Global Conference on Aquaculture Millennium +20
Registration number: 10961237

Growth-related SNP development in Penaeus vannamei by the next-generation sequencing and DNA pools sequencing

Yongyu Huang[1], Lili Zhang[1], Shiyu Huang[1], Guodong Wang[1,*], Zhangwu Yang[2,*]
1 Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture and Rural Affairs, Fisheries College of Jimei University, Xiamen 361021, China
2 Fisheries Research Institute of Fujian, 7 Shanhai Road, Huli, Xiamen 361000, China
* Correspondence

Highlights

- Hypothetical SNPs were called from the transcriptome of RG and SG in *P. vannamei*.
- P-value and AFI were used as the main index to screen out growth-related SNPs.
- 104/216,015 high-quality SNPs were gained using p-value<4.97e-7 and AFI>4 or <0.25.
- Verified by DNA pools-seq, the positive rate of high-quality SNPs was up to 72.22%.

Introduction

- There is little transcriptome data about the growth of *P. vannamei* and fewer reports about growth-related SNPs.
- A precise screening procedure is needed to maximize accuracy and prevent false-positive SNP detection.
- Next-generation resequencing of DNA pools is an efficient method for the identification of SNPs.
- In this work, SNPs were detected using transcriptome data, to identify high-quality SNPs that may be related to growth performance and validate these SNPs by next-generation resequencing of DNA pools.

Materials

- Sixteen shrimp strains with different genetic backgrounds (produced by four strains interbred each other, Figure 2) were used in the present study.
- After 120 days of breeding, the five heaviest and lightest individuals were collected from each cage in the first pond, with eighty shrimps in total, as sample RG1 and SG1. Their eyestalks, hepatopancreas, and intestinal tract tissues were collected to extract total RNA, respectively.
- Similarly, in the second and third ponds, samples RG2, SG2, and RG3, SG3 were obtained.

Figure. 1 Rapid-growing (RG) and slow-growing (SG) shrimps

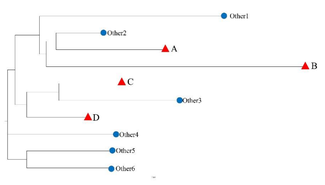

Figure. 2 The genetic background of shrimps. The four different strains (A, B, C, and D) interbred each other to produce the 16 crossed families.

Methods

- Using the software GATK to call SNPs by default parameters.
- Used two-tailed Fisher's exact test to determine the significance of allele frequency difference of each SNP between RG and SG.
- AFI (allele frequency imbalances, the ratio between the allele frequencies of the RG and that of the SG) were defined and computed.
- A total of 35 target SNPs were selected from 100,633 putative SNPs for further validation
- 240 RG and 240 SG shrimps were adopted to extract DNA respectively, and make two DNA pools.

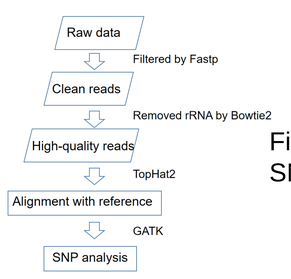

Figure. 3 Flow chart of SNP analysis

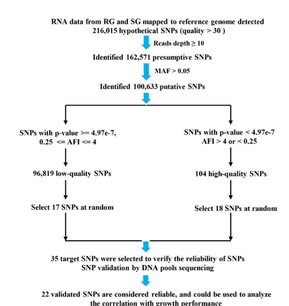

Figure. 4 Workflow used for identifying SNPs in the *P. vannamei* transcriptome

Results

Table1. The information of 35 target SNPs

Gene id	Position	Ref	Alt	Read depth	Pvalue	AFI	Pool-seq	Name
C7M84_022026	27796	G	A	287.5	0.0028	0.84	A	G27796A
C7M84_023984	36958	C	T	326	0.0279	1.09	T	C36958T
LOC113805038	349955	G	C	16.5	0.0403	0.75	NA	
C7M84_004438	204734	G	A	138	0.0276	0.85	A	G204734A
C7M84_005801	133937	T	C	2182	0.0008	1.04	C	T133937C
C7M84_007144	137866	A	G	1134	0.0371	1.11	G	A137866G
C7M84_009716	107815	T	C	659.5	0.0112	1.12	NA	
C7M84_017766	148891	C	T	106	0.0070	0.33	NA	
C7M84_021354	328130	A	G	53.5	0.0065	3.19	G	A328130G
C7M84_021883	50172	T	C	32	0.0391	0.26	NA	
LOC113828755	914967	C	T	314	1.17E-06	0.31	NA	
C7M84_022385	231745	G	A	139	0.0099	2.74	NA	
C7M84_022682	221463	T	A	40.5	0.0095	2.26	A	T221463A
C7M84_023278	233420	G	A	5404	0.0062	0.86	NA	
C7M84_023424	1031552	A	T	377.5	0.0330	1.13	T	A1031552T
C7M84_020628	42707	C	T	146	0.2135	1.33	NA	
C7M84_024169	1182178	T	A	241	0.0003	0.70	A	T1182178A
After filtering								
C7M84_013676	55042	G	A	118.5	2.93E-07	0.20	NA	
C7M84_022166	785995	C	T	172.5	2.14E-22	0.10	NA	
C7M84_025140	59957	T	C	1659.5	4.37E-71	0.22	C	T59957C
C7M84_000346	559390	A	G	86.5	4.58E-10	0.24	G	A559390G
C7M84_000503	110344	A	G	390	2.31E-54	6.21	G	A110344G
C7M84_000990	656281	G	A	323.5	7.39E-38	0.24	A	G656281A
C7M84_001073	1317191	T	A	164	1.37E-07	0.25	NA	
C7M84_003252	917813	T	C	54	9.89E-13	0.10	C	T917813C
C7M84_004254	362599	C	T	830	5.10E-30	4.59	T	C362599T
C7M84_006107	134086	G	C	128.5	2.12E-20	8.08	C	G134086C
C7M84_011240	585697	T	C	2094	6.46E-84	4.77	C	T585697C
C7M84_012141	195878	T	C	79.5	9.74E-09	8.45	C	T195878C
C7M84_012205	93082	T	C	214.5	1.67E-07	4.36	C	T93082C
C7M84_013033	54546	C	T	61.5	1.72E-10	7.38	T	C54546T
C7M84_014204	246976	A	G	74	2.72E-07	0.22	G	A246976G
C7M84_016823	230068	A	T	615	2.96E-28	4.57	NA	
C7M84_018561	236704	C	T	223	4.18E-11	5.30	T	C236704T
C7M84_014903	247322	C	T	859	1.24E-58	0.17	NA	

Note: The former 17 SNPs were randomly chosen from 96,819 low-quality SNPs. After filtering means these eighteen SNPs were selected at random from 104 high-quality SNPs. NA represents no SNP was detected in target position

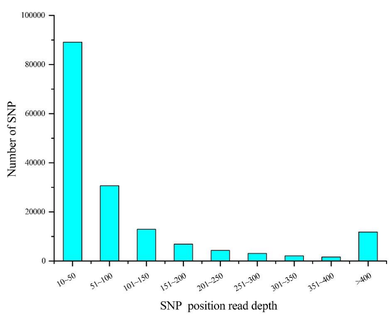

Figure. 5 Statistics of 162,571 presumptive SNPs position read depth

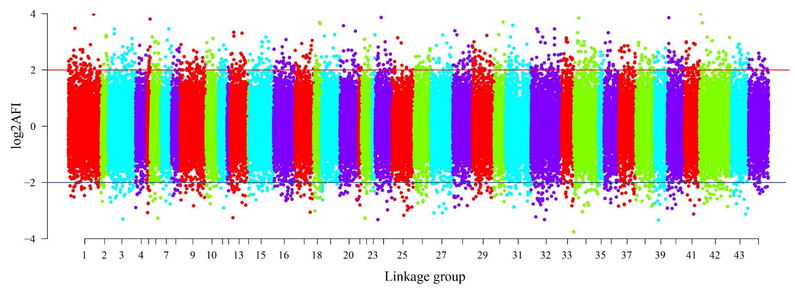

Figure. 6 The AFI distribution of 100,633 putative SNPs in linkage groups

Figure. 7 Venn diagram. DEGs mean differently expressed genes between RG and SG. FGs represent functional genes containing candidate SNPs (p-value < 0.05, AFI > 4 or < 0.25).

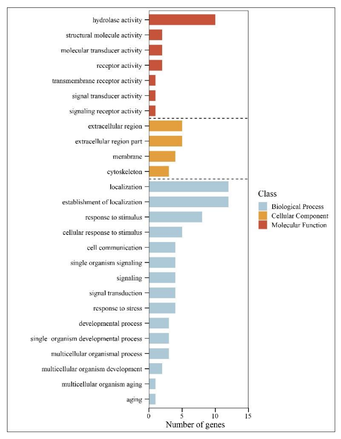

Figure. 8 Gene Ontology classifications for 566 functional genes (FGs).

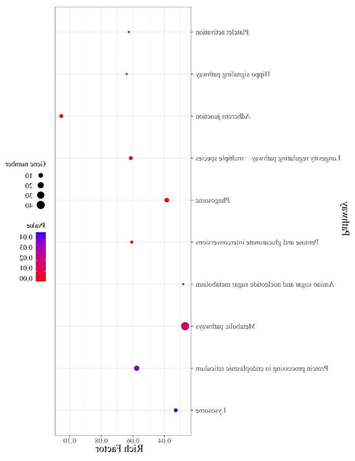

Figure. 9 The top 10 of KEGG pathway for 566 functional genes (FGs)

Discussion

- The feasibility of the workflow for screening SNPs
- SNP validation by DNA pools sequencing
- SNPs functional analysis

Conclusion

- A total of 216,015 hypothetical SNPs were detected from RNA-seq data.
- Read depth and MAF are important but not critical factors. P-value and AFI can predict SNP sites effectively.
- Twenty-two SNPs were validated in total in this experiment.
- Our workflow has enormous potential for SNPs development of economic species in aquaculture, and enables us to efficiently find some SNP sites of interest in the presence of thousands of SNPs and facilitate subsequent genotyping.

Funding

- National Natural Science Foundation of China (grant number 31702339)
- Seed Industry Innovation and Industrialization Engineering Project of Fujian Province (grant number 2017FJSCZY02)
- Foreign Cooperation Project of Fujian Provincial (grant number 2018I0013)

- Yongyu Huang
- yongyuhuang10390@163.com