

**Genome Wide Analysis Reveals the Extrinsic Cellulolytic and Biohydrogen
Generating Abilities of *Neocallimastigomycota* Fungi**

Ayyappa Kumar Sista Kameshwar ^a, Wensheng Qin ^{a*}

^a*Department of Biology, Lakehead University, 955 Oliver Road, Thunder Bay,
Ontario, P7B 5E1, Canada.*

Wensheng Qin * Email: wqin@lakeheadu.ca, Tel: 807-343-8467

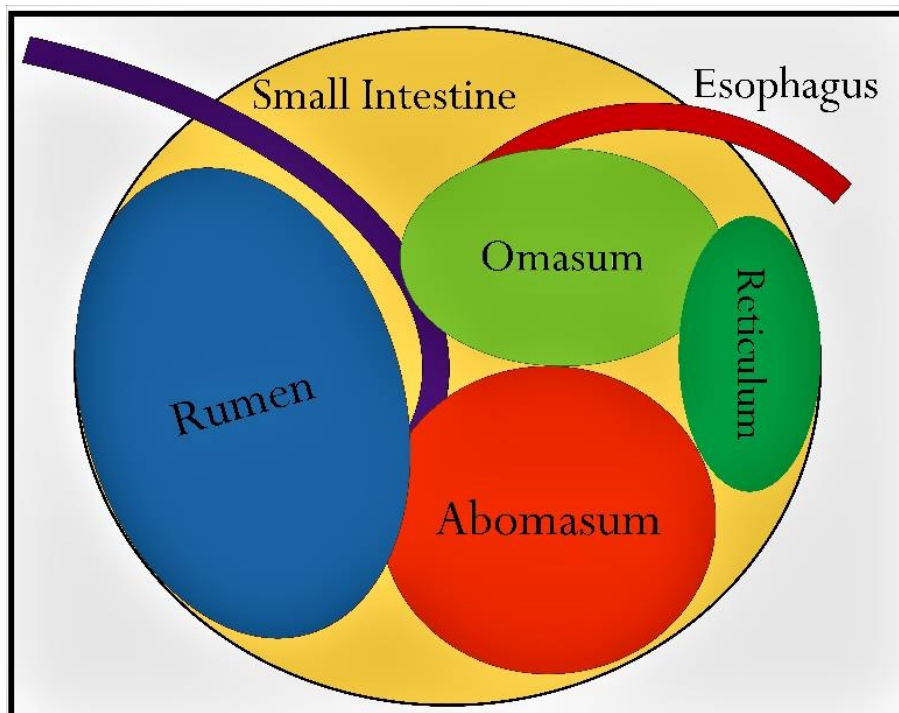


Figure S1: Pictorial representation of four chambered gut of ruminating animals a) rumen b) omasum c) abomasum d) small intestine.

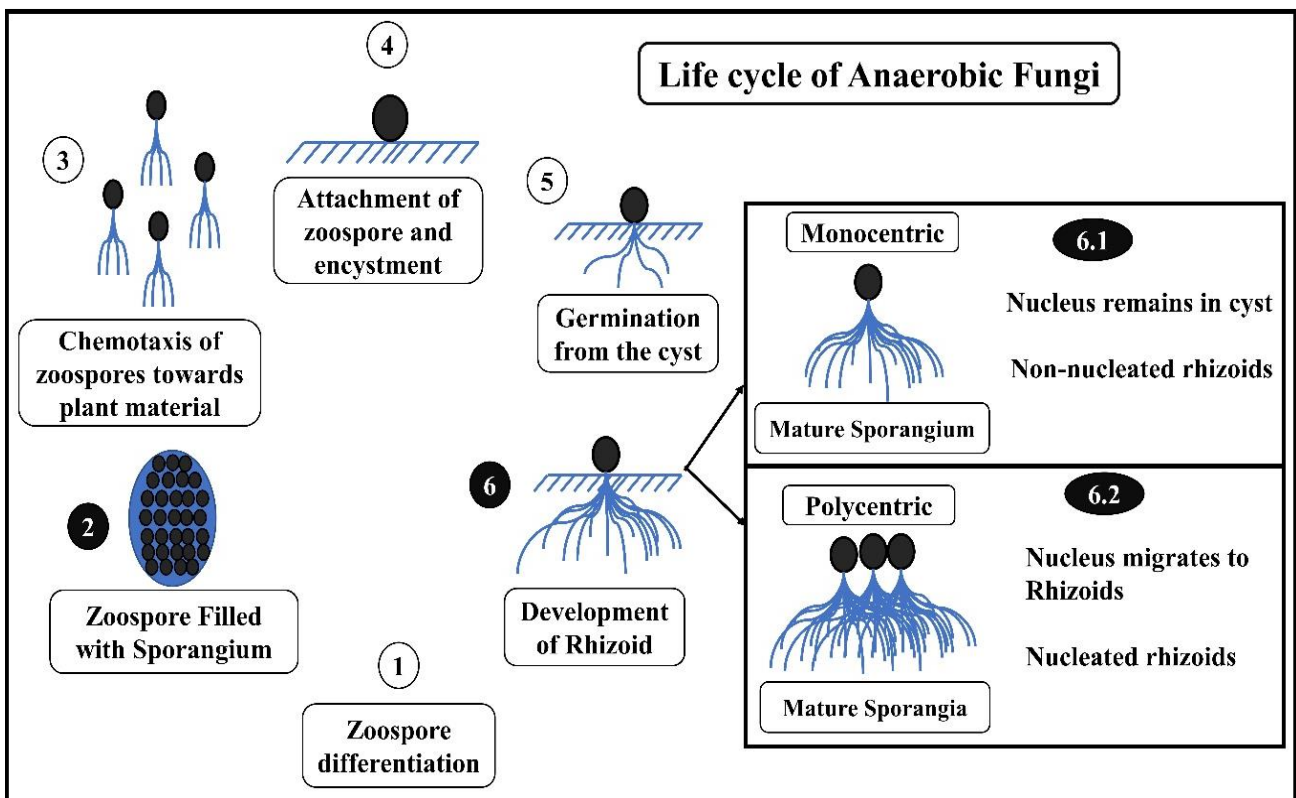


Figure S2: Different stages of anaerobic fungi life cycle from differentiation of zoospores to development of rhizoids (monocentric and polycentric based on the type of genus), lifecycle stages represented in dark (2 and 6) were the highly reported structural features.

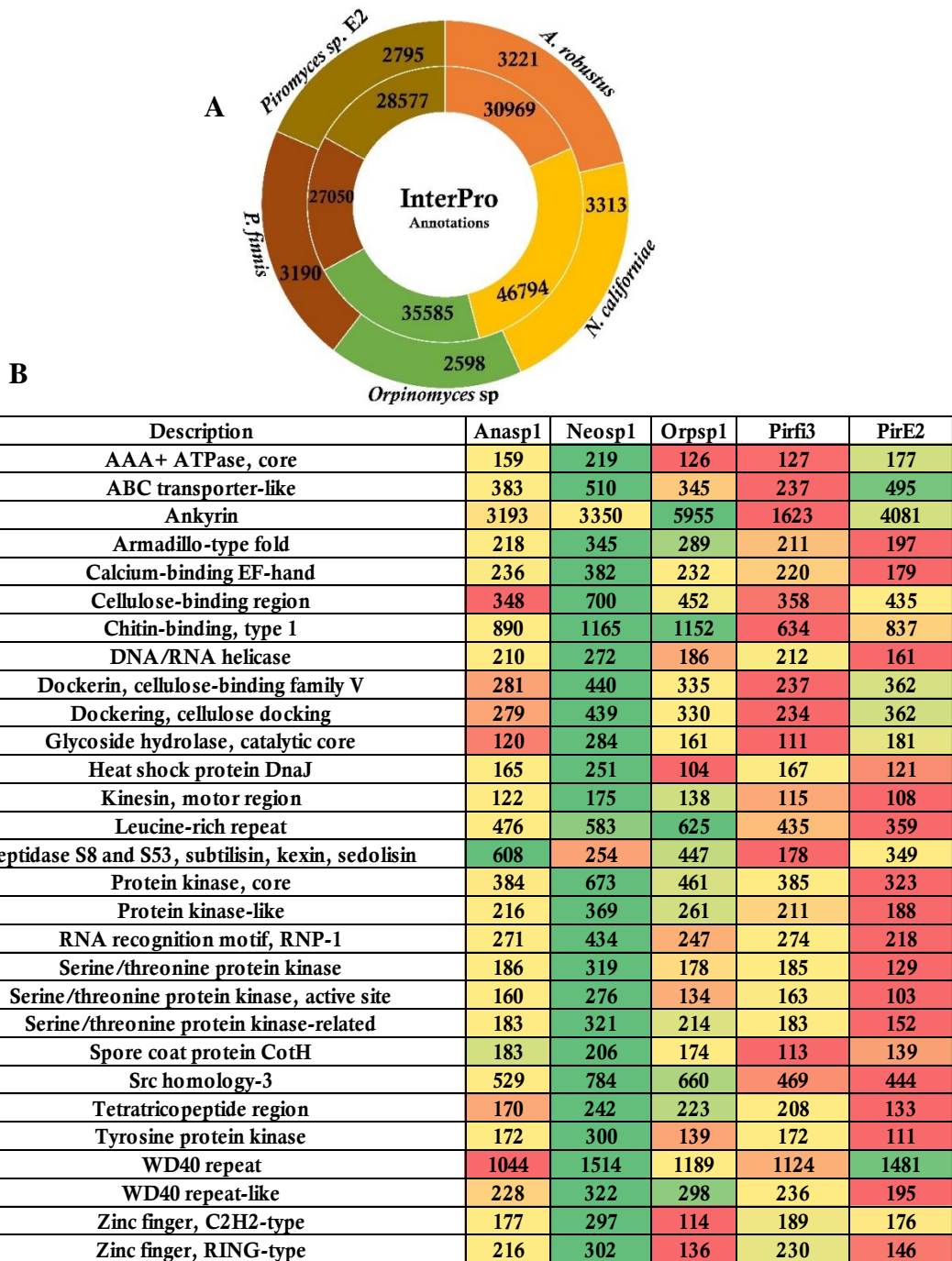


Figure S3: A) Genome wide distribution of InterPro annotations, where inner circle represents total number of InterPro predicted protein domains and outer circle represents the single copy number of InterPro predicted protein domains B) highly distributed genes among the genomes of the sequenced

Carbohydrate transport and metabolism	Anasp1	Neosp1	Orsp1	Pirf3	PirfE2
Chitinase	358	639	360	293	434
beta-1,6-N-acetylglucosaminyltransferase	36	77	65	38	61
Gluconate transport-inducing protein	32	47	60	25	41
Beta-glucosidase	7	16	10	10	16
Golgi mannosyltransferase	7	10	7	8	2
Maltase glucoamylase GH-31	6	9	23	2	5
Alpha-amylase	5	5	5	1	5
Fucosyltransferase	5	4	2	2	2
Glycosyl hydrolase, GH-47	5	6	1	2	3
Hexokinase	5	9	5	4	4
Pyrophosphate-dependent phosphofructo-1-kinase	5	9	5	4	4
Trehalose-6-phosphate synthase TPS1	5	7	8	5	7
1,4-alpha-glucan branching enzyme	4	7	6	4	5
Glycosyl transferase, GH-8	4	10	2	5	5
Endo-1,3-beta-glucanase	4	8	2	2	3
Mutarotase	4	3	2	5	5
Glyceraldehyde 3-phosphate dehydrogenase	3	7	4	3	3
Phosphoglycerate mutase	3	5	1	3	3
Sugar kinase	3	4	4	3	3
UDP-galactose transporter	3	5	1	3	4

Figure S4: Heatmap showing the distribution of genes encoding for the carbohydrate transport and metabolism

KOG ID	Anasp1	Neosp1	Orsp1	Pirfi3	PirE2
M	273	378	619	225	355
N	11	18	20	11	10
O	835	1323	1042	695	729
T	1050	1548	1246	881	1010
U	481	882	658	552	461
V	170	201	177	113	191
W	96	132	109	90	168
Y	106	229	181	124	198
Z	281	506	433	330	313
ALL (CSP)	3303	5217	4485	3021	3435
A	319	469	413	336	294
B	157	271	236	174	200
J	272	478	282	301	272
K	819	1098	900	674	773
L	225	287	324	245	225
ALL (ISP)	1792	2603	2155	1730	1764
C	119	206	103	104	103
D	353	525	480	331	323
E	236	344	295	196	227
F	70	112	76	76	74
G	578	1037	676	492	693
H	57	87	52	56	50
I	233	359	271	175	279
P	213	302	250	195	184
Q	121	105	296	45	170
ALL (M)	1980	3077	2499	1670	2103
R	1873	2601	2751	1411	1955
S	608	951	709	630	549
ALL (PC)	2481	3552	3460	2041	2504

Figure S5: Heatmap showing the highly distributed genes among the genomes of the sequenced anaerobic fungi Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orsp1 (*Orpinomyces sp.*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*)

Type of Peptidase	Anasp1	Neosp1	Orpsp1_1	PirE2_1	Pirfi3	Clan
Pepsin A	1	2	24	7	1	AA
Signal Peptidase II	0	0	0	2	0	AC
presenilin 1	2	4	1	2	2	AD
Papain	18	30	42	16	20	CA
separase	11	21	10	13	14	CD
Ulp1 peptidase	0	3	1	0	1	CE
pyroglutamyl-peptidase I	1	0	0	1	1	CF
sortase A	0	0	0	2	1	CL
dipeptidyl-peptidase VI	0	1	0	0	0	CO
dipeptidyl-peptidase VI	0	0	1	1	0	CP
Peptidase Inhibitors not assigned	12	24	6	6	5	I-
alpha-1-peptidase inhibitor	13	13	3	13	0	ID
Survivin	0	0	1	0	0	IV
Peptidase B Inhibitor	1	0	1	0	1	JC
serine carboxypeptidase Y inhibitor	1	2	1	0	1	JE
chagasin	0	0	0	1	0	JL
Oprin	0	1	0	0	0	JM
Metallopeptidase not assigned	11	8	0	8	4	M-
zincin	48	45	116	66	30	
Carboxypeptidase A1	0	1	1	2	2	MC
zinc D-Ala-D-Ala carboxypeptidase	1	0	0	0	0	MD
pitrylsin	2	1	4	2	1	ME
leucine aminopeptidase 3	3	5	5	4	3	MF
methionyl aminopeptidase 1	5	10	8	6	5	MG
aminopeptidase I	8	15	7	6	5	MH
membrane dipeptidase	1	1	3	1	1	MJ
site 2 peptidase	1	1	1	0	1	MM
RPN11 peptidase	5	7	4	5	5	MP
Ssy5 peptidase	0	8	4	1	0	PA
Ntn-hydrolase	15	31	17	15	17	
dipeptidase E	1	3	9	1	3	PC
ornithine acetyltransferase precursor	0	1	1	2	1	PE
subtilisin	43	16	42	38	17	SB
alpha/beta hydrolase	91	125	118	127	58	
D-Ala-D-Ala carboxypeptidase A	0	0	2	0	0	SE
signal peptidase I	6	8	2	2	3	SF
peptidase Clp	1	2	0	1	1	SK
nucleoporin	0	0	0	2	0	SP
rhomboid-1	10	15	6	4	5	ST
Unassigned	0	0	1	0	0	U-

Figure S6: Heatmap showing the highly distributed genes among the genomes of the sequenced anaerobic fungi Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orpsp1 (*Orpinomyces sp.*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*)

KEGG Pathway Classes in the Genomes	Anasp1	Neosp1	Orpsp1_1	Pirfi3	PirE2_1
Amino sugar and nucleotide sugar metabolism	91	156	65	76	65
Ascorbate and aldarate metabolism	16	37	13	13	12
Butanoate metabolism	35	50	31	32	26
C5-Branched dibasic acid metabolism	4	9	2	6	5
Citrate cycle (TCA cycle)	21	38	12	19	16
Fructose and mannose metabolism	41	73	38	44	32
Galactose metabolism	35	57	25	24	23
Glycolysis / Gluconeogenesis	36	68	35	32	37
Glyoxylate and dicarboxylate metabolism	11	17	9	10	10
Inositol phosphate metabolism	86	147	107	81	80
Pentose and glucuronate interconversions	38	143	27	27	25
Pentose phosphate pathway	51	92	57	62	56
Propanoate metabolism	6	10	21	5	10
Pyruvate metabolism	34	64	23	27	29
Starch and sucrose metabolism	87	139	93	88	88
Carbon fixation in photosynthetic organisms	17	38	20	18	22
Methane metabolism	4	7	6	4	3
Nitrogen metabolism	27	25	15	17	24
Oxidative phosphorylation	26	50	24	27	23
Photosynthesis	19	34	20	21	17
Photosynthesis - antenna proteins	0	1	1	1	0
Reductive carboxylate cycle (CO2 fixation)	13	27	21	11	15
Sulfur metabolism	6	11	1	4	5
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	1	1	0	2	0
1,2-Dichloroethane degradation	0	0	0	0	0
1,4-Dichlorobenzene degradation	10	17	7	8	4
1- and 2-Methylnaphthalene degradation	35	48	19	30	26
2,4-Dichlorobenzoate degradation	3	4	2	3	3
3-Chloroacrylic acid degradation	2	4	2	3	2
Atrazine degradation	1	1	0	1	0
Benzoate degradation via CoA ligation	45	61	33	33	34
Benzoate degradation via hydroxylation	2	3	0	3	1
Biphenyl degradation	1	1	0	1	0
Bisphenol A degradation	23	30	12	12	18
Caprolactam degradation	12	20	14	8	6
Carbazole degradation	0	0	0	0	0
Drug metabolism - cytochrome P450	7	15	5	5	7
Drug metabolism - other enzymes	15	32	12	17	17
Ethylbenzene degradation	34	45	19	27	24
Fluorene degradation	1	1	0	1	0
Fluorobenzoate degradation	2	1	1	0	0
Gamma-Hexachlorocyclohexane degradation	25	30	11	13	16
Geraniol degradation	3	1	8	0	2
Metabolism of xenobiotics by cytochrome P450	7	16	5	5	7
Naphthalene and anthracene degradation	30	38	18	21	25
Styrene degradation	0	0	0	0	1
Tetrachloroethene degradation	5	9	6	3	3
Trinitrotoluene degradation	6	6	4	6	4

Figure S7: Heatmap showing the highly distributed genes among the genomes of the sequenced anaerobic fungi Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orpsp1 (*Orpinomyces sp.*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*)

GH	Anasp1	Neosp1	Orpsp1_1	PirE2_1	Pirfi3
GH1	7	16	10	17	10
GH2	1	7	1	1	1
GH3	15	53	18	26	15
GH5	26	72	51	46	29
GH6	13	28	49	35	22
GH8	2	2	1	2	1
GH9	9	14	13	12	12
GH10	15	60	32	29	21
GH11	33	30	52	72	41
GH13	12	13	13	10	6
GH16	12	21	6	6	8
GH18	12	17	9	15	9
GH20	1	2	1	1	1
GH24	2	1	3	3	7
GH25	2	2	4	3	4
GH26	3	15	1	9	4
GH28	0	1	0	1	0
GH30	2	4	3	3	1
GH31	7	10	19	5	2
GH32	2	11	1	1	0
GH36	1	1	1	1	1
GH37	1	1	2	1	1
GH38	2	1	2	2	1
GH39	5	9	3	5	2
GH43	18	48	32	31	14
GH45	14	29	16	20	15
GH47	5	6	1	3	2
GH48	7	21	14	14	13
GH53	1	3	0	2	1
GH64	1	4	2	1	1
GH67	0	0	0	1	0
GH73	0	1	0	0	0
GH74	1	4	2	4	3
GH78	1	1	2	1	0
GH95	2	2	0	1	1
GH97	0	1	0	0	0
GH105	0	2	0	2	1
GH114	21	22	12	79	24
GH115	1	9	3	4	3
GH120	1	0	0	1	0
GH124	2	3	0	1	1
GH133	1	1	0	1	1

GT	Anasp1	Neosp1	Orpsp1_1	PirE2_1	Pirfi3
GT1	5	5	4	2	4
GT2	37	65	30	25	27
GT3	2	4	0	2	2
GT4	6	10	7	7	7
GT6	0	1	0	2	0
GT8	4	10	2	5	5
GT10	5	4	2	2	2
GT11	0	1	1	0	0
GT15	1	1	1	1	1
GT17	12	10	4	15	6
GT20	5	7	6	5	5
GT22	4	6	5	3	4
GT24	0	1	1	1	1
GT31	1	1	0	0	1
GT32	2	3	1	1	2
GT33	1	1	0	0	1
GT34	14	18	15	5	14
GT35	1	2	3	1	1
GT39	8	14	11	6	9
GT41	1	2	1	1	1
GT48	1	0	0	0	0
GT50	1	1	1	1	1
GT58	1	1	1	1	1
GT66	1	1	1	1	1
GT69	0	0	1	1	0
GT71	13	17	8	13	8
GT76	1	1	1	1	1
GT101	0	3	2	2	1
MM	4	6	2	4	4

Figure S8: Genome wide distribution of glycosyl hydrolases (GH) and glycosyl transferases (GT) of the sequenced anaerobic fungi Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orpsp1 (*Orpinomyces sp.*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*)

CBM	Anasp1	Neosp1	Orpsp1_1	PirE2_1	Pirfi3
CBM1	94	145	104	102	103
CBM6	6	15	7	14	7
CBM10	12	33	29	2	20
CBM12	1	2	2	3	6
CBM13	20	48	23	12	17
CBM18	447	521	500	640	288
CBM21	4	8	4	2	5
CBM22	1	1	0	0	1
CBM25	11	4	13	1	1
CBM26	20	10	26	15	1
CBM29	2	15	12	7	18
CBM32	0	5	1	2	0
CBM35	5	8	2	6	4
CBM48	8	17	5	5	9
CBM50	5	4	4	4	5
CBM52	5	11	5	3	5
CBM61	1	3	1	1	1
CBM63	6	7	3	2	3
CBM66	2	0	3	0	0

PL	Anasp1	Neosp1	Orpsp1_1	PirE2_1	Pirfi3
PL1	6	33	19	17	10
PL3	1	30	8	6	3
PL4	3	17	3	9	2
PL9	1	0	0	0	0
PL11	0	2	0	0	0

CE	Anasp1	Neosp1	Orpsp1_1	PirE2_1	Pirfi3
CE1	28	48	33	37	17
CE2	11	8	6	10	1
CE3	1	1	1	3	0
CE4	47	88	45	41	43
CE6	11	14	8	18	11
CE8	5	14	9	8	5
CE12	6	8	2	7	6
CE15	2	6	0	3	2
CE16	10	26	15	12	7

Figure S9: Genome wide distribution of carbohydrate binding domains (CBM), polysaccharide lyases (PL) and carbohydrate esterases (CE) of the sequenced anaerobic fungi Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orpsp1 (*Orpinomyces sp.*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*)

<i>Neocallimastix californiae</i>	
Alcohol dehydrogenase, class IV	14
Prohibitins and stomatins of the PID	14
Phosphoenolpyruvate carboxykinase	8
Voltage-gated shaker-like K ⁺ channel	8
Aconitase/homoaconitase	6
NADH:flavin oxidoreductase	6
NADP ⁺ -dependent malic enzyme	6
Mitochondrial carnitine-acylcarnitine carrier protein	5
NADH dehydrogenase subunits 2, 5	5
Vacuolar H ⁺ -ATPase V0 sector, subunit a	5
5'-AMP-activated protein kinase, γ subunit	4
Mitochondrial ADP/ATP carrier proteins	4
NAD-dependent malate dehydrogenase	4
NADPH oxidase	4
Predicted mitochondrial carrier protein	4
Vacuolar H ⁺ -ATPase V0 sector, subunits c/c'	4

<i>Anaeromyces robustus</i>	
Prohibitins and stomatins of the PID	8
Alcohol dehydrogenase, class IV	7
Phosphoenolpyruvate carboxykinase	7
NADH:flavin oxidoreductase	4
Vacuolar H ⁺ -ATPase V0 sector, subunit a	4
Voltage-gated shaker-like K ⁺ channel	4
5'-AMP-activated protein kinase, γ subunit	3
Aconitase/homoaconitase	3
Glycolate oxidase	3

<i>Orpinomyces sp</i>	
Vacuolar H ⁺ -ATPase V0 sector, subunit a	9
NADPH oxidase	4
Alcohol dehydrogenase, class IV	4
Aconitase/homoaconitase	4
UDP-glucuronosyl and UDP-glucosyl transferase	3
5'-AMP-activated protein kinase, γ subunit	3
NADP-dependent flavoprotein reductase	3
NADH:flavin oxidoreductase	3
Predicted mitochondrial carrier protein	3
NADP/FAD dependent oxidoreductase	3

<i>Piromyces finnis</i>	
Alcohol dehydrogenase, class IV	6
Aconitase/homoaconitase	4
Mitochondrial carnitine-acylcarnitine carrier	4
Vacuolar H ⁺ -ATPase V0 sector, subunit a	4
Glycolate oxidase	3
Phosphoenolpyruvate carboxykinase	3
NADH:flavin oxidoreductase	3

<i>Piromyces sp. E2</i>	
Alcohol dehydrogenase, class IV	9
Prohibitins and stomatins of the PID	5
NADH:flavin oxidoreductase	5
Voltage-gated shaker-like K ⁺ channel	4
Aconitase/homoaconitase	4
Phosphoenolpyruvate carboxykinase	4
Succinate dehydrogenase, flavoprotein subunit	3
Vacuolar H ⁺ -ATPase V0 sector, subunit a	3
NADP-dependent flavoprotein reductase	3

Figure S10: Genome wide classification of KOG category energy production and conversion among the sequenced anaerobic fungi Anasp1 (*Anaeromyces robustus*), Neosp1 (*Neocallimastix californiae*), Orpsp1 (*Orpinomyces sp*), PirE2 (*Piromyces sp. E2*) and Pirfi3 (*Piromyces finnis*).