Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically-linked capacity for the degradation of lignocellulose

Adam J. Book*[1,2], Gina R. Lewin*[1,2], Bradon R. McDonald[1,2], Taichi E. Takasuka[1,3], Drew T. Doering[1,2], Aaron S. Adams[4], Joshua A. V. Blodgett[5], Jon Clardy[5], Kenneth F. Raffa[4], Brian G. Fox[1,3], and Cameron R. Currie[1,2]

SUPPLEMENTAL MATERIAL

Dataset S1 – TIGERFAM protein families used to generate phylogenetic tree in Figure 3
Dataset S2 – CAZy annotations for SACTE, SDPB6, SACTG and SPB74
Dataset S3 – LC-MS/MS of SACTE, S2a protein report, S2b peptide report
Dataset S4 – LC-MS/MS of SDPB6, S2a protein report, S2b peptide report
Dataset S5 – RNAseq DEseq ActE
Dataset S6 – RNAseq DEseq SDPB6
Dataset S7 – KEGG and CAZy DE analysis

Table S1 – RNA-Seq statistics

Strain	Growth Condition	Sample ID	Total Reads	Mapped Reads	% Mapped	% 16S rRNA	% 23S rRNA	Genome Coverage (reads per CDS)
SACTE	Glucose	AB-70	15154581	7954329	52%	0.2%	2.7%	1197
		AB-32	16291889	10289877	63%	0.2%	2.3%	1548
		AB-18	8213541	4849260	59%	1.5%	3.5%	730
	AFEX-CS	AB-39	7777610	3273638	42%	0.2%	0.9%	492
		AB-57	21935958	13832930	63%	0.1%	0.5%	2081
		AB-44	11087724	4933583	44%	0.2%	1.1%	742
SDPB6	Glucose	AB-58	26068114	10134764	39%	0.1%	0.0%	1593
		AB-35	17600584	5195048	30%	0.5%	0.2%	817
		AB-60	17661709	5757524	33%	0.2%	0.1%	905
	AFEX-CS	AB-26	11205491	3246117	29%	52.6%	10.3%	510
		AB-29	18020392	4908959	27%	0.7%	0.7%	772
		AB-27	11538238	4930499	43%	0.7%	0.7%	775

Coleoptera (beetles)
Dendroctonus ponderosae
Dendroctonus frontalis

Hymenoptera (wasps, ants, bees and sawflies)
Sirex noctilio

98
49
23
75 *Mecoptera* (scorpion flies)
51
95
100 *Trichoperta* (caddisflies)
36
100 *Lepidoperta* (moths and Butterflies)
100 *Diperta* (flies)
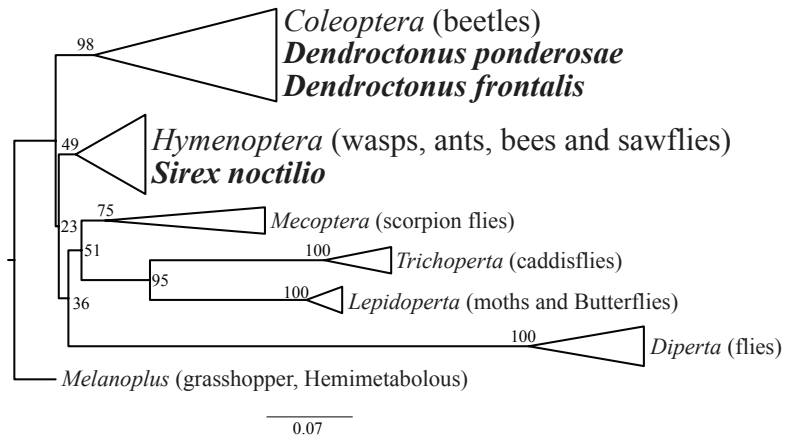
Melanoplus (grasshopper, Hemimetabolous)

0.07

FIG S1. 18S phylogenetic tree of select holometabolous insect phyla. The maximum likelihood tree was generated from an alignment containing greater than three species per phyla and the hemimetabolous outgroup. Phyla clades were collapsed for clarity. Bootstrap support for the maximum-likelihood analysis is given at each node.
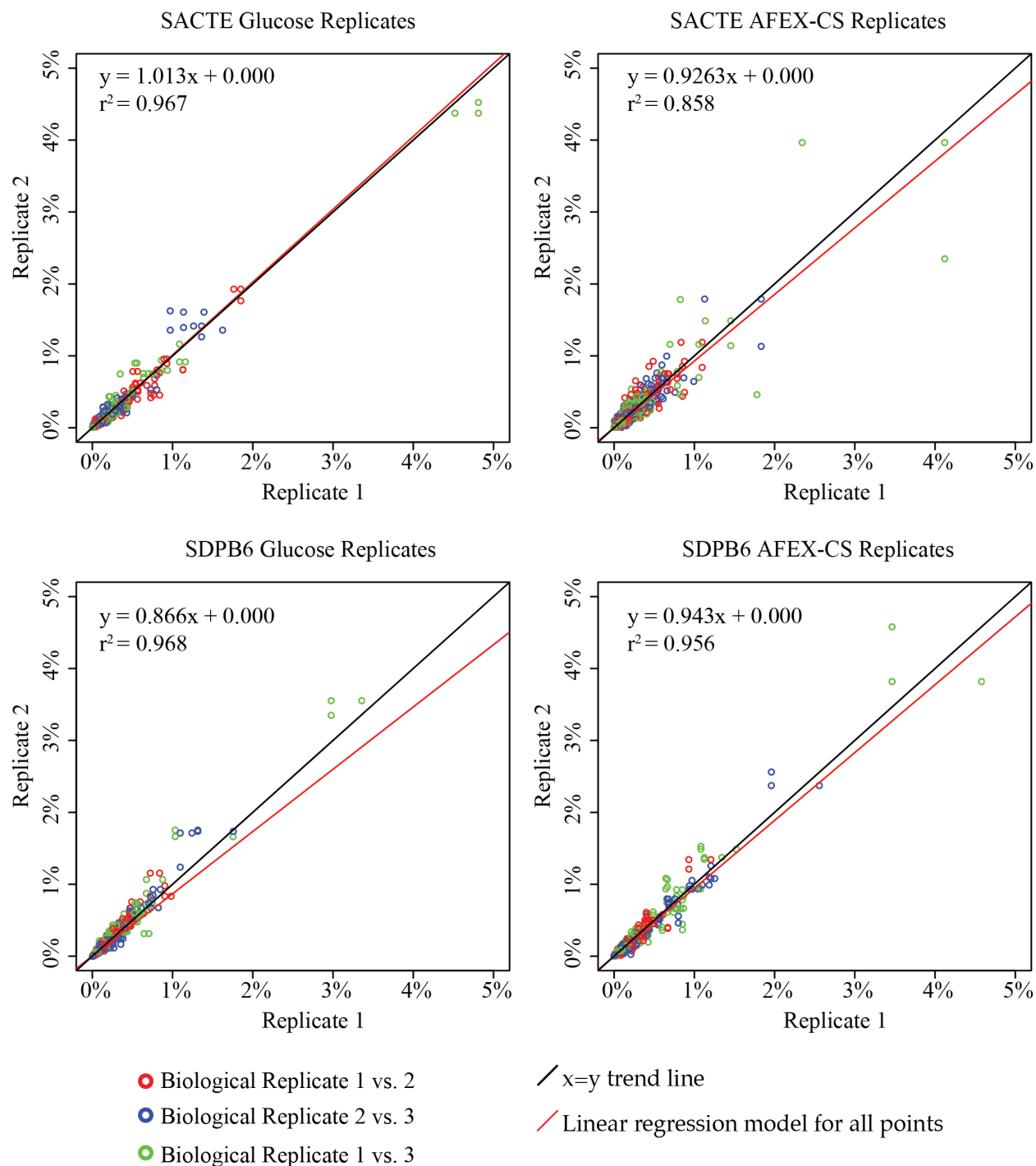
FIG S2. Pairwise comparison of normalized read counts for all protein coding genes between biological replicates. Read counts were normalized by dividing the number of reads for each gene by the total reads for that biological replicate. Red circles represent replicate 1 versus replicate 2; blue, replicate 2 versus replicate 3; green, replicate 1 versus replicate 3. Black line indicates y=x, the theoretical line for perfect match between replicates. The red line indicates the linear regression model for all points, the model equation is provided along with the r^2 value.

SACTE

Color Key
and Histogram

Count

8
6
4
2
0

0 20 40 60
Value

ACS2
ACS1
ACS3
G3
G2
G1

G1 G2 G3 ACS3 ACS1 ACS2

SDPB6

Color Key
and Histogram

Count

6
4
2
0

0 20 40 60 80
Value

ACS1_AB26
ACS2_AB27
ACS3_AB29
G1_AB58
G2_AB35
G3_AB60

G3_AB60 G2_AB35 G1_AB58 ACS3_AB29 ACS2_AB27 ACS1_AB26

FIG S3. Sample to sample variance of RNA-Seq data. Heatmap showing the Euclidean distances between samples as calculated from the variance-stabilizing transformation of the RNA-Seq count data. Red indicates strong correlation. Cluster-tree indicates relatedness of samples.