LBNL-58995 Spanish

Presentación del DOE-Joint Genome Institute

Martha Lucia Posada-Buitrago, Ph.D.
Molecular Biologist
Mission:

To develop and exploit new sequencing and other high-throughput, genome-scale and computational technologies as a means for discovering and characterizing the basic principles and relationships underlying the organization, function, and evolution of living systems.
JGI Partnerships

U.S. D.O.E. JOINT GENOME INSTITUTE

Stanford
Finishing

LANL
Los Alamos National Laboratory
R&D, Finishing, Informatics

LBNL
Lawrence Berkeley National Laboratory
Informatics, Mapping, R&D

LLNL
Lawrence Livermore National Laboratory
Informatics, Finishing, Mapping

PNNL
Proteomics
The Pacific Northwest National Laboratory

ORNL
Oak Ridge National Laboratory
Mouse genetics, Annotation

JGI-PGF
Walnut Creek, CA
Sequencing, Informatics, Comparative and Functional Genomics, R&D
DNA Sequencing Production at the JGI

- April 2002: 1 gb/month
- January 2004: 2 gb/month
- July 2004: 2.5 gb/month
- March 2005: 3.1 gb/month

(equivalent to 1 human genome/month)

Total (3/99-4/05) 82.893 gb
(equivalent of sequencing 27 human genomes)
1. Shear DNA
2. Ligate into pUC18
3. Transform
4. Plate
5. Pick colonies
6. Grow overnight
DNA sequencing process
Rolling circle amplification of plasmid clones for sequencing template

1. PlateMate adds lysis buffer to small amount of culture

2. Cells are heat-lysed

3. Hydra 384's with Twister arms add RCA reagents.

4. ON incubation
DNA sequencing process
Sequencing Chemistry

F and R reactions are separated with hydra 384’s with twister arms

Sequencing reagents are added with Cavro Dispense System
Sequence reactions with Quad-head PCR machines, then clean-up using BioMek robots and SPRI
DNA sequencing process
Capillary Group

35 MegaBACE
4000
60 ABI 3730
Q20 / month = 3.1 Gb
Online tracking of progress

LIMS uses barcode readers at every step and allows real-time tracking of all reagents, personnel, and processes.
Informatics team assembles, verifies, annotates genomes

Best assemblies come from end sequences from a mixture of clone sizes.

Typically, the JGI makes 3 libraries: 3-4 Kb in plasmids, 8-10 Kb in plasmids, 40 Kb in fosmids.
Genome annotation and visualization tools

<table>
<thead>
<tr>
<th>Genome Browser</th>
<th>Search</th>
<th>Filter</th>
<th>GO</th>
<th>Browse</th>
<th>Annotation</th>
<th>Clone Home</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Scaffold</th>
<th>Position</th>
<th>Size</th>
<th>Name</th>
<th>Type</th>
<th>Predicted</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaffold_1</td>
<td>1-67247</td>
<td>30000</td>
<td>DNA</td>
<td>Gene</td>
<td>Protein</td>
<td>Protein</td>
</tr>
<tr>
<td>Scaffold_2</td>
<td>100000</td>
<td>100000</td>
<td>RNA</td>
<td>Gene</td>
<td>RNA</td>
<td>RNA</td>
</tr>
<tr>
<td>Scaffold_3</td>
<td>150000</td>
<td>150000</td>
<td>miRNA</td>
<td>Gene</td>
<td>miRNA</td>
<td>miRNA</td>
</tr>
</tbody>
</table>

**Annotation Details**
- **Gene Ontology (GO):** Gene Ontology terms associated with each annotation.
- **Function:** Predicted function based on annotation.

**Search Options**
- **Base Position:** Search by base position.
- **File:** Download or export data.
- **GO:** Search by Gene Ontology terms.
- **Clone:** Search by clone name.

**Visualization Options**
- **Zoom:** Adjust zoom level.
- **Display:** Customize display settings.

**Annotation Tools**
- **Gene Finder:** Find genes within specified positions.
- **Sequence Viewer:** View sequence details.
- **Expression Profile:** Analyze expression levels across different conditions.
11% of Human Genome Sequenced by JGI
Selected JGI Microbes

- Burkholderia cepacia
- Cytophaga hutchinsonii
- Desulfitobacterium halfniese
- Enterococcus faecium
- Ferroplasma acidarmanus
- Magnetospirillum magnetotacticum
- Nitrosomonas europaea
- Prochlorococcus marinus
- Pseudomonas fluorescens
- Rhodobacter sphaeroides
- Rhodopseudomonas palustris
- Sphingomonas aromaticivorans
- Thermomonospora fusca
- Trichodesmium erythraeum
- Xylella fastidiosa
- Nostoc punctiforme
- Marine synechococcus
- Magnetococcus MC-1
• Close relative of the well-studied X. laevis, a major model organism for developmental biology
• Favorite system for toxicology (EPA)
• Coordinated with
  • WashU BAC map project
  • cDNA projects at NIH, Sanger
  • other projects from international frog research community
• 7x coverage by early ‘05
Trichoderma reesei possesses a host of carbohydrate degrading enzymes and is used extensively in industrial processes.

White rot fungi like *P. chrysosporium* are uniquely able to degrade lignin, the second most abundant natural polymer and a major component of biomass.

Phakopsora pachyrhizi & *P. meibomiae*

Soybean rust was recently found in US.

Highly repetitive sequence
Sudden Oak Death

- In partnership with the USDA, NSF, VBI, California Oak Mortality Taskforce, County Ag Commissioners, City of Walnut Creek, WC Chamber of Commerce.
- 4 TV Stations; various print media
40,000-year-old DNA

DNA Extraction (tooth)

DNA Comparisons

94% Microbe/Other DNA

6% Cave Bear DNA

High Throughput DNA Sequencing
• Allocation of ~15 gigabases/year for sequencing to advance the frontiers of science supported by DOE

• 56 Proposals received in Feb. ‘04 totaling 100Gb in requested sequencing (equivalent to the current WW sequencing capacity)

• 150 Proposals received in Feb. ’05.

• SAC approved 23 projects beginning Fall ‘04

• New RFP Spring ‘06
Roya de la soja: Una Nueva Amenaza

Martha Lucía Posada-Buitrago Ph.D.
Bióloga Molecular

DOE- Joint Genome Institute
Lawrence Berkeley National Laboratory
Causada por dos especies de hongos:

*Phakopsora pachyrhizi*

aka aislamiento del “viejo mundo” o “Asiático”

Patógeno muy agresivo.

*Phakopsora meibomiae*

aka aislamiento del “Nuevo Mundo” o “Americanos”

No muy agresivo
Hospederos de la roya de la soya

LEGUMINOSAS
(Papilionoideae)

Especies cultivables:
Glycine max
Phaseolus lunatus
Phaseolus vulgaris
Vigna unguiculata
Cajanus cajan
Pachyrhizus erosus

Plantas ornamentales:
Lupinus, poinciana real

Hospederos salvajes:
Kudzu, clavo dulce

Kudzu infectado con la roya de la soja

Photos by Reid D. Frederick
<table>
<thead>
<tr>
<th>País</th>
<th>Año</th>
<th>Nota</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japón</td>
<td>1904</td>
<td>Se piensa que fue transportada por las corrientes de aire desde Asia</td>
</tr>
<tr>
<td>Kenia</td>
<td>1997/1998</td>
<td></td>
</tr>
<tr>
<td>Sudáfrica</td>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>Paraguay</td>
<td>2001/2002</td>
<td>Se piensa que fue transportada por las corrientes de aire desde Africa</td>
</tr>
<tr>
<td>Brasil</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>Bolivia</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>Junio 2004</td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>Oct 2004</td>
<td>Huracán Iván</td>
</tr>
</tbody>
</table>
Situacion actual de la roya de la soja

P. pachyrhizi
P. meibomiae
Efectos de la roya

Defoliación prematura

Disminución de la producción caracteriza por:
- Aumento en el número de vainas vacías por planta
- Disminución en el número de vainas normales por planta
- Disminución en el número de semillas/planta
- Disminución en el peso de semillas/planta
- Disminución en la germinabilidad de las semillas
Campos de soja infectada (Zimbabwe)
Síntomas de la roya
Why Containment?

- Biological activity of organisms
- Potential benefit to agriculture
- Need for research
- Existing regulations
Double-pane tempered safety glass
Negative air pressure
HEPA Air filtration
Waste water treatment
Steam and ethylene oxide sterilization
Precautions by personnel

- Authorized personnel only
- Clean suits
- Shower out
15-17 days old plants

Inoculation

ca 250,000 spores/pot

Dew chamber. 20°C, 16 - 24 hours
Spore harvesting (14 – 21 dpi)
Sintomas de la roya

Cotiledones Infectados

Tallo infectado

Vainas infectadas

Photos by Christine Stone
Hojas Infectadas

9 días

12 días

15 días

18 días

Photos by Christine Stone
Estudios de expresión génica en planta infectada: Real Time RT PCR, ESTs (genotecas de cDNA en pSPORT1 y genotecas sustractivas –Clontech–). (USDA/ARS/FDWSRU y DOE-JGI).

• Secuenciación del genoma de *Phakopsora pachyrhizi* (USDA/ARS/FDWSRU y DOE-JGI).

• Proteínas de esporas dormantes, en germinación y de planta infectada (reacciones susceptible y resistente) (USDA/ARS/FDWSRU).

• Diseño de marcadores genéticos para identificación de aislamientos (USDA/ARS/FDWSRU, DHS).
• Búsqueda de resistencia en todas las variedades conocidas de soja (25000) y algunos huéspedes alternativos (Illinois University, USDA/ARS/FDWSRU).

• Tratamientos con fungicidas (Illinios University, USDA/ARS/FDWSRU, Cultivadores de Zimbabwe y Suráfrica, EMBRAPA Brasil).

• Educación y preparación de personal y trabajadores en campo para identificación del patógeno y acción de emergencia (USDA/ARS/FDWSRU, USDA/APHIS, Universidades estatales de los estados productores de soja). Estudios satelitales climáticos y de movimiento de esporas (Iowa State University - ISU).

• Anticuerpos (ISU, US Navy).
Estudios de Expresión Génica
Comprensión de las bases moleculares y bioquímicas del proceso de infección de la roya de la soja que pueda conducir a nuevos métodos de control de la enfermedad.
**Glycine max cvar Wiliams – Phakopsora pachyrhizi**

**Interacción Susceptible**

- **2 h** Desarrollo de los apresorios
- **5 h** Expansión de los apresorios
- **7-12 h** Penetración a través de la cutícula
- **12-16h** Aumento en el diámetro
- **24 h** Hifas primarias emergen desde la tev
- **48 h** Crecimiento Intercelular de las hifas (60µm desde el sitio de penetración)
- **3-8 días** Crecimiento Intercelular de las hifas (75-450 m desde el sitio de penetración)
- **9 días** Esporulación
- **14 días** Máxima esporulación

(Basado en Koch et al. 1983; Keogh et al. 1980)
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.
Real Time RT-PCR Studies
Advantages of Real-time PCR

- Combines the sensitivity of PCR along with hybridization
- Post-PCR confirmation steps are eliminated
  - No agarose gels & Southern blots
- Assays are very rapid
- Potential for multiplexing
  - Several probes of a particular pathogen
  - Specific probes from several pathogens
Polymerization

Forward Primer

Probe

Reverse Primer

5' 3' 5'

5' 3' 5'

5' 3' 5'

Reporter

Quencher
Strand Displacement

Forward Primer

Probe

Reverse Primer

5' 3' 5'

5' 3' 5'

Reporter

Quencher

5' 3' 5'

5' 3' 5'
Polymerization Completed

- Forward Primer
- Probe
- Reverse Primer
- Reporter
- Quencher
Advantages of Real Time PCR

- Combines the sensitivity of PCR along with hybridization.
- Post-PCR confirmation steps are eliminated
  - No agarose gels & Southern blots
- Assays are very rapid
- Potential for multiplexing
  - Several probes of a particular pathogen
  - Specific probes from several pathogens
Smart Cycler® XC

External Case

- Ruggedized
- Air Tight
- Water Proof

Desktop Smart Cycler for reference
32 capillary chamber

R.A.P.I.D
**Gene Selection**

*P. pachyrhizi* putative Heat-induced catalase, ATP-binding cassette (ABC) transporter, Plasma membrane (H+) ATPase and two constitutive genes, putative Alpha and Beta-tubulin, were selected from the ESTs from the germinated spores to study their gene expression during the infection cycle of *P. pachyrhizzi* on soybean.

**RNA**

Total RNA (40ng) from non-infected plant, germinating spores, infected leaf tissue from 1, 2, 4, 6, 8, 10, 12 and 14 dpi were used as template. Positive controls were performed using fungal DNA (25ng), while RNase treated RNA samples and no template were used as negative controls.

**Real Time RT-PCR**

Real Time RT-PCR was performed in the ABI Prism 7700 (Perkin Elmer) with 40ng of total RNA using the SuperScript One-Step RT-PCR with Platinum Taq Kit (Invitrogen), following the manufacturer’s protocol for a 25µl reaction.
Primers and probes designed for Real Time RT-PCR assays

<table>
<thead>
<tr>
<th>Putative Gene</th>
<th>Forward Primer</th>
<th>Fluorogenic Probe</th>
<th>Reverse Primer</th>
<th>Amplicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat-induced Catalase</td>
<td>CCTGGTGAGAGCCGTTCTGCA</td>
<td>FAM- ACCCAGTCCTTAATCGAGGCTATTTCC-TAMRA</td>
<td>TGACGATGGGTGCAGGGT</td>
<td>70</td>
</tr>
<tr>
<td>ABC Transporter</td>
<td>GAAACATTGGATGTACAACCTGGA</td>
<td>FAM- CCCTATACCTCGGGATTGGAAGTCTCTT-TAMRA</td>
<td>TCGAGTGTCGAGCCTCATTT</td>
<td>76</td>
</tr>
<tr>
<td>Plasma membrane (H+) ATPase</td>
<td>TCGTTACACGGCTGTGTTT</td>
<td>FAM- TTATAGAGAGACCATCGGCGGCTTT-TAMRA</td>
<td>AGCAATCAGAAAAGCGCCC</td>
<td>68</td>
</tr>
<tr>
<td>Alpha-tubulin</td>
<td>CCAAGGGCTTCTTCTGTGTTCA</td>
<td>FAM- TCGTGGAGGAGCCGACTGACTGCTTTCA-TAMRA</td>
<td>CAAGAGAGAGCGCCAAACC</td>
<td>65</td>
</tr>
<tr>
<td>Beta-tubulin</td>
<td>CCCGTGCCAGTTTGTATTG</td>
<td>FAM- TTGGAAACCCGAAATCGGATTCCG-TAMRA</td>
<td>CCAAGGATCCCGGATCGA</td>
<td>64</td>
</tr>
</tbody>
</table>

Putative genes of *P. pachyrhizi* selected from the germinating spores cDNA library. Primers and probes were designed using Primer Express 1.0 (Perkin Elmer). Primers (Operon); Probes (Synthegen).

**Thermal cycling conditions (ABI 7700):**

- 50°C for 15 min (hold)
- 95°C for 10 min (hold)
- 40 cycles: 95°C for 15 s
- 60°C for 1 min
Real Time RT-PCR spectra for 10dpi. Fungal DNA (positive control), RNase treated sample and no template (negative controls).
Real Time RT-PCR spectra for Alpha-tubulin
Expression patterns of five putative genes over the infection cycle of *P. pachyrhizi* on *G. max* generated using Real Time RT-PCR. $C_T$ (threshold cycle) is the cycle in which a significant increase in $\Delta R_n$ is detected. Germinating spores (GS) were used as a positive control. dpi: days post inoculation.
heat-induced catalase - 60C, 1 min

528 2mM MgSO4, 60C, 1 min: 10-7-03; FDNA, NIP, NT

528 3mM MgSO4, 60C, 1 min: 10-6-03; NT, NIP, FDNA

528 3mM MgSO4, 60C, 1 min: 10-7-03; FDNA, NIP, NT

528 4mM MgSO4, 60C, 1 min: 10-7-03; FDNA, NIP, NT

528 5mM MgSO4, 60C, 1 min: 10-7-03; FDNA, NIP, NT

528 5mM MgSO4, 60C, 1 min: 10-7-03; FDNA, NIP, NT
**heat-induced catalase – 62 C, 30 sec**

528 4mM MgSO4, 62C, 30 sec:
10-9-03; FDNA, NIP, NT

528 5mM MgSO4, 62C, 30 sec:
10-9-03; FDNA, GS RNA, NT

528 6mM MgSO4, 62C, 30 sec:
10-9-03; FDNA, NIP, NT

528 5mM MgSO4/BSA, 62C, 30 sec:
10-9-03; FDNA, GS RNA, NT
heat-induced catalase – 64 C, 30 sec

528 4mM MgSO4, 64C, 30 sec: 10-9-03; FDNA, NIP, NT

528 5mM MgSO4, 64C, 30 sec: 10-9-03; FDNA, GS RNA, NT

528 6mM MgSO4, 64C, 30 sec: 10-9-03; FDNA, NIP, NT

528 6mM MgSO4/BSA, 64C, 30 sec: 10-9-03; FDNA, GS RNA, NT
Plasma membrane (H+) ATPase – 60C, 1 min

1340 3mM MgSO₄, 60C, 1 min: 10-6-03

1340 2mM MgSO₄, 60C, 1 min: 10-8-03; FDNA, NIP, NT

1340 3mM MgSO₄, 60C, 1 min: 10-8-03; FDNA, NIP, NT

1340 4mM MgSO₄, 60C, 1 min: 10-8-03; FDNA, NIP, NT

1340 5mM MgSO₄, 60C, 1 min: 10-8-03; FDNA, NIP, NT

1340 6mM MgSO₄, 60C, 1 min: 10-8-03; FDNA, NIP, NT
Plasma membrane (H+) ATPase – 4mM

1340 4mM MgSO4, 62C, 30 sec:
10-9-03; FDNA, NIP, NT

1340 4mM MgSO4, 62C, 30 sec:
10-9-03; FDNA, GS RNA, NT

1340 4mM MgSO4, 64C, 30 sec:
10-9-03; FDNA, NIP, NT
**Alpha-tubulin, 60C, 1 min**

A-tub 3mM MgSO4, 60C, 1 min: 10-6-03; FDNA, NIP, NT

A-tub 2mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

A-tub 3mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

A-tub 4mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

A-tub 5mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

A-tub 6mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT
Alpha-tubulin 4mM MgSO4, 30 sec

**A-tub** 4mM MgSO4, 62C, 30 sec: 10-9-03; FDNA, NIP, NT

**A-tub** 4mM MgSO4, 64C, 30 sec: 10-9-03; FDNA, NIP, NT
**ß-tub - 60C, 1 min**

**ß-tub** 2mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

**ß-tub** 3mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

**ß-tub** 4mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

**ß-tub** 5mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT

**ß-tub** 6mM MgSO4, 60C, 1 min: 10-8-03; FDNA, NIP, NT
**B-tub 4mM MgSO4, 30 sec:**

10-9-03; FDNA, NIP, NT
ABC transporter (ATP-binding cassette transporter) - 62C, 30 sec

917 2mM MgSO4, 62C, 30 sec: 10-10-03; FDNA, NIP, NT

917 3mM MgSO4, 62C, 30 sec: 10-10-03; FDNA, NIP, NT

917 4mM MgSO4, 62C, 30 sec: 10-10-03; FDNA, NIP, NT

917 5mM MgSO4, 62C, 30 sec: 10-10-03; FDNA, NIP, NT

917 6mM MgSO4, 62C, 30 sec: 10-10-03; FDNA, NIP, NT
ABC transporter (ATP-binding cassette transporter) - 64C, 30 sec

917 2mM MgSO₄, 64C, 30 sec: 10-10-03; FDNA, NIP, NT

917 3mM MgSO₄, 64C, 30 sec: 10-10-03; FDNA, NIP, NT

917 4mM MgSO₄, 64C, 30 sec: 10-10-03; FDNA, NIP, NT

917 5mM MgSO₄, 64C, 30 sec: 10-10-03; FDNA, NIP, NT

917 6mM MgSO₄, 64C, 30 sec: 10-10-03; FDNA, NIP, NT
Specific objectives

• Develop a suppression subtractive hybridization (SSH) library of the resistant interaction and identify transcripts/ESTs (Expressed Sequence Tags)
Suppression subtractive hybridization (SSH) cDNA library

Two libraries:
- pooled RNA from t = 1, 6, 12, 24, 48hpi (each from first trifoliate, from 4 plants)

- **Forward Subtraction:**
  - **Tester** = Komata/HW94  [Resistant/immune]
  - **Driver** = Komata/TW72  [Susceptible]

- **Reverse Subtraction:**
  - **Tester** = Komata/TW72  [Susceptible]
  - **Driver** = Komata/HW94  [Resistant/immune]
Suppression subtractive hybridization (SSH) cDNA library

Resistant rxn “Tester”

Suppress “common” transcripts & pcr-select “uncommon” from tester

Susceptible rxn “Driver” (overabundance)

Activator or up-regulated gene in a resistant reaction
Suppression subtractive hybridization (SSH)
cDNA library

- Our unique approach:
  "driver" = susceptible interaction

- This should identify not just the general "defense-related" genes of typical pathogen invasion, also genes that are differentially turned on that prevents the disease from progressing

- Suppression should allow for the identification of unique, rare gene expression
Results:

Forward Subtraction Library:

– 1056 clones sent for single-pass sequencing
  [Nucleic Acid Facility (NAF) at USDA-ARS-ERRC in Wyndmoor, PA]
– 45 clones did not sequence
– Due to method (blunt-digest) clones with multi-inserts (~15%), 1182 ESTs
– Insert sizes of EST ranged from 52nt to >600nt, no full-length transcripts were identified
– A low-redundant subset of 979 EST
Subtractive suppressive cDNA library
Data Analysis

- Comparative genomics using sequence-similarity tool BLAST (Basic local alignment search tool)
  - BLASTx = protein database
  - EST = dbEST
  - Unigene = compiled cluster of sequences from ESTs/mRNA/genomics projects
- Further analysis into Functional Categories
  (MIPS- Munich Information Center for Protein Sequences)
Results: Forward Subtraction
Functional Categories

- Unknown/ Unclassified: 26%
- Metabolism: 12%
- Secondary Metabolism: 2%
- Energy: 8%
- Energy - Photosynthesis: 5%
- Cell Cycle: 1%
- Transcription: 5%
- Protein Synthesis: 5%
- Protein Fate: 7%
- Protein w/ binding/regulation: 5%
- Cellular Transport: 11%
- Cellular Communication/ Signal Transduction: 4%
- Biogenesis (structural): 2%
- Subcellular Localization: 1%
- Cell Rescue, Defense, Stress Response: 6%
- Metabolism: 12%
- Secondary Metabolism: 2%
- Unknown/ Unclassified: 26%
Suppression subtractive hybridization (SSH)
cDNA library

Reverse subtraction

Susceptible rxn “Tester”

Resistant rxn “Driver” (overabundance)

Suppress “common” transcripts & pcr-select “uncommon” from tester

Repressor/other regulated gene in a susceptible reaction
Results:
Reverse Subtraction Library:
- 672 clones sequenced (NAF)
- 12% with multi-inserts, 590 ESTs sequence edited & computer analyzed
- 555 low redundant ESTs further analyzed and grouped into functional categories
Results: Reverse Subtraction
Functional Categories

- Unknown/Unclassified: 27%
- Metabolism: 11%
- Secondary Metabolism: 2%
- Protein Synthesis: 6%
- Energy: 10%
- Energy - Photosynthesis: 7%
- Protein Fate: 11%
- Protein w/ binding/regulation: 4%
- Cellular Transport: 7%
- Cellular Communication/Signal Transduction: 3%
- Biogenesis (structural): 1%
- Subcellular Localization: 1%
- Cell Rescue, Defense, Stress Response: 3%
- Transcription: 5%
- Cell Cycle: 1%
- Subcellular Localization: 1%
- Biogenesis (structural): 1%
Specific objectives

- Develop a suppression subtractive hybridization library of the resistant interaction and identify transcripts/ESTs (Expressed Sequence Tags)
- Protein profiling of germinating and resting urediniospores from *P. pachyrhizi*
Enriched extracellular proteins from germinating and resting urediniospores

- Vacuum infiltrate leaflets
- Low spin, collect infiltrate - 45µm filter
- Concentrate, dialysis, acetone-precipitation

USDA-ARS-FDWSRU & USDA-ARS-ERRC )
Enriched extracellular proteins from germinating and resting urediniospores

- 2-D protein gel
- Pick spots for MALDI
- In-gel trypsin-digestion
- MALDI/TOF-TOF mass spectrometry

ABI4700

USDA-ARS-FDWSRU & USDA-ARS-ERRC

Protein profiling with 2D-gel and MALDI/TOF-TOF mass spec 12hpi

Germinating spores

Resting spores

Selected spots were blasted against the “nr” database and the EST database (six reading frames)

USDA-ARS-FDWSRU & USDA-ARS-ERRC)
Enriched extracellular proteins from soybean leaves from resistant and susceptible interaction

Time points: 0h, 12h, 24h, 48h, 72h, 6dpi
Treatments: Mock, HW94-1, TW72-1
Soybean cv.: G. max cv. Komata
  - Vacuum infiltrate leaflets
  - Low spin, collect infiltrate - 45µm filter
  - Concentrate, dialysis, acetone-precipitation

USDA-ARS-FDWSRU & USDA-ARS-ERRC )
Enriched extracellular proteins from soybean leaves from resistant and susceptible interaction

- 2-D protein gel
- Pick spots for MALDI
- In-gel trypsin-digestion
- MALDI/TOF-TOF mass spectrometry

ABI4700

USDA-ARS-FDWSRU & USDA-ARS-ERRC
Enriched extracellular proteins from soybean leaves from resistant and susceptible interaction

Time points: 0h, 12h, 24h, 48h, 72h, 6dpi
Treatments: Mock, HW94-1, TW72-1
Soybean cv.: G. max cv. Komata
- Vacuum infiltrate leaflets
- Low spin, collect infiltrate - 45µm filter
- Concentrate, dialysis, acetone-precipitation

USDA-ARS-FDWSRU & USDA-ARS-ERRC}
Enriched extracellular proteins from soybean leaves from resistant and susceptible interaction

- 2-D protein gel
- Pick spots for MALDI
- In-gel trypsin-digestion
- MALDI/TOF-TOF mass spectrometry

ABI4700

USDA-ARS-FDWSRU & USDA-ARS-ERRC
GENOME SEQUENCING PROJECT

Phakopsora pachyrhizi
Phakopsora meibomiae
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
<table>
<thead>
<tr>
<th>Library (Insert size)</th>
<th>Bases sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. pachyrhizi</strong></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>416.63 MB</td>
</tr>
<tr>
<td><strong>P. Meibomiae</strong></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>131.17</td>
</tr>
</tbody>
</table>
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>
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.
Random fosmids

Stanford University:

Finished 87
Incomplete 28

Selected fosmids

Lawrence Livermore National Laboratory:

Probes designed for 120
Selected 50
To go 70
Sequencing 24
Finished 0

Probes designed based on ESTs selected by high similarity to “interesting” genes from other fungi, unknown genes highly expressed in germinating spores from *P. pachyrhizi* and represented in the “sequenced” proteins from germinating and resting spores.
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**
The complete nucleotide sequence of the mitochondrial (mt) genome was determined for *Phakopsora pachyrhizi* and *P. meibomiae*.

These 32 kb genomes contain the genes encoding 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 (rrns and rrnl) and tRNAs for all amino acids.

<table>
<thead>
<tr>
<th></th>
<th><em>P. Pachyrhizi</em></th>
<th><em>P. meibomiae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>31.82 Kb</td>
<td>32.52 Kb</td>
</tr>
<tr>
<td>G+C</td>
<td>34.6 %</td>
<td>34.9 %</td>
</tr>
</tbody>
</table>
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 four mitochondrial-encoded proteins from 16 taxa. The genes encoding cob, cox1, nad1 and nad5 are present in all organisms compared. Parsimony-bootstrap support was calculated from 100 replicates using Paup 4.0b10. *Monosiga brevicollis*, *Phytophthora infestans* and *Reclinomona americana* were included as outgroups.
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. 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

USDA/ARS/FDWSRU

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.