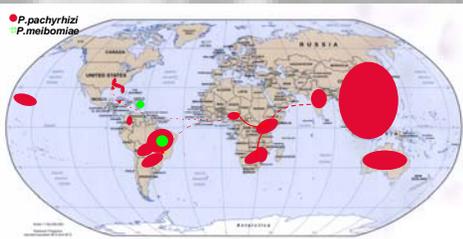


## INTRODUCTION

Soybean Rust is caused by two closely related species of fungi: *Phakopsora pachyrhizi* (aka "Old World" or the Asian isolate; more aggressive pathogen) and *Phakopsora meibomia* (aka "New World" or the American isolate; not as aggressive). *P. pachyrhizi* is responsible for significant losses of soybean crop in Africa, Asia, Australia and South America and it is considered a major threat to world soybean production. Recent findings of this pathogen in the Southeast continental U.S.A have triggered an increasing interest in the scientific community. No commercially available soybean cultivars are resistant to all isolates of rust and very little is known about the molecular mechanisms involved in the soybean-rust interaction. Currently, in the areas where the soybean rust is present, fungicides are used to control the disease, but they can be expensive or potentially hazardous. In order to develop new strategies to control the disease, it is crucial to increase our understanding of the biology of the pathogen and the infection process.

Here, we present strategies and preliminary results from the *P. pachyrhizi* Genome Sequencing Project.



Geographical distribution of soybean rust.

## RATIONALE

### Comprehensive understanding of host-pathogen interactions to develop effective and durable resistance

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 2 summarizes the results for each method.

Estimation Method	Genome Size
cDNA Coverage	720 Mb
All-Pairs Read Alignment	500-800 Mb
Gene Density	300-700 Mb
Shotgun Fosmid Coverage	600-950 Mb

Table 2. Summary of *P. pachyrhizi* genome size estimates

### Fosmid sequencing strategy

#### Random fosmids

Finishing at Stanford University:

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
Sequencing	24
Finished	0

Probes were designed based on ESTs selected by high homology to pathogenicity related genes, important metabolic pathway genes or highly expressed genes during spore germination from *P. pachyrhizi*.

### MITOCHONDRIAL GENOMES

Known mitochondrial genome sequences were blasted against the entire set of reads. Potential mitochondrial sequences were assembled with the Phred Phrap Package. This resulted in single contig assemblies for both fungal mitochondrial genomes.

#### 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

#### Phylogenetic analysis

1582 amino acid position from seven mitochondrial-encoded proteins were analyzed for 21 taxa. The genes encoding *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.

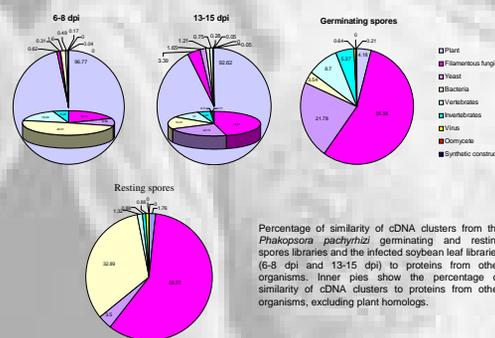
	<i>P. Pachyrhizi</i>	<i>P. meibomia</i>
Size	31.82 Kb	32.52 Kb
G+C	34.6 %	34.9 %

## cDNA libraries

Germinating Spores	Resting spores	Hyphal growth	High sporulation
16 Hours on water surface	Kept at -80°C	6 Days after inoculation	13 Days after inoculation
		7	14
		8	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).

Description	ESTs	cDNAs	Libraries	Clusters	Consensus	Singlets
6-8 dpi	6100	5374	1	1154	1278	1827
13-15 dpi	6023	4610	1	1291	1387	1356
Resting urediniospores	2295	1762	1	293	455	335
Germinating urediniospores	29601	18638	1	2686	3394	2142
<i>Phakopsora pachyrhizi</i> v2.1	44019	30244	4	5105	6165	4961



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 (ApoEacitin, 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.



## Initial Genome Project Strategy

### Random shotgun libraries:

General 3kb insert size in vector pUC18,  
Mid-size 8-10kb insert in vector p21  
Fosmid (40kb insert size) in pCC1FOS

### cDNA libraries from different stages of *P. pachyrhizi* (in pSPORT1)

#### Sequencers:

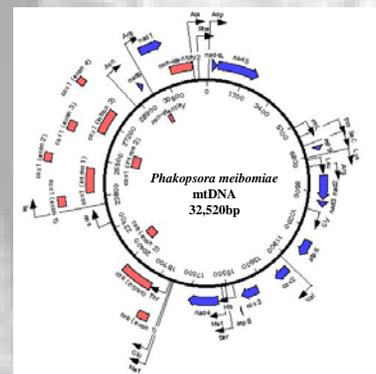
ABI3730  
MegaBACE 4000

#### Informatics:

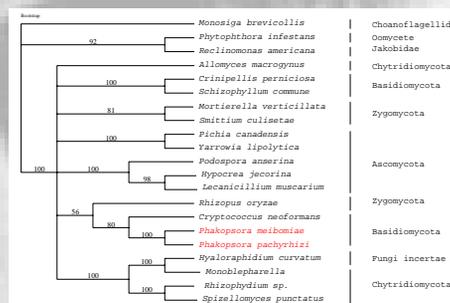
Reads processing by Phred  
Reads assembly by Phrap  
Verification  
Genome annotation

	Library (Insert size)	Bases sequenced
<i>P. pachyrhizi</i>	3 Kb	146.60 Mb
	8 Kb	264.28 Mb
	40 Kb	5.75 Mb
<b>Total</b>		<b>416.63 Mb</b>
<i>P. Meibomia</i>	3 Kb	125.20 Mb
	8 Kb	5.97 Mb
<b>Total</b>		<b>131.17</b>

Table 1. Summary of *P. pachyrhizi* and *P. meibomia* sequencing at the JGI.



*Phakopsora meibomia* mitochondrial functional map



Phylogenetic tree of 1296 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.

## Sequencing is in progress with funding through a joint agreement between the USDA-ARS and the DOE-JGI

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-AC02-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36. <http://www.jgi-psf.org> is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable. The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, sex, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write USDA, Office of Civil Rights, Room 326-W, Whitten Building, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.