **Brassica oleracea** (n=9)
    - cabbage
    - broccoli
    - cauliflower
    - Brussel sprouts
    - kale
    - Chinese cab.
    - pak choi
    - turnip
    - annual oilseed
    - biennial oilseed
  - **Brassica rapa** (n=10)
    - annual oilseed
    - biennial oilseed
    - rutabaga
  - **Brassica napus** (n=19)
    - annual oilseed
    - biennial oilseed

- **Arabidopsis**
  - **Arabidopsis thaliana** (n=5)

**Collaboration:**
- *B. oleracea* diversity: Graham Teakle, Guy Barker, et al.
- *B. rapa* diversity: Graham King et al.
- *B. napus* diversity: Michael Gore (Cornell)
Acknowledgments - People

MU Collaborators/Coauthors:
Gavin Conant
Corey Hudson
Michael Bekaert
Pat Edger  Zhiyong Xiong
Michelle Tang
Dustin Mayfield

External Collaborators/Coauthors:
Eric Schranz  Mike Barker
Marcus Koch  Klaus Mummenhoff
Martin Lysak  Mark Beilstein
Ihsan Al-Shehbaz

Team Butterfly: Chris Wheat, Heiko Vogel,
Hanna Heidel-Fischer, David Heckel,
Goran Hellekant
Acknowledgments - Funding

**NSF Plant Genome:**
Functional Genomics of Plant Polyploids.

**NSF Plant Genome Comparative Sequencing:**
Toward unravelling the morphological plasticity and genome redundancy of *Brassica oleracea*.

**NSF Tree of Life Program:**
From Acorus to Zingiber: Assembling the Phylogeny of Monocots

**MU Life Sciences Fellowships, LSUROP, UMEB**

**Society for Systematic Biology**
**Botanical Society of America**
**American Society of Plant Taxonomists**
Thank You!