



Supplementary Material

De novo Assembly and Annotation of the Whole Transcriptome of *Penaeus penicillatus*

Binbin Shan, Yan Liu, Changping Yang, Shengnan Liu and Dianrong Sun*

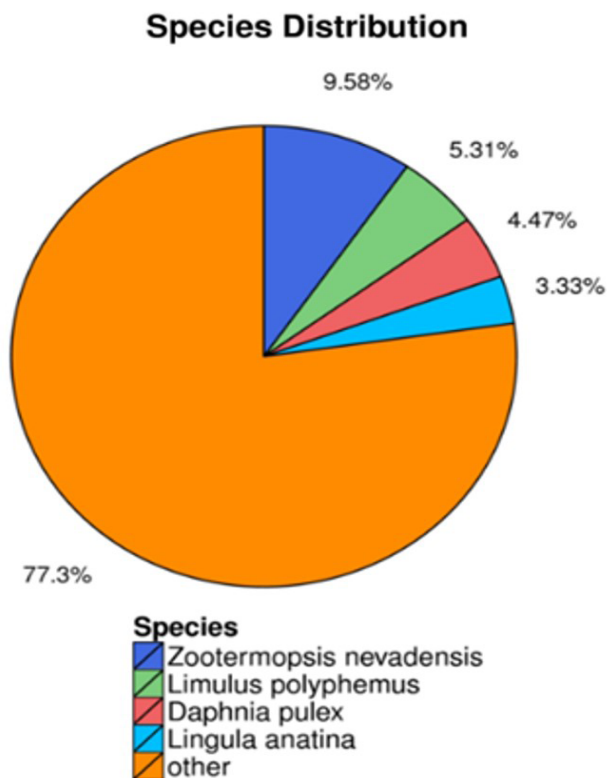
Key Laboratory of South China Sea Fishery Resources Exploitation and Utilization,
Ministry of Agriculture, Guangzhou 510000, P.R. China

* Corresponding author: drsun73@163.com
0030-9923/2018/0006-2273 \$ 9.00/0
Copyright 2018 Zoological Society of Pakistan

Supplementary Table I.- The top 30 statistically significant KEGG classifications in *P. penicillatus* transcriptome.

Pathway	Pathway ID	Unigenes in pathway	Level 1	Level 2
1 Metabolic pathways	ko01100	1302 (11.96%)	Metabolism	Global and overview maps
2 PI3K-Akt signaling pathway	ko04810	437 (4.01%)	Environmental Information Processing	Signal transduction
3 Phospholipase D signaling pathway	ko05146	429 (3.94%)	Environmental Information Processing	Signal transduction
4 Neuroactive ligand-receptor interaction	ko03013	418 (3.84%)	Environmental Information Processing	Signaling molecules and interaction
5 Glutamatergic synapse	ko05200	413 (3.79%)	Organismal Systems	Nervous system
6 RNA transport	ko05110	353 (3.24%)	Genetic Information Processing	Translation
7 Spliceosome	ko04510	349 (3.21%)	Genetic Information Processing	Transcription
8 Amoebiasis	ko03040	347 (3.19%)	Human Diseases	Infectious diseases: Parasitic
9 Vibrio cholerae infection	ko00230	300 (2.76%)	Human Diseases	Infectious diseases: Bacterial
10 Pathways in cancer	ko03015	297 (2.73%)	Human Diseases	Cancers: Overview
11 Endocytosis	ko05169	290 (2.66%)	Cellular Processes	Transport and catabolism
12 Regulation of actin cytoskeleton	ko05166	255 (2.34%)	Cellular Processes	Cell motility
13 Epstein-Barr virus infection	ko04144	232 (2.13%)	Human Diseases	Infectious diseases: Viral
14 cGMP-PKG signaling pathway	ko05016	232 (2.13%)	Environmental Information Processing	Signal transduction
15 Huntington's disease	ko04270	230 (2.11%)	Human Diseases	Neurodegenerative diseases

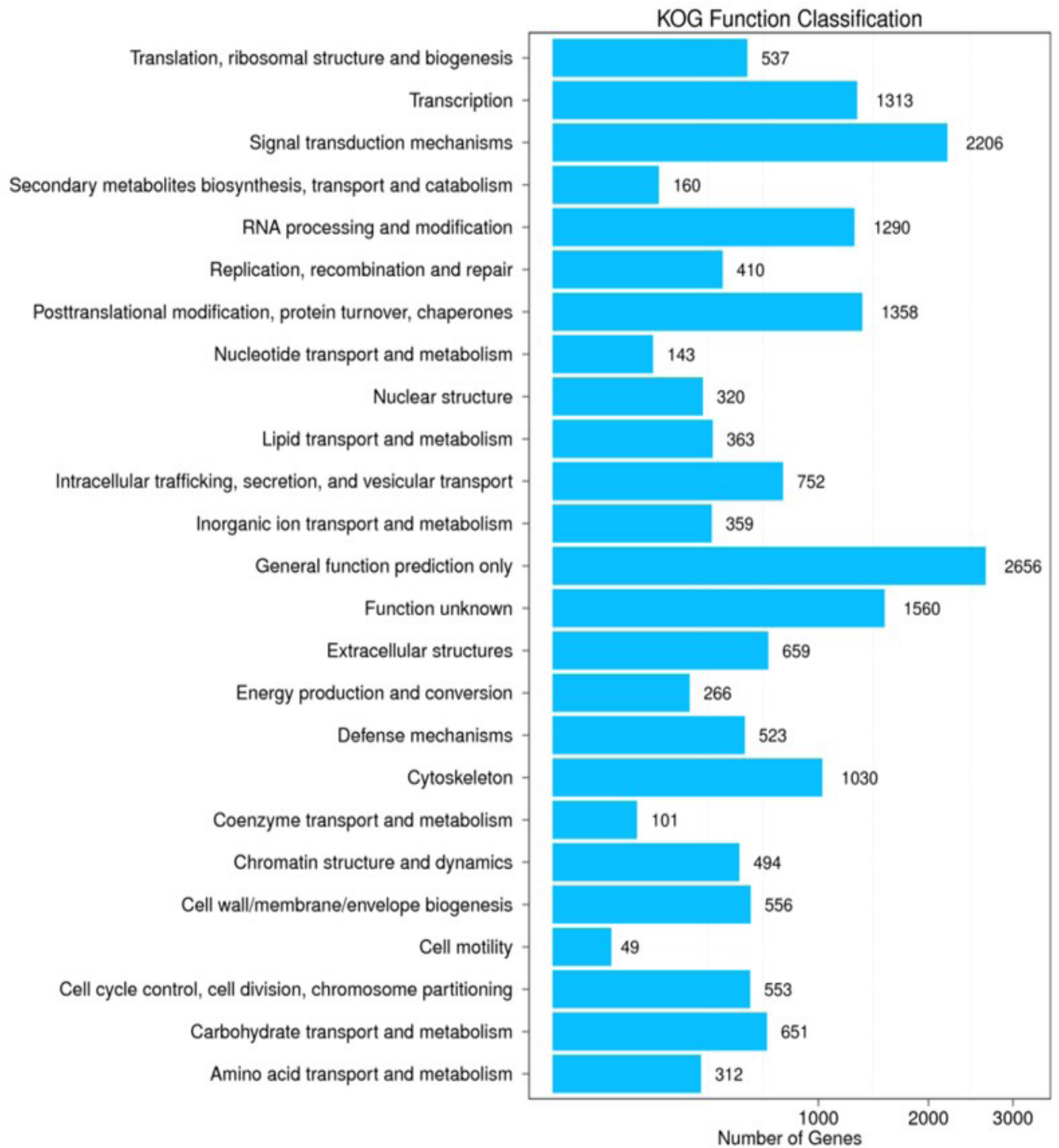
Pathway	Pathway ID	Unigenes in pathway	Level 1	Level 2
16 Focal adhesion	ko05132	229 (2.1%)	Cellular Processes	Cellular community
17 mRNA surveillance pathway	ko04530	227 (2.09%)	Genetic Information Processing	Translation
18 Protein processing in endoplasmic reticulum	ko04010	226 (2.08%)	Genetic Information Processing	Folding, sorting and degradation
19 cAMP signaling pathway	ko05168	219 (2.01%)	Environmental Information Processing	Signal transduction
20 Thyroid hormone signaling pathway	ko00240	215 (1.97%)	Organismal Systems	Endocrine system
21 Purine metabolism	ko05202	208 (1.91%)	Metabolism	Nucleotide metabolism
22 Alzheimer's disease	ko04520	193 (1.77%)	Human Diseases	Neurodegenerative diseases
23 Herpes simplex infection	ko00310	186 (1.71%)	Human Diseases	Infectious diseases: Viral
24 Transcriptional misregulation in cancer	ko05414	185 (1.7%)	Human Diseases	Cancers: Overview
25 Proteoglycans in cancer	ko05410	185 (1.7%)	Human Diseases	Cancers: Overview
26 Pyrimidine metabolism	ko04145	176 (1.62%)	Metabolism	Nucleotide metabolism
27 Dilated cardiomyopathy	ko05130	172 (1.58%)	Human Diseases	Cardiovascular diseases
28 MAPK signaling pathway	ko04666	172 (1.58%)	Environmental Information Processing	Signal transduction
29 MicroRNAs in cancer	ko04310	171 (1.57%)	Human Diseases	Cancers: Overview
30 Tight junction	ko05164	170 (1.56%)	Cellular Processes	Cellular community



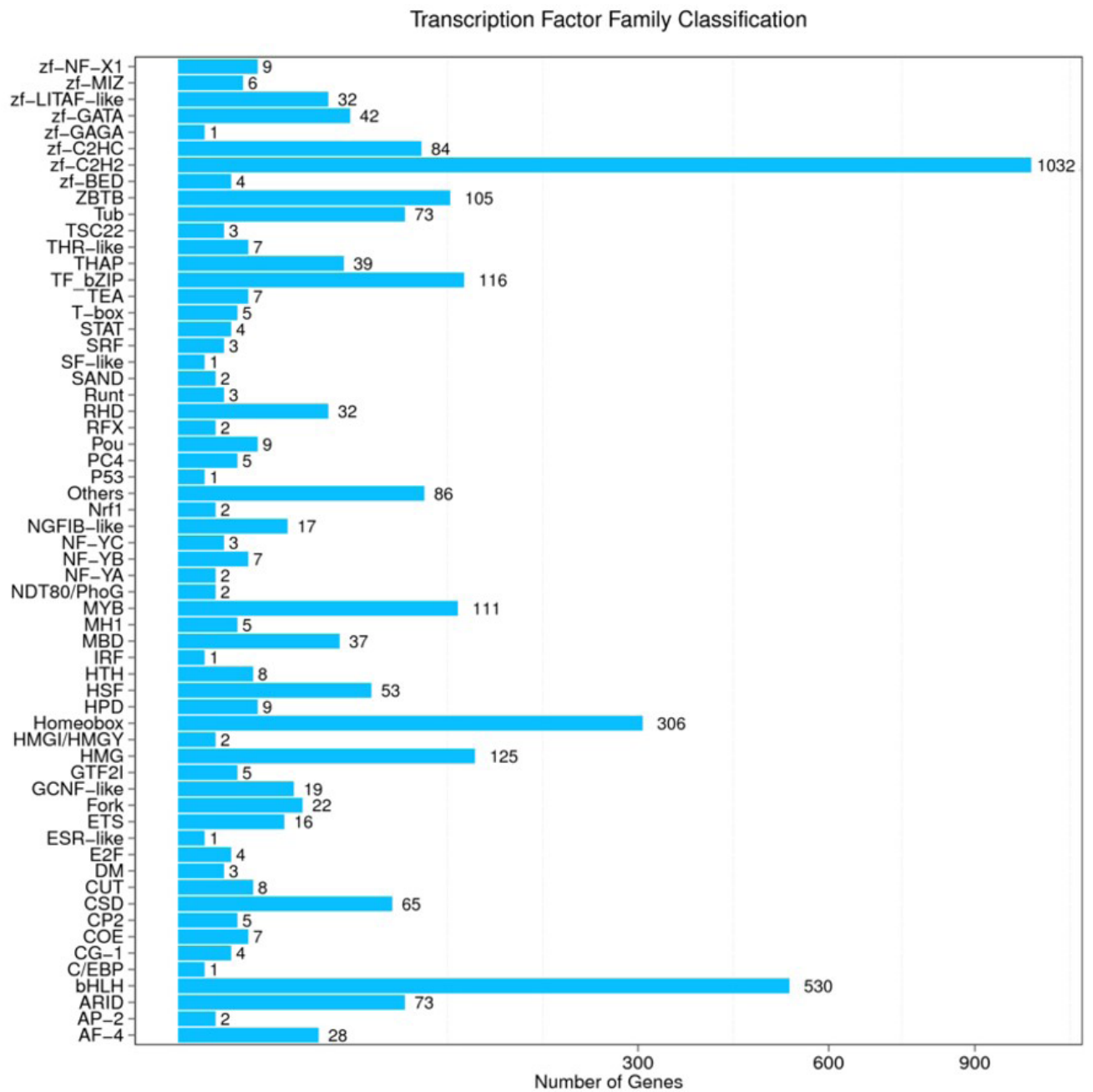
Supplementary Table II.- Summary of the EST-SSRs in *P. penicillatus* unigenes.

	Total number
Total number of sequences examined	27,850
Total size of examined sequences (bp)	25,571,120
Total number of identified SSRs	9222
Number of SSR containing sequences	6337
Number of sequences containing more than one SSR	1925
Mononucleotide	2219
Dinucleotide	4034
Trinucleotide	2194
Tetranucleotide	376
Pentanucleotide	114
Hexanucleotide	285

Supplementary Fig. S1. Species distribution of Blast top-hits against the NCBI Nr database.



Supplementary Fig. S2. Cluster of eukaryotic groups (KOG) classification of putative proteins



Supplementary Fig. S3. The proportion of *P. penicillatus* transcripts belonging to different major TF families.