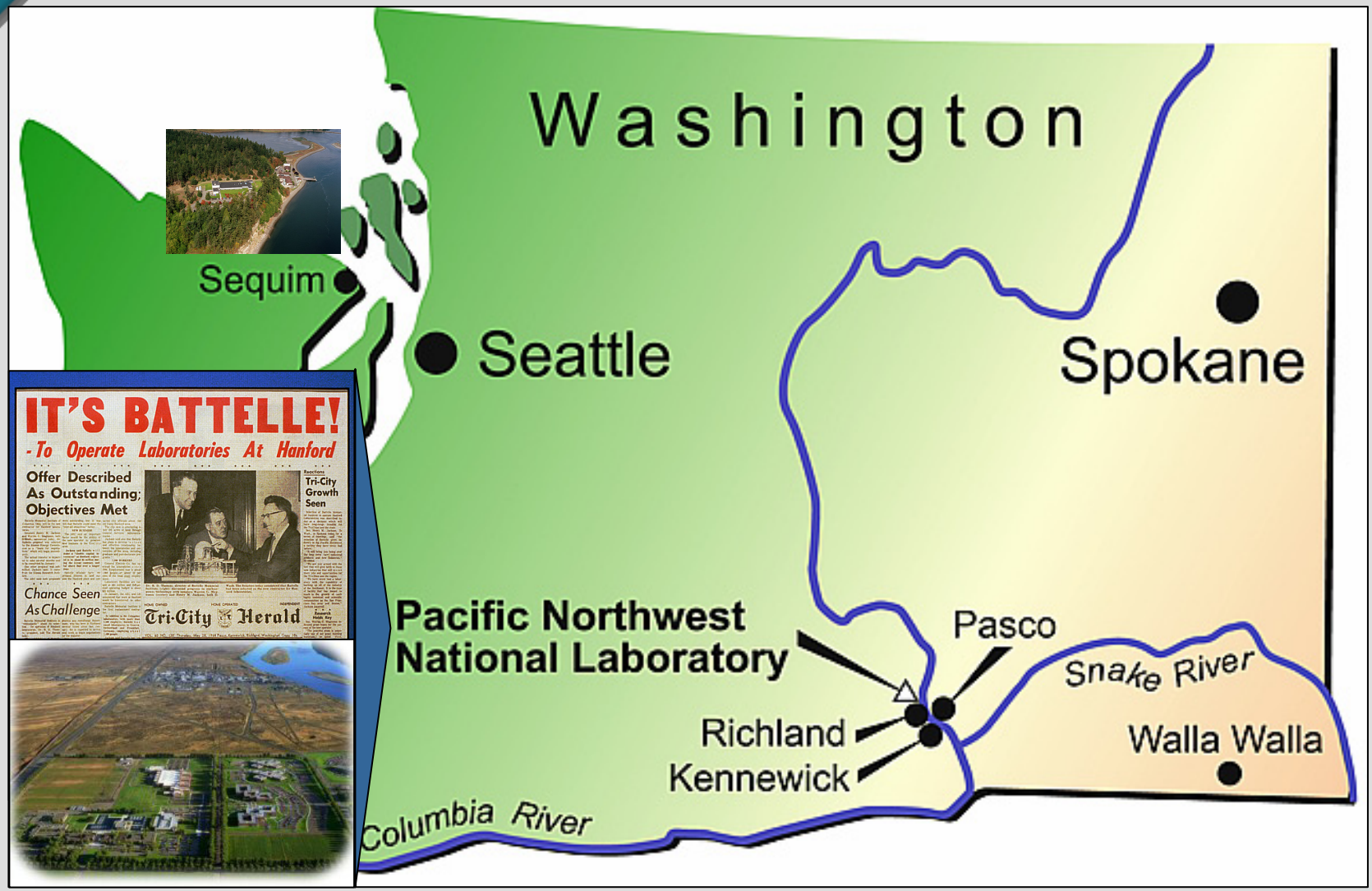
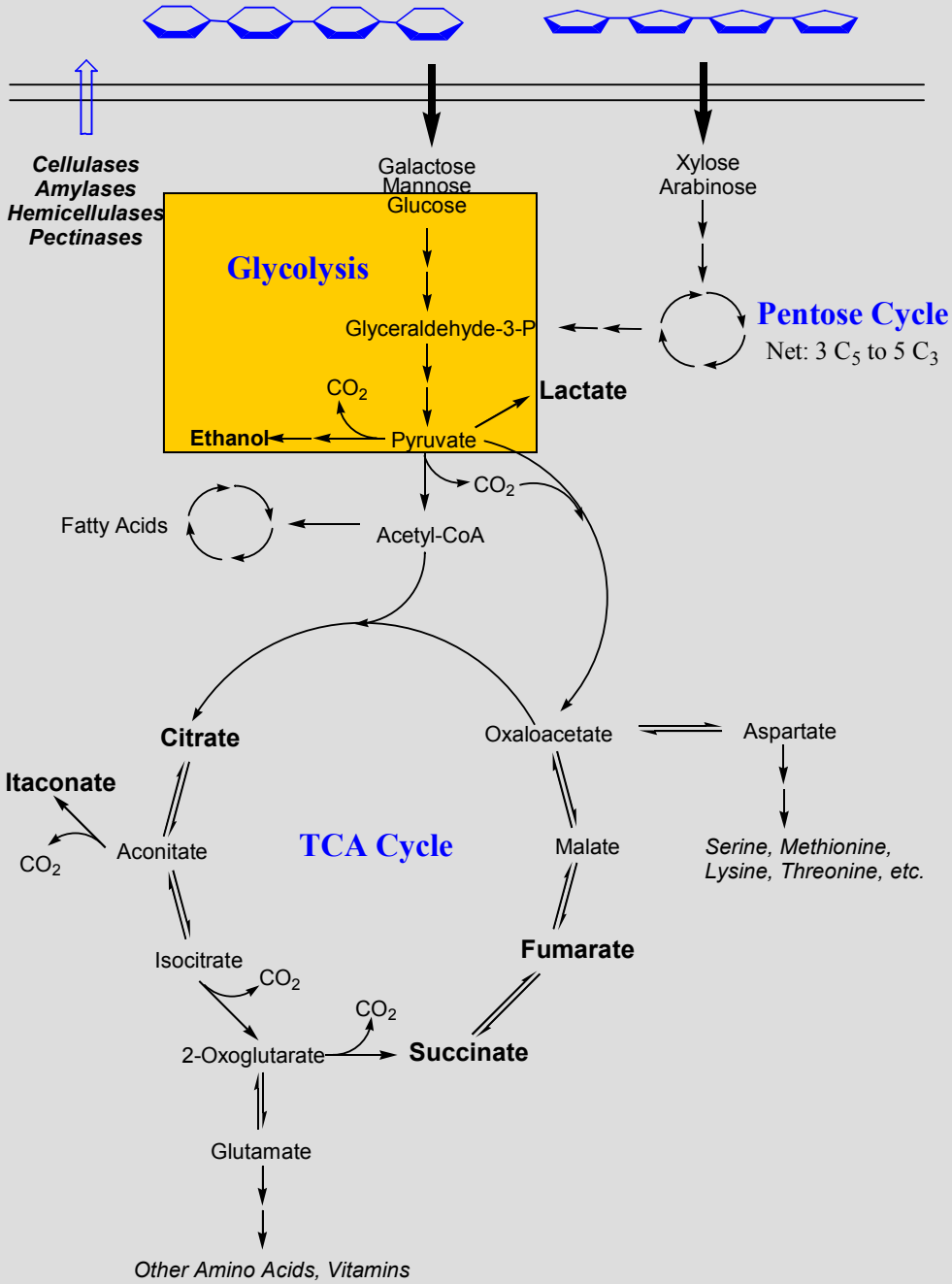
# What's new in the *Aspergillus niger* genome?

Scott E. Baker
Pacific Northwest National Laboratory
Advances Against Aspergillosis 2006
Athens, Greece
February 25, 2006


Washington
Sequim
Seattle
Spokane
IT'S BATTELLE!
- To Operate Laboratories At Hanford
Offer Described As Outstanding; Objectives Met
Sections Tri-City Growth Seen
Chance Seen As Challenge
Tri-City Herald
Pacific Northwest National Laboratory
Pasco
Snake River
Richland
Kennewick
Walla Walla
Columbia River

# Filamentous Fungi

- Nature's versatile biomass conversion factories
- 100's of enzymes for plant polymer hydrolysis
- Utilize both $C_6$ and $C_5$ sugars
- Produce 10s to 100s of grams/liter enzyme protein, organic acids



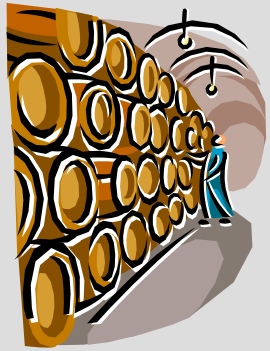

Battelle

# Current and future routes to fuels and chemicals

**Petroleum products**

Petroleum

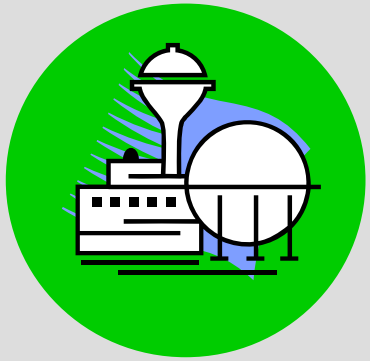
Petroleum refinery



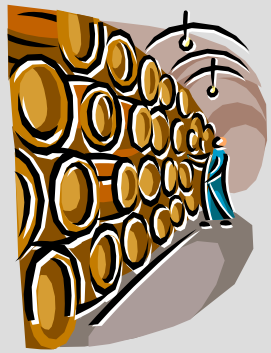
Products: Fuels and chemicals

**Biobased products**

Complex biomass: Agricultural products and "waste"

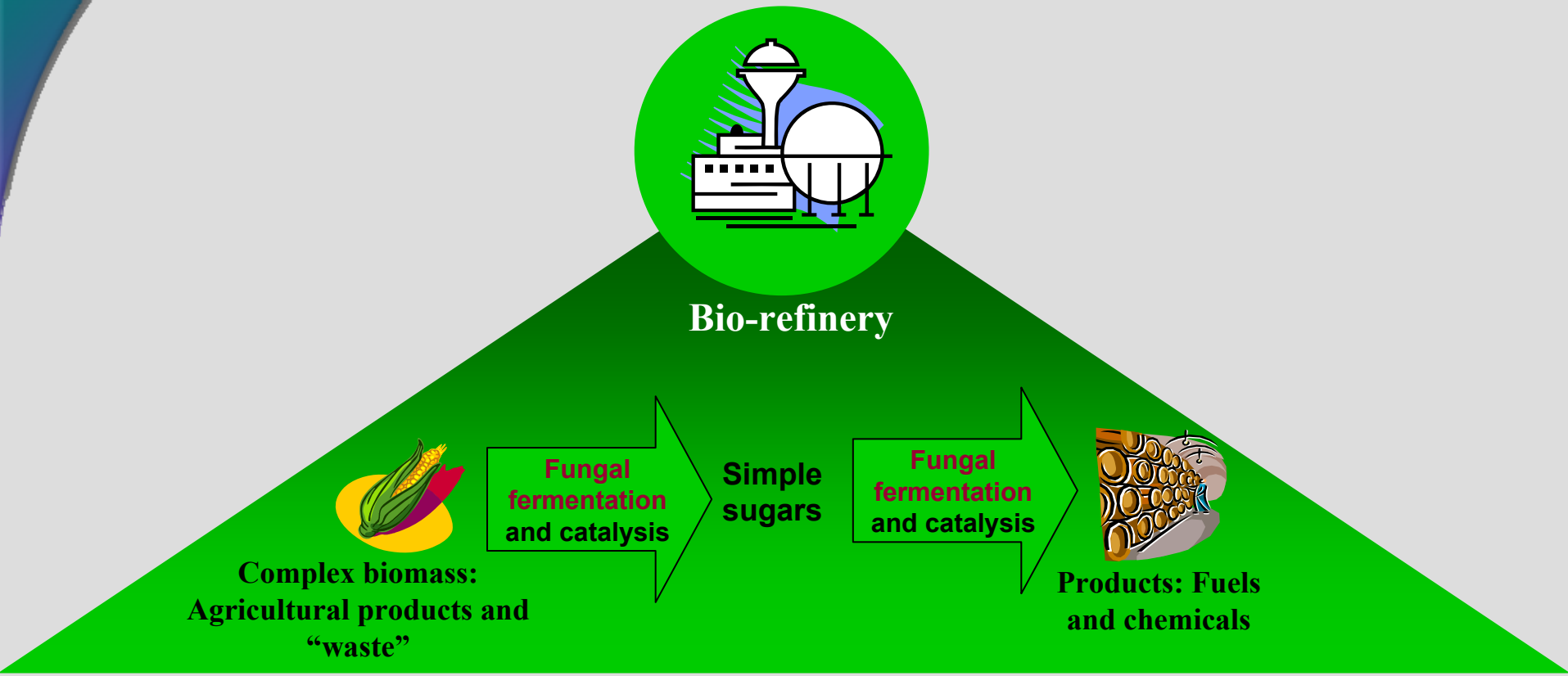


Bio-refinery



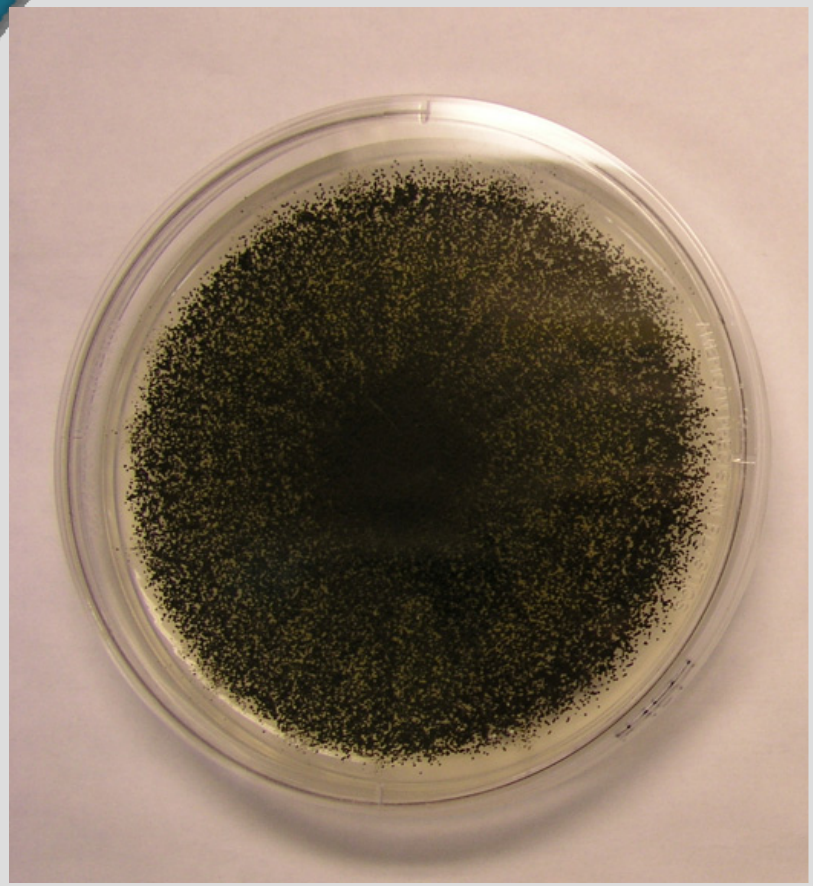
Products: Fuels and chemicals

# Filamentous fungi inside the Bio-refinery

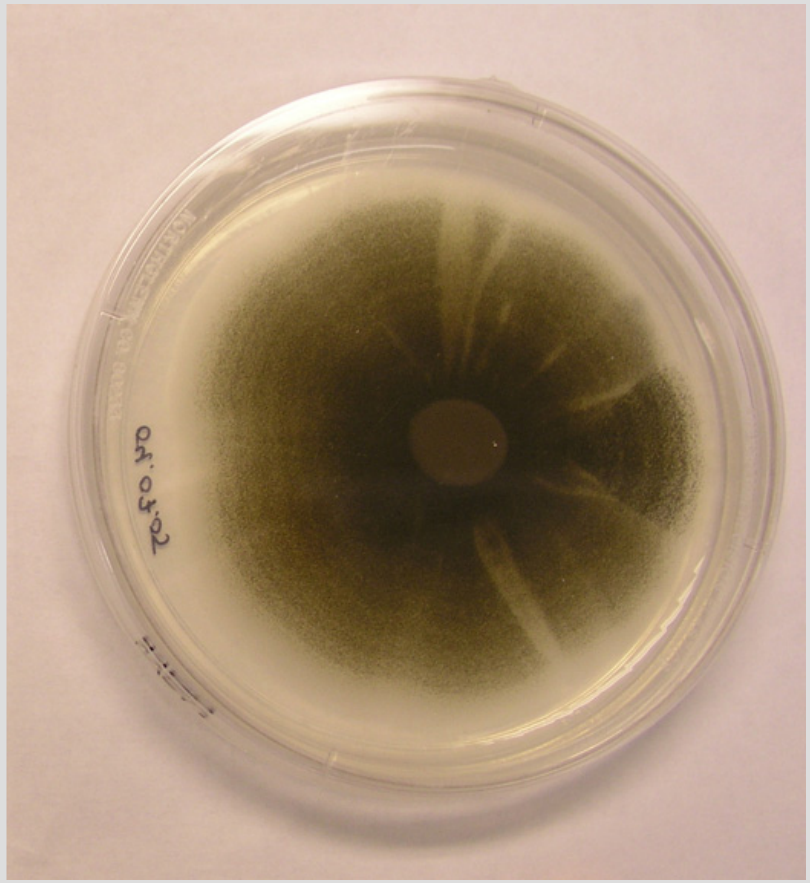


# Why a public *Aspergillus niger* genome project?

- A bioprocess organism
  - First citric acid process reported in 1917 with wildtype ATCC 1015 *Aspergillus niger* strain
  - Highly efficient fermentation of glucose to citric acid
- A protein production organism
  - Source of important enzymes
  - Industrial protein producer
- Sequenced twice by industry
  - Public access to sequence with restriction
- Large economic footprint
- Pathogen of immunocompromised humans



Wild-type (ATCC 9029)

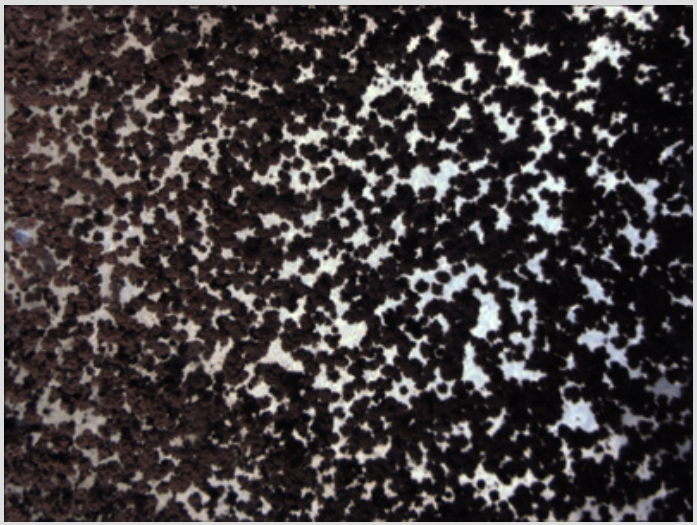


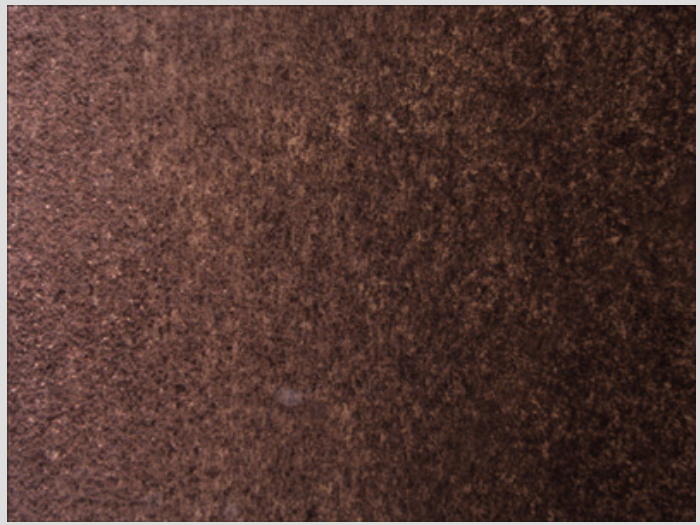
NRRL 3122

Wild-type (ATCC 9029)

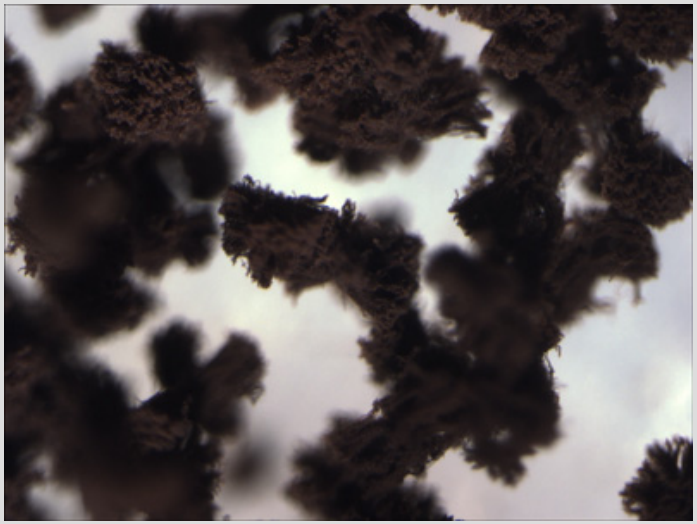
NRRL 3122

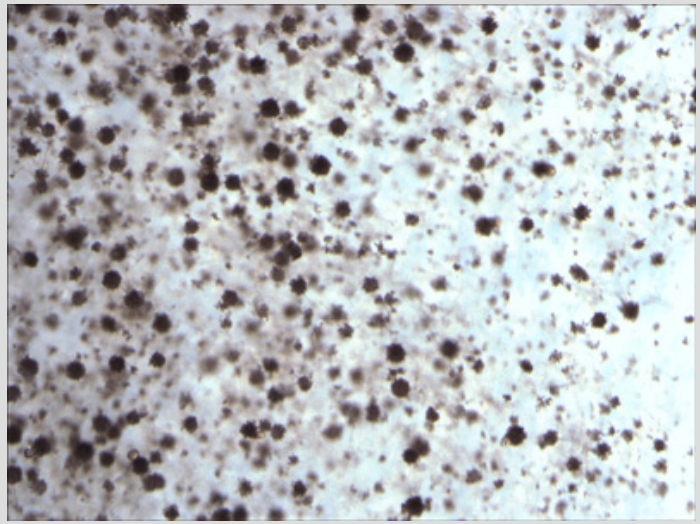
~10X





~80X





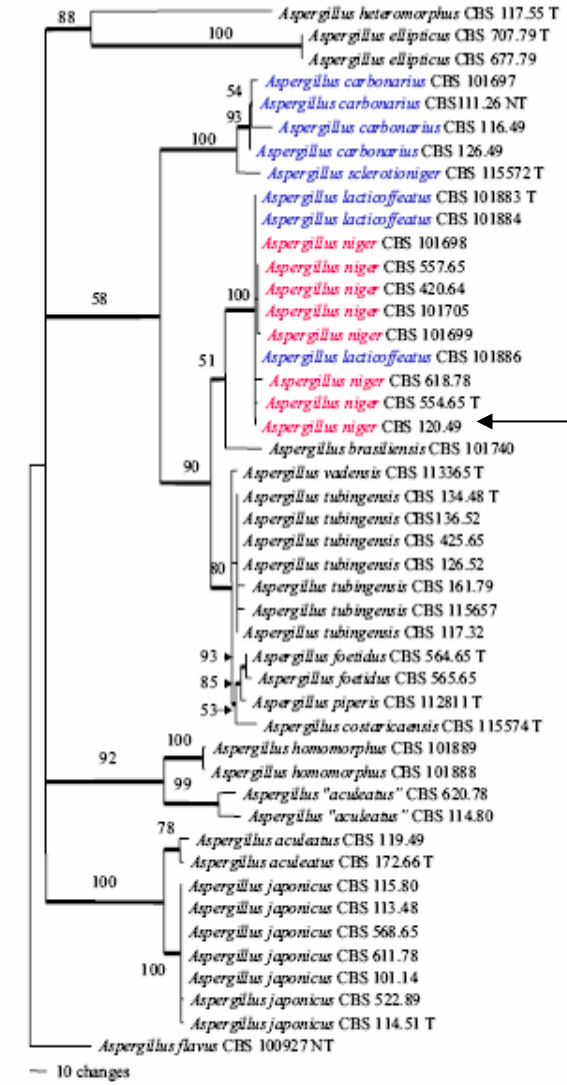



Fig. 1. One of the 5000 equally MPT of 719 steps based on heuristic search partial β-tubulin sequences with *A. flavus* as an outgroup. The branches in bold are 100 % in the 70 % majority-rule consensus of equally parsimonious trees. The numbers represent bootstrap percentages > 50 % (CI = 0.701, RI = 0.898 RC = 0.630, HI = 0.299). Names in blue are ochratoxin producing taxa. Taxa in red contain isolates which can produce ochratoxin.

Phylogeny from Robert A. Samson, Jos A.M.P. Houbraken, Angelina F.A. Kuijpers, J. Mick Frank and Jens C. Frisvad. 2004. New ochratoxin A or sclerotium producing species in Aspergillus section Nigri. Studies in Mycology. 50:45-61.

# The DOE *Aspergillus niger* genome project

- Proposed to the US Department of Energy Microbial Genome Program by the PNNL Fungal Biotechnology team
- Collaboration with DOE's Joint Genome Institute
- Current status
  - Final draft coverage: ~8X shotgun
  - EST libraries constructed from RNA isolated from citric acid production and complex biomass digestion conditions~30,000 sequenced
  - Annotation: In collaboration with JGI/LANL
  - Public release: Soon
- Other *Aspergillus niger* genomes
  - ATCC 9029: low coverage, sequenced by Integrated Genomics – Sequence available on request. Contact Scott Baker or Jon Magnuson (scott.baker@pnl.gov or jon.magnuson@pnl.gov)
  - CBS 513.88 (derivative of NRRL 3122/ATCC 22343): DSM – announced public release at Asilomar FGC

| | Integrated Genomics | JGI |
|---|---|---|
| •Date | 2000 | 2005 |
| •Strain | Wild-type (ATCC 9029) | Wild-type (ATCC 1015) |
| •Method | Shotgun<br>No finishing | Shotgun<br>Plus finishing |
| •Coverage | ~4-6X | ~8X |
| •Genomic library insert size | 1-2kb | 3kb<br>8kb<br>40kb |
| •Contigs or scaffolds | >9000 contigs | <100 scaffolds |

# The "QC" *A. niger* ATCC 1015 assembly – 4X coverage

- total number of scaffolds: 118
- total length of scaffolds: 35634017
- N50 scaffold number: 6
- N50 scaffold size: 1931570
- total number of contigs: 1646
- total length of contigs: 34162656
- N50 contig number: 215
- N50 contig size: 47636
- Total: 243,688 reads
  - 3 kb: 105,065 = 43.1%
  - 8 kb: 118,655 = 48.7%
  - 40 kb: 19,968 = 8.2%

# The "draft" *A. niger* ATCC 1015 assembly – 8X coverage

- Over 40 that encode ketosynthase and acyl-transferase domains(i.e. PKSs and FASs)
- Mat-1-1(alpha box)
- ~95% of the of the genome is found in 24 scaffolds – 1.5 scaffolds/chromosome arm
- EST coverage/annotation
  - Genencor to release ~7500 EST sequences from several different growth condition libraries
  - JGI sequenced 30,000 ESTs
  - The Fungal Genomics program at Concordia University will contribute ~12,000 ESTs
- Annotation jamboree scheduled for April 2006, following the European Conference on Fungal Genetics
- Limited gap closure or "finishing" is planned by JGI-LANL

# What's next? A multi-gene phylogeny and more genomes from Aspergillus section Nigri

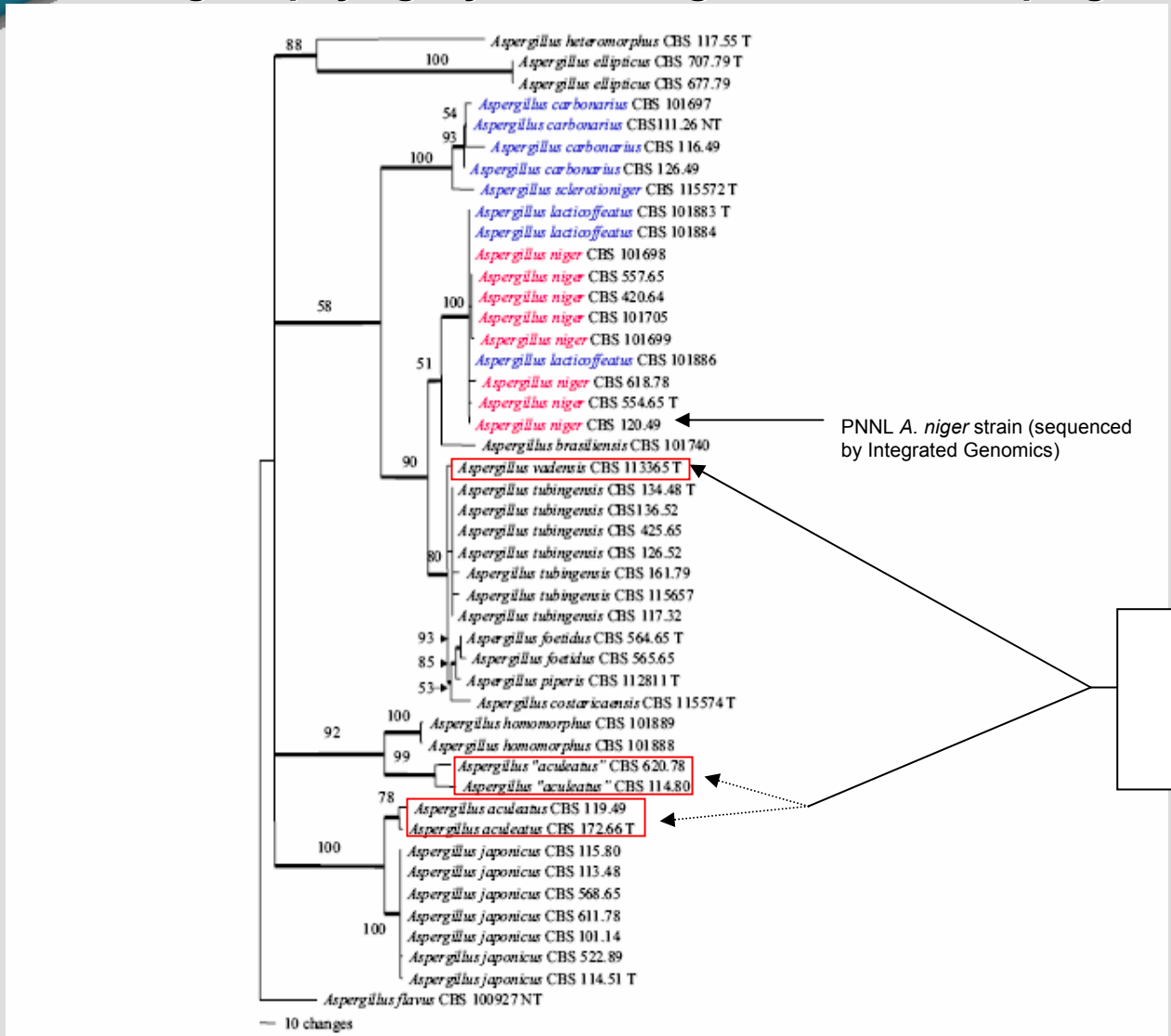
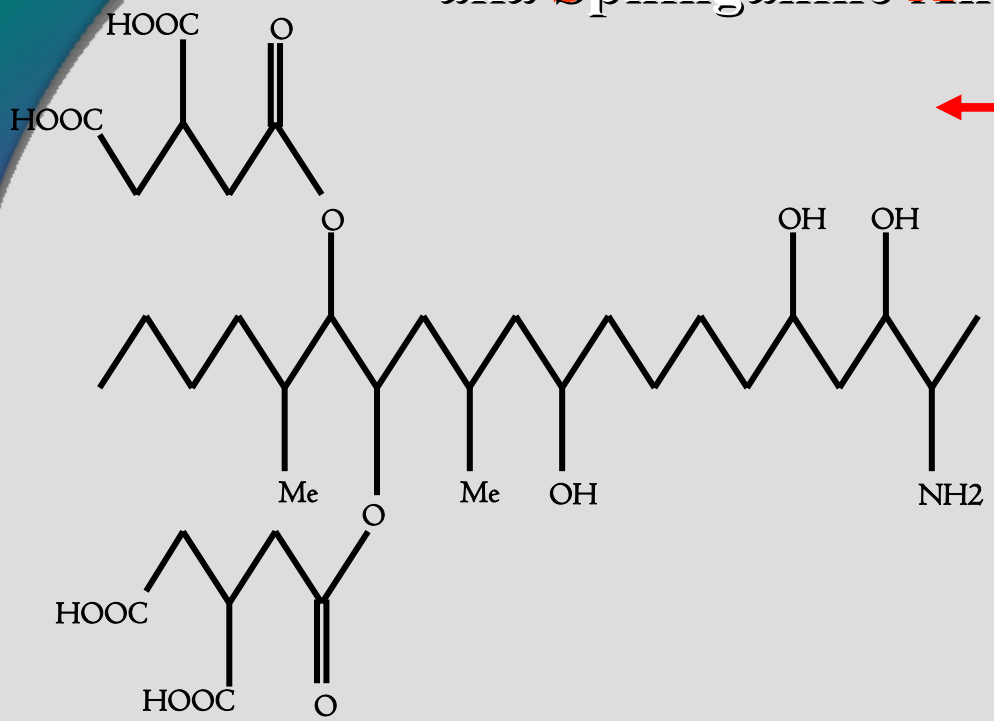



Fig. 1. One of the 5000 equally MPT of 719 steps based on heuristic search partial β-tubulin sequences with *A. flavus* as an outgroup. The branches in bold are 100 % in the 70 % majority-rule consensus of equally parsimonious trees. The numbers represent bootstrap percentages > 50 % (CI = 0.701, RI = 0.898 RC = 0.630, HI = 0.299). Names in blue are ochratoxin producing taxa. Taxa in red contain isolates which can produce ochratoxin.

Phylogeny from Robert A. Samson, Jos A.M.P. Houbraken, Angelina F.A. Kuijpers, J. Mick Frank and Jens C. Frisvad. 2004. New ochratoxin A or sclerotium producing species in Aspergillus section Nigri. *Studies in Mycology*. 50:45-61.

# SAMs

## aka Sphinganine Analog Mycotoxins

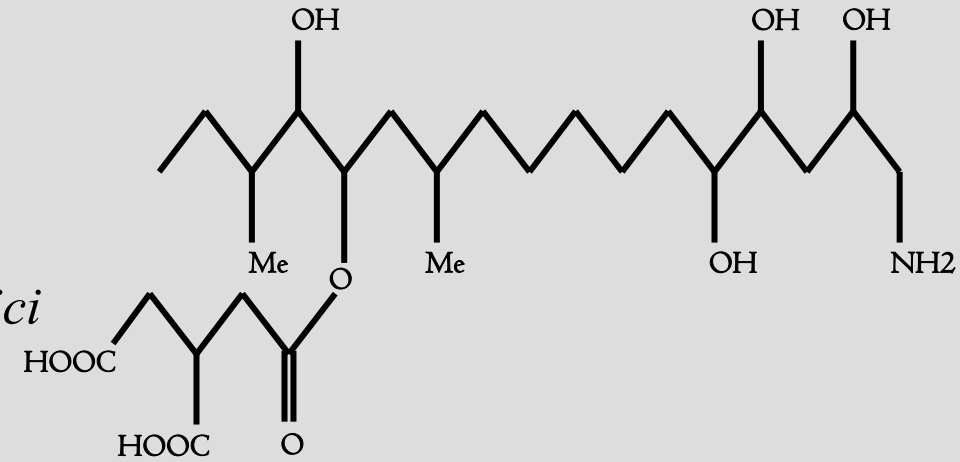



**Fumonisin B1**

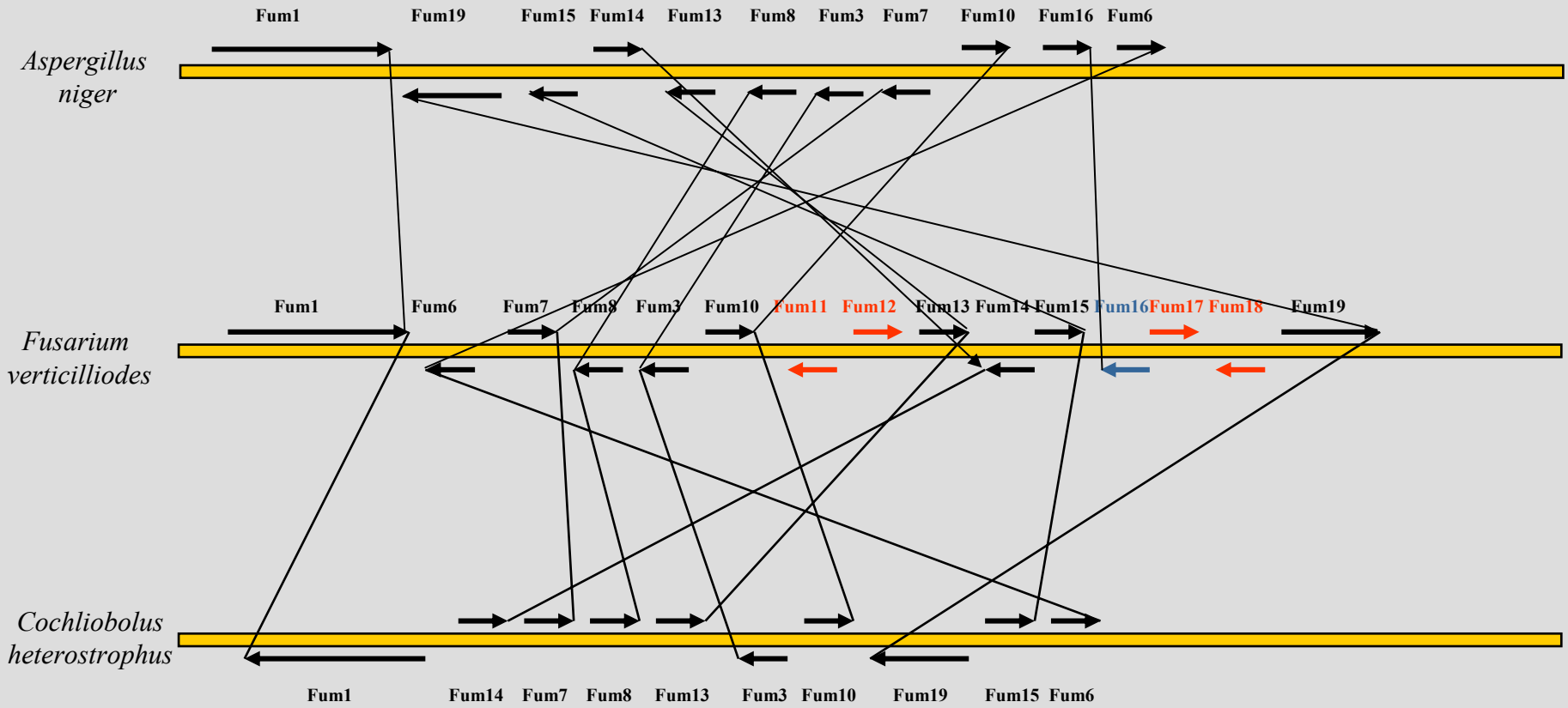
- Mycotoxin
- Produced by *Fusarium* spp.
- Ceramide synthesis inhibitor

**AAL toxin**

- Virulence factor
- Produced by *A. alternata lycopersici*
- Ceramide synthesis inhibitor



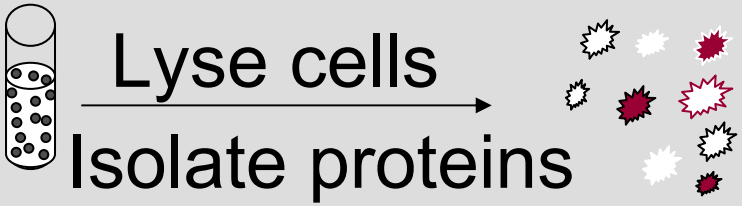
# Conserved fumonisin-like gene clusters

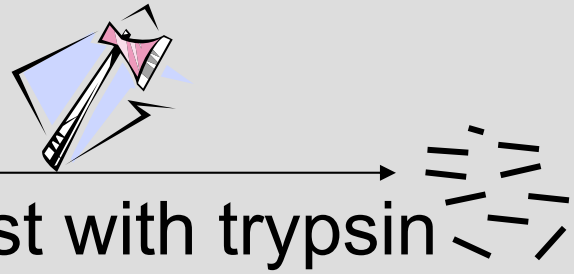


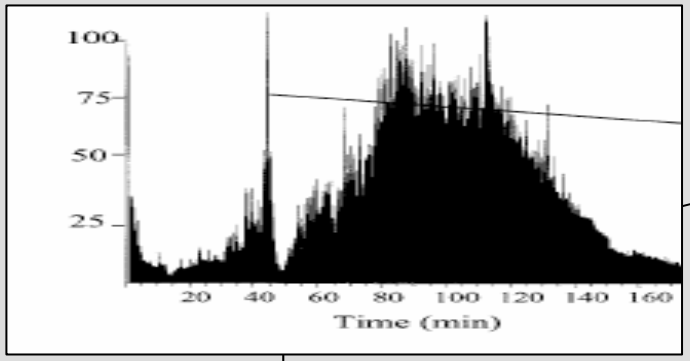
# Future directions

- Cell biological studies of wild-type vs. “hyper producers”
- Comparative genomics of wild-type vs. “hyper producers”
- Phylogenetic studies of *Aspergillus niger* and relatives
- Medium throughput deletion of putative *A.niger* morphology genes

# General Procedure for Proteomics



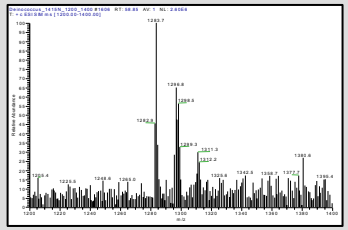




Separate in one or more dimensions reverse phase, ion exchange

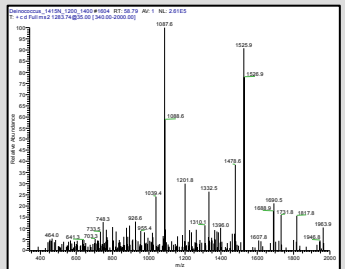
- Perform
- LCQ
- analysis

Raw Data

MS



MS/MS



Run data through peptide identifying program (SEQUEST)

Identify unique peptide = identify parent ORF

# Quantitative proteomics

- Used for comparison of biological samples generated by two or more different experimental conditions
- Current technologies utilize isotopic labeling strategies
  - ICAT
  - Metabolic labeling
  - Pairwise comparison
- Our goal: Generate a quantitative proteomic methodology using statistical analysis of raw MS abundance data and that does not use isotopic labeling

# Acknowledgements


PNNL Fungal Biotechnology Team

Ellen Panisko
Ken Bruno
Jon Magnuson
Ziyu Dai
Chris Wend
Beth Hofstad
Mark Butcher
Andy Zwoster
Debbie Lee
Kyle Fowler
Kelly Vincent
Katie Panther
Dennis Stiles
Linda Lasure

PNNL Proteomics QC Analysis Team

Don Daly
Kevin Anderson
Matt Monroe

DOE JGI *A. niger* ATCC 1015 genome

Dan Drell
Diego Martinez
Dan Rokhsar
Chris Detter
David Bruce
Erika Lindquist
Paul Richardson
…the list continues to grow!