The work presented here is supported by Dr. Hess’ Start-up Grant and Analysis of enzyme diversity added to the pool of known cellulases described by Zhang and colleagues (9).

To overcome the limitations of current annotation algorithms and to increase the diversity of biomass degrading enzymes that are currently known, we developed a machine-learning program that uses the global sequence fingerprint of gene clusters containing known cellulases. This global sequence fingerprint was used to build a sequence model to identify novel cellulases from biomass degrading enzymes that are currently known, we developed a machine-learning program to identify novel cellulases from microbial genomes, and searched for novel cellulases. We identified 62 hypothetical proteins without assigned function that possess a cellulase-specific fingerprint and which have not been identified as "cellulases" by currently available annotation algorithms.

**Methodology & Results**

**Amplification of cellulase candidates from genomic DNA and cloning strategy**

**Gene-specific primers**

**Gene clusters containing known cellulases**

**Microbes resources of genes possessing a cellulase-specific fingerprint**

**Analysis of putative cellulase genes from genomic DNA**

**Expression of the cellulase candidates**

**Ongoing Process of Research**

- Test enzymatic activity of constructs expressed in BL21 host strains using carboxymethylcellulose (CMC) plates as described by Zhang and colleagues (9).
- Purification of cellulase candidates using a one-step purification strategy facilitated by H1-tags located at the N-terminus of the protein as described previously (10).
- Analyze to identify recombinant proteins with biomass degrading activity.
- Analysis of enzyme diversity added to the pool of known cellulases.

**References**

1. Alistipes finegoldii AHN 2437, DSM 17242
2. Acidobacteria bacterium Ellin345
3. Actinosynnema mirum DSM 43827
4. Acidimicrobiiales bacterium PH20
5. Acidimicrobiiales bacterium PH14
6. Acidimicrobiiales bacterium PH13
7. Acidimicrobiiales bacterium PH12
8. Acidimicrobiiales bacterium PH11
9. Acidimicrobiiales bacterium PH9
10. Acidimicrobiiales bacterium PH8
11. Acidimicrobiiales bacterium PH7
12. Acidimicrobiiales bacterium PH5
13. Acidimicrobiiales bacterium PH4
14. Acidimicrobiiales bacterium PH3

**Acknowledgments**

The work presented here is supported by Dr. Hess’ Start-up Grant and Analysis of enzyme diversity added to the pool of known cellulases.

**Figures**

Fig. 1: General steps required to clone and express PCR product. A) PCR amplification using genomic DNA as a template. B) Taq Polymerase. C) Gene-specific primers. D) pET 102/D-TOPO 5700 bp. E) Transform into TOP10 E.coli cells. F) Select and analyze colonies. G) Choose a positive transformant and isolate plasmid DNA. H) Transform BL21 Star (DE3) for expression genes.

**Table 1:** Test the efficiency of our sequence model to identify novel cellulases, we downloaded all hypothetical proteins without assigned function from the Integrated Microbial Genomes (IMG) database (9). The data repository contains all finished microbial genomes, and searched for novel cellulases. We identified 62 hypothetical proteins without assigned function that possess a cellulase-specific fingerprint and which have not been identified as "cellulases" by currently available annotation algorithms.

**Table 2:** Gene-specific primers were designed for 62 newly identified cellulase candidates.