SANDIA REPORT
SAND2009-6963
Unlimited Release
Printed November 2009

Efficient Breakdown of Lignocellulose Using Mixed-Microbe Populations for Bioethanol Production

Amy J. Powell, Bryce Ricken and Jaclyn Murton

Prepared by Amy J. Powell
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2009-6963
Unlimited Release
Printed November 2009

Efficient Breakdown of Lignocellulose Using Mixed-Microbe Populations for Bioethanol Production

Amy J. Powell, Bryce Ricken and Jaclyn Murton
Bioenergy and Defense Technology Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1413

Abstract

This report documents progress in discovering new catalytic technologies that will support the development of advanced biofuels. The global shift from petroleum-based fuels to advanced biofuels will require transformational breakthroughs in biomass deconstruction technologies, because current methods are neither cost effective nor sufficiently efficient or robust for scaleable production. Discovery and characterization of lignocellulolytic enzyme systems adapted to extreme environments will accelerate progress. Obvious extreme environments to mine for novel lignocellulolytic deconstruction technologies include aridland ecosystems (ALEs), such as those of the Sevilleta Long Term Ecological Research (LTER) site in central New Mexico (NM). ALEs represent at least 40% of the terrestrial biosphere and are classic extreme environments, with low nutrient availability, high ultraviolet radiation flux, limited and erratic precipitation, and extreme variation in temperatures. ALEs are functionally distinct from temperate environments in many respects; one salient distinction is that ALEs do not accumulate soil organic carbon (SOC), in marked contrast to temperate settings, which typically have large pools of SOC. Low productivity ALEs do not accumulate carbon (C) primarily because of extraordinarily efficient extracellular enzyme activities (EEAs) that are derived from underlying communities of diverse, largely uncharacterized microbes. Such efficient enzyme activities presumably reflect adaptation to this low productivity ecosystem, with the result that all available organic nutrients are assimilated rapidly. These communities are dominated by ascomycetous fungi, both in terms of abundance and contribution to ecosystem-scale metabolic processes, such as nitrogen and C cycling. To deliver novel, robust, efficient lignocellulolytic enzyme systems that will drive transformational advances in biomass deconstruction, we have: 1) secured an award through the Department of Energy (DoE) Joint Genome Institute (JGI) to perform metatranscriptomic functional profiling of eukaryotic microbial communities of blue grama grass (*Bouteloua gracilis*) rhizosphere (RHZ) soils and 2) isolated and provided initial genotypic and phenotypic characterization data for thermophilic fungi. Our preliminary results show that many strains in our collection of thermophilic fungi frequently outperform industry standards in key assays; we also demonstrated that this collection is taxonomically diverse and phenotypically compelling. The studies summarized here are being performed in collaboration with University of New Mexico and are based at the Sevilleta LTER research site.

Contents

1.0 Introduction

Developing efficient, environmentally benign and affordable technologies to produce advanced lignocellulosic biofuels sustainably will require transformational breakthroughs in biomass deconstruction. Applied, mission-oriented research into biotechnology-based strategies to produce biofuels from lignocellulosic feedstocks is critical to national security and economic recovery. A significant bottleneck in scalable lignocellulosic biofuels production is the need for novel, efficient enzyme systems that are robust to a diverse range of biomass pretreatment strategies [2, 3]. Aridland ecosystems (ALEs), such as those of the Sevilleta Long Term Ecological Research (LTER) site in central New Mexico (NM) represent obvious extreme environments to mine for novel lignocellulolytic deconstruction technologies because the prevailing abiotic conditions are harsh and little organic C accumulates in soils.

Soil organic carbon (SOC) in ALEs is ephemeral, primarily due to the extraordinary efficiency of microbially-secreted nutrient scavenging proteins [4-6]. This efficiency reflects community-level adaptations to C limited soils [4, 7]. Given high temperatures, low water availability and extreme soil alkalinity of these environments, it is reasonable to expect enrichment for novel, robust lignocellulolytic enzyme systems capable of efficient, prolonged catalysis under biorefinery-approximating conditions. Application of novel glycosyl hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases and their associated carbohydrate binding modules that show stability and high rates of catalysis at elevated temperatures, low water tension, high ionic strength and extreme pH is required to optimize biomass deconstruction processes [8]. Novel lignin-degrading enzymes are also essential for optimal polysaccharide breakdown of heavily lignified feedstocks, such as *Populus trichocarpa* and species of *Eucalyptus*. To date, all industrially relevant lignases are derived from basidiomycete species, such as *Phanerochaete chrysosporium*, that are endemic to temperate ecosystems. ALEs, in contrast, will yield novel enzyme systems and diverse carbohydrate-active gene products that are robust to biorefinery conditions.

Culture-independent metatranscriptomic profiling of two key primary-producer associated microhabitats[1] of Sevilleta soils will yield a vast new repertoire of lignocellulolytic enzymes, import/export motifs, carbohydrate binding modules, as well as environmentally-deployed small signaling molecules and metabolic pathways optimized by natural selection for function in extreme environments. Our recently awarded JGI CSP grant will provide the first important dataset in this area of work. Our main research goal is to identify eukaryotic expressed sequences that encode novel, robust lignocellulolytic enzyme systems that are known to occur in the Sevilleta ALE, a classic extreme environment [4]. These lignocellulolytic enzyme systems show significantly greater catalytic efficiency than those of mesic environments [4, 6]. Novel enzyme systems will be captured by sequencing microbiotic soil crust (MSC)[2] and blue grama (*Bouteloua gracilis*) grass rhizosphere soil (RHZ)[3] in the Sevilleta ALE.

In addition to our culture-independent efforts, we also employed a culture-based approach. ALEs are known to harbor true fungal thermophiles, such as *Chaetomium thermophilum* [1, 9-16]. This, and taxonomically allied organisms, have important histories in industrial applications and modern military campaigns; for instance, thermostable xylanases and cellulases (enzymes that degrade plant cell walls) are derived from species of *Chaetomium* and *Thermomyces* [11, 13, 17-19]; further, thermophilic fungi exacted significant losses for the military during the Second World War. In the Pacific theater,

[1] A small area with physical and ecological characteristics that distinguish it from its immediate surrounding area

[2] Microbiotic soil crusts (MSCs) occur in open-canopy arid ecosystems, such as Sevilleta grassland, and feature highly specialized communities of cyanobacteria, mosses, terrestrial algae, lichens and fungi

[3] Rhizosphere soils; the zone immediately surrounding plant root systems; these soils are known to possess extensive detoxification and biomass breakdown activities

thermophilic/thermotolerant fungi degraded hemp and rubber-based military assets, such as tents, rope, uniforms and tires with alarming efficiency. Thermophilic/extremophilic fungi have obvious utility in engineering processes relevant to advanced biofuels production and bioenergy co-product development. We have isolated thermophilic fungi from two different extreme environments, and completed preliminary phenotyptic screens and taxonomic characterization.

2.0 Functional metatranscriptomic profiling of eukaryotic microbial communities of a key ALE microhabitat- a culture independent approach

The advent of next generation sequencing technology has enabled relatively unbiased investigation of microbial communities and their associated metabolic activities at a scope that was not previously possible, due to high sequencing costs [20, 21]. Further, new high throughput sequencing platforms partially address the problem of unculturable and/or uncultured microorganisms. Discovery and characterization of lignocellulolytic enzyme systems adapted to extreme environments will accelerate advanced biofuels development.

2.1 Overview of methodology for functional metatranscriptomic profiling of microbial communities

We have been awarded at least 600 megabases (Mb) of high-throughput sequence for the ALE eukaryotic microbial metatranscriptome of blue grama RHZ soil. We will submit high quality total RNA preparations for metatranscriptomic analyses. A 454-GS-FLX sequencing platform will be employed for the proposed study, which will start with polyadenylated (eukaryotic) RNA isolated from total RNA.

At least 50 fungal genomes have been sequenced and annotated, and tens of thousands of expressed sequence tags (ESTs) are associated with these sequenced genomes [22-24]. The dozens of well-curated fungal genomes typically also have physical, genetic and/or optical maps [23, 25, 26]. Numerous public databases and resources will support detailed functional annotation and metabolic pathway reconstruction [27-29]. Results of the ALE metatranscriptomic survey will be incorporated into the JGI's Integrated Microbial Genomes analytical framework [30]. The incorporation of our results will extend the IMG/M database, which currently emphasizes prokaryotic metagenomic communities. All available relevant resources will be leveraged to support annotation and analyses of metatranscriptomic data obtained in the proposed work.

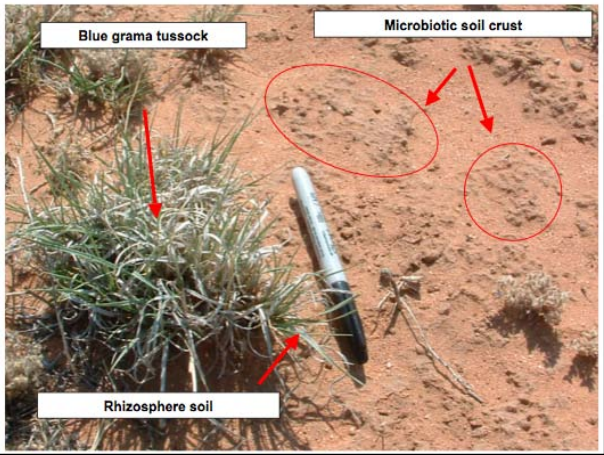


Fig. 1: Microtopography of an Aridland Ecosystem (ALE). Rhizosphere soil (RHZ) of blue grama grass harbors diverse, largely undescribed assemblages of interacting microbes. These microbes secrete efficient, robust lignocellulolytic enzymes to assimilate the maximum amount of organic nutrients, which are severely limited in this, and similar environments.

2.2 Experimental plan

Our goal is to discover novel, robust enzyme systems that have utility in industrial-scale biomass deconstruction. To achieve this goal, we will perform a eukaryotic microbial metatranscriptomic inventory of blue grama grass (*Bouteloua gracilis*) rhizosphere (RHZ) soils at the Sevilleta LTER site in

central NM (**Figure 1**). Blue grama RHZ soil samples will be processed to exclude roots and microfauna above a 2mm threshold.

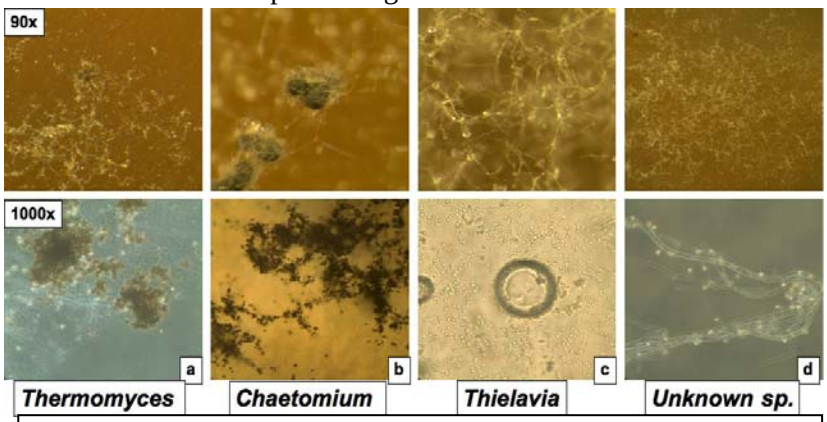
2.3 Preliminary and expected results

Blue grama RHZ soils are readily obtained from our experimental plots at the Sevilleta LTER research site. High quality, abundant total RNA can be reproducibly isolated from soil samples within a day in our laboratories; we have successfully isolated (polyadenylated) eukaryotic microbial total RNA from black grama grass (*B. eriopoda*) RHZ soils. Sequencing will begin during the first quarter of 2010.

Microbes endemic to ALEs have evolved highly efficient scavenging systems. This fact is evident in the significantly higher degradative extracellular enzyme activities (EEAs) in ALE soils, as compared to those of temperate settings [4]. These high EEAs reflect community-level adaptation to limited nutrient availability and extreme abiotic conditions (**Figure 1**) [4-6]. In sum, extreme conditions of ALEs, together with demonstrated, measureable community-level adaptations, such as efficient EEAs, signal the presence of novel, robust enzyme systems that have great potential utility for industrial bioconversion processes.

3.0 Isolation and characterization of thermophilic fungi endemic to an ALE grassland-a culture-based approach

Initial experiments from our laboratory showed that the thermophilic fungus *Thermomyces lanuginosus* rapidly hydrolyzed the lignin component of Timothy hay (Roberto Rebeil, 2008, personal communication). Isolation and examination of additional thermophilic fungal taxa endemic to extreme environments should be performed because: 1) thermophilic fungi have saprophytic (plant biomass decomposer) life histories; 2) thermophilic fungi show key adaptations, such as xerotolerance (*i.e.*, tolerance for- and/or adaptation to- low water activity conditions) and robustness to high temperatures (≥50°C); 3) thermophilic fungi can be grown at temperatures (≥50°C) that increase metabolic activity and reduce contamination by other microbes; and 4) thermophilic fungi produce thermostable gene products that have demonstrated lignocellulolytic activities in industrial contexts.



Figs. 2a-d: Morphotaxa of thermophilic fungi endemic to the Sevilleta ALE. Four morphotaxa were frequently observed in ALEs soils and lignocellulose-rich materials. Our collection includes species of *Chaetomium*, *Thermomyces* and *Thielavia*, among others. Species in these genera are currently used in industrial settings [1].

To date, we have established an SNL-based collection of 48 strains of thermophilic fungi, representing one of the largest and most taxonomically diverse collections in the world. Many strains in this collection have already shown great potential utility in the realm of plant biomass deconstruction.

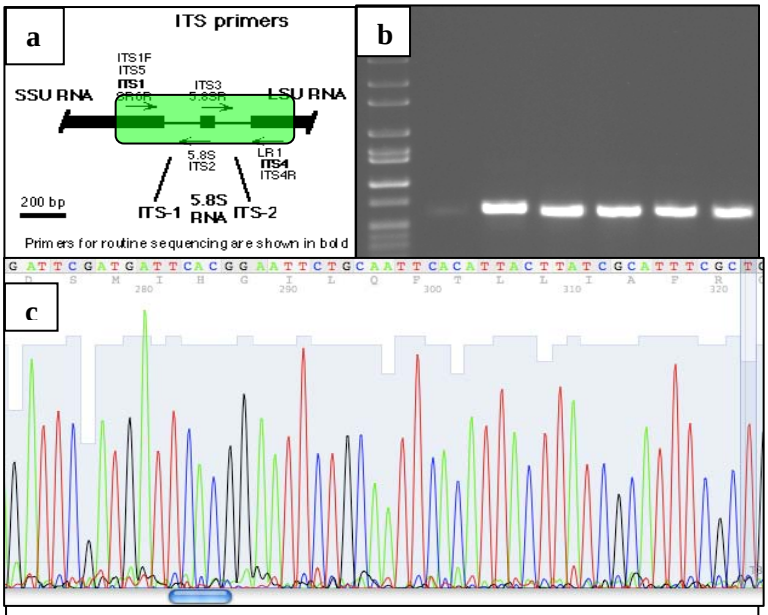
3.1 Overview of methodologies for isolation and characterization of thermophilic fungi

Soil and microbiotic crust samples (0.5 to 1.5 g) were plated on Petri dishes with the appropriate media (hay agar, hay infusion agar and/or malt). Plates contained antibiotics (50 mg/mL of ampicillin) to inhibit bacterial growth. Culture dishes were incubated at 50°C for at least one week. This high incubation temperature selected against non-thermophilic fungi. Plates were monitored for the appearance of fungal growth, and pure subcultures were obtained by serial transfer of spores and/or mycelial material to new plates [31]. Replicate transfers were incubated at room temperature to confirm thermophily. Strains that grew well at 50°C but poorly, or not at all, at room temperature were designated putative thermophiles.

A preliminary three-year study of soil samples from the Sevilleta National Wildlife Refuge revealed that more than 50% of such plates (several hundred examined) possessed growth of thermophilic fungi. The three most common species were *Sporotricum spp*., *Talaromyces thermophilus* and *Thermomyces lanugiosus*. These preliminary results confirmed the presence of thermophilic fungi at the proposed research site.

3.2 Morphological taxonomic typing results

Initial provisional taxonomic assignments were made for thermophilic fungal strains in pure culture by microscopic examination. We determined that at least four distinct morphotaxa were represented among strains sampled from the Sevilleta ALE (**Figures 2a-d**).



Figs. 3a-c: Phylogenetically/taxonomically relevant locus targeted for amplification and sequence analyses. The internal transcribed spacer (ITS; **3a**, highlighted in green) region of the ribosomal DNA repeat was amplified *via* polymerase chain reaction (**3b**) and sequenced (**3c**) for all thermophilic fugal strains. These results are foundational to definitive taxonomic and phylogenetic characterization, and refine and extend morphology-based taxonomic assignments. Definitive taxonomic classification and phylogenetic contextualization are the necessary prerequisites for whole genome analyses, which will be performed in collaboration with the Joint BioEnergy Institute (JBEI), the Joint Genome Institute (JGI), Novozymes and UNM.

3.3 Molecular genotyping results

Well-characterized evolutionarily-conserved loci were sequenced to make definitive taxonomic assignments for ALE thermophilic fungal strains [32, 33]. Genomic DNA was isolated from all strains, and phylogenetically/taxonomically-relevant loci, such as the internal transcribed spacer (ITS) region of the ribosomal DNA repeat (**Figure 3**) and single-copy nuclear protein coding genes such as elongation factor-1α (*ef1-α*), as well as the largest subunit of the RNA polymerase II (RPB1), were amplified *via* polymerase-chain reaction (**PCR**). PCR products (**Figures 3a-c**) were purified and sequenced using commercially available reagents and kits. We generated a total of 71 sequences for the ITS region and 44 sequences for *ef1-α.* Initial bioninformatic analyses of the molecular results indicate that our collection has at least seven distinct taxonomic lineages. As indicated by ITS sequences, the seven well-resolved taxa are: 1) *Aspergillus fumigatus*, 2) *Thermomyces lanuginosa*, 3) *Scytalidium thermophilum* 4) *Talaromyces themophilus* 5) *Thermoascus sp*. 6) *Chaetomium sp.* and 7) *Humicula fuscoatra*. In addition to these taxa, our collection also includes twenty-four strains that are most closely related to sequences for "uncultured

compost fungi," a GenBank *ad hoc* designation for undescribed lineages. Sequences for all three loci will be submitted to GenBank (http://www.ncbi.nlm.nih.gov/) upon completion of the analyses for all strains.

3.4 Phenotypic characterization

Semi-quantitative phenoptypic screens were employed to identify strains of elite performing thermophilic fungi. Organismal traits relevant to industrial-scale plant biomass deconstruction were examined for all isolates. These traits include acidophily (*i.e.*, growth at pH 4.0), alkalophily (*i.e.*, growth at pH 10.0), hyperthermophily (*i.e.*, growth at ≥ 60 °C) and halophily (*i.e.*, growth at 5%, 7.5% and 10% NaCl). These traits were chosen for analyses because pH (as a function of the particular biomass pretreatment strategy employed [2, 3]), ionic strength and temperature are key variables in industrial biorefinery settings. Key organismal phenotypes were assessed in terms of radial growth for all thermophilic strains relative to comparatively well-characterized standard strains of biomass degrading fungi (*i.e.*, *Trichoderma reesei*, *Thielavia terrestris* and *Chaetomium globosum* [34-36]). Initial screening indicates that many of our strains perform better than the standards (**Table 1**). Our results are significant because these organismal phenotypes are possibly, if not likely, underpinned by gene products that will advance plant biomass breakdown catalytic technologies [8]. In addition to elite performance over single traits, such as growth at extreme pH values or under high ionic strength, some strains show robust growth over combined biorefinery-relevant traits, signaling the presence of gene products supporting extremophilic physiologies.

Number of strains of thermophiles endemic to extreme environments that perform better than standards	***Biorefinery-relevant trait screen***
9	**Hyperthermophlic growth ≥ 70 °C**
23	**Halophilic growth at 7.5% NaCl**
28	**Acidophilic growth at pH 4.0**
37	**Alkalophilic growth at pH 10.0**

Table 1: Our collection of thermophilic fungi includes strains that perform significantly better than industry standards (*e.g.*, *Trichoderma reesei*) in screens that measure grow in biorefinery-like conditions. Some strains even grow more robustly than standards over combinations of traits, such as hyperthermophily and halophily.

Targeted screens were performed on a subset of thermophilic strains for which morphology-based taxonomic assignments could be made (**Figs. 2a-d**). These screens were designed to assess strains' growth rates on recalcitrant plant cell wall components, such as different forms of cellulose (*i.e.*, avicel, or crystalline cellulose; carboxy-methyl cellulose, a soluable form of this molecule), xylan and pectin [8]. Results from these preliminary targeted screens suggest that strains in our collection will metabolize plant cell wall components at least as well as, if not more readily than current industry standards, *Trichoderma reesei* and *Thielavia terrestris* (**Figure 4**). For example, a *Chaetomium* strain from our collection (**Fig. 4**, yellow bar) has a similar or greater growth rate *vs*. both standards (**Fig. 4**, bars under yellow stars) on plant cell wall components. Our *Chaetomium* and the other three thermophilic strains (*Thermomyces sp.*, *Thielavia sp.* and the *unknown sp.*) performed better than current industry standard *Thielavia terrestris*, a thermophilic ascomycete in all instances, and at least as well as *T. reesei* in the growth assays on avicel, glucose, hay agar and xylan.

4.0 Summary

Results from the studies summarized here promise to deliver novel, robust, efficient lignocellulolytic enzyme systems to drive the development of transformational biomass deconstruction technologies. Transformational advances in biomass deconstruction technologies are essential for advanced biofuels production. Our recent JGI award will have far-reaching scientific and technical impacts because it has set the stage for metatranscriptomic functional profiling of extreme environments, such as ALEs, for the larger related purposes of climate mitigation and renewable fuels development, both of which will require biotechnology-based solutions. Culture-based studies of extreme environments will identify elite-performing strains that are predicted to possess high-value traits (*i.e.*, genes & pathways) that will have great potential utility in plant biomass breakdown applications. Genomic analyses of elite-performing thermophiles will be employed to identify genes and pathways of potential value to advanced biofuels production. Genomes of closely-related thermophiles and non-thermophiles will be compared to elucidate the genetic and/or molecular bases of fungal thermophily. Genome-enabled studies of thermophiles endemic to ALEs will be performed in collaboration with scientists at the Joint BioEnergy Institute (JBEI), Novozymes and the JGI.

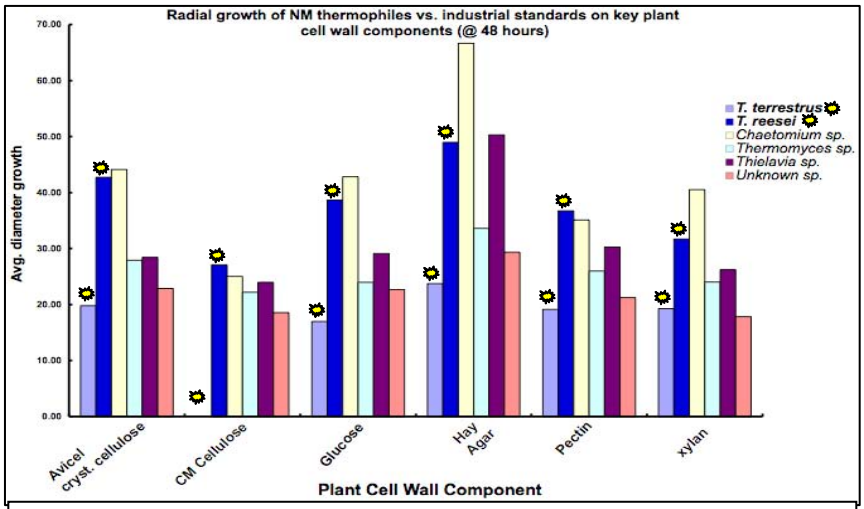


Fig. 4: Targeted phenotypic screens to measure radial growth on recalcitrant plant cell wall components. Radial growth of the first thermophilic fungi isolated from an extreme environment was measured and compared to industry standards *Thielavia terrestris* and *Trichoderma reesei*. Strains were grown on different recalcitrant plant cell wall components to identify candidates that potentially possessed novel lignocellulolytic machinery. Transformational advances in enzymatic efficiencies and robustness are needed to significantly reduce the cost of biomass deconstruction to make advanced biofuels production feasible. Enzyme systems derived from thermophilic fungi endemic to extreme environments show great promise in this realm because these microorganisms' lignocellulolytic machinery is continually optimized by natural selection to maximally exploit limited organic nutrients.

5.0 References

1. Maheshwari R, Bharadwaj G., Bhat M.K.: **Thermophilic fungi: their physiology and enzymes**. ***Microbiology and Molecular Biology Reviews*** 2000, **64**(3):461-488.
2. Lau MW, Dale BE: **Cellulosic ethanol production from AFEX-treated corn stover using Saccharomyces cerevisiae 424A(LNH-ST)**. ***Proceedings of the National Academy of Sciences*** 2009:-.
3. Rubin EM: **Genomics of cellulosic biofuels**. ***Nature*** 2008, **454**(7206):841-845.
4. Scott L. Collins RLS, Chelsea Crenshaw, Laura Green, Andrea Porras-Alfaro, Martina Stursova, Lydia H. Zeglin: **Pulse dynamics and microbial processes in aridland ecosystems**. ***Journal of Ecology*** 2008, **96**(3):413-420.
5. Stursova M, Crenshaw C, Sinsabaugh R: **Microbial Responses to Long-Term N Deposition in a Semiarid Grassland**. ***Microbial Ecology*** 2006, **51**(1):90-98.
6. Stursova M, Sinsabaugh RL: **Stabilization of oxidative enzymes in desert soil may limit organic matter accumulation**. ***Soil Biology and Biochemistry*** 2008, **40**(2):550-553.
7. Sponseller RA: **Precipitation pulses and soil CO₂ flux in a Sonoran Desert ecosystem**. ***Global Change Biology*** 2007, **13**(2):426-436.
8. United States Department of Energy (DOE) OoSaEEaRE: **Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda**. ***A Research Roadmap Resulting from the Biomass to Biofuels Workshop*** 2006.
9. Andersen B, Nissen AT: **Evaluation of media for detection of Stachybotrys and Chaetomium species associated with water-damaged buildings**. ***International Biodeterioration & Biodegradation*** 2000, **46**(2):111-116.
10. Jonny Eriksen JG: **Cellulases from Chaetomium thermophile var. dissitum**. ***European Journal of Biochemistry*** 1977, **77**(3):445-450.
11. Katapodis P, Christakopoulou V, Kekos D, Christakopoulos P: **Optimization of xylanase production by Chaetomium thermophilum in wheat straw using response surface methodology**. ***Biochemical Engineering Journal*** 2007, **35**(2):136-141.
12. Lee S, Hanlin RT: **Phylogenetic relationships of Chaetomium and similar genera based on ribosomal DNA sequences**. ***Mycologia*** 1999, **91**(3):434-442.
13. Soni R, Sandhu DK, Soni SK: **Localisation and optimisation of cellulase production in Chaetomium erraticum**. ***Journal of Biotechnology*** 1999, **73**(1):43-51.
14. Suyanto, Suyanto, Ohtsuki, Ohtsuki T, Yazaki, Yazaki S, Ui, Ui S, Mimura, Mimura A: **Isolation of a novel thermophilic fungus Chaetomium sp. nov. MS-017 and description of its palm-oil mill fiber-decomposing properties**. ***Applied Microbiology and Biotechnology*** 2003, **60**(5):581-587.
15. Tansey MR: **Effect of Temperature on Growth Rate and Development of the Thermophilic Fungus Chaetomium thermophile**. ***Mycologia*** 1972, **64**(6):1290-1299.
16. Wijeratne EMK, Paranagama PA, Gunatilaka AAL: **Five new isocoumarins from Sonoran desert plant-associated fungal strains Paraphaeosphaeria quadriseptata and Chaetomium chiversii**. ***Tetrahedron*** 2006, **62**(36):8439-8446.
17. Singh S, Reddy P, Haarhoff J, Biely P, Janse B, Pillay B, Pillay D, Prior BA: **Relatedness of Thermomyces lanuginosus strains producing a thermostable xylanase**. ***Journal of Biotechnology*** 2000, **81**(2-3):119-128.

18. Li XT, Jiang ZQ, Li LT, Yang SQ, Feng WY, Fan JY, Kusakabe I: **Characterization of a cellulase-free, neutral xylanase from Thermomyces lanuginosus CBS 288.54 and its biobleaching effect on wheat straw pulp**. ***Bioresource Technology*** 2005, **96**(12):1370-1379.
19. Chadha BS, Jaswinder K, Rubinder K, Saini HS, Singh S: **Xylanase production by Thermomyces lanuginosus wild and mutant strains**. ***World Journal of Microbiology and Biotechnology*** 1999, **15**(2):217-221.
20. Mardis ER: **The impact of next-generation sequencing technology on genetics**. ***Trends in Genetics*** 2008, **24**(3):133-141.
21. Hugenholtz P, Tyson GW: **Microbiology: Metagenomics**. ***Nature*** 2008, **455**(7212):481-483.
22. Fitzpatrick DA, Logue M.E., Stajich J.E., Butler G: **A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis**. ***BMC Evolutionary Biology*** 2006, **6**.
23. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B: **Genomics of the fungal kingdom: Insights into eukaryotic biology**. ***Genome Research*** 2005, **15**(12):1620-1631.
24. Wolfe KH: **Evolutionary Genomics: Yeasts Accelerate beyond BLAST**. ***Current Biology*** 2004, **14**:R392-R394.
25. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan HQ *et al*: **The genome sequence of the rice blast fungus *Magnaporthe grisea***. ***Nature*** 2005, **434**(7036):980-986.
26. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S *et al*: **The genome sequence of the filamentous fungus *Neurospora crassa***. ***Nature*** 2003, **422**(6934):859-868.
27. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome**. ***Nucleic Acids Res*** 2004, **32**:D277-D280.
28. Zdobnov EM, Apweiler R: **InterProScan - an integration platform for the signature-recognition methods in InterPro**. ***Bioinformatics*** 2001, **17**:847 - 848.
29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. ***Nat Genet*** 2000, **25**(1):25 - 29.
30. Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, Chen IMA, Grechkin Y, Dubchak I, Anderson I *et al*: **IMG/M: a data management and analysis system for metagenomes**. ***Nucl Acids Res*** 2008, **36**(suppl_1):D534-538.
31. Cooney DG, Emerson R.: **Thermophilic Fungi: an Account of Their Biology, Activities, and Classification**. 1964.
32. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J *et al*: **Reconstructing the early evolution of Fungi using a six-gene phylogeny**. ***Nature*** 2006, **443**(7113):818-822.
33. Porras-Alfaro A, Herrera J, Sinsabaugh RL, Odenbach KJ, Lowrey T, Natvig DO: **Novel Root Fungal Consortium Associated with a Dominant Desert Grass**. ***Appl Environ Microbiol*** 2008, **74**(9):2805-2813.
34. Khaldi N, Collemare J, Lebrun M-H, Wolfe K: **Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi**. ***Genome Biology*** 2008, **9**(1):R18.

35. Yang S-W, Mierzwa R, Terracciano J, Patel M, Gullo V, Wagner N, Baroudy B, Puar M, Chan T-M, Chu M: **Sch 213766, A Novel Chemokine Receptor CCR-5 Inhibitor from Chaetomium globosum**. ***The Journal of Antibiotics*** 2007, **60**(8):524-528.
36. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D *et al*: **Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina)**. ***Nat Biotech*** 2008, **26**(5):553-560.

6.0 Appendices

6.1 Appendix A: Preliminary results for FY07-09

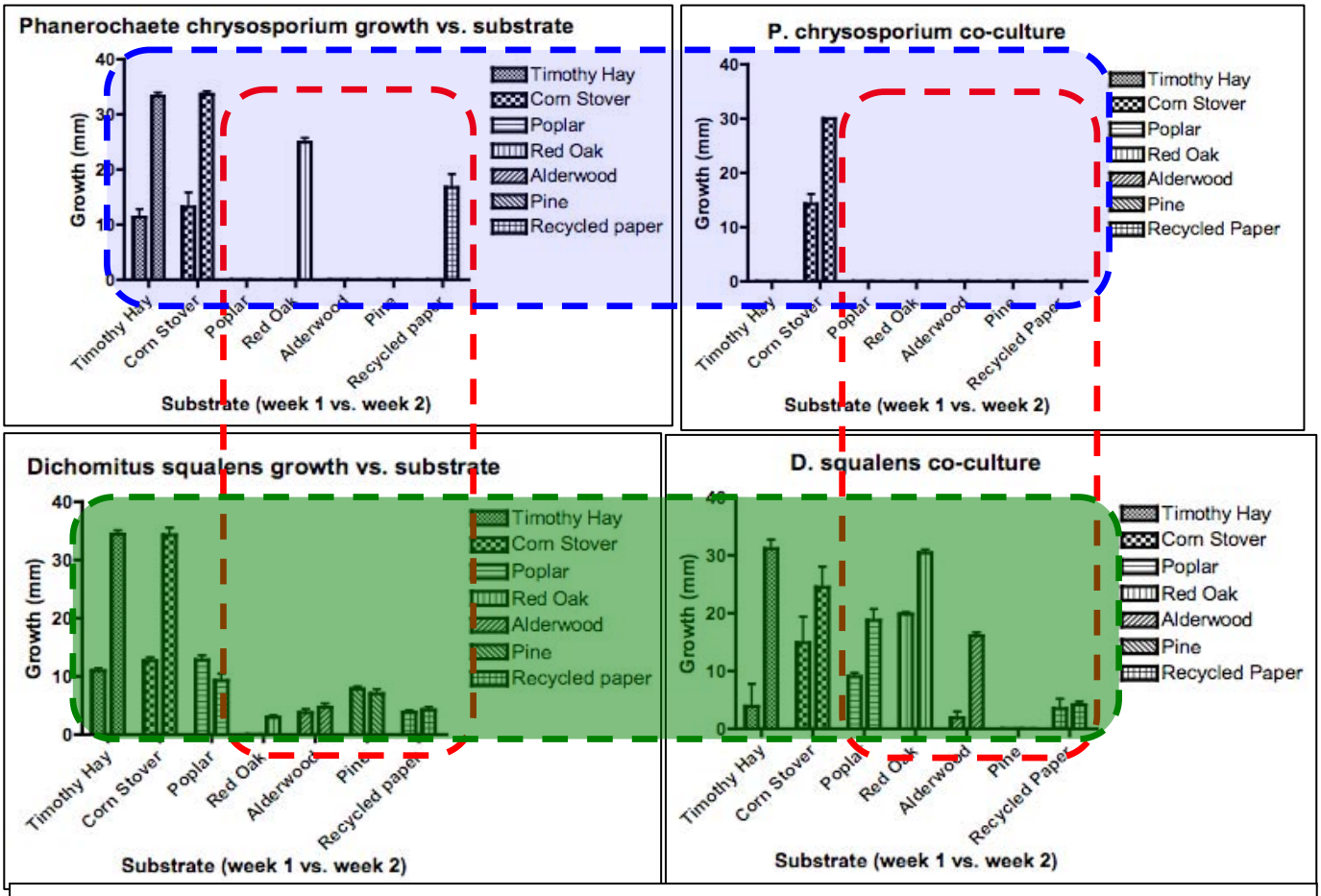


Figure: Radial growth of biomass degrading fungi on different lignocellulosic substrates. Monocultures have altered degradation spectra *vs.* co-cultures (blue and green boxes), as indicated by radial growth experiments & HPLC data (not shown). *Phanerochaete chrysosporium*, a white rot fungus, shows evidence for growth inhibition in co-culture because this organism does not assimilate red oak and recycled paper in the presence of other fungi (*Dichomitus squalens* and *Fusarium oxysporum*). By contrast, *Dichomitus squalens*, a red rot basidiomycetous fungus, grew significantly more robustly in the presence of *Fusarium oxysporum* (a causal agent of wilt disease on cereal crops), and *Phanerochaete chrysosporium* on different lignocellulosic substrates. These fungi are native to temperate settings and are thus likely not robust to industrial biorefinery conditions neither at an organismal nor a gene products level. These results highlight the need for a comprehensive genetic/genomic/-*omics* surveys and molecular phenotypic screens. Such comprehensive surveys and screens promise to deliver biotechnology-based solutions for biomass deconstruction challenges.

6.2 Appendix B:

Joint Genome Institute Community Sequencing Program proposal

Community Sequencing Program: **Project Proposal**

Proposer's Name: **The Sevilleta Genomics Consortium (SGC); Amy J. Powell (Principle Investigator); Blake Simmons, Donald O. Natvig, Scott Collins, Robert Sinsabaugh, Andrea Porras-Alfaro, Diego A. Martinez, Chris Detter, Ralph A. Dean, Jon Magnuson, Randy Berka (Co-Principle Investigators)** _____

Project Title: *An aridland ecosystem eukaryotic microbial metatranscriptome of blue grama grass rhizosphere soils as a tool to deliver transformational biomass deconstruction technologies and to gauge the impact of key global environmental change drivers* ____

Proposal ID: 164_____

A) Brief description: (Limit 1 page)

Abstract:

The global shift from fossil fuels to biofuels will require transformational breakthroughs in biomass deconstruction technologies, because current biofuel methods are neither cost effective nor sufficiently efficient or robust for scaleable production. Characterization of lignocellulolytic enzyme systems adapted to extreme environments will accelerate progress. Obvious extreme environments to mine for novel lignocellulolytic deconstruction technologies include aridland ecosystems (ALEs), such as those of the Sevilleta Long Term Ecological Research (LTER) site in central New Mexico. ALEs represent at least 35% of the terrestrial biosphere and are classic extreme environments, wherein low nutrient availability, high UV flux, limited and erratic precipitation, and extreme variation in temperatures represent the prevailing abiotic conditions. ALEs are functionally distinct from mesic environments in many respects; one salient distinction is that ALEs do not accumulate soil organic carbon (SOC), in marked contrast to mesic settings, which typically have large pools of SOC. Low productivity ALEs do not accumulate carbon (C) primarily because of extraordinarily efficient extracellular enzyme activities (EEAs) that are derived from underlying communities of diverse, largely uncharacterized microbes. Such efficient enzyme activities presumably reflect adaptation to this low productivity ecosystem, with the result that all available organic nutrients are assimilated rapidly. These communities are dominated by ascomycetous fungi, both in terms of abundance and contribution to ecosystem-scale metabolic processes, such as nitrogen (N) and C cycling. To deliver novel, robust, efficient lignocellulolytic enzyme systems that will drive transformational advances in biomass deconstruction, we propose a eukaryotic microbial metatranscriptomic inventory of blue grama grass (*Bouteloua gracilis*) rhizosphere (RHZ) soils. Blue grama is a keystone species in this and other North American grassland ecosystems. By sampling experimental plots with manipulated temperature and N, the proposed metagenomic inventory will also address the impacts of factors being evaluated in ongoing ecosystem studies that examine the effects of increased anthropogenic N deposition and climate change. Results obtained in the proposed metatranscriptomic inventory shall deliver novel, robust, efficient lignocellulolytic enzyme systems that will transform the biomass deconstruction technological landscape. At the same time, our work will give functional insight into community-level responses to the separate and combined effects of global-environmental change drivers.

Scope of Work:

Samples will be prepared from eukaryotic microbes that are endemic to blue grama RHZ soils. We request 600 Mb of high throughput sequence, representing replicated samples across four experimental treatments. We will require support from the JGI for expressed sequences clustering, functional annotation and human validation of predictions. We also seek collaboration with JGI in the areas of analyses and interpretation. Annotation jamborees that include key members of the fungal and environmental *–omics* communities, as well as, participants from the recently established DoE BioEnergy centers, are requested. Additional sequencing beyond the proposed 600 Mb may be required, should significant problems with clustering and annotation of expressed sequences become evident.

B) Background information

Technical Information:

The Sevilleta Genomics Consortium (SGC) requests 600 megabases (Mb) of high-throughput sequence for the aridland ecosystem (ALE) eukaryotic microbial metatranscriptome. We will submit high quality total RNA preparations for metatranscriptomic analysis from four experimental ecosystem manipulation treatments (detailed below) with five replicate samples per treatment, resulting in approximately 30 Mb of sequence per replicate. The Joint Genome Institute's (JGI) recommended sequencing platform will be employed for the proposed study, starting with polyadenylated (eukaryotic) RNA isolated from total RNA.

The ALE eukaryotic microbial communities that are the focus of the proposed study are from the Sevilleta Long Term Ecological Research (LTER; http://sev.lternet.edu/) site in central New Mexico. This program, associated with the Sevilleta National Wildlife Refuge, is part of the National Science Foundation (NSF)-sponsored LTER network (http://www.lternet.edu/). Metatranscriptomic inventory of rhizosphere (RHZ) soil of blue grama (*Bouteloua gracilis*), a C4 perennial keystone grass in this and other ecosystems, will be performed [2]. Eukaryotic transcripts from blue grama RHZ soils will be drawn from four existing experimental treatments within the ongoing Nighttime Warming (NW) project. These treatments are: i) ambient (no treatment), ii) increased temperature, iii) increased nitrogen (N) and iv) both increased temperature and increased N [3]. The proposed sampling scheme thus represents four conditions from which polyadenylated RNA will be isolated for expressed sequence analyses. Blue grama RHZ soil samples will be processed to exclude roots and microfauna above a 2mm threshold.

The requested deep sequencing coverage is necessary for enzyme and pathway discovery and to develop a comprehensive understanding of eukaryotic microbial metabolic processes of ALEs, which remain essentially uncharacterized, despite the fact that such environments represent at least 35% of terrestrial biomes [4, 5]. Replicate sampling within and among treatments is necessary for robust comparisons and conclusions.

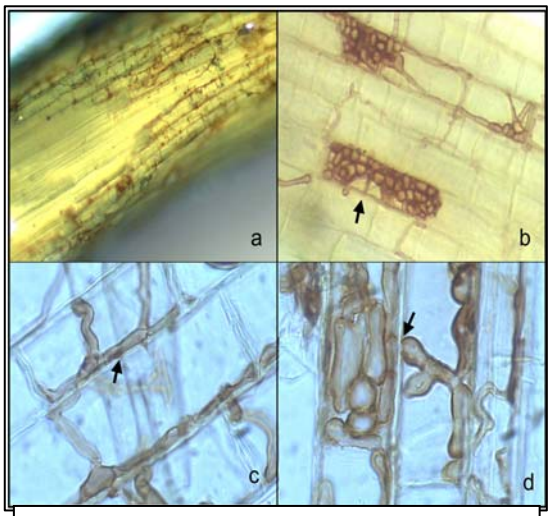


Fig. 1 Dark Septate Fungi (DSF) colonizing *B. gracilis* roots. Ascomycetous DSF infect grasses and colonize RHZ soils in ALEs. Extensive, melanized fungal biomass is ramified in root tissues (arrows). DSF belong to the order *Pleosporales*, a group that contains mutualists and numerous pathogens. Fungal communities in roots and RHZ soils are significantly different. Blue grama RHZ soils show three-fold greater phylogenetic diversity than roots.

Blue grama grass roots and rhizosphere soil communities: a tale of two microhabitats. Recent studies provide insight into the demography of eukaryotic microbial residents of ALE blue grama roots and RHZ soils [2, 6, 7]. Porras-Alfaro *et al* showed that a largely novel consortium fungal endophytes is associated with blue grama roots, with the majority of phylotypes being identified as dark-septate fungi (DSF), which are taxonomically allied with the ascomycetous order Pleosporales (**Figure 1**); 630 fungal internal transcribed spacer (ITS) sequences representing 51 phylotypes were obtained in this survey. Analyses of blue grama RHZ soils indicate that the magnitude of phylogenetic diversity is significantly greater than that observed in blue grama roots, and more than one-third of fungal taxa that occur in this microhabitat likely

represent novel species [2, 7]. In both blue grama roots and RHZ soils, most taxa belong to phylum Ascomycota (83%), with Pleosporales being the most frequently observed taxonomic order. While Pleosporales is the most commonly observed higher-level taxonomic group in both blue grama roots and RHZ soils, the composition of these communities is significantly different [7]. In sum, these molecular taxonomic inventories establish that: 1) most eukaryotic microbial taxa in blue grama RHZ soils are ascomyceteous fungi, with many phylotypes belonging to order Pleosporales, a group known to contain numerous important plant pathogens [8] and 2) RHZ soils are colonized by multiple clades of undescribed fungi.

Genome size, G+C content, polymorphism level, repeat structure and expected features of the ALE metatranscriptome. The proposed ALE eukaryotic microbial metatranscriptome will capture two important categories of genes and pathways: 1) constitutively expressed core metabolic genes and pathways and 2) genes and pathways that are expressed as a function of specific environmental conditions and/or experimental manipulation. Both categories are of broad interest ecologically, in terms of landscape-scale, below-ground metabolic processes such as nitrogen (N) and carbon (C) cycling; and they are of particular interest to biomass deconstruction researchers, as we expect to identify novel, robust lignocellulolytic enzyme systems that show adaptation to extreme, C starved environments. We expect to capture the most sequences from the most abundant eukaryotic microbes in blue grama RHZ soils, particular members of the Pleosporales. In addition to this fungal group, basidiomycete, arbuscular mycorrhizal (AM), and flagellated fungi have been observed in blue grama RHZ soils, albeit at lower frequencies, and it is expected that protein-coding sequences from these latter groups will be observed [7]. Genes and pathways of interest in the core metabolic category include those involved in fungal nitrification/denitrification and carbon metabolism, [9]. In the differential-expression category, we expect to capture diverse, novel lignocellolytic enzyme systems encoding glycosyl hydrolases and transferases, peptidases, phenol oxidases, peroxidases, lignases and laccases that will be of immediate utility to feedstocks deconstruction applications [4, 10]. We also expect to observe sequences that encode products with predicted roles in heat shock and chronic physiologic stress management responses [11].

Species	Size (Mb)	No. of genes	% Coding	% GC
T. reesei	33.9	9,129	40.4	52
F. graminearum	36.1	11,640	56.24	48.3
N. crassa	38.7	10,620	38.5	49.6
M. grisea	39.4	12,841	50.4	52
A. nidulans	30.1	10,701	58.8	50.3
S. cerevisiae	12.0	5,885	72.55	38.3
P. chrysosporium	34.5	10,048	42.22	56.8

Table 1: Summary statistics for fungal genomes. Genome sizes, the numbers of protein-coding genes and GC content of representative fungal genomes provide a benchmark estimate for expected features of the ALE eukaryotic microbial metatranscriptome. This table is adapted from Martinez *et al*, 2008 [1].

Summary statistics for fully sequenced fungal genomes provide insight into what can be expected for GC content, within-genome polymorphism and the protein coding potential of the largely ascomycetous fungal residents of RHZ soils (**Table 1**). Fungal genomes contain varying amounts of repeat sequences, ranging from 3%-21% of total genomic sequence [12, 13]. We expect the ALE eukaryotic microbial metatranscriptome to fall within a comparable range of repeat sequences and to have similar GC content (*ca.* 50 %) for protein-coding genes (**Table 1**). Given the predominance of Pleosporalean taxa and their inferred close-relationships with known

plant pathogens, it is possible that within-genome polymorphism of ALE blue grama grass RHZ fungal residents will be substantial; our laboratories have shown that within-genome polymorphism can be relatively high within *Magnaporthe*, the rice blast fungus, and we have speculated that this genomic feature is associated with phytopathogenic life histories [2, 7, 14, 15].

Available Resources:

At least 50 fungal genomes have been sequenced and annotated, and tens of thousands of expressed sequence tags (ESTs) are associated with these sequenced genomes [12, 16, 17]. The dozens of well-curated fungal genomes typically also have physical, genetic and/or optical maps [12, 14, 18]. Numerous public databases and resources will support detailed functional annotation and metabolic pathway reconstruction [19-21]. Results of the ALE metatranscriptomic survey will be incorporated into the JGI's Integrated Microbial Genomes analytical framework [22]. The incorporation of our results will extend the IMG/M database, which currently emphasizes prokaryotic metagenomic communities. In sum, all available relevant resources will be leveraged to support annotation and analyses of metatranscriptomic data obtained in the proposed work.

Potential Technical Challenges:

Anticipated technical challenges include short sequence reads that preclude reliable clustering and annotation of expressed sequences. Short sequence reads can be a consequence of the specific sequencing technology that is employed. Short sequence reads can also stem from poor environmental RNA preparations. Assuming 454 FLX/Roche technology and sufficient quantities ($\geq$ 20 µg) of high quality total environmental RNA, we expect sequence reads that are at least 300 bases, which will enable reliable clustering and annotation, given the proposed level of sequence coverage. A different, but related problem of globally short sequence reads is the spurious assignment of two or more reads to more than one locus, thus confounding functional annotation. Our experimental design mitigates many problems associated with short sequence reads, in that we will examine only eukaryotic protein-coding genes, which are derived from mature transcripts. Rational annotation strategies can also be implemented to minimize spurious functional attribution. A recent study of eukaryotic microbial soil communities of a pine-dominated forest confirms the basic feasibility of the proposed work [23]. In our own laboratories, we have successfully isolated eukaryotic microbial total RNA from black grama grass (*B. eriopoda*) RHZ soils.

Starting Materials:

Blue grama RHZ soils are readily obtained from our experimental plots at the Sevilleta LTER NW research site. High quality, abundant total RNA can be reproducibly isolated from soil samples within a day in our laboratories.

C) Project Description:

Importance of the Proposed Research:

Developing efficient, environmentally benign and affordable technologies to produce advanced lignocellulosic biofuels sustainably will require transformational breakthroughs in biomass deconstruction. Applied, mission-oriented research into biotechnology-based strategies to produce biofuels from lignocellulosic feedstocks is critical to national security and economic recovery. A significant bottleneck in scalable lignocellulosic biofuels production is the need for novel, efficient enzyme systems that are robust to a diverse range of biomass pretreatment strategies [24, 25].

Our first research goal is to identify eukaryotic expressed sequences that encode novel, robust lignocellulolytic enzyme systems that are known to occur in an aridland ecosystem (ALE), a classic extreme environment [4]. These lignocellulolytic enzyme systems show significantly greater catalytic efficiency than those of mesic environments [4].

Our second goal is to functionally characterize the impacts of the key global environmental change factors (increased temperature and increased nitrogen deposition) upon ALEs. Increased anthropogenic N deposition, even at sites that are distant from large conurbations, is a documented result of global fossil fuel dependency and modern agricultural practices. Higher temperatures are the known effect of continually rising greenhouse gas emissions, and credible climate change models predict accelerating near term global temperature increases [26]. Thus, our ancillary goal is to develop a detailed, mechanistic model of microbially-mediated carbon (C) and nitrogen (N) fluxes through ALEs in the context of an ongoing global environmental change experiment at the Sevilleta LTER site. These areas of research are crucial for understanding global C and N budgets, and for formulating rational green house gas management strategies.

Scientific Questions:

Goal 1: Identifying novel, robust, efficient lignocellulolytic enzyme systems in a functionally uncharacterized community of eukaryotic microbes. Our first goal is to discover novel, robust enzyme systems that have utility in industrial-scale biomass deconstruction. To achieve this goal, we propose a eukaryotic microbial metatranscriptomic inventory of blue grama grass (*Bouteloua gracilis*) rhizosphere (RHZ) soils at the Sevilleta LTER site (**Figure 2**). Aridland ecosystems (ALEs), such as those of the Sevilleta, represent at least one-third of terrestrial biomes and are subject to extreme abiotic conditions [4, 5]. Despite their global significance, these ecosystems have not been the focus of functional *-omics* studies [4, 5]. The proposed metatranscriptomic inventory will provide important insight into global ALE function and will yield novel gene products that encode lignocellulolytic enzymes with immediate relevance to industrial scale bioconversion processes.

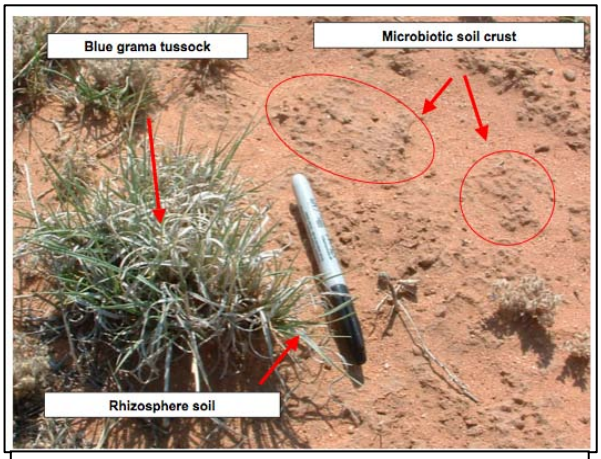


Fig 2 Microtopography of an Aridland Ecosystem (ALE). Rhizosphere soil (RHZ) of blue grama grass harbors diverse, largely undescribed assemblages of interacting microbes. Pulsed patterns of metabolic activity reflect community-level adaptation to extreme abiotic conditions.

Blue grama is a keystone grass species in this and other North American ecosystems [2]. Recent molecular surveys show that 70% of fungal taxa that occur in ALE blue grama RHZ soils are novel and represent functionally uncharacterized eukaryotic microbial communities [2]. Fungi are the dominant microbes in ALEs in terms of their abundance and contribution to ecosystem-scale metabolic processes, such as C and N cycling, and they drive lignocellulosic biomass breakdown so effectively that C turnover can exceed net primary productivity [4, 10, 27]. In contrast to temperate systems with large soil organic carbon (SOC) pools, fungi and other microbes in ALEs aggressively scavenge for recalcitrant C and other resources in nutrient poor soils [4, 10]. The evolution of efficient microbial scavenging systems is evident in the significantly higher degradative extracellular enzyme activities (EEAs) in ALE soils, as compared to those of temperate settings [4]. These high EEAs reflect community-level adaptation to limiting nutrient conditions. Thus, *in situ,* gene products from ALEs are necessarily stable at high temperatures, low water tension and in alkaline soil conditions [4, 9, 10]. In sum, extreme abiotic conditions of ALEs, together with demonstrated, measureable community-level adaptations such as high lignocellulose degradative EEAs, signal the presence of novel, robust enzyme systems that have great potential utility for industrial bioconversion processes.

Goal 2: Understanding the effects of key global environmental change factors upon eukaryotic microbial metabolic processes in an ALE. Our second goal is to understand the separate and combined effects of increased temperature and N on the RHZ metatranscriptome. Increased temperatures and N deposition are having unprecedented impacts upon global ecosystems over short time scales. Dramatic examples include the rapid melting of polar ice sheets and acidification of oceans [28, 29]. In the case of ALEs, fundamental questions remain regarding N and C cycling, and developing a comprehensive understanding of metabolic processes of this and other extreme environments will require a genes and pathways-centered mechanistic account. Developing such an account is crucial, given intensifying environmental degradation, which short circuits normal ecosystem function and compromises delivery of essential ecosystem services [30-33]. Extreme abiotic conditions of ALEs give rise to pulsed patterns of metabolic activity [4, 34]. Precipitation events activate temperature-dependent metabolic activities, including nutrient assimilation, respiration and growth, as well as ecosystem-scale exchange processes, such as N and C cycling [4]. Pulsed metabolic processes can be mapped onto important microtopographic landscape features that determine microbial distribution and activity (**Figures 2 and 3**) [2, 4, 7]. These microbial metabolic processes show spatial and temporal compartmentalization [4].

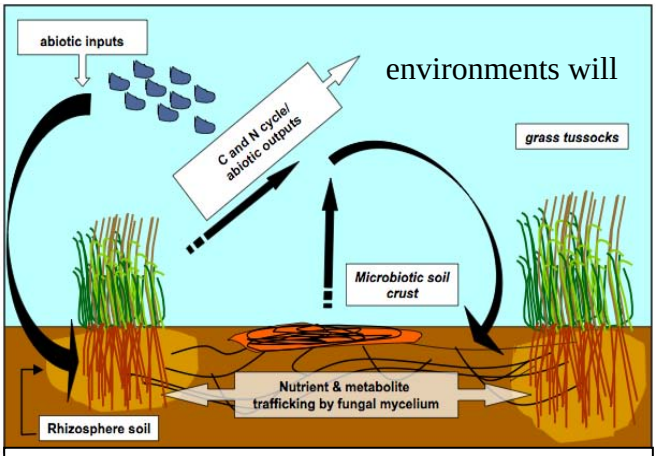


Fig. 3 *M*icrobial *A*ridland *M*etabolic *L*oop (MAML). The loop is metabolically integrated by fungal mycelial networks. We propose a metatranscriptomic inventory of blue grama grass rhizosphere (RHZ) soils to identify novel lignocellulolytic enzyme systems. The proposed work will also elucidate the impacts of key global environmental change factors upon ALE microbial metabolic processes.

Spatial compartmentalization is evident in the occurrence of distinct RHZ and microbiotic soil crust (MSC) communities (**Figures 2 and 3**) [2, 7]. These communities are linked to primary producers by fungal mycelial networks (**Figure 3**) [35]. Temporal partitioning in these communities is apparent in shorter periods of metabolic activity for bacteria and plants, *versus* those of fungi [4]. A new model, the microbial aridland metabolic loop (MAML), integrates these compartments, and extends the pulse-reserve paradigm to include microbial metabolism (**Figure 3**) [4, 35]. The MAML model provides a compelling framework to interpret findings and to test hypotheses for the proposed metatranscriptomic study.

Preliminary Data Produced in Support of the Proposed Project:

Our laboratories have made rapid, substantial progress towards preliminary characterization of the organismal and metabolic complexity of eukaryotic microbial communities of the Sevilleta ALE [2, 4, 6, 7, 9, 10, 35-38]. Of immediate importance to the proposed work is our ability to isolate high quality RNA in useful quantities in a timely manner from diverse environmental samples, including RHZ soils of endemic grasses, herbivore dung and microbiotic soil crust. Isolation of high quality RNA in sufficient quantities is central to the success of the proposed work.

Size and Nature of the Community of Users:

Interest and utility of ALE metatranscriptomic data. The proposed work will deliver results that are of broad interest and utility to researchers involved in advanced biofuels production, functional microbial -*omics*, landscape ecology and global environmental change studies. Chief among potential end users are the recently created Department of Energy (DoE) Bioenergy Research Centers. The three centers, the DoE BioEnergy Science Center, DoE Great Lakes Bioenergy Science Center and the DoE Joint Bioenergy Institute, are tasked with engineering solutions to bottlenecks in lignocellulosic fuels production. Current and previous industrial collaborators in the DoE bioenergy mission will also profit from our results. Novozymes and Genencor, collaborators on the recently published genome analysis for the model cellulase-producing fungus *Trichoderma reesei,* and recipients of DoE biomass awards, are actively engaged in solving temperature and pH stability problems inherent to current enzymatic technologies used in biorefinery settings. Further, Novozymes (in collaboration with the JGI and Pacific Northwest National Laboratory) have recently sequenced the genome of the thermophilic ascomycetous fungus *Thielavia terrestris*. The aforementioned efforts will complement the proposed work and synergistically strengthen our comparative analysis. The SGC is committed to collaborating with scientists at Novozymes on this, and the current *T. terrestris* project, as well as near term future research that will investigate fungal extremophily, broadly.

Separate groups inside and outside the DoE National Laboratory system, including the Chemical and Biological Process Development Group at the Pacific Northwest National Laboratory (PNNL) and National Renewable Energy Laboratory (NREL), are also potential end users of our data. Independent researchers at other DoE laboratories, Los Alamos National Laboratory (LANL) for example, are likely end users and collaborators, as well. We will work with curators of the Carbohydrate Active Enzymes Database (CAZyDB) to provide enhanced functional annotation of lignocellulolytic enzymes and carbohydrate-active gene products discovered in the proposed study. To facilitate the broader research community's assimilation of our results, all data will be made available through the JGI's public portal and will also be incorporated into the Integrated Microbial Genomes framework (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi).

From a purely ecological perspective, it is impossible to overstate the importance of the proposed work. For decades, scientists have attempted to understand ecological processes that determine the distribution of key plant species in ALEs. The need to understand the relationship between above and below-ground biodiversity and ecosystem functioning in this, and similar settings, has taken on tremendous urgency because of intensifying global environmental change pressures [5, 26, 39]. The proposed work highlights the crucial need to develop a detailed mechanistic account of how below-ground microbial metabolic processes shape above-ground landscapes.

Relevance of the ALE Metatranscriptome to DoE Missions and its Broader Economic and Societal Value:

The proposed metatranscriptomic inventory will have substantial scientific and technical significance for 1) alternative energy production and 2) global carbon cycling and sequestration because the target microbial communities possess adaptations that have great potential utility in biomass deconstruction applications and because the flow of C and N through ALEs is poorly understood [4, 27, 40, 41].

Mission area 1: alternative energy production. Soil organic carbon (SOC) in ALEs is ephemeral, primarily due to the extraordinary efficiency of microbially-secreted nutrient scavenging proteins. This microbial arsenal of secreted enzymes has exceptionally high lignocellulolytic activities [4, 10]. Given high temperatures, low water activity and extreme soil alkalinity of these environments, it is reasonable to expect enrichment for novel, robust lignocellulolytic enzyme systems capable of efficient, prolonged catalysis under conditions approximating those of biorefineries. Application of novel glycosyl hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases and their associated carbohydrate binding modules that show stability and high rates of catalysis at elevated temperatures, low water tension and extreme pH is required to optimize biomass deconstruction processes. Novel lignin-degrading enzymes are also essential for optimal polysaccharide breakdown of heavily lignified feedstocks, such as *Populus trichocarpa* and species of *Eucalyptus*. To date, all industrially relevant lignases are derived from basidiomycete species, such as *Phanerochaete chrysosporium*, that are endemic to temperate ecosystems. ALEs, in contrast, will yield novel enzyme systems and diverse carbohydrate-active gene products that are robust to biorefinery conditions.

Mission area 2: carbon cycling and sequestration. Given that ALEs currently represent at least 35% of terrestrial biomes and that desertification processes are intensifying globally, it is critical to develop a mechanistic understanding of carbon flux in existing and emerging aridland environments [5, 27]. ALEs accumulate little carbon because extraordinarily efficient microbial EEAs are supplemented by abiotic processes, including photodegradation and stabilized enzyme reserves. Together, these conditions generate selective pressures for tight coupling of symbiotic C and nutrient cycles that integrate fungal and phototrophic communities [4]. Understanding mechanisms of C flow through ALEs is essential for developing cogent global environmental change models that are useful in formulating mitigation strategies, particularly for regions such as the American Southwest, that are expected to warm at even greater rates, and thus are likely more vulnerable to environmental degradation processes [42]. In sum, inventory of an ALE eukaryotic microbial metatranscriptome will yield a novel repertoire of lignocellulolytic enzymes, import/export motifs, carbohydrate binding modules, as well as environmentally-deployed small signaling molecules and metabolic pathways optimized by natural selection for function in extreme environments.

Economic and societal value. Novel lignocellulolytic enzyme systems will significantly enhance biomass deconstruction efficiency, thus providing a critical advancement towards scalable biofuels production. Increased component enzyme half life and catalytic activity signal reduced input costs for biomass pre-treatment cocktails. Enhanced component enzyme efficiency predicts transformational gains in catalytic synergism of cocktails. Significantly increased catalytic efficiency possibly also translates into reduced duration and/or severity in pre-treatment steps, such that fewer inhibitory compounds result and maximal yields of fermentable sugars are obtained.

Lignocellulolytic enzyme systems identified in the ALE metatranscriptomic survey represent numerous novel engineering targets for potential commercial partners, such as Novozymes and Genencor. Engineering and production of novel biomass deconstruction enzymes will thus accelerate the "greening" and recovery of the national economy. Novel enzyme systems discovered in the proposed research are readily integrated into existing biorefinery infrastructure.

Optimized enzymatic catalysis at biomass pre-treatment stages will reduce negative environmental impacts of biofuels production. Increased enzymatic efficiency at pre-treatment stages is critical for recovering the maximum amount of fermentable sugars and will thus streamline the complexity of downstream biofuels production processes. Deployment of novel enzyme systems will also contribute to more efficient land use by reducing dedicated acreage to energy crops. This reduction translates into less intensive use of arable lands and water resources. Less intensive land use also brings the accompanying diminution of chemical inputs, further decreasing dependency on petroleum-based agricultural amendments. In sum, novel enzyme systems identified and characterized in the proposed work will deliver transformational biomass deconstruction technologies. Such transformative advances can plausibly be brought to market in the near term and will be necessary to achieve federally-mandated alternative fuels production targets.

Conclusion:

The proposed ALE eukaryotic microbial metatranscriptome will deliver novel, robust, efficient lignocellulolytic enzyme systems to drive the development of transformational biomass deconstruction technologies. Transformational advances in biomass deconstruction technologies are essential for scaleable biofuels production. The proposed work will also elucidate the influence of increased temperature and nitrogen deposition upon ALE function. Significantly, the proposed work represents the first microbial metatranscriptomic study to be performed as part of an ongoing field-based, NSF and DoE-supported global environmental change experiment.

D) References

1. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D *et al*: **Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina)**. ***Nat Biotech*** 2008, **26**(5):553-560.
2. Porras-Alfaro A, Herrera J, Sinsabaugh RL, Odenbach KJ, Lowrey T, Natvig DO: **Novel Root Fungal Consortium Associated with a Dominant Desert Grass**. ***Appl Environ Microbiol*** 2008, **74**(9):2805-2813.
3. Collins SC, et al: **Rapid Plant Community Responses During the Summer Monsoon to Nighttime Warming in a Northern Chihuahuan Desert Grassland**. ***Submitted*** 2009.
4. Scott L. Collins RLS, Chelsea Crenshaw, Laura Green, Andrea Porras-Alfaro, Martina Stursova, Lydia H. Zeglin: **Pulse dynamics and microbial processes in aridland ecosystems**. ***Journal of Ecology*** 2008, **96**(3):413-420.
5. Reynolds JF, Smith DMS, Lambin EF, Turner BL, II, Mortimore M, Batterbury SPJ, Downing TE, Dowlatabadi H, Fernandez RJ, Herrick JE *et al*: **Global Desertification: Building a Science for Dryland Development**. ***Science*** 2007, **316**(5826):847-851.
6. Porras-Alfaro A, Herrera J, Natvig D, Sinsabaugh R: **Effect of long-term nitrogen fertilization on mycorrhizal fungi associated with a dominant grass in a semiarid grassland**. ***Plant and Soil*** 2007, **296**(1):65-75.
7. Porras-Alfaro Aea: **Diversity and Distribution of Soil Fungal Communities in a Semiarid Grassland**. ***Submitted*** 2008.
8. Berbee ML: **The phylogeny of plant and animal pathogens in the Ascomycota**. ***Physiological and Molecular Plant Pathology*** 2001, **59**:165-187.
9. Stursova M, Crenshaw C, Sinsabaugh R: **Microbial Responses to Long-Term N Deposition in a Semiarid Grassland**. ***Microbial Ecology*** 2006, **51**(1):90-98.
10. Stursova M, Sinsabaugh RL: **Stabilization of oxidative enzymes in desert soil may limit organic matter accumulation**. ***Soil Biology and Biochemistry*** 2008, **40**(2):550-553.
11. Wijeratne EMK, Paranagama PA, Gunatilaka AAL: **Five new isocoumarins from Sonoran desert plant-associated fungal strains Paraphaeosphaeria quadriseptata and Chaetomium chiversii**. ***Tetrahedron*** 2006, **62**(36):8439-8446.
12. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B: **Genomics of the fungal kingdom: Insights into eukaryotic biology**. ***Genome Research*** 2005, **15**(12):1620-1631.
13. Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al*: **The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis**. ***Nature*** 2008, **452**(7183):88-92.
14. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan HQ *et al*: **The genome sequence of the rice blast fungus *Magnaporthe grisea***. ***Nature*** 2005, **434**(7036):980-986.
15. Powell A, Conant G, Brown D, Carbone I, Dean R: **Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens**. ***BMC Genomics*** 2008, **9**(1):147.

16. Fitzpatrick DA, Logue M.E., Stajich J.E., Butler G: **A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis**. ***BMC Evolutionary Biology*** 2006, **6**.
17. Wolfe KH: **Evolutionary Genomics: Yeasts Accelerate beyond BLAST**. ***Current Biology*** 2004, **14**:R392-R394.
18. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S *et al*: **The genome sequence of the filamentous fungus *Neurospora crassa***. ***Nature*** 2003, **422**(6934):859-868.
19. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome**. ***Nucleic Acids Res*** 2004, **32**:D277-D280.
20. Zdobnov EM, Apweiler R: **InterProScan - an integration platform for the signature-recognition methods in InterPro**. ***Bioinformatics*** 2001, **17**:847 - 848.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. ***Nat Genet*** 2000, **25**(1):25 - 29.
22. Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, Chen IMA, Grechkin Y, Dubchak I, Anderson I *et al*: **IMG/M: a data management and analysis system for metagenomes**. ***Nucl Acids Res*** 2008, **36**(suppl_1):D534-538.
23. Bailly J, Fraissinet-Tachet L, Verner M-C, Debaud J-C, Lemaire M, Wesolowski-Louvel M, Marmeisse R: **Soil eukaryotic functional diversity, a metatranscriptomic approach**. ***ISME J*** 2007, **1**(7):632-642.
24. Lau MW, Dale BE: **Cellulosic ethanol production from AFEX-treated corn stover using Saccharomyces cerevisiae 424A(LNH-ST)**. ***Proceedings of the National Academy of Sciences*** 2009:-.
25. Rubin EM: **Genomics of cellulosic biofuels**. ***Nature*** 2008, **454**(7206):841-845.
26. United Nations Intergovernmental Panel on Climate Change I: **Climate Change 2007, the Fourth Assessment Report (AR4)**. 2007.
27. Stone R: **ECOSYSTEMS: Have Desert Researchers Discovered a Hidden Loop in the Carbon Cycle?** ***Science*** 2008, **320**(5882):1409-1410.
28. Steffensen JP, Andersen KK, Bigler M, Clausen HB, Dahl-Jensen D, Fischer H, Goto-Azuma K, Hansson M, Johnsen SJ, Jouzel J *et al*: **High-Resolution Greenland Ice Core Data Show Abrupt Climate Change Happens in Few Years**. ***Science*** 2008, **321**(5889):680-684.
29. Orr JC, Fabry VJ, Aumont O, Bopp L, Doney SC, Feely RA, Gnanadesikan A, Gruber N, Ishida A, Joos F *et al*: **Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms**. ***Nature*** 2005, **437**(7059):681-686.
30. Seager R, Ting M, Held I, Kushnir Y, Lu J, Vecchi G, Huang H-P, Harnik N, Leetmaa A, Lau N-C *et al*: **Model Projections of an Imminent Transition to a More Arid Climate in Southwestern North America**. ***Science*** 2007, **316**(5828):1181-1184.
31. Costanza R, d'Arge R, de Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O'Neill RV, Paruelo J *et al*: **The value of the world's ecosystem services and natural capital**. ***Nature*** 1997, **387**(6630):253-260.
32. Carly J. Stevens NBDDJGGJOM: **Loss of forb diversity in relation to nitrogen deposition in the UK: regional trends and potential controls**. ***Global Change Biology*** 2006, **12**(10):1823-1833.

33. Stevens CJ, Dise NB, Mountford JO, Gowing DJ: **Impact of Nitrogen Deposition on the Species Richness of Grasslands**. ***Science*** 2004, **303**(5665):1876-1879.
34. Belnap J, Welter JR, Grimm NB, Barger N, Ludwig JA: **LINKAGES BETWEEN MICROBIAL AND HYDROLOGIC PROCESSES IN ARID AND SEMIARID WATERSHEDS**. ***Ecology*** 2005, **86**(2):298-307.
35. Laura E. Green AP-ARLS: **Translocation of nitrogen and carbon integrates biotic crust and grass production in desert grassland**. ***Journal of Ecology*** 2008, **9999**(9999).
36. Crenshaw C, Lauber C, Sinsabaugh R, Stavely L: **Fungal control of nitrous oxide production in semiarid grassland**. ***Biogeochemistry*** 2008, **87**(1):17-27.
37. Moorhead DL, Sinsabaugh RL: **A THEORETICAL MODEL OF LITTER DECAY AND MICROBIAL INTERACTION**. ***Ecological Monographs*** 2006, **76**(2):151-174.
38. Zeglin L, Stursova M, Sinsabaugh R, Collins S: **Microbial responses to nitrogen addition in three contrasting grassland ecosystems**. ***Oecologia*** 2007, **154**(2):349-359.
39. Kintisch E: **GLOBAL WARMING: Projections of Climate Change Go From Bad to Worse, Scientists Report**. ***Science*** 2009, **323**(5921):1546-1547.
40. Georg Wohlfahrt LFFJAA, III: **Large annual net ecosystem CO₂ uptake of a Mojave Desert ecosystem**. ***Global Change Biology*** 2008, **14**(7):1475-1487.
41. Xie J, Li Y, Zhai C, Li C, Lan Z: **CO2 absorption by alkaline soils and its implication to the global carbon cycle**. ***Environmental Geology*** 2009, **56**(5):953-961.
42. Kerr RA: **GLOBAL WARMING: Climate Change Hot Spots Mapped Across the United States**. ***Science*** 2008, **321**(5891):909-.

Distribution

1	MS0123	Donna L. Chavez, 1011
1	MS0899	Technical Library, 9536 (Electronic Copy)
1	MS0895	Anthony Martino, 8622
1	MS1413	Bryce Ricken, 8622
1	MS1413	Jaclyn Murton, 8622
1	MS1413	Amy Powell, 8622