SOYBEAN RUST, A RISING STAR IN PHYTOPATHOLOGY

Martha Lucia Posada-Buitrago Ph.D.
Molecular Biologist

DOE- Joint Genome Institute
Lawrence Berkeley National Laboratory
Soybean Rust

Caused by two species of fungi:

*Phakopsora pachyrhizi*
aka “Old World” or “Asian” isolate
More aggressive pathogen.

*Phakopsora meibomiae*
aka “New World” or “American” isolate
Not as aggressive
**LEGUMES** (Papilionoideae)

**Cultivated Crops:**
- *Glycine max* (soybeans)*
- *Phaseolus lunatus* (lima and butter beans)*
- *Phaseolus vulgaris* (green beans, kidney beans)
- *Vigna unguiculata* (cowpeas)*
- *Cajanus cajan* (pigeon peas)
- *Pachyrhizus erosus* (yam bean, jicama)*

**Ornamental plants:**
- Hyacinth bean, lupine, royal poinciana

**Wild hosts:**
- Kudzu, sweet clover

Photos by Reid D. Frederick
<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>How Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>1904</td>
<td>Thought to be windborne from Asia</td>
</tr>
<tr>
<td>Kenya</td>
<td>1997/1998</td>
<td>Thought to be windborne from Africa</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>1997/1998</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>Paraguay</td>
<td>2001/2002</td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>Argentine</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>Bolivia</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>2004</td>
<td>Hurricane Ivan</td>
</tr>
<tr>
<td>USA</td>
<td>Oct 2004</td>
<td></td>
</tr>
</tbody>
</table>
Soybean Rust in the World

- **P. pachyrhizi**
- **P. meibomiae**
Premature defoliation

Yield decrease characterized by:
- Increase in number of unfilled pods/plant
- Decrease in number of normal pods/plant
- Decrease in number of seeds/plant
- Decrease in weight of seed/plant
- Decrease in 1000-seed weight
- Decrease in germinability of seed
Soybean fields (Zimbabwe)

Photos by Reid D. Frederick
### Glycine max cv. Williams – Phakopsora pachyrhizi

**Interacción Susceptible**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 h</td>
<td>Appresoria begin developing</td>
</tr>
<tr>
<td>5 h</td>
<td>Appresoria expansion</td>
</tr>
<tr>
<td>7-12 h</td>
<td>Penetration through cuticle</td>
</tr>
<tr>
<td>12-16 h</td>
<td>Increase in diameter</td>
</tr>
<tr>
<td>24 h</td>
<td>Primary hyphae emerging from tev</td>
</tr>
<tr>
<td>48 h</td>
<td>Intercellular hyphal growth (60 µm from penetration site)</td>
</tr>
<tr>
<td>3-8 days</td>
<td>Intercellular hyphal growth (75-450 µm from penetration site)</td>
</tr>
<tr>
<td>9 days</td>
<td>Sporulation</td>
</tr>
<tr>
<td>14 days</td>
<td>Sporulation peak</td>
</tr>
</tbody>
</table>

*(Based on Koch et al. 1983; Keogh et al. 1980)*
Symptoms

Infected leaves

9 dpi

12 dpi

15 dpi

18 dpi

Photos by Christine Stone
Genome Sequencing Project

Funded by

the U.S. Department of Agriculture/
Agricultural Research Service (USDA/ARS)

the U.S. Department of Energy/
Joint Genome Institute (DOE-JGI)
Random shotgun libraries:
  3kb insert size in vector pUC18,
  Mid-size insert 8-10kb in vector p21
  36-40kb insert size in pCC1FOS (Fosmids)
cDNA libraries from different stages of *P. pachyrhizi*

**Sequencers:**
- ABI3730
- MegaBACE 4000

**Informatics:**
- Reads processing by Phred
- Reads assembly by Phrap
- Verification
- Genome annotation
<table>
<thead>
<tr>
<th>Library</th>
<th>Bases sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(Insert size)</em></td>
<td></td>
</tr>
<tr>
<td><em>P. pachyrhizi</em></td>
<td></td>
</tr>
<tr>
<td>3 Kb</td>
<td>146.60 Mb</td>
</tr>
<tr>
<td>8 Kb</td>
<td>264.28 Mb</td>
</tr>
<tr>
<td>40 Kb</td>
<td>5.75 Mb</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>416.63 MB</strong></td>
</tr>
<tr>
<td><em>P. Meibomiae</em></td>
<td></td>
</tr>
<tr>
<td>3 Kb</td>
<td>125.20 Mb</td>
</tr>
<tr>
<td>8 Kb</td>
<td>5.97 Mb</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>131.17 MB</strong></td>
</tr>
</tbody>
</table>

DOE-JGI Data by 5.31.05
The mean G+C content in *P. pachyrhizi* and *P. meibomiae* is 34-35%, estimated with the “G+C content program” (Chapman) on sequences from three different genomic libraries.
Several independent methods were used to estimate the genome size. Although there were considerable uncertainties associated with most of the methods, they consistently yielded a genome size above 500 MB.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Genome Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA Coverage</td>
<td>720 Mb</td>
</tr>
<tr>
<td>All-Pairs Read Alignment</td>
<td>500-800 Mb</td>
</tr>
<tr>
<td>Gene Density</td>
<td>300-700 Mb</td>
</tr>
<tr>
<td>Shotgun Fosmid Coverage</td>
<td>600-950 Mb</td>
</tr>
</tbody>
</table>
Random fosmids

Finishing at Stanford:

Finished 87 (approx. 3.48 Mb x 8)
Incomplete 28 (approx. 1.12 Mb x 8)

Selected fosmids

Lawrence Livermore National Laboratory (LLNL):

Probes designed for 120 “genes”

Selected 50
To go 70
Sequenced 15
Finished 0

Probes designed based on ESTs selected by high similarity to “interesting” genes from other fungi and unknown genes highly expressed in germinating spores from *P. pachyrhizi*. 
Germinating Spores | Resting spores | Hyphal growth* | High sporulation*
---|---|---|---
16 Hours on water surface | Kept at – 80°C | Days after inoculation | Days after inoculation
6 | 7 | 8 | 13 | 14 | 15

* :mRNA was extracted from infected leaf at each time point and pooled together for the construction of the cDNA libraries. Unidirectional cDNA libraries constructed in plasmid pSPORT1 (Invitrogen).

<table>
<thead>
<tr>
<th>Description</th>
<th>ESTs</th>
<th>cDNAs</th>
<th>Libraries</th>
<th>Clusters</th>
<th>Consensus</th>
<th>Singlets</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-8 dpi</td>
<td>6100</td>
<td>5374</td>
<td>1</td>
<td>1154</td>
<td>1278</td>
<td>1827</td>
</tr>
<tr>
<td>13-15 dpi</td>
<td>6023</td>
<td>4610</td>
<td>1</td>
<td>1291</td>
<td>1387</td>
<td>1356</td>
</tr>
<tr>
<td>Resting urediniospores</td>
<td>2295</td>
<td>1762</td>
<td>1</td>
<td>393</td>
<td>455</td>
<td>335</td>
</tr>
<tr>
<td>Germinating urediniospores</td>
<td>29601</td>
<td>18638</td>
<td>1</td>
<td>2686</td>
<td>3394</td>
<td>2142</td>
</tr>
<tr>
<td>Phakopsora pachyrhizi v2.1</td>
<td>44019</td>
<td>30244</td>
<td>4</td>
<td>5105</td>
<td>6165</td>
<td>4961</td>
</tr>
</tbody>
</table>
Percentage of similarity of cDNA clusters from the *Phakopsora pachyrhizi* germinating and resting spores libraries and the infected soybean leaf libraries (6-8 dpi and 13-15 dpi) to proteins from other organisms. Inner pies show the percentage of similarity of cDNA clusters to proteins from other organisms, excluding plant homologs.
The cDNA clusters were classified into functional categories based on the BlastX hits and the Pfam hits, according to the Expressed Gene Anatomy database (EGAD, TIGR, Rockville, MD).

Approximately 23% of the cDNA clusters from the 6-8 dpi and 13-15 dpi libraries and 40% from the germinating and resting spores libraries show similarity to hypothetical proteins or proteins of unknown function.

Several homologs to pathogenesis related proteins (PR proteins) and defense proteins were identified in the infected leaf tissue libraries (Apidaecin, Beta defensin, Thaumatin, etc). In the GS library several homologs to pathogenicity proteins were identified. All the libraries show a high percentage of metabolism related proteins.
Known mitochondrial genome sequences were blasted against the entire set of reads from the genome project. Potential mitochondrial sequences were assembled with the Phred Phrap Package. This resulted in single contig assemblies for both fungal mitochondrial genomes, *P. pachyrhizi* and *P. meibomiae*.

**Genome analysis and annotation:**

**DOGMA** Dual Organellar GenoMe Annotator (http://bugmaster.jgi-psf.org/dogma).

**tRNAscan-SE 1.21** (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/)

**MacVector 7.1** (Accelrys)

**Blast algorithm**
These genomes contain:

- ATP synthase subunits 6, 8, and 9 (*atp6, atp8, and atp9*)
- cytochrome oxidase subunits I, II, and III (*cox1, cox2, and cox3*)
- apocytochrome b (*cob*)
- reduced nicotinamide adenine dinucleotide ubiquinone oxireductase subunits (*nad1, nad2, nad3, nad4, nad4L, nad5, and nad6*)
- the large and small mitochondrial ribosomal RNAs (*rnl* and *rns*).
- t*RNAs* for all amino acids.
Phakopsora meibomiae
mtDNA
32,520bp
Comparison of mitochondrial genomes from the four phyla of fungi. Protein-coding and rRNA genes are represented by boxes; arrows indicate the direction of transcription. Lines within genes represent presence of intron(s).
Phylogenetic tree of 1582 amino acid position from seven mitochondrial-encoded proteins from 21 taxa, including 18 species from all fungal phyla and *Monosiga brevicollis*, *Phytophthora infestans* and *Reclinomonas americana* as outgroups. The genes encoding cob, cox1, cox2, cox3, nad1, nad4 and nad5 are present in all organisms compared. Parsimony-bootstrap support was calculated from 100 replicates using Paup 4.0b10.
Acknowledgements

Reid D. Frederick
Jane J. Choi
Christine L. Stone
Craig Austin
Laura Ewing
Jeffrey L. Boore
Peter Brokstein
Nick Putman
Harris Shapiro
Jarrod Chapman

Lawrence Livermore National Laboratory
Laurie Gordon
• This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.