Supplementary Data

Table S1: Isolation of *Fusarium oxysporum*

Test sample	Media	Organism	CFU/ml	On potato dextrose agar
7 (ESI)	Potato dextrose agar	*Fusarium*	231	
9 (ISI)			Excess growth	

Table S2: Isolation of *Aspergillus niger*

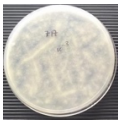
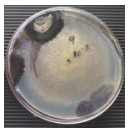
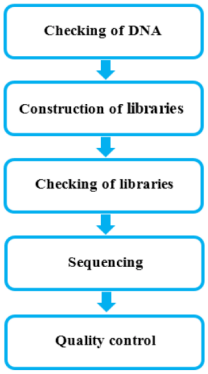
Test sample	Media	Organism	CFU/ml	On potato dextrose agar
7A (ESI)	Potato dextrose agar	*Aspergillus*	Excess growth	
9 (ISI)			Excess growth	

Table S3: DNA Sequencing Information

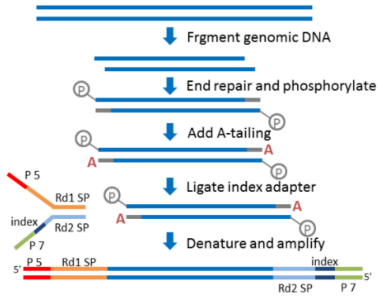
1.	Library Preparation and Sequencing
2.	Sample Quality Control
3.	Library Construction and Quality Control
4.	Sequencing Results and Instructions
5.	Data Quality Control
6.	Distribution of Sequencing
7.	Quality Distribution of Sequencing
8.	Error Rate Distribution of A/T/G/C Base

9. Results of Raw Data Filtering
10. Summary of Sequencing Data Information

A. Library Preparation and Sequencing from the DNA samples to the final data, each step, including sample test, library preparation, and sequencing, influences the quality of the data, and data quality directly impacts the analysis results. To guarantee the reliability of the data, quality control (QC) is performed at each step of the procedure. The workflow is as follows:



B. Sample Quality Control
There are three main methods of QC for DNA samples:
(1) Nanodrop: tests DNA purity (OD260/OD280).
(2) Agarose Gel Electrophoresis: tests DNA degradation and potential contamination.
(3) Qubit 2.0: quantifies the DNA concentration precisely.
2. Library Construction and Quality Control Qualified DNA is cut into fragments by restriction enzyme. The construction of the DNA libraries is through the processes of end repairing, adding A to tails, purification, PCR amplification and etc. Libraries were sequenced by Illumina high-throughput sequencer with paired-end sequencing strategy. The principle of library construction is as follows:



C. Sequencing The qualified libraries are fed into sequencers after pooling according to its effective concentration and expected data volume.

Results and Instructions
Data Quality Control
Distribution of Sequencing Quality The “e” represents the sequence error rate and Qphred represents the base quality value,Qphred=-10log10 (e). The relationship between sequencing error rate (e) and sequencing base quality value (Qphred) is as below:

Data Quality Control

Phred score	error base	right base	Q-score
10	1/10	90%	Q10
20	1/100	99%	Q20
30	1/1000	99.9%	Q30
40	1/10000	99.99%	Q40

The distribution of quality score (sample ISI):

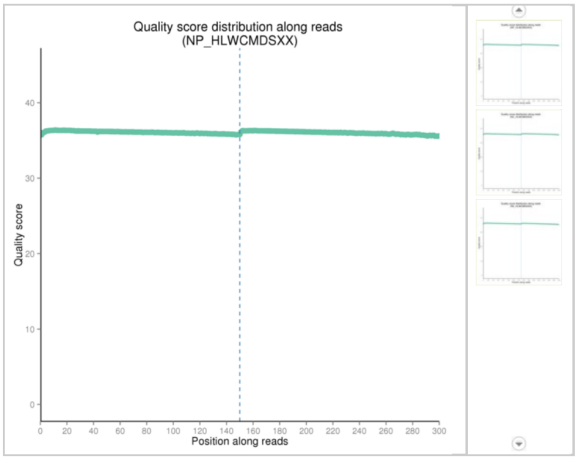


Fig.1 Distribution of Sequencing Quality

The base position is on the horizontal axis and the sequencing quality is on the vertical axis
The first half part of the distribution is for reads1 and the latter half part is for reads2

Distribution of Sequencing Error Rate for Illumina SBS technology, the distribution of sequencing error rate has two features:

(1) Error rate grows with sequenced reads extension because of the consumption of sequencing reagent.
(2) The phenomenon is common in the Illumina high-throughput sequencing platform (Erlich Y. et al. 2008; Jiang et al. 2011).
(3) The first several bases have higher sequencing error rate than others. At the beginning of sequencing, the focusing of the sequencer's fluorescence image sensor sensing element is not sensitive enough, thus, the quality of acquired fluorescence image is low. Generally, single base error rate should be lower than 1%.

The error rate of this project (sample ISI):

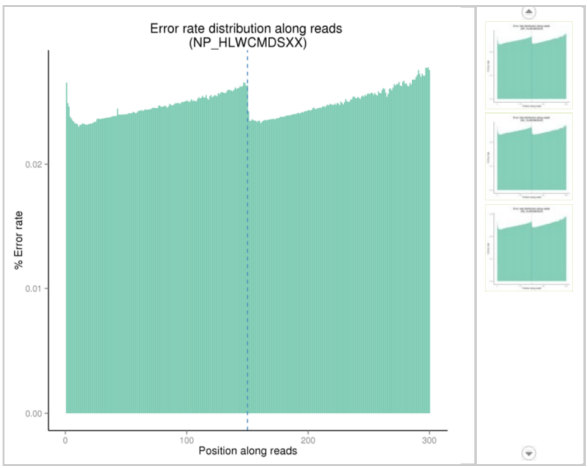


Fig.2 Error Rate Distribution

The base position is on the horizontal axis and the single base error rate is on the vertical axis

The first half part of the distribution is for reads1 and the latter half part is for reads2

Distribution of A/T/G/C Base

It is used to identify the separation situation of AT and GC by checking the distribution of GC content. According to the principle of complementary bases, the content of AT and GC should be equal at each sequencing cycle and be constant and stable in the whole sequencing procedure. But in practical measurement, due to the primer amplification bias and some other reasons, the first 6 to 7 nucleotides will fluctuate which is normal and reasonable.

The distribution of GC content (sample ISI):

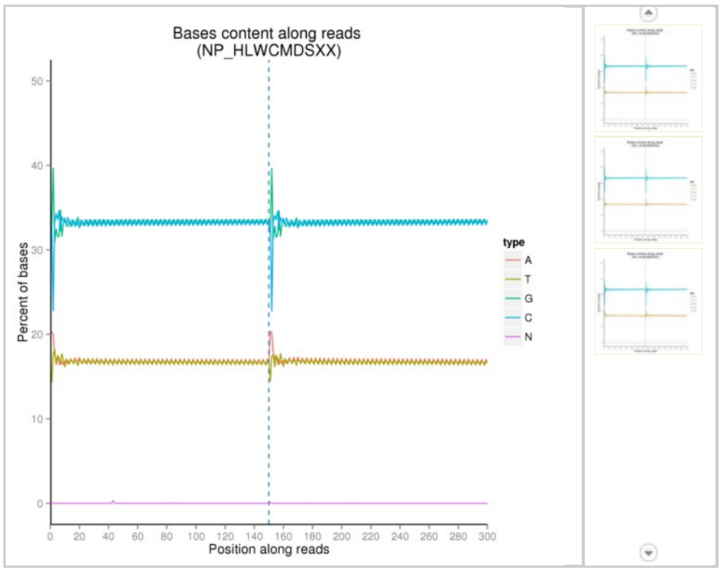


Fig.3 A/T/G/C Distribution

The base position is on the horizontal axis and the single base percentage is on the vertical axis

The first half part of the distribution is for reads1 and the latter half part is for reads2

Results of Raw Data Filtering The sequenced reads (raw reads) often contain low quality reads and adapters, which will affect the analysis quality. So it's necessary to filter the raw reads and get the clean reads. The filtering process is as follows:
(1) Remove reads containing adapters.
(2) Remove reads containing N > 10% (N represents the base cannot be determined).
(3) Remove reads containing low quality (Qscore<= 5) base which is over 50% of the total base. Adapter sequences:
5' Adapter : 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
3' Adapter (The underlined 6bp bases is Index)
5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG-3'
The Sequencing data filtration of this project can be seen in Fig.4 (Sample ISI):

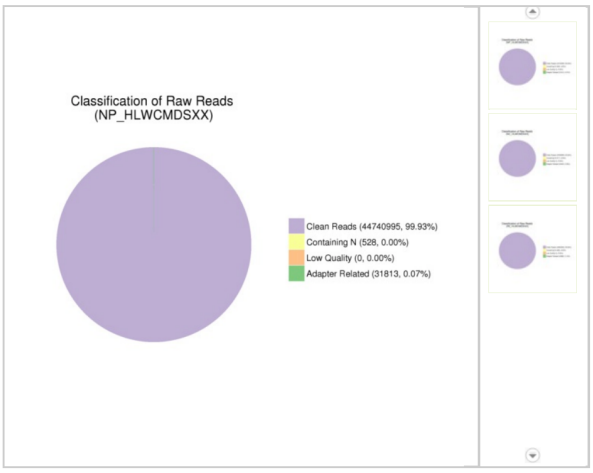


Fig.4 Composition of Raw Data

Different color for different components:

(1)Adapter related: (reads containing adapter) / (total raw reads)

(2)Containing N: (reads with more than 10% N) / (total raw reads)

(3)Low quality: (reads of low quality) / (total raw reads)

(4)Clean reads: (clean reads) / (total raw reads)

Summary of Sequencing Data Information
The total output of data on the sequencer: Raw data 40.1 G.

Table S4: Data Quality Summary (R2 – ESI, NP – ISI and NC - ASI)
The detail statistics for the quality of sequencing data.

Sample	Library	Flowcell/Lane	Raw reads	Raw data(G)	Effective(%)	Error(%)	Q20(%)	Q30(%)	GC(%)
R2	FDME190668638-1a	HLWCMDSXX_L1	44869321	13.5	99.90	0.03	97.81	93.96	66.33
NP	FDME190668639-1a	HLWCMDSXX_L1	44773336	13.4	99.93	0.03	98.00	94.41	66.43
NC	FDME190668640-1a	HLWCMDSXX_L1	44063752	13.2	99.92	0.02	98.12	94.64	66.45

Sample: sample name
Raw reads: four rows are taken as a unit to calculate the total amount of read1 and read2 in raw data files
Raw bases: (total raw reads) * (sequence length), calculating in G Error rate: base error rate
Q20, Q30: (Base count of Phred value > 20 or 30) / (Total base count) GC content: (G & C base count) / (Total base count)

Table S5: Physicochemical and Microbial Analysis of the soil samples

Sample	pH	EC	N	P	K	OC	Cl	Fe	Cu	Mn	Zn	B	Microbial Count/g	
		(µs/cm)	%			ppm							Bacterial (cfu)	Fungal (cfu)
ESI	7.73	135	0.20	0.0084	0.011	0.92	15	0.98	26.9	9.5	24.8	3.4	1968	154
ASI	6.35	139	0.191	0.010	0.011	0.93	18	0.93	29.4	9.1	30.9	4.1	2240	170
ISI	6.63	180	0.20	0.011	0.014	0.97	21	0.98	31.4	9.6	33.2	4.3	2126	154

EC-Electrical Conductivity;

Table S6: Relative Abundance of Microbial species

Species			Parent Sequence Count		Relative Frequency %		P-values		Effect size
	ESI	ISI	ESI	ISI	ESI	ISI	PVal	corrected	
Achromobacter sp. 2789STDY5608625	299	8	2975	588	10.05	1.36	5.90E-16	3.00E-12	8.69
Achromobacter sp. K91	236	4	2975	588	7.93	0.68	9.17E-15	3.62E-11	7.25
Achromobacter aegrifaciens	233	5	2975	588	7.83	0.85	1.33E-13	3.94E-10	6.98
Microbacterium sp. SUBG005	196	6	19463	13719	1.01	0.04	1.22E-37	2.17E-33	0.96
Agrobacterium larrymoorei	174	14	691	741	25.18	1.89	6.24E-44	2.22E-39	23.29
Curtobacterium sp. MR_MD2014	153	6	2534	695	6.04	0.86	8.66E-11	1.62E-07	5.17
Pseudomonas sp. T	140	7	4318	3374	3.24	0.21	2.75E-27	1.63E-23	3.03
Moraxella osloensis	132	9	146	26	90.41	34.62	2.90E-09	4.69E-06	55.80
Rheinheimera sp.	113	1	329	54	34.35	1.85	2.55E-08	3.63E-05	32.49
Chryseobacterium arthrosphaerae	111	2	689	230	16.11	0.87	6.00E-13	1.42E-09	15.24
Aspergillus arachidicola	55	16	5460	30850	1.01	0.05	1.12E-31	1.33E-27	0.96
Aspergillus candidus	50	16	5460	30850	0.92	0.05	4.06E-28	2.89E-24	0.86
Aspergillus campestris	48	11	5460	30850	0.88	0.04	1.32E-29	1.17E-25	0.84
ISI vs ASI	ISI	ASI	ISI	ASI	ISI	ASI	PVal	Corrected	ES
Streptomyces sp. FxanaC1	89	18	268202	200210	0.03	0.01	9.95E-09	5.89E-05	0.02
Streptomyces sp. F12	63	9	268202	200210	0.02	0.00	2.74E-08	0.00012189	0.02
Rhizobium sp. NFACC06-2	27	2	6383	5120	0.42	0.04	1.07E-05	0.02533065	0.38
Nocardia sp. Root136	26	17	9801	27406	0.27	0.06	2.54E-06	0.00645715	0.20
Variovorax sp. 369	24	9	4257	7685	0.56	0.12	1.56E-05	0.03264587	0.45
Aspergillus nomius	24	18	30850	4648	0.08	0.39	1.88E-06	0.00514648	-0.31

Aspergillus ochraceoroseus	20	15	30850	4648	0.06	0.32	1.35E-05	0.03006614	-0.26
Mycobacterium sp. DL90	18	51	43318	36512	0.04	0.14	1.87E-06	0.00553966	-0.10
Pseudomonas sp. LFM046	17	160	3374	5016	0.50	3.19	4.00E-20	1.42E-15	-2.69
Streptomyces sp. SN-593	16	52	268202	200210	0.01	0.03	1.60E-08	8.13E-05	-0.02

Table S7: Pathway predictions

	ASI	ISI	ESI	Name	Pathway
Unmapped	66201187	63530411	67052468		
Unclassified	8120933	9418508	7584133	Unclassified	Unclassified
K03088	103582	115430	90207	RNA polymerase sigma-70 factor, ECF subfamily	Brite Hierarchies; Protein families: genetic information processing; Transcription machinery
K12132	101294	107393	91587	eukaryotic-like serine/threonine-protein kinase [EC:2.7.11.1]	Brite Hierarchies; Protein families: metabolism; Protein kinases
K01990	72289	84001	73808	ABC-2 type transport system ATP-binding protein	Brite Hierarchies; Protein families: signaling and cellular processes; Transporters
K00059	62612	75406	69969	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	Metabolism; Lipid metabolism; Fatty acid biosynthesis ǀ Metabolism; Metabolism of cofactors and vitamins; Biotin metabolism ǀ Metabolism; Biosynthesis of other secondary metabolites; Prodigiosin biosynthesis ǀ Brite Hierarchies; Protein families: metabolism; Lipid biosynthesis proteins
K06994	60365	63022	61404	putative drug exporter of the RND superfamily	Not Included in Pathway or Brite; Poorly characterized; General function prediction only
K01992	59730	68818	59492	ABC-2 type transport system permease protein	Brite Hierarchies; Protein families: signaling and cellular processes; Transporters
K00249	58830	76269	74103	acyl-CoA dehydrogenase [EC:1.3.8.7]	Metabolism; Carbohydrate metabolism; Propanoate metabolism ǀ Metabolism; Lipid metabolism; Fatty acid degradation ǀ Metabolism; Amino acid metabolism; Valine, leucine and isoleucine degradation ǀ Metabolism; Metabolism of other amino acids; beta-Alanine metabolism ǀ Organismal Systems; Endocrine system; PPAR signaling pathway
K02035	53373	80130	60833	peptide/nickel transport system	Cellular Processes; Cellular community - prokaryotes; Quorum sensing ǀ Brite Hierarchies; Protein

	substrate-binding protein	families: signaling and cellular processes; Transporters

Table S8: PFAM Abundance Overview

	ASI	ISI	ESI
Unmapped	66201187	63530411	67052468
Unclassified	8836439	10084893	8234898
PF00005 [ABC transporter]	253554	318934	273226
PF07690 [Major Facilitator Superfamily]	195590	263909	255603
PF00528 [Binding-protein-dependent transport system inner membrane component]	159904	226135	201665
PF00501 [AMP-binding enzyme]	114937	147778	135954
PF00171 [Aldehyde dehydrogenase family]	104552	123499	120812
PF00072 [Response regulator receiver domain]	82514	90451	80950
PF00106 [short chain dehydrogenase]	64910	80596	74544

Table S9: Variant Analysis

Fusarium oxysporum
Variants rate details

Chromosome	Length	Variants	Variants rate
NC_030986.1	6,854,980	12	571,248
NC_030987.1	5,577,357	26	214,513
NC_030989.1	5,212,762	7	744,680
NC_030990.1	4,914,260	14	351,018
NC_030992.1	4,347,182	7	621,026
NC_030993.1	3,984,410	7	569,201
NC_030994.1	3,304,701	4	826,175
NC_030995.1	2,896,840	2	1,448,420
NC_030996.1	2,337,134	2	1,168,567
Total	**39,429,626**	**81**	**486,785**

Table S10: Number variants by type

Type	Total
SNP	81
MNP	0
INS	0
DEL	0
MIXED	0
INV	0
DUP	0
BND	0
INTERVAL	0
Total	**81**

Number of effects by impact

Type (alphabetical order)	Count	Percent
LOW	125	21.295%
MODERATE	26	4.429%
MODIFIER	436	74.276%

Number of effects by functional class

Type (alphabetical order)	Count	Percent
MISSENSE	26	17.219%
SILENT	125	82.781%

Missense / Silent ratio: 0.208

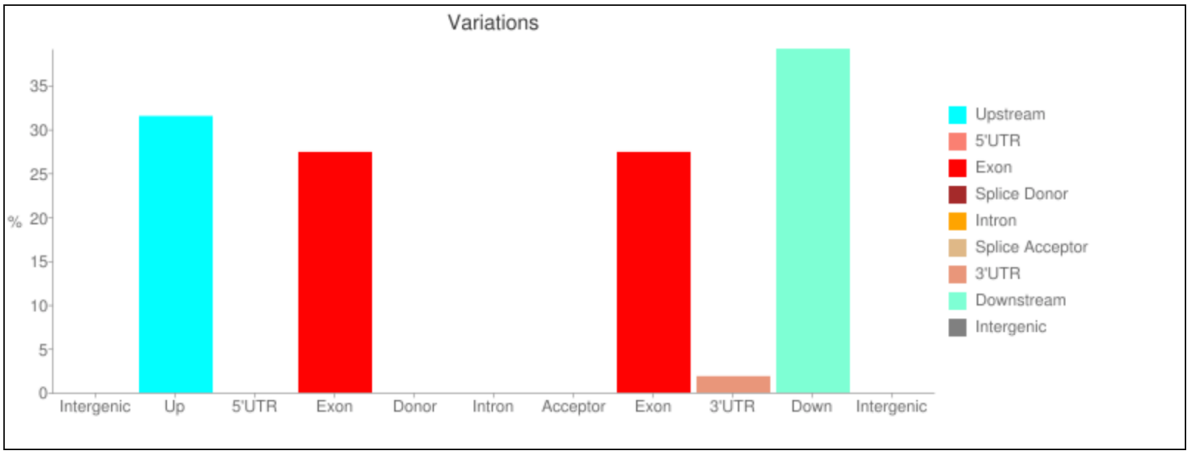
Number of effects by type and region

Type

Type (alphabetical order)	Count	Percent
DOWNSTREAM	230	39.182 %
EXON	10	1.704%
NON_SYNONYMOUS_CODING	26	4.429%
SYNONYMOUS_CODING	125	21.295 %
UPSTREAM	185	31.516 %
UTR_3_PRIME	11	1.874%

Region

Type (alphabetical order)	Count	Percent
DOWNSTREAM	230	39.182 %
EXON	161	27.428 %
UPSTREAM	185	31.516 %
UTR_3_PRIME	11	1.874%



Base changes (SNPs)

	A	C	G	T
A	0	7	12	1

C	0	0	6	7
G	7	10	0	3
T	4	20	4	0

Ts/Tv (transitions / transversions)

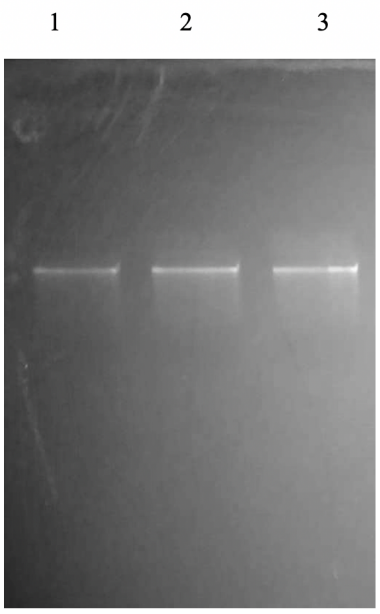
Note: Only SNPs are used for this statistic.
Note: This Ts/Tv ratio is a 'raw' ratio (ratio of observed events).

Transitions	90
Transversions	70
Ts/Tv ratio	1.2857

All variants:

Sample, Total
Transitions ,90,90
Transversions ,70,70
Ts/Tv ,1.286,1.286

Supplementary figures
Figure S1.DNA Extraction



Genomic DNA loaded on 1% Agarose gel

Lane Description:
Lane 1: gDNA from ISI sample (2ul loaded)
Lane 1: gDNA from ASI sample (2ul loaded)
Lane 1: gDNA from ESI sample (2ul loaded)

Figure S2: Rarefaction curves from experimental data sets ESI, ISI and ASI

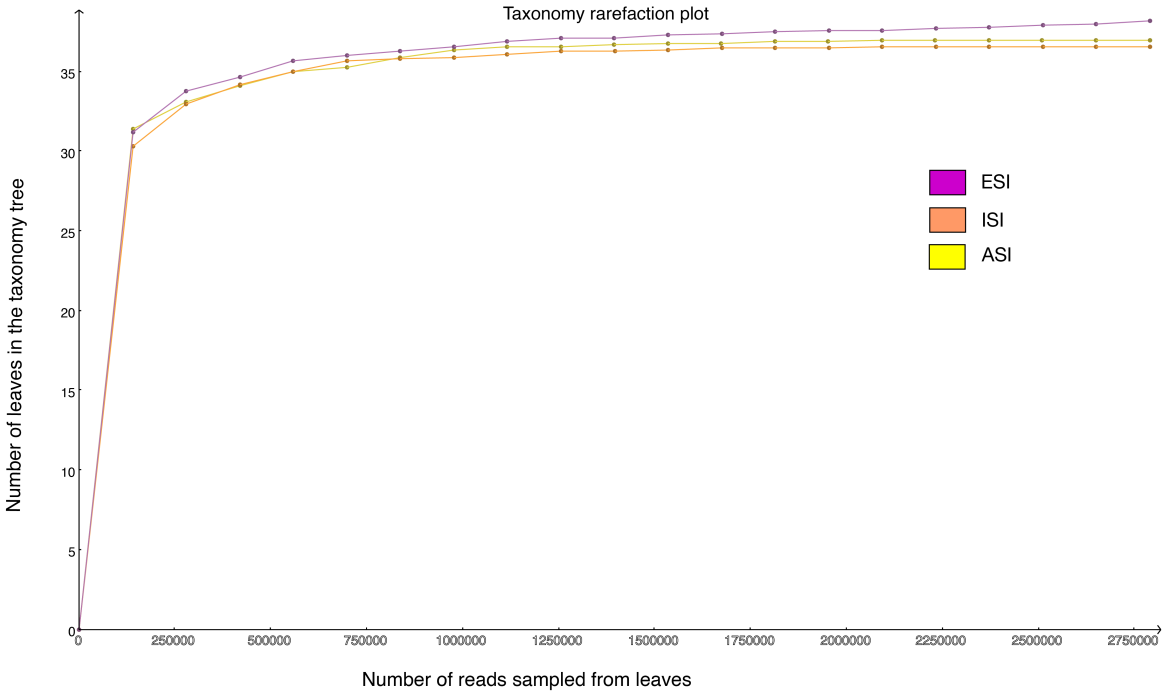


Figure S3: Pathway predictions

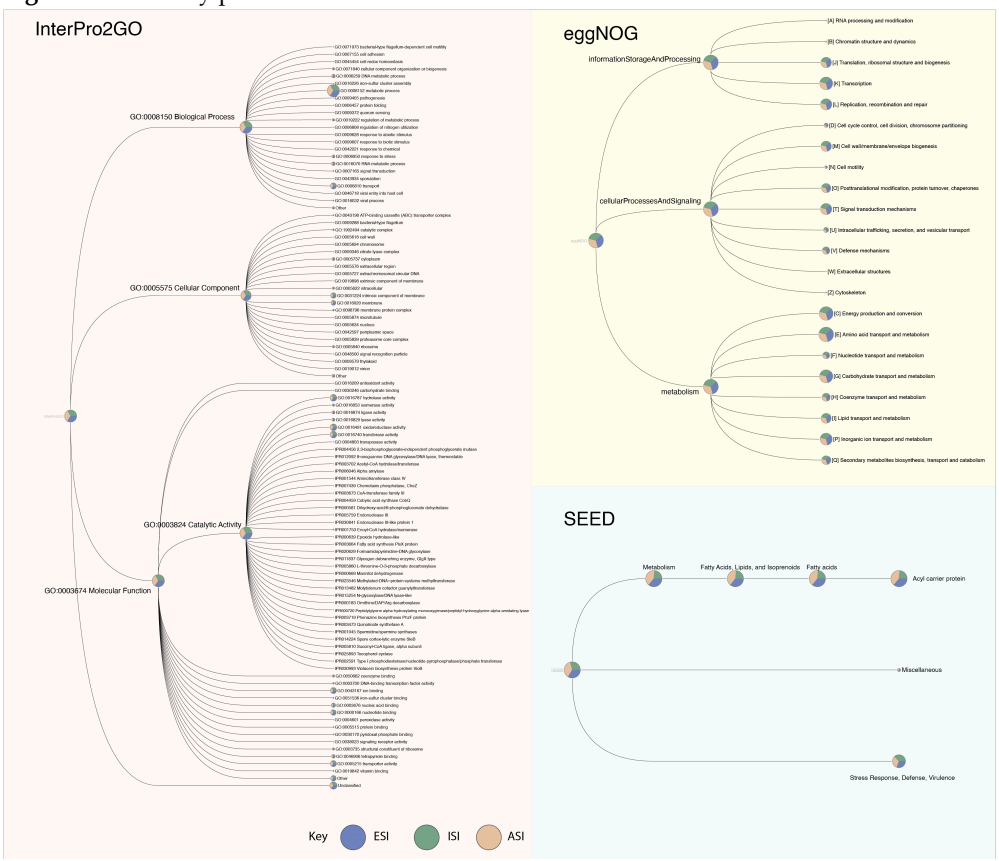


Figure S4: KEGG Annotations

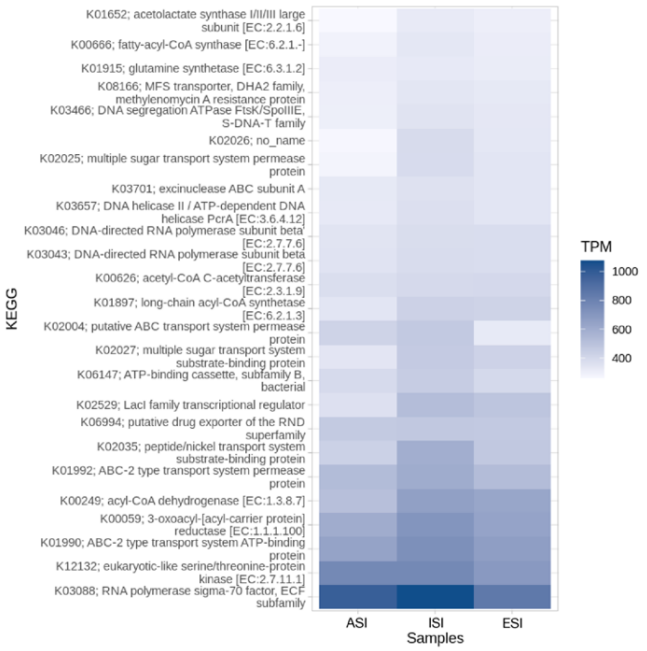


Figure S5: COG Annotations

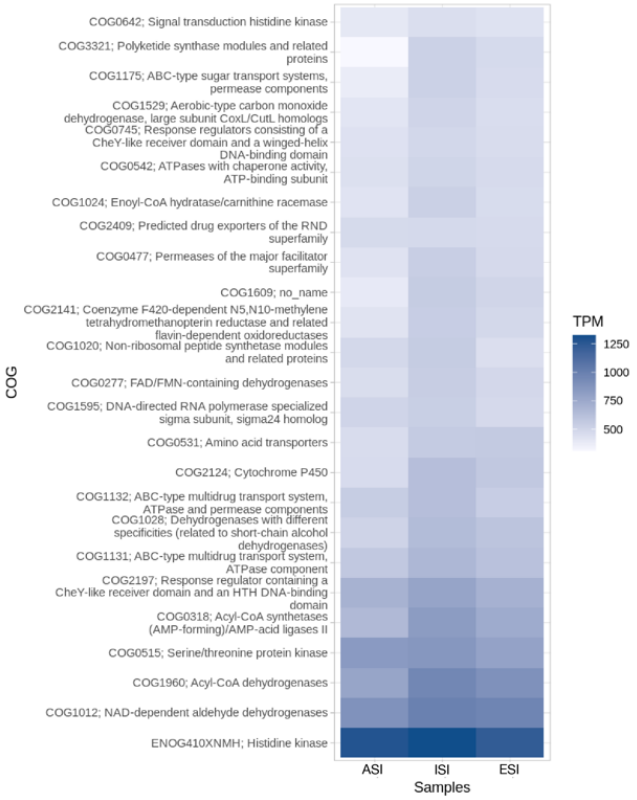


Figure S6: PFAM Annotations