The forgotten fungi: Phylogenetic diversity, lignocellulolytic capabilities, and genomic analysis of the Neocallimastigomycota

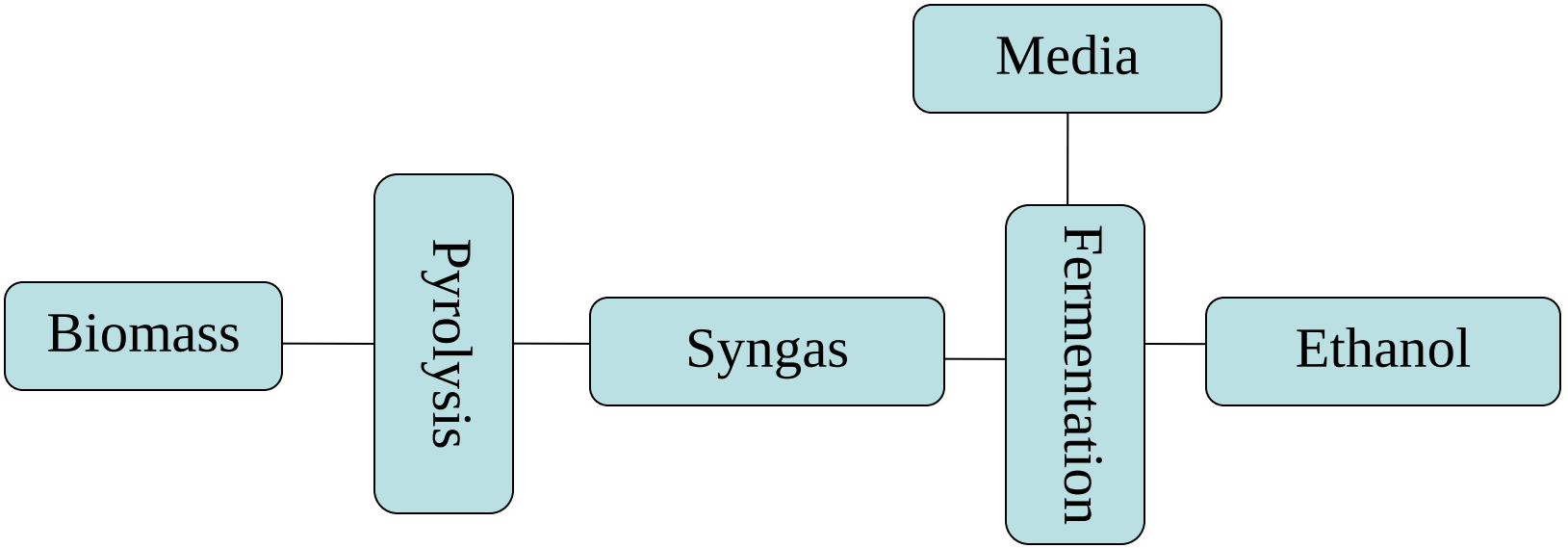
Mostafa Elshahed

Assistant Professor, Department of Microbiology and Molecular Genetics

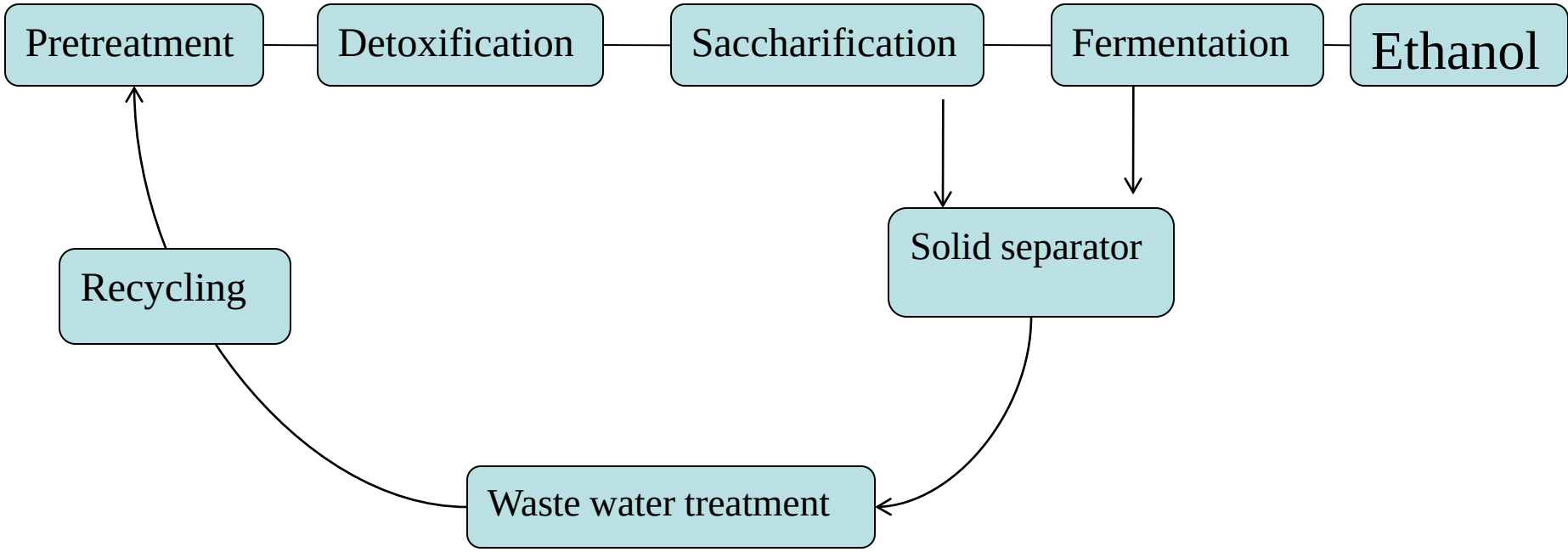
Oklahoma State University

Lignocellulosic biomass conversion to ethanol: two strategies

I. Indirect fermentation



II. Direct fermentation



I. Direct fermentation

- **Challenges**: Cost.

  Could be overcome by:

  - ☞Overcoming recalcitrance
    - Lignocellulosic biomass
    - Fungal strains
    - Saccharification enzymes
  - ☞Overcoming complexities (CBP)
    - Simpler pretreatments
    - Simultaneous saccharification and fermentation
    - Simplifying waste treatment

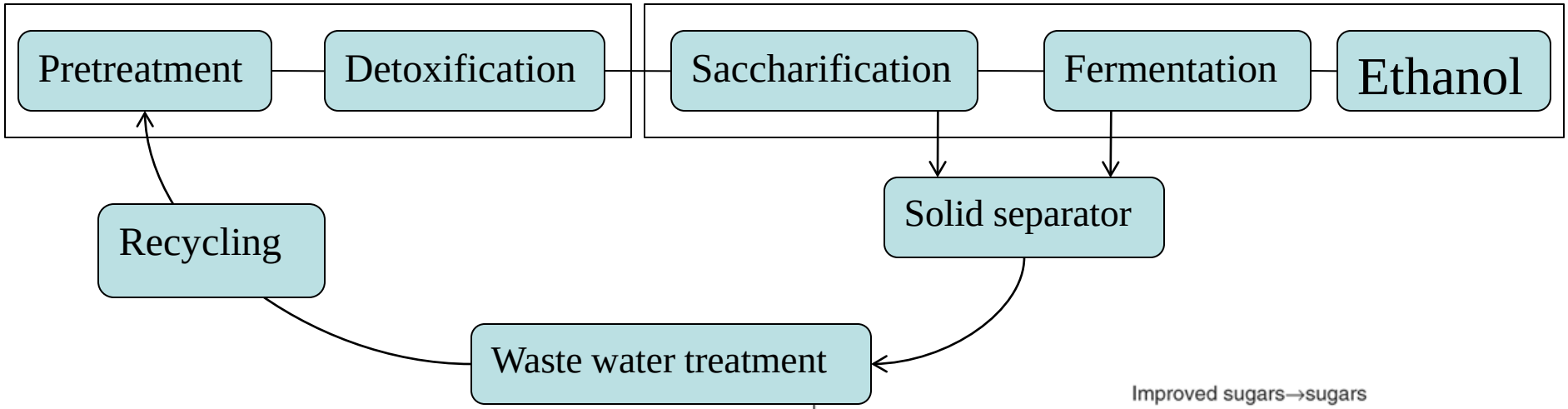
- **Advantages**
  - ☞High theoretical yield
  - ☞Sugar fermentation to ethanol well researched, wealth of knowledge on *S. cerevisiae*, *E. coli*
  - ☞Production of longer chain alcohols could be achieved in a high yield

Where would a microbiologist fit in such efforts?

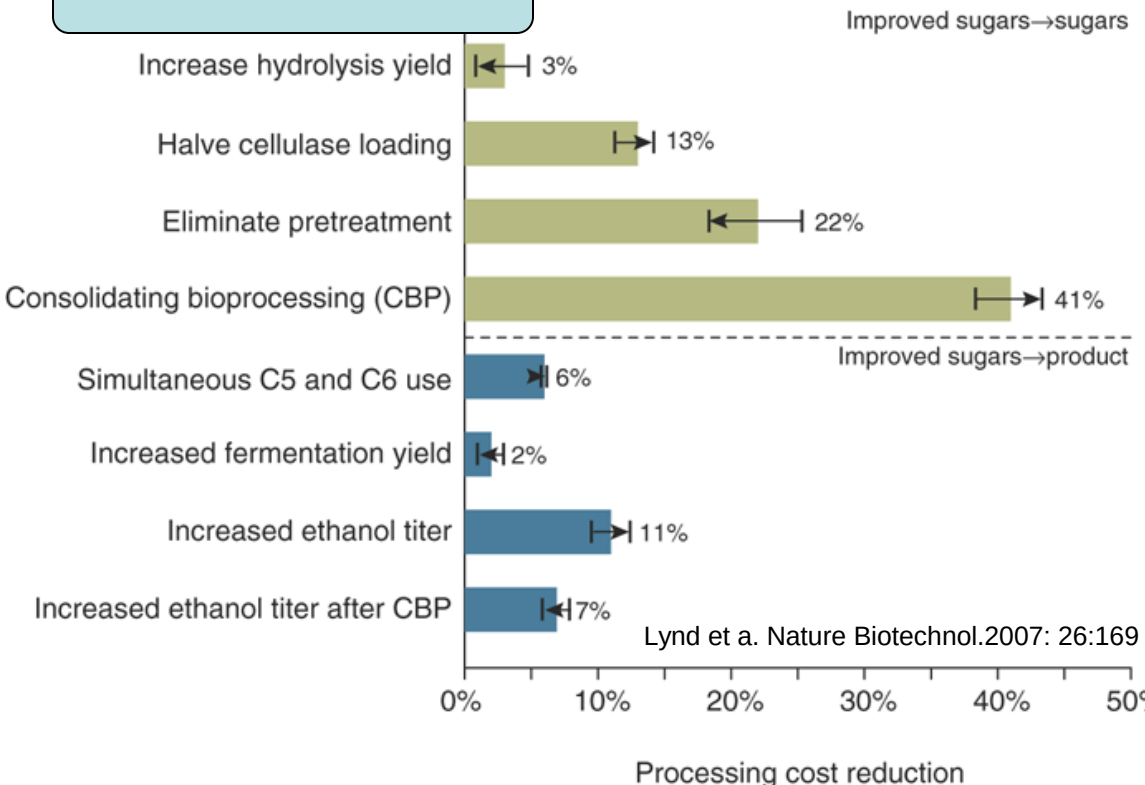
✸ Bioprospecting and Strain development

- Bioprospecting: Identification, isolation of novel microorganisms and novel enzymes
- Strain development
  - ☞ Directed evolution
  - ☞ Genetic manipulation
  - ☞ Cloning and expression, synthetic biology approaches
- **Goal**: novel ***AND*** better microorganisms, enzymes
- Critical evaluation of ***capability*** and ***potential***
- Clear role in a larger scheme necessary

Bioprospecting and strain development aiming at eliminating pretreatment, consolidating bioprocessing



Microorganisms that require no or little pretreatment and/or could conduct simultaneous saccharification and fermentation, deliver the largest cost reduction



Lynd et a. Nature Biotechnol.2007: 26:169

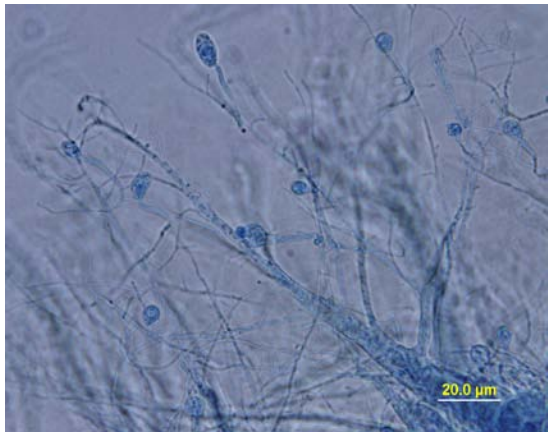
Nature's way in breaking lignocellulosic biomass

- Fungi
  - ☞Invasive, produce multiple extracellular enzymes, expose lignocellulosic materials to microbes
  - ☞Primary colonizers and degraders of dead plant biomass in soil.
  - ☞Under aerobic conditions, biomass is completely degraded to CO_2 (Aerobic fungi)
  - ☞Aerobic fungal enzymes are used for saccharification

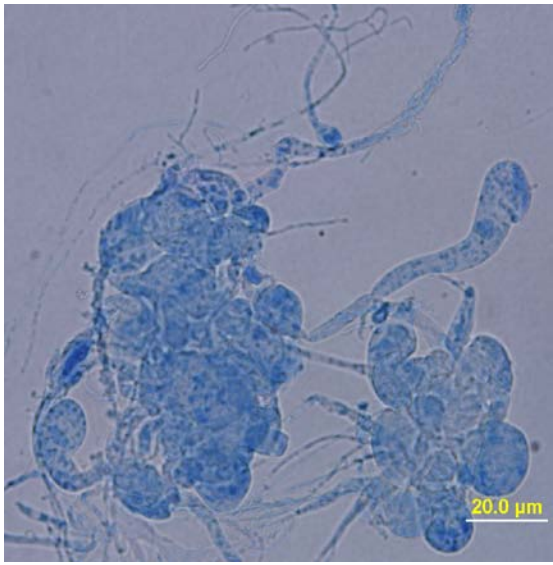
- *S. Cerevisiae* / fermentative bacteria are used for sugar fermentation

- CBP efforts with engineered Enteric bacteria, *C. thermocellum* are underway

Anaerobic fungi combine the invasiveness of fungi with the fermentative capabilities of yeast/ anaerobic bacteria



- Discovered in sheep in 1975*
- Present in the gut, alimentary tract of many ruminants, herbivores
- Initial colonizers of ***fresh*** plant materials in cow rumen
- Separate fungal phylum (Neocallimastigomycota)
- Six Genera, 20 species currently described.



*Orpin GC: Studies on the rumen flagellate Neocallymastix frontalis J. Gen. Microbiol. 91:215-218

Evaluation of the role of anaerobic fungi in biofuel research

- *Theoretically*, anaerobic fungi are ideal candidates for use in direct fermentation schemes
  - ☞ Highly invasive
  - ☞ No or minimum pretreatment necessary
  - ☞ Capable of metabolizing cellulose, hemicellulose, pectin
  - ☞ Capable of metabolizing C6, C5 sugars, ethanol is one of the products

- "The anaerobic fungi also attract attention as a new group of cellulase- and hemicellulase-producing microorganisms. The challenge of adapting this group of microorganisms in biotechnology will undoubtedly be accepted by scientists in the near future". Tom Bauchop. Biosystems 1989 23:53-64

Extensive research, Critical evaluation needed

Information on anaerobic fungi is sparse after 35 years of their discovery

- Few interested research groups (Only a handful of laboratories are (semi)-active around the world).
- No genome, partial genome sequenced, few genes known
- No genetic system available
- Researched mainly by animal scientists, ecologists, evolutionary biologists, and protein chemists
- Few, if any, sustained biofuel-related research had been conducted

Research goals

1. Gauge the global diversity of anaerobic fungi
2. Isolate multiple, diverse isolates
3. Evaluate abilities to grow and metabolize switch grass
4. Identify enzyme activities, quantify products
5. Genome sequencing, cloning and expression studies
6. Formulate an ideal role for these fungi in direct fermentation approaches
7. Critical evaluation of their strengths and weaknesses)
8. Commercialization

Bioprospecting

Strain development

Research thrusts

Ecology/Microbiology

- Diversity of anaerobic fungi in herbivores
- Isolation of robust strains

Metabolic potential

- Growth on model plant polymer
- Growth on switch grass
- Effect of Pretreatment
- Co-culturing effort
- Lignin and Coal

Genomes

- Lingcellulolytic gene repetoire and comparative genomics
- Central metabolic pathways
- Sugar/oligo transporters
- Evolutionary aspects: hydrogenosome structure and moasaic genome

Critical assessment

- Bioconversion agents in pure culture
- Bioconversion agents in Coculture
- Pretreatment requirements
- Source of Lignocellulolytic enzymes

Post genomic

- Experimental validation of activity of novel GH, CBMs
- Proteomic/transcriptomic analysis of anaerobic fungal gene expression on switch grass
- Cloning/ expression of key enzymes, site directed mutagenesis

I.A. Ecology of the Neocallimastigomycota

- Six genera, 20 species currently described.
- Presence documented in at least 50 ruminant and non-ruminant herbivorous mammals.
- Prevalence identified through isolation of a single/few strains or through microscopic observations based on phenotypic characters.
  - These culture-based and microscopic studies have provided many valuable insights.
- Molecular methods utilizing the internal transcribed spacer regions:
  - Identify isolates
  - Identify anaerobic fungal community composition using fingerprinting techniques (DGGE, T-RFLP, ARISA, size-based selections (Spreadex), HMM-based typing).

However, little is known regarding:

1. Phylogenetic Diversity within the Neocallimastigomycota.
2. Complexity of the Anaerobic Fungal Community within a Single or Multiple Hosts.
3. Influence of Various Ecological and Environmental Factors on Anaerobic Fungal Diversity.

Diversity of the Neocallimastigomycota in Herbivores

- A high-throughput pyrosequencing approach was used to gauge the diversity of anaerobic fungi in 33 herbivorous mammals and reptiles.
- PCR products using primers specific for the ribosomal ITS1 region with additional barcodes and adapters were used for the sequencing.
  - Total sequences: 267,287
  - Avg. # of seqs per animal: 8100
  - Genbank accession numbers: GQ576478-GQ843764
- Cutoff values of 5% and 17% for species and genera.

Hindgut fermenters	**Foregut fermenters (nonruminants)**	**Foregut fermenters (ruminants)**	
Equidae	*Macropodidae*	*Giraffidae*	*Bovidae*
Horse (3)	White-fronted wallaby	Rothschild's giraffe	Bontebok
Miniature donkey	Red kangaroo	Okapi	Grant's gazelle
Somali wild ass		*Cervidae*	Southern gerenuk
Grants zebra		Indo-Chinese sika deer	American bison
Grevy's zebra (2)		Indian hog deer	Greater kudu
Rhinocerotidae	**Foregut fermenters (pseudoruminants)**	American elk	Goral
Black rhinoceros	*Hippopotamidae*	Pere David's deer	Sable antelope
	Pygmy hippopotamus	Western tufted deer	Nile lechwe
Reptilian			Domestic cattle
Iguanidae	*Camelidae*	*Antilocapridae*	Domestic sheep
Green iguana	Llama	Pronghorn	Domestic goat

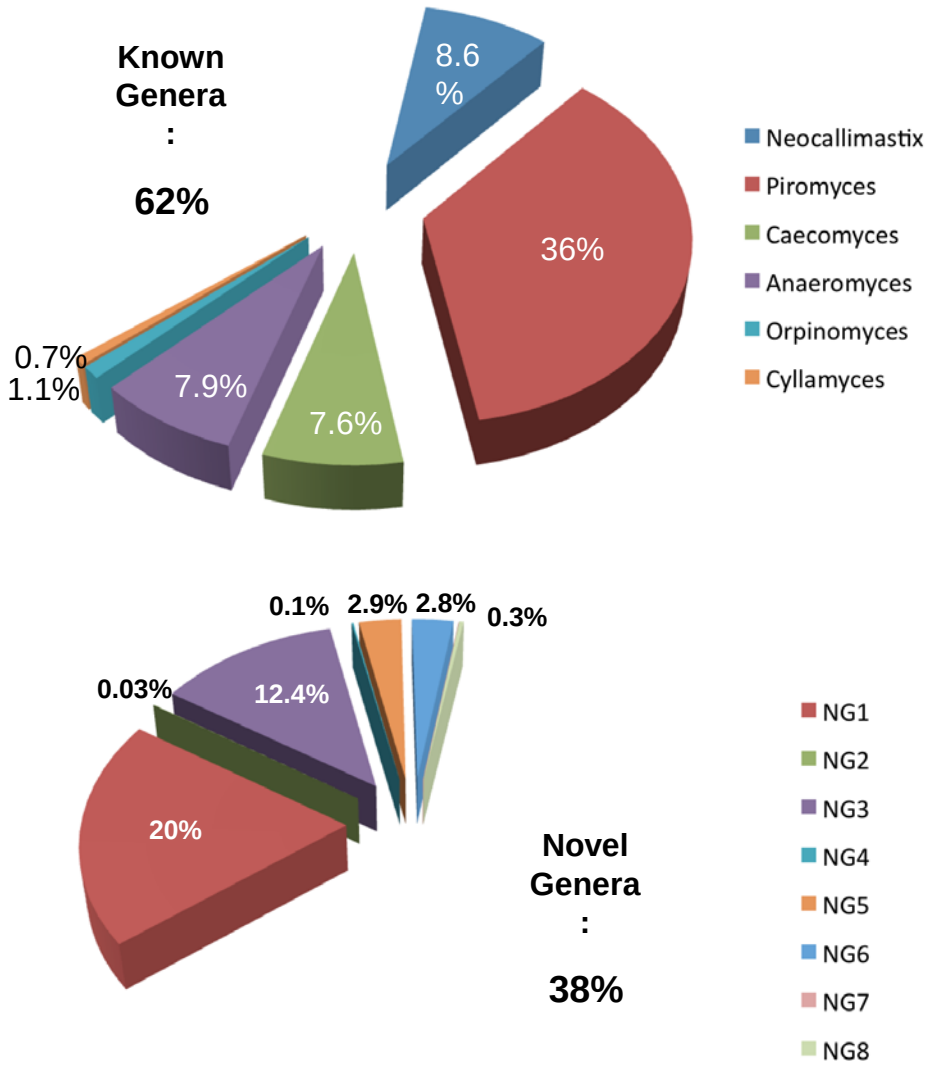
Phylogenetic diversity of the Neocallimastigomycota

•The 6-recognized AF genera constituted only 61.75% of the sequences generated in this study.

- •*Piromyces*: 96,122
- •*Neocallimastix, Caecomyces* and *Anaeromyces*: each ~21,500
- •*Orpinomyces* and *Cyllamyces*: <3,000

•Remaining, 38.25% of sequences showed less than 83% similarity to known AF genera sequences and clustered into 8 phylogenetically distinct groups.

- •NG1 and NG3 constituted nearly 20% and 12% of the entire dataset, respectively.
- •NG2 and NG5 were present in multiple animals (8 and 14, respectively), but typically were present in low abundance in data sets where they were encountered.
- •Others e.g. NG4, NG5, NG6, and NG8 had an extremely limited abundance and distribution and abundance, although NG6 constituted all the microbial community in greater Kudu.



Liggenstoffer et al. The ISME J. **4:**1225–1235

Abundance and distribution of anaerobic fungal genera in herbivores

- Abundance and Distribution of Monocentric Genera?
  - *Piromyces* (28)
  - *Neocallimastix* (18)
  - *Caecomyces* (14)

- Abundance and Distribution of Polycentric Genera?
  - *Anaeromyces* (26)
  - *Orpinomyces* (8)
  - *Cyllamyces* (2)

- Abundance and Distribution of Novel Groups (NG)?
  - NG1 (23) and NG3 (23)
  - NG2 (8) and NG5 (14)
  - NG4 (2) and NG6 (5) and NG8(2)
  - NG7 (1)

Complexity of anaerobic fungal communities within a single host:

1 Genus : 7
2 Genera :1
3 Genera: 6
4 Genera: 8
5 Genera: 5
6 Genera: 3 sheep, Cattle, Somali Wild ass
7 Genera: 3 hosts Bison, Sable Antelope, Nile Lechwe

Are there Measureable Factors that Influence the Anaerobic Fungal Diversity in Herbivores

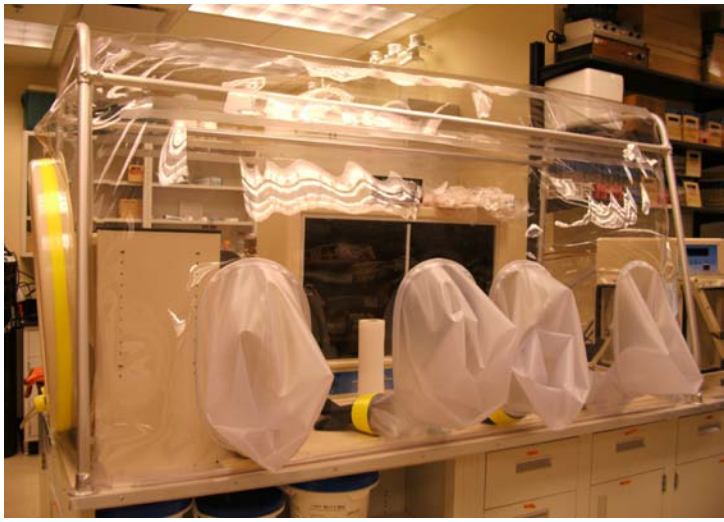
- Animal datasets were grouped using 4 different measureable factors.
- Each animal dataset was ordered according to diversity using three different parameters.
  - No of genera at the 0.05 level (quantitative)
  - Rarefaction curve rankings (ordinal)
  - Diversity rankings using Renyi and Hulbert methods (ordinal).
- Rankings were used to examine the impact of measureable factors on the diversity of anaerobic fungi in the animals sampled.
  - Gut type, ruminant ability, and feed showed low correlation (r=0.20-0.37) with all three diversity ranking schemes, a higher correlation (r=0.56-0.63) was observed when correlating to animal family.

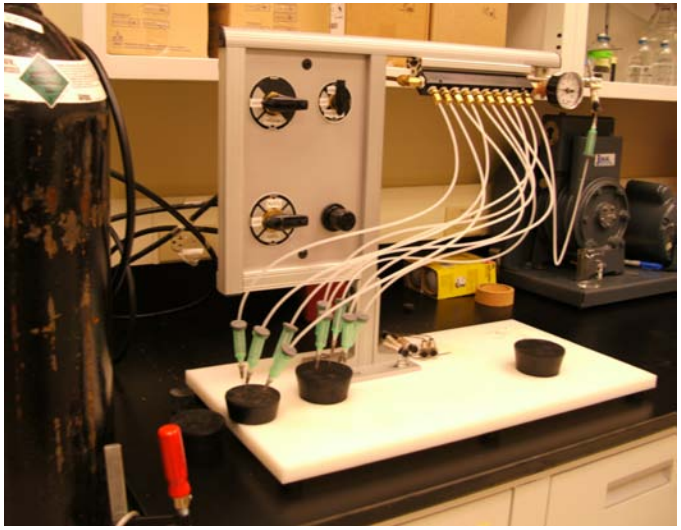
Families	Gut Types	Ruminance	Feed Types
Equidae	Hindgut	Non ruminant	Praire
Rhinocerotidae	Foregut	Pseudo ruminant	Alfalfa
Macropodidae	Other (Iguana)	Ruminant	Prairie and Alfalfa
Hippopotamidae		Other (Iguana)	Fresh/Green
Camelidae			
Giraffidae			
Cervidae			
Antilocapridae			
Bovidae			
Iguanidae			

Factor	Div Ordering	Rarefaction	No of genera
Family	0.63	0.60	0.56
Gut Type	0.37	0.30	0.21
Ruminance	0.28	0.37	0.32
Feed Type	0.20	0.29	0.31

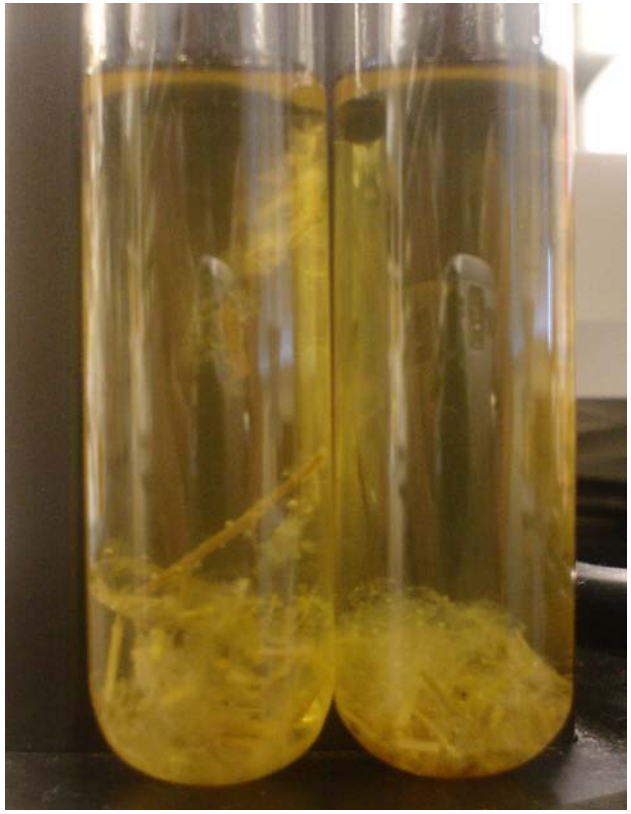
I.B. Isolation of anaerobic fungi on switch grass

- Culture-dependent methods using Hungate-type anaerobic techniques, modified Orpin's medium and roll tubes were used to obtain isolates from fecal samples.
- Identification of the isolates through microscopic examination and sequencing of the ITS spacer region.
- Isolation conducted on samples from hosts with a diverse community (e.g. domestic cattle, sheep, goat), or hosts with a high proportion of novel groups (horse, minature donkey)

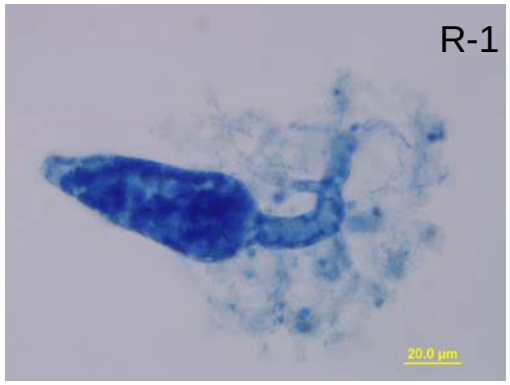


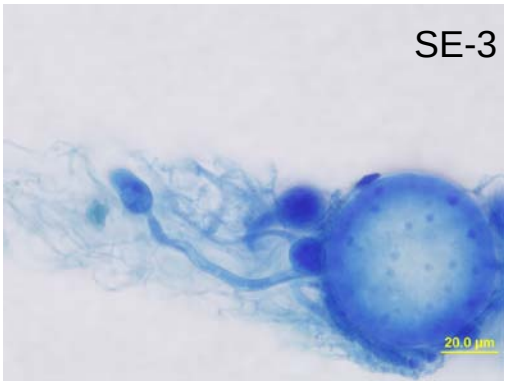


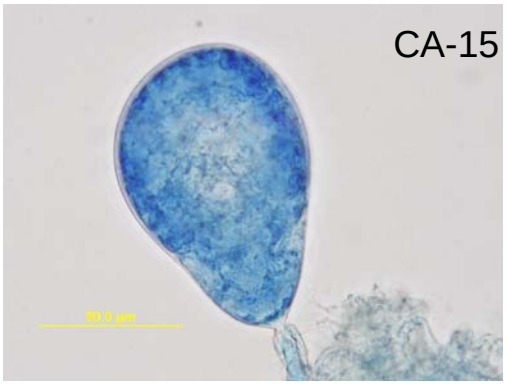
Anaerobic fungal isolates on switch grass

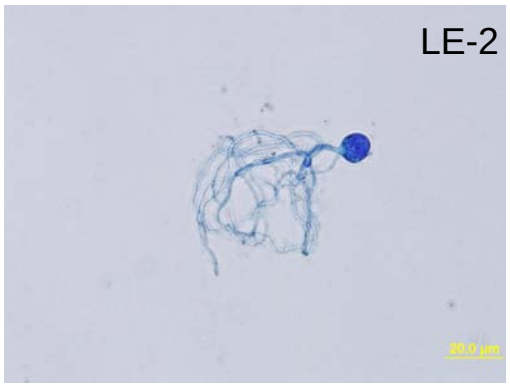


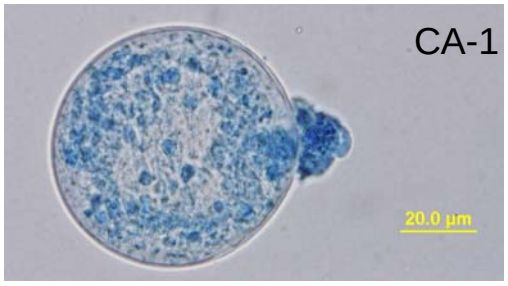
Anaerobic fungal isolates

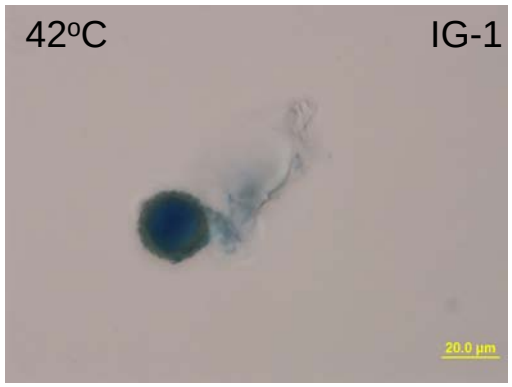


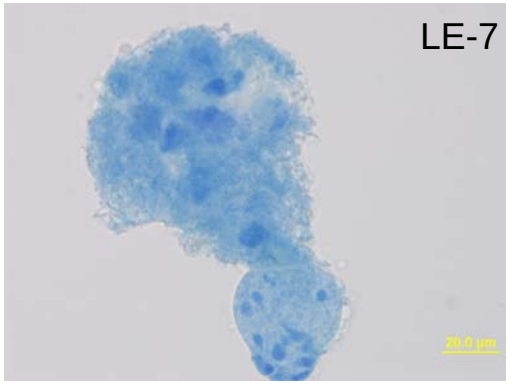


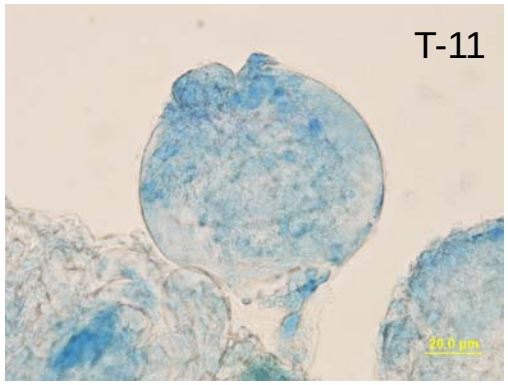


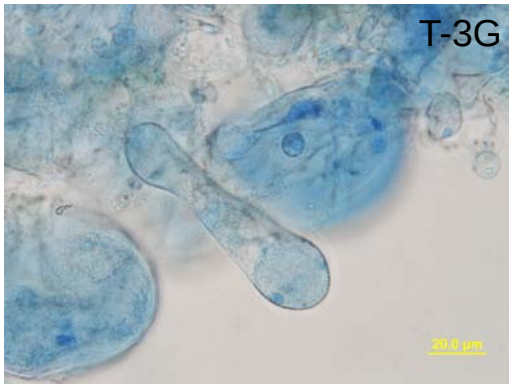






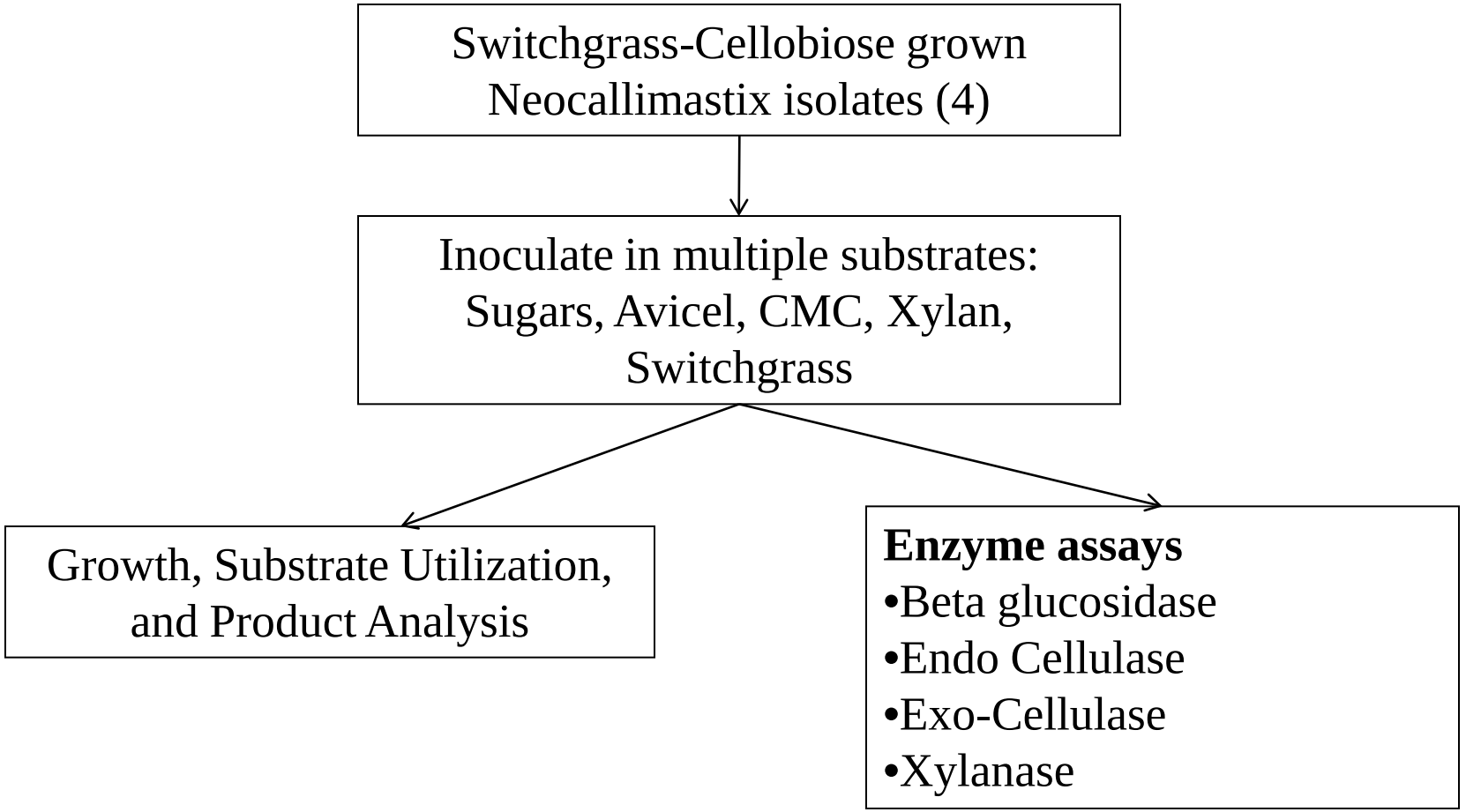






- Multiple monocentric and polycentric isolates obtained (55)
- Senescence occured in the majority of isolates obtained
- Neocallimastix-associated isolates from cow feces are the most resilient, robust
- Chosen for further metabolic studies

II. Exploring the lignocellulolytic potential of anaerobic gut fungi



Metabolic capabilities, preliminary results

	Anaerobic Fungal Isolate							
	Cattle-1A (C1A)		Cattle-1C (C1C)		Cattle-2A (C2A)		Cattle-2C (C2C)	
	Growth	ΔpH	Growth	ΔpH	Growth	ΔpH	Growth	ΔpH
Substrate								
Glucose	✓	-0.52	✓	-0.61	✓	-0.54	✓	-0.52
Xylose		-0.08		-0.03		-0.01		0.04
Arabinose		-0.06		-0.04		-0.08		-0.02
Mannose	?	-0.11	✓	-0.27	?	-0.14	✓	-0.42
Galactose		-0.05		-0.01		-0.03		0.02
Cellobiose	✓	-0.68	✓	-0.69	✓	-0.66	✓	-0.69
Avicel	✓	-0.24	✓	-0.17	✓	-0.18	✓	-0.15
CMC	?	-0.16		-0.08	?	-0.12		-0.07
Xylan	✓	-0.29	✓	-0.25	✓	-0.24	✓	-0.20
Switchgrass	✓	-0.69	✓	-0.68	✓	-0.67	✓	-0.68

- All isolates grew well on Switchgrass, Xylan, Avicel, Cellobiose and Glucose.
- Varying ability of isolates to grow on Mannose and CMC.
- No significant growth of isolates on Xylose, Arabinose or Galactose.

Current effort

- Other substrates: Oligosaccharides, model hemicellulases
- Fermentation balance/rate measurments
- Loss of Cellulose, hemi/ pectin/lignin in Switchgrass experiments
- Optimizatioin of Lignocellulolytic degradation
- Comparison to other direct fermentation system

- Further out:
  - Effect of various pretreatment procedures on switchgrass metabolism
  - Co-culturing with *S. Cerevisiae*
  - Lignin metabolism

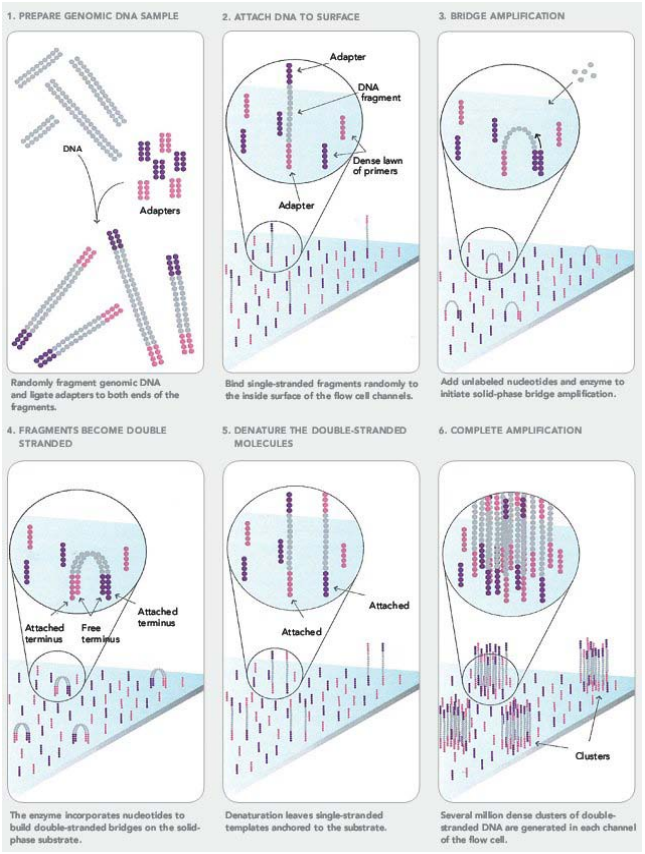
III. Genome of Neocallimastix sp strain S4

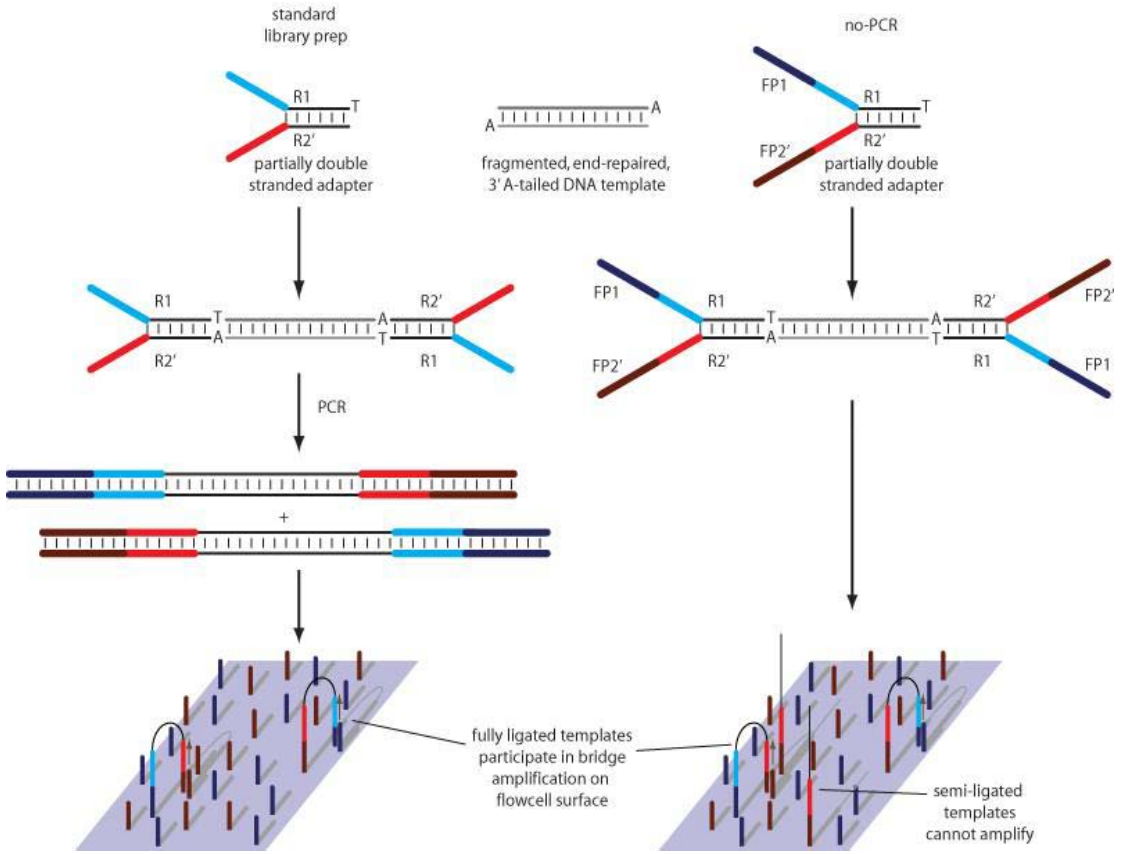
- **Neocallimastigomycota genomics**
  - No complete or partial genomes available
  - Few genomic fragments sequenced
  - cDNA-based identification and cloning of few lignocellulolytic cellulosomal enzymes

- **Challenges**
  - Unknown genome size
  - Extremely high A+T content
  - Preliminary project abandoned by the DOE-JGI

- **Potential**
  - ***Novel lignocellulolytic reperetoire***
  - Novel secondary metabolites
  - Evolutionary insights

A modified Illumina sequencing approach



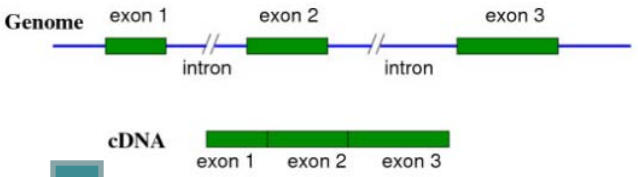


- Illumina's Sequence by synthesis approach allows for sequencing of homopolymer regions with very little bias.
- Illumina sequencing was coupled with minimal PCR amplification library generation to get the most even coverage possible on very high AT% anaerobic fungal genomic material.

Gene Calling Training Optimization

GMAP

Neocallimastix cDNA Sequences : 996
Piromyces cDNA Sequences: 18076
Total : 19072 cDNA Sequences



cDNA alignment

Eukaryotic Gene Calling

GLIMMERHMM
Eukaryotic Gene-Finding System

FGENESH geneid

e Prediction in Eukaryotes

For eukaryotic gene prediction you can use the parallel combination of GeneMark-E* and GeneMark.hmm-E.
For a novel genome (the one whose name is not in the list of available models) you can install and run locally GeneMark.hmm-ES, the self-training program (just 10MB sequence is needed for training).

Augustus [gene prediction]

Gene module prediction

Gene Module Finalization

EVidenceModeler (EVM)

EVM

Exogean

PASA

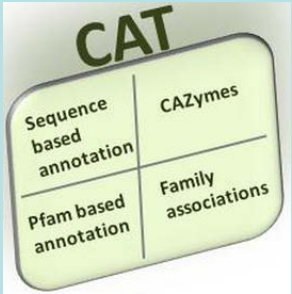
Genewise

Genomic Functional Annotation

Gene module comparison

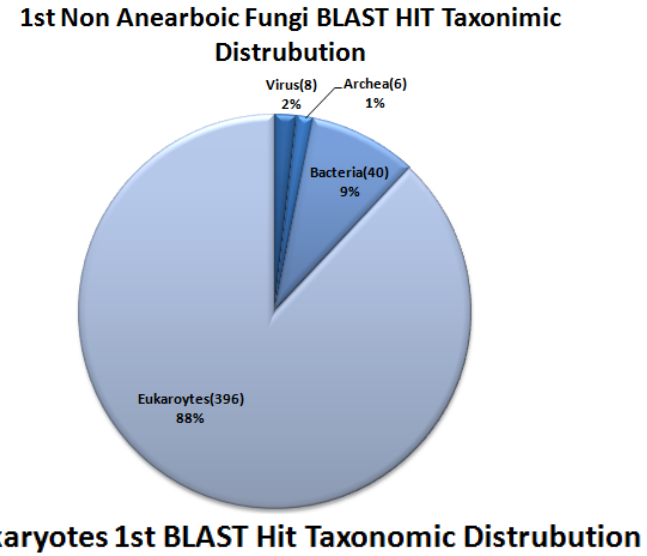
Metabolic reconstruction

CAZY-identification

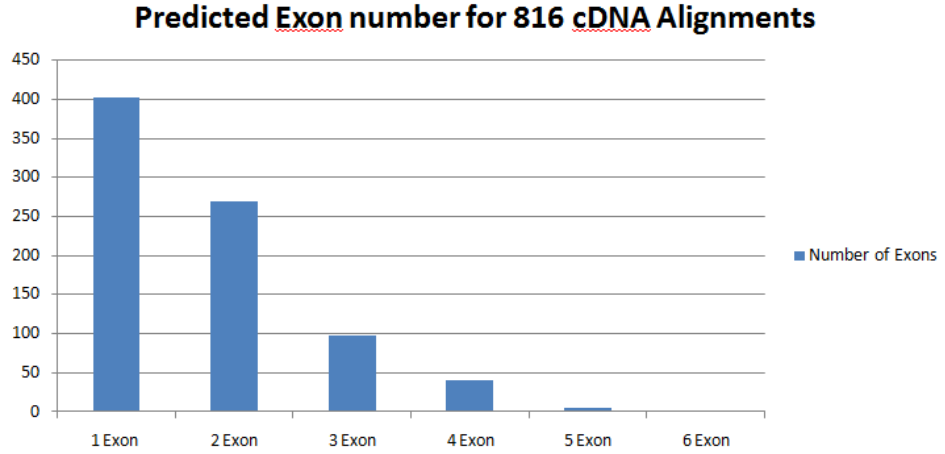


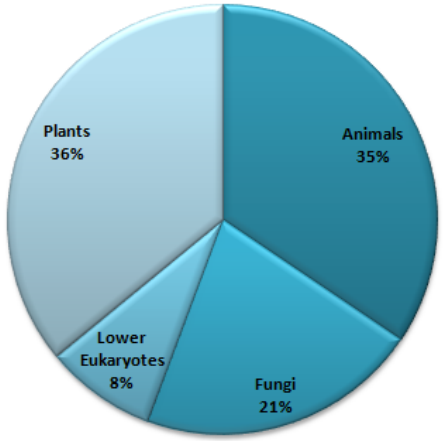
Genome sequencing:Preliminary results

- Total Length of Assembled Contigs 53,821,617 BP (~ 53.5 Million)
- Largest Assembled Contig: 40,857BP
- N50 = 1932 BP
- G+C content = 26%
- Paired end reads, N content = 18.7%



Eukaryotes 1st BLAST Hit Taxonomic Distrubution





Low number of introns/gene compared to other fungi

No particular affiliation with a specific genome sequenced species

Lignocellulolytic gene repertoire

Activity Group	Family	Function	Open Reading Frames
CELLUASES			
	GH45	*Endogluconase*	1
	GH48	*Endogluconase, Cellobiohydrolase*	24
	GH5	*Cellulase, and many other activities, majority of sequences in database are cell*	4
	GH6	*Endoguconases, cellobiohydrolases*	26
	GH7	*Endogluconase, Cellobiohydrolase,*	1
	GH9	*Endogluconase, Cellobiohydrolase, Beta Glucosidase*	1
HEMI CELLUASES			
	GH10	*Xylanases*	16
	GH11	*Xylanases*	35
	GH26	*Xylanase, Mannanase*	24
	GH43	*Xylanase, arabifuranosidase, galactase, xylocidase*	16
OLIGOSACCRIDE DEGRADATION			
	GH1	*Diverse, majoity belonging to beta glucosidase beta galactosidae*	17
	GH2	*Beta galactosidase, B mannosidase, beta glucuronosidase, but not beta glucosi*	7
	GH3	*Mainly Beta Glucosidases*	29
	GH4	*Maltose 6 phosphate glucosidase, alpha glucosidase, alpha galactosidase*	4
	GH31	*Alpha glucosidase, alpha xylosidase*	26
	GH32	*Invertase (Sucrose to fruc and glucose), also acts on other fructose bond, a fr*	3
	GH38	*Alpha Mannosidae*	23
	GH43	*Galactase, xylocidase; few Xylanase, arabifuranosidase,*	16
	GH57	*Amylase, pullulanase, etc*	1
	GH92	*Apha Mannosidases*	1
	GH97	*Alpha glucosidase, alpha galactosidase*	1
Other Activies			
	GH18	*Chitinase, class III, class IV*	1
	GH19	*Chitinase, class I, II, IV)*	2
	GH20	*β-hexosaminidase (Breaks down N-acetylglucosamine polymers)*	1
	GH25	*Lysozyme, hydrolysis Nacetlyglucosamine and N-acetylmuramic acid bonds*	1
	GH13	*Amylase, pullulanase, etc*	5
	GH57	*Amylase, pullulanase, etc*	1
	GH66	*Cycloisomaltooligosaccharide glucanotransferase, dextranase*	1
	GH89	*α-N-acetylglucosaminidase*	1
	GH109	*N-acetylgalactosaminidase (not glucose but galactose)*	1
	GH115	*Xylan α-1,2-glucuronosidase (Hydrolysis of (1→2)-α-D-(4-O-methyl)glucurono*	7
Total	**30**		**297**

Other genes of interest

Gene Name	Signifigance	BLAST Realative	e value
Swollen (expansin realtive)	Disrutps Plant Cell Wall, orthologus to plant gene	Swollenin [Aspergillus fumigatus Af293] >gi|66845375|gb|EAL85710.1|	4.92E-116
Tomatinase	Degradation of plant product anti fungal secondary metobolite	Tomatinase [Fusarium oxysporum f. sp. lycopersici]	1.63E-14
Cellulosome Dockerin	Evidence of Extracellular Cellulosome	Cellulosome protein dockerin type I [Clostridium cellulovorans 743B]	1.66E-65
Polysaccride lyase	Beta-Ellimantion mechanism for polysaccraides	Polysaccharide lyase, putative [Phytophthora infestans T30-4]	1.50E-35
Feruoryl Estrase	Plant Cell Wall Degradation	Feruloyl esterase A [Orpinomyces sp. PC-2]	1.11E-78
Secreted GDSL/acylhydrolase	Complex Polysaccharide Degradation	extracellular GDSL-like lipase/acylhydrolase, putative [Neosartorya fischeri NRRL 181]	1.26E-50
Rhamnosidase	Plant Cell Wall Degradation/Industrial Uses	alfa-L-rhamnosidase [Rhodopirellula baltica SH 1]	1.63E-119

Current & Future genomic/post genomic efforts

- Second round of Illumina sequencing
- Comparative glycogenomics
- Cellulosomal reconstruction
- Novel CAZY enzymes, CBMs identification
- Cloning, expression, enzyme characterization
- ***Further out:***
  - •Proteomic/transcriptomic studies
  - •Site-directed mutagenesis of selected CAZY enzymes
  - •Gene knockout, pathway engineering towards solventogenesis

Acknowledgments

Audra Liggenstoffer

Brian Couger

Noha Youssef

OKLAHOMA EPSCoR
Experimental Program to Stimulate Competitive Research

Oklahoma Bioenergy Center