OPEN ACCESS Freely available online

PLOS ONE

SNP Discovery in the Transcriptome of White Pacific Shrimp *Litopenaeus vannamei* by Next Generation Sequencing

Yang Yu[1,2], Jiankai Wei[1,2], Xiaojun Zhang[1], Jingwen Liu[1,2], Chengzhang Liu[1], Fuhua Li[1]*, Jianhai Xiang[1]

1 Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, **2** University of Chinese Academy of Sciences, Beijing, China

Abstract

The application of next generation sequencing technology has greatly facilitated high throughput single nucleotide polymorphism (SNP) discovery and genotyping in genetic research. In the present study, SNPs were discovered based on two transcriptomes of *Litopenaeus vannamei* (*L. vannamei*) generated from Illumina sequencing platform HiSeq 2000. One transcriptome of *L. vannamei* was obtained through sequencing on the RNA from larvae at mysis stage and its reference sequence was *de novo* assembled. The data from another transcriptome were downloaded from NCBI and the reads of the two transcriptomes were mapped separately to the assembled reference by BWA. SNP calling was performed using SAMtools. A total of 58,717 and 36,277 SNPs with high quality were predicted from the two transcriptomes, respectively. SNP calling was also performed using the reads of two transcriptomes together, and a total of 96,040 SNPs with high quality were predicted. Among these 96,040 SNPs, 5,242 and 29,129 were predicted as non-synonymous and synonymous SNPs respectively. Characterization analysis of the predicted SNPs in *L. vannamei* showed that the estimated SNP frequency was 0.21% (one SNP per 476 bp) and the estimated ratio for transition to transversion was 2.0. Fifty SNPs were randomly selected for validation by Sanger sequencing after PCR amplification and 76% of SNPs were confirmed, which indicated that the SNPs predicted in this study were reliable. These SNPs will be very useful for genetic study in *L. vannamei*, especially for the high density linkage map construction and genome-wide association studies.

Citation: Yu Y, Wei J, Zhang X, Liu J, Liu C, et al. (2014) SNP Discovery in the Transcriptome of White Pacific Shrimp *Litopenaeus vannamei* by Next Generation Sequencing. PLoS ONE 9(1): e87218. doi:10.1371/journal.pone.0087218

Editor: Peng Xu, Chinese Academy of Fishery Sciences, China

Received September 26, 2013; **Accepted** December 18, 2013; **Published** January 30, 2014

Funding: This work is supported by National High Technology Research and Development Program (863 Program) of China (2012AA10A404, 2012AA092205). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fhli@qdio.ac.cn

Introduction

Litopenaeus vannamei (*L. vannamei*) is widely cultured in Asia, South and North America. According to the statistics on the global aquaculture production by FAO in 2011, the whole production of *L. vannamei* occupied 76% of the world penaeid shrimp production. In order to understand the molecular mechanism for desired traits, molecular genetics research such as linkage map construction [1–4], quantitative trait loci (QTL) analysis, trait-genotype association study [5] have been conducted to allow marker-assisted selection. Although *L. vannamei* is an important and worldwide shrimp species in aquaculture, available genetic markers in public database for this species was limited.

Single nucleotide polymorphisms (SNPs) are the most abundant type of DNA sequence polymorphism which has been proved to be useful in genetic studies [6]. It has been applied in quantitative trait loci (QTL) mapping and genome wide association studies (GWAS) in model organisms and human [7–10]. In aquaculture species, SNP markers are becoming more important for linkage map construction and association studies [11–14]. In recent years, more efforts have been made for SNP discovery in *L. vannamei* [15–17]. However, these SNPs were still insufficient for high density SNP chip construction and genome wide association studies.

Next generation sequencing technologies have made high throughput SNP discovery feasible for non-model species [18–20]. Recently, transcriptome sequencing has become the major method for SNP discovery [18]. Through transcriptome sequencing, functional genes could be sequenced at high coverage, which ensured full-scale SNP discovery in coding genes with high accuracy. Large amount of SNPs have been developed by next generation transcriptome sequencing in aquaculture species such as catfish [21], Atlantic Cod [11], oyster [22], half-smooth tongue sole [23], Atlantic Herriing [18], silver carp [24], common carp [25] and Altantic Salmon [26]. In *L. vannamei*, transcriptome sequencing was applied in shrimp larvae, TSV-infected and non-infected individuals using next-generation sequencing technique [27,28]. It supplied a large amount of gene information related to development and disease resistance. However, SNPs discovery in transcriptome data by next generation sequencing was not reported in *L. vannamei* till present.

In the present study, SNPs were predicted by analysis of reads from two transcriptomes of *L. vannamei* and the characterization of these SNPs was analyzed. The predicted SNPs in this study will be very helpful for further genetic studies in *L. vannamei*, especially for the high density linkage map construction and genome wide association studies.

Materials and Methods

Transcriptome Sequencing and Reads Collection

Hundreds of *L. vannamei* larvae at developmental stage of mysis were collected from Guangtai Marine breeding company in Hainan province, China. They were frozen in liquid nitrogen immediately and stored in −80°C before RNA isolation. Total RNA was isolated using RNAiso Plus (Takara, Japan) following manufacturer's protocol. The RNA amounts and quality was estimated using NanoDrop 1000 spectrophotometer (Nano-Drop Technologies, USA). Sequencing of RNA extracted from mysis larvae was conducted in BGI (Shenzhen, China) using the paired-end RNA-Seq method [29]. The detail procedure was the same as described previously [30]. The reads obtained through this way were defined as M transcriptome in the present study.

The reads of 100 larvae of *L. vannamei* at 20 days post spawning, which was deposited in NCBI with the Sequence Read Archive (SRA) accession number of SRR346404 [28], were downloaded and defined as P transcriptome. All reads were filtered with NGS QC Toolkit using default settings before further analysis.

SNP Detection

The reads from M transcriptome were *de novo* assembled using Trinity [31] and the assembled sequence was used as reference in this study. Firstly, SNPs were detected using reads of M transcriptome and P transcriptome separately. Short reads of M and P transcriptome were separately mapped to the reference using BWA version 0.5.9 (http://bio-bwa.sourceforge.net/) with the default settings except for no gap tolerance. The software package SAMtools (http://samtools.sourceforge.net/) was used to convert sequence alignment/map (SAM) file to sorted binary alignment/map (BAM) file [32]. The command Rmdup was used to remove duplicates and SNPs were detected using mpileup in SAMtools using the following parameter: −6 (Illumina 1.3+ encoded quality score) -g -u (Compute genotype likelihoods and generate binary call format) -C 50–D (Output per-sample read depth). SNPs were called by Bcftools. SNPs with quality score more than 20 and read depth over 10 were filtered as high quality SNPs. Since prediction accuracy of SNPs is dependent on sequence coverage [33], we combined the reads of the two transcriptomes (M+P transcriptome) in order to improve sequence coverage. SNPs detection was conducted again using the combined reads with the same method. The further characteristic analysis was based on the SNPs predicted in the reads of M+P transcriptome.

Statistics of SNP Information

Mapped reads ratio (MRR) to the reference in each dataset was calculated by applying flagstat command of SAMtools software to the BAM file. SNP frequency was calculated by dividing the total length of reference by total number of SNPs. Transition verse transversion ratio was calculated by analyzing each type of DNA substitution. SNP depth (DP) and Minor Allele Frequency were also extract from the result file of SAMtools. SNP classification was obtained by a Perl script used in previous study [17].

Functional Annotation of SNPs

The unigenes containing SNPs were annotated using Basic Local Alignment Search Tool (BLAST) against to NCBI non-redundant (nr) database by BLASTX(E-value cut off $<1.0e-5$). The BLAST results were utilized by Blast2GO to annotate the unigenes with GO terms of biological processes, molecular functions, and cellular components [34]. Annotated information was imported into BGI WEGO program (http://wego.genomics.org.cn) in WEGO native format to plotting GO annotation results. KEGG pathways were assigned to unigenes containing SNPs using the online KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/tools/kaas/) [35]. KEGG Orthology (KO) assignment was applied using Bi-directional Best Hit (BBH) method.

SNP Validation

In order to validate the accuracy of SNPs prediction, 50 SNPs were randomly selected for SNP validation using DNA as templates. Primers were designed to amplify the flanking sequence of selected SNPs using Primer 3 with fragment length of 200 bp. Primers were synthesized in Sangon Biotech (Shanghai, China). Eight DNA pools made by 96 individuals were used as templates for amplification. The amplified PCR products were sequenced by ABI Prism 3730 sequencer (Applied Biosystem, USA) and sequencing results were analyzed by BioEdit version 7.0.5.3 (http://www. mbio.ncsu.edu/BioEdit/bioedit.html).

Results

Generation of Expressed Short Reads

After filter using NGS QC Toolkit, a total of 53,902,786 high quality reads with 90 bp were generated from M transcriptome (NCBI Sequence Read Archive accession number SRR1039534). A total of 24,680,276 high quality reads with 90 bp were obtained from P transcriptome (Table 1).

Reads Mapping and SNP Detection

De novo assembly of the independent short reads from M transcriptome generates 66,215 unigenes with an average length of 705 bp. A total of 86.83% and 71.70% of the short reads from M and P transcriptome were mapped to the reference separately (Table 1). The alignment file was used for SNP detection using SAMtools [32]. In order to get more reliable SNPs, those with quality score over 20 and the reads depth over 10 were regarded as high quality SNPs. A total of 58,717 and 36,277 putative SNPs with high quality were predicted in the M transcriptome and P transcriptome, respectively. Using the two sets of data together (M+P transcriptome), a total of 96,040 SNPs with high quality were predicted. Comparison of SNPs predicted in M transcriptome, P transcriptome and M+P transcriptomes were shown in Figure 1. It showed that 49% of the SNPs in P transcriptome were found to be the same as that in P transcriptome, 98% of the SNPs predicted in M transcriptome or P transcriptome could be found in the predicted SNPs in M+P transcriptomes. In the M+P transcriptome, other 20,225 SNPs were predicted besides the common SNPs to those in M transcriptome or P transcriptome. We analyzed the heterozygosity and homozygosity of those 20,225 SNPs in the two transcritpomes. It showed that most of the SNPs were heterozygous in each transcriptome. As the SNPs

Table 1. Statistics of raw reads with high quality and mapped reads ratio of M transcriptome and P transcriptome.

Dataset	Mapped reads	Raw reads	Mapped reads ratio
M[a]	46,803,406	53,902,786	86.83%
P[b]	17,694,976	24,680,276	71.70%

M[a] Reads from M transcriptome.
P[b] Reads from P transcriptome.
doi:10.1371/journal.pone.0087218.t001

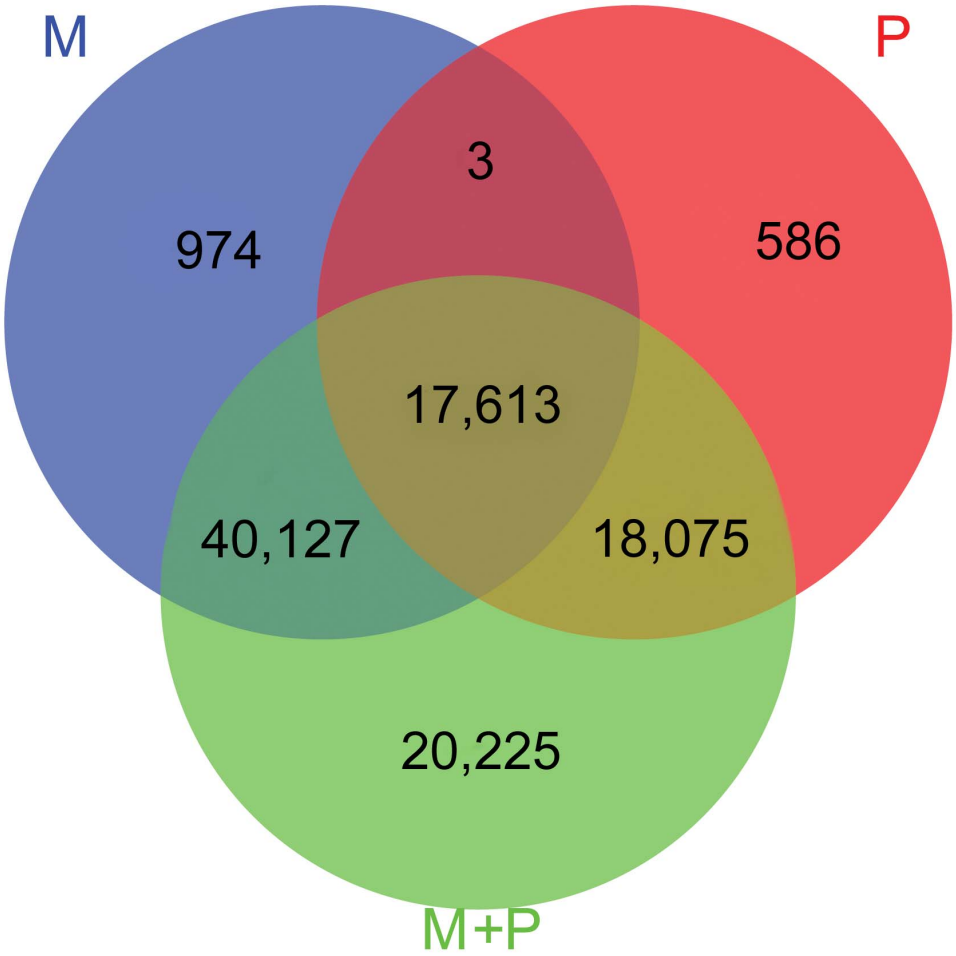

Figure 1. Venn diagram of SNPs discovered using reads of the three datasets. M represented SNPs discovered using reads from transcriptome of L.vannmei at developmental stage of mysis, P represented SNPs discovered using reads from transcriptome of L.vannmei at developmental stage of post larva. M+P represented SNPs discovered using the reads of the two transcriptomes together.
doi:10.1371/journal.pone.0087218.g001

discovered by combined data supplied more information, further analysis was conducted on the SNPs predicted in the M+P transcriptomes.

The estimated SNP frequency was 0.21% (one per 476 bp). Within the identified SNPs, more transitions substitution (66.8%) were found than transversion substitution (33.2%) (Table 2). In terms of transition substitution, the amount of A/G transitions was similar to that of C/T transition. In terms of transversion substitution, the frequency of four types (G/C, G/T, A/C and A/T) was equal. The estimated ratio for transition to transversion was 2.0.

Table 2. Statistics of transition and transversion type in the total SNPs.

Type	Transition		Transversion			
	GA	CT	GC	GT	AC	AT
Number	32,209	32,695	6,404	7,191	7,388	10,153
Percentage	33.5%	33.3%	6.9%	7.9%	7.7%	10.8%

doi:10.1371/journal.pone.0087218.t002

Read Depth and Minor Allele Frequency Distribution

As the read depth in SNPs position was closely related to the prediction accuracy of SNPs [30], the statistics of read depth for each SNP was calculated and plotted (Figure 2). It showed that the estimated average read depth was 44. SNPs with read depth between 10 and 50 account for the majority (72%) while SNPs with read depth range from 51–100 account for nearly 20%.

Minor allele frequency (MAF) was calculated based on the sequence data. As only SNPs with MAF more than 0.05 are regarded as true SNPs, those with more than 0.05 MAF were analyzed. SNPs with MAF range from 0.45–0.5 accounts for 60% of the total SNPs (Figure 3).

SNP Classification

As shown in Table 3, totally 57,793 SNPs were annotated. Among these, 34,371 SNPs were located in the open reading frame including 5,242 non-synonymous SNPs and 29,129 synonymous SNPs, 23,422 SNPs were located in 5′ or 3′ UTR region. As there were limited data of related species, 38,247 SNPs were not annotated.

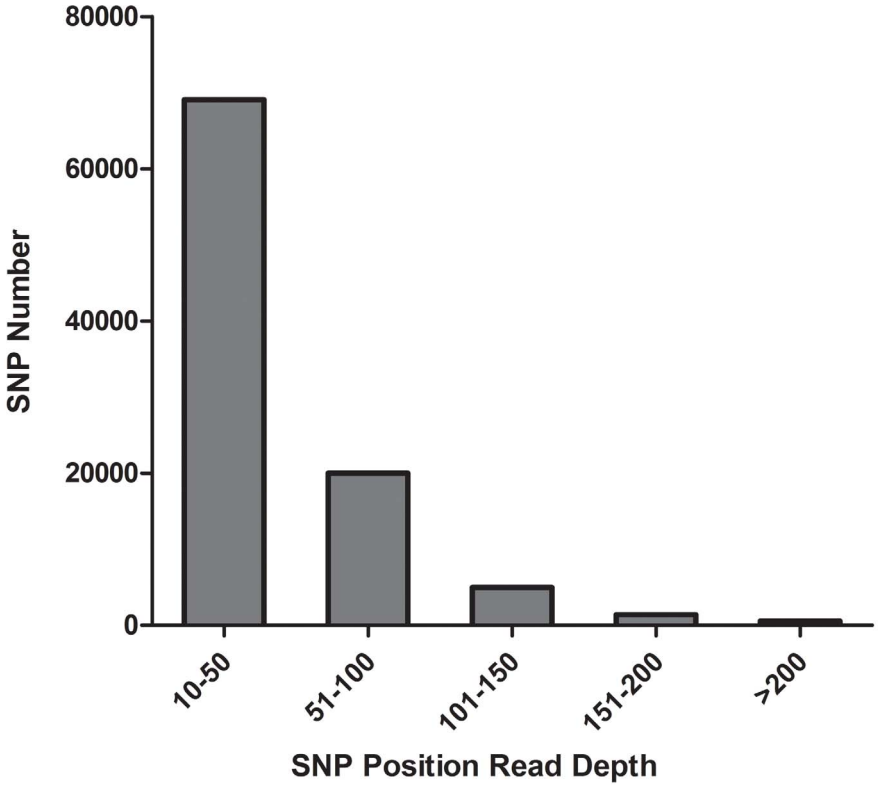

Figure 2. Statistics of read depth in each SNP in *L. vannamei*. The horizontal axis represented the read depth of SNPs, the vertical axis represented the number of SNPs with the corresponding read depth. The average read depth was 44.
doi:10.1371/journal.pone.0087218.g002

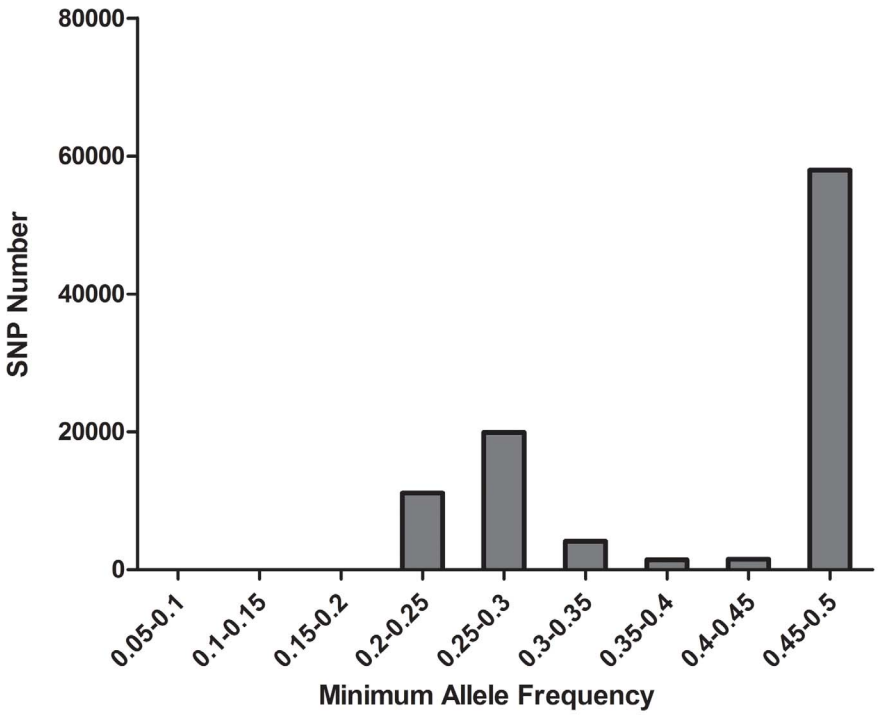

Figure 3. Statistics of minor allele frequency of total discovered SNPs in *L. vannamei*. SNP number with MAF below 0.2 was too small to be observed in the column diagram.
doi:10.1371/journal.pone.0087218.g003

Table 3. Classification of identified SNPs using a manual Perl script.

SNP classification	SNP number
Non-synonymous SNP	5,242
Synonymous SNP	29,129
5′ or 3′ UTR	23,422
Not annotated	38,247
Total	96,040

doi:10.1371/journal.pone.0087218.t003

Assessment of SNP Distribution

SNP distribution among unigenes is important when considering the marker density and genome coverage using SNP marker [21], especially when these SNPs were used for linkage map construction. In this study, we found that all the SNPs were distributed in 25,071 unigenes (38% of the total unigenes). Among these 25,071 unigenes, unigenes with 1 SNP were more common and those with no more than 10 SNPs occupied 93% of total unigenes. A total of 1,740 unigenes containing more than 10 SNPs were observed. The detailed SNPs distribution among those unigenes was shown in Figure 4. In order to investigate the mutation rate among unigenes, the SNP frequency within unigenes was calculated and shown in Figure 5. The top twenty annotated unigenes with the highest SNP frequency were listed in Table 4.

SNP Annotation and Functional Analysis

Among the 25,071 unigenes containing SNPs, 13,591 unigenes (54%) had significant hit to the protein in the non-redundant (nr) database. The unigenes were annotated by the corresponding top best BLASTx hit. After the Gene Ontology annotation, 12,978 unigenes (52%) were assigned with one or more GO ID. The plotted GO annotations of these annotated unigenes were shown in Figure 6.

The unigenes with higher SNP frequency (more than 0.014) were separately extracted from the annotated file and their GO annotation were plotted together with those of total SNPs (Figure 6). Genes in synapse, synapse part, virion and virion part in cellular component category tended to be less polymorphic. Genes in auxiliary transport protein, metallochaperone, protein tag and translation regulator in molecular function category tended to be less polymorphic. Genes in cell killing, immune system process, locomotion, rhythmic process and viral reproduction process in biological process category tended to be less polymorphic.

KEGG analysis showed that 10,803 (43% of total) unigens containing 23,553 SNPs could be annotated by KEGG database. These unigenes could be assigned to 254 KEGG pathways (Table S1). The top 20 assignment KEGG pathways were shown in Figure 7.

SNP Validation

Among the 50 primer pairs designed for SNP validation, 34 could amplify target sequences. Within these amplified sequences, 26 SNPs were validated (Table S2). The estimated predicting accuracy reached 76%. For SNPs with quality score over 90, the predicting accuracy could be up to 87.5%.

Discussion

The purpose of the present study is to develop a large amount of convinced and gene specific SNPs for *L. vannamei* through transcriptome sequencing. As next generation sequencing could enable the deep and efficient probing of transcriptome [36], most functional genes at the corresponding developmental stage could be involved in the transcriptome. It ensures a sufficient resource for gene-associated SNPs discovery. To generate more SNP information of *L. vannamei*, two sets of transcriptome data were

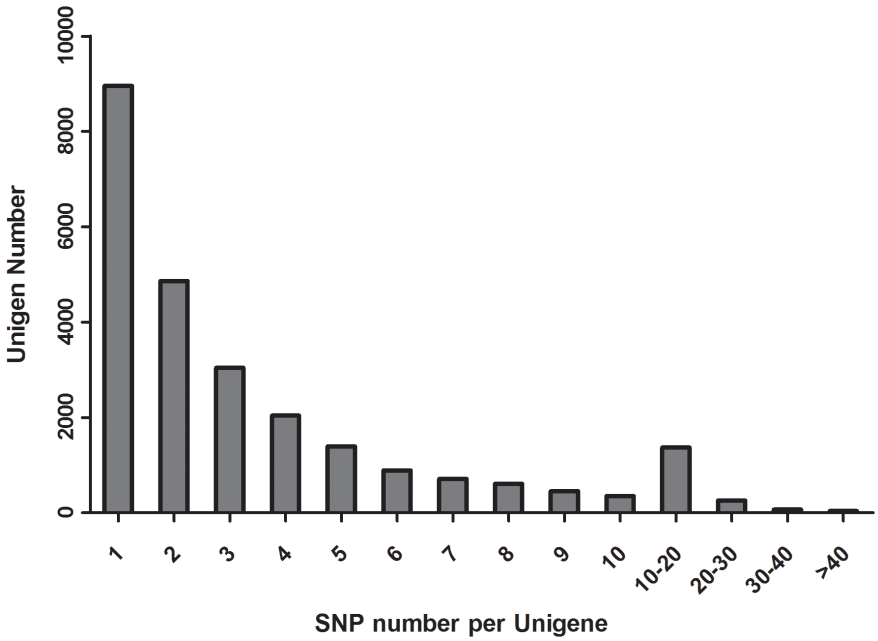

Figure 4. The distribution of SNPs in unigenes. The horizontal axis represented SNP numbers per unigene. The vertical axis represented the number of unigenes.
doi:10.1371/journal.pone.0087218.g004

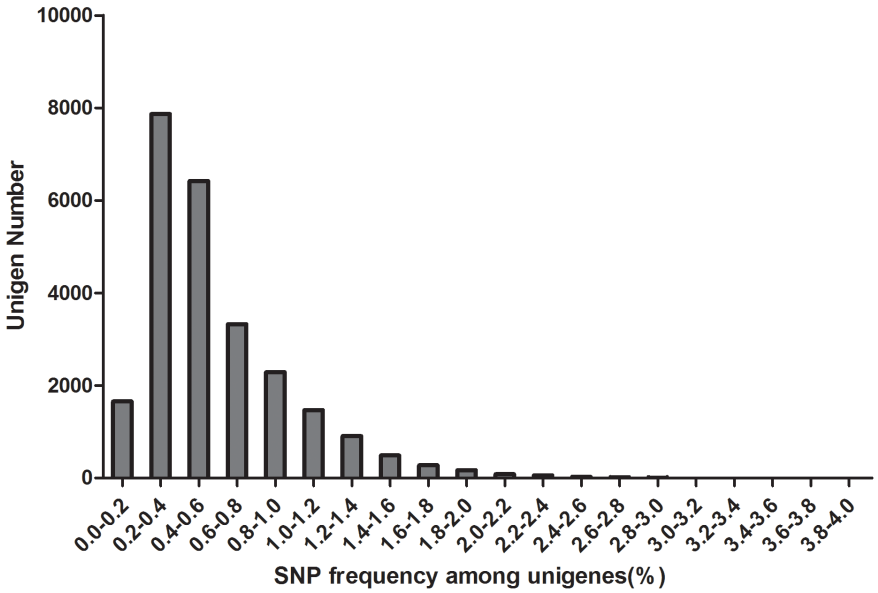

Figure 5. The frequencies of SNPs in unigenes. The frequencies of SNPs in each unigene was calculated by dividing unigene length by SNPs number per unigene. Number of unigenes with SNP frequency over 0.030 was too small to be observed in the column diagram.
doi:10.1371/journal.pone.0087218.g005

used in this study. A total of 58,717 and 36,277 SNPs with high quality were predicted by using these two sets of data separately. Using the reads of two transcriptomes together, we detected 96,040 SNPs including 20,225 SNPs which was not detected by using the two transcriptomes separately. Heterozygosity and homozygosity analysis of these 20,225 SNPs indicate most of these SNPs were polymorphic in each transcriptome.

The overlap analysis of SNPs discovered by two transcriptomes separately showed that only half of the SNPs were identical. The identical SNPs may refer to SNPs which could be easily transferred in this species. This result indicated that *L. vannamei* may endow a moderately high genetic diversity. Similar reports were published in *Streblospio benedicti* where population differentiation were observed between different populations [37].

Table 4. Annotation of the top twenty unigenes with the highest SNP frequency.

Unigenes	SNP frequency	Nr Annotation
Unigene15570	0.02253	viral A-type inclusion protein [*Trichomonas vaginalis G3*]
Unigene14752	0.022634	PREDICTED: protein lethal(2)essential for life-like [*Acyrthosiphon pisum*]
Unigene1532	0.02268	early cuticle protein 5 [*Callinectes sapidus*]
Unigene3466	0.023529	hypothetical protein [*Monosiga brevicollis MX1*]
Unigene2230	0.023564	hypothetical protein DDB_G0288123 [*Dictyostelium discoideum AX4*]
Unigene2037	0.024038	anti-lipopolysaccharide factor [*Macrobrachium rosenbergii*]
Unigene17231	0.024096	hypothetical protein [*Monosiga brevicollis MX1*]
Unigene27996	0.02446	dimethylglycine dehydrogenase, isoform CRA_a [*Homo sapiens*]
Unigene6364	0.024476	GD25285 [*Drosophila simulans*]
Unigene18738	0.024725	hypothetical protein [*Monosiga brevicollis MX1*]
CL1560.Contig3	0.026201	Spz3 [*Litopenaeus vannamei*]
Unigene7779	0.026201	GM24425 [*Drosophila sechellia*]
Unigene14736	0.026415	PREDICTED: uncharacterized protein C15orf39-like isoform 1 [*Pongo abelii*]
CL1752.Contig1	0.026477	PREDICTED: hypothetical protein [*Strongylocentrotus purpuratus*]
Unigene27125	0.02649	hypothetical protein, conserved in P. knowlesi [*Plasmodium knowlesi strain H*]
Unigene6916	0.028302	Hypothetical protein CBG18078 [*Caenorhabditis briggsae*]
CL3267.Contig1	0.028871	hypothetical protein TcasGA2_TC002455 [*Tribolium castaneum*]
Unigene7603	0.030303	hypothetical protein BRAFLDRAFT_187367 [*Branchiostoma floridae*]
CL3620.Contig1	0.03125	crustacyanin-A, partial [*Cherax quadricarinatus*]
Unigene893	0.034392	heat shock protein 21 [*Macrobrachium rosenbergii*]

doi:10.1371/journal.pone.0087218.t004

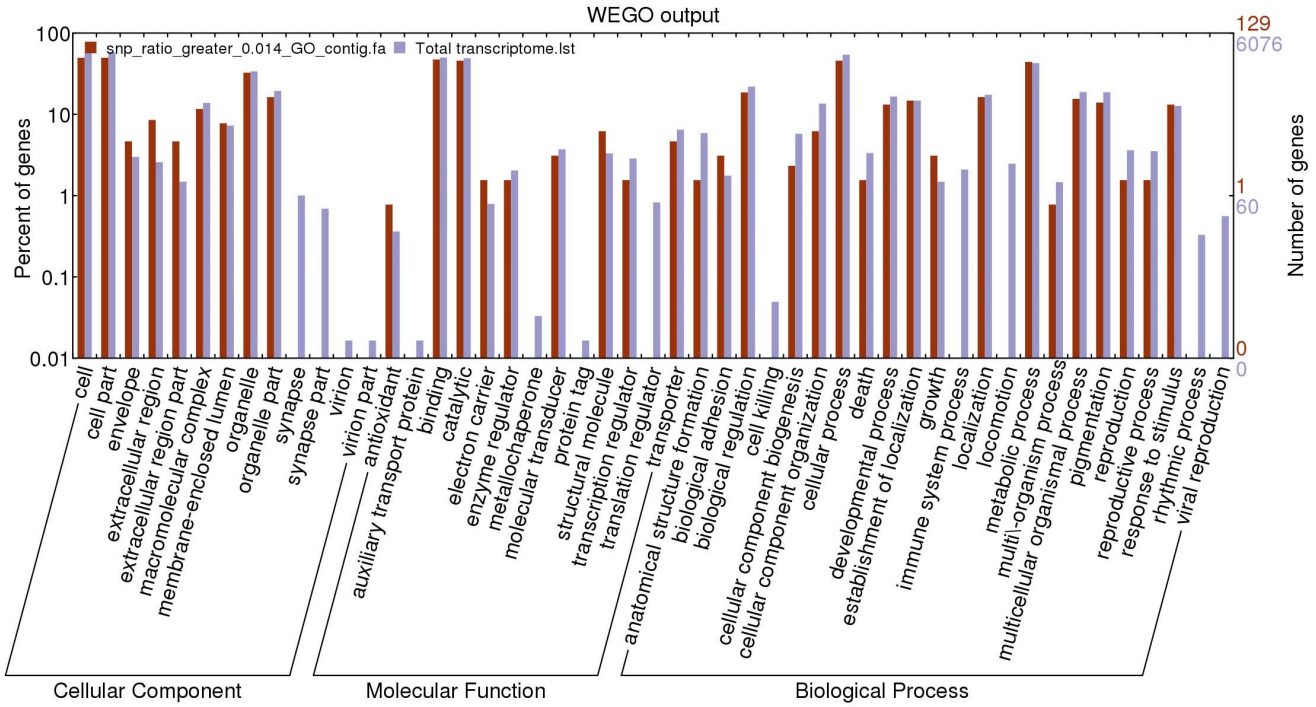

Figure 6. Gene ontology of all the annotated unigenes and unigenes with SNP frequency more than 0.014. The blue column represented gene ontology of all unigenes containing SNPs. The red column represented gene ontology of unigenes with SNP frequency more than 0.014.
doi:10.1371/journal.pone.0087218.g006

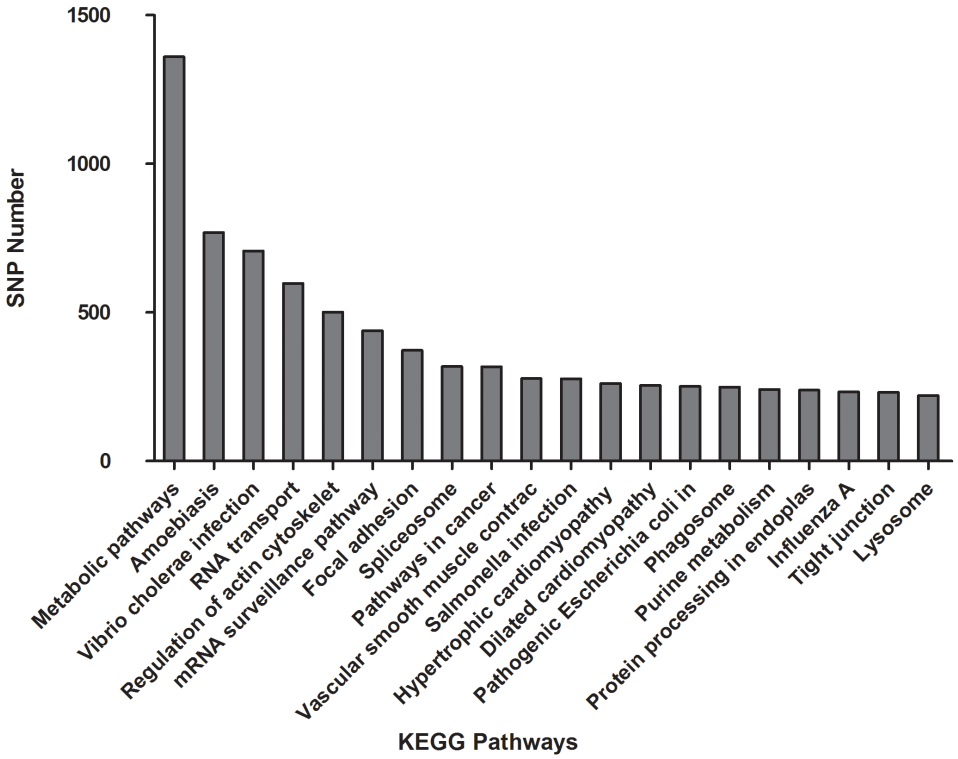

Figure 7. The top 20 KEGG pathway classification of assigned SNPs. The horizontal axis represented KEGG pathway annotation. The vertical axis represented the number of SNPs assigned to the corresponding KEGG pathway.
doi:10.1371/journal.pone.0087218.g007

MAF is a relative estimation of true allele frequency in the population [38]. The distribution of minor allele frequencies in this study deviated obviously from a uniform distribution, with an excess of alleles at high frequency. SNPs with MAF range from 0.45 to 0.5 accounts for 60% of the total SNPs, which is different from the data reported previously [39]. It inferred that the individuals used for transcriptome sequencing contained a high number of heterozygous loci. It is consistent with the results discovered by genome sequencing [40]. The estimated SNP frequency in *L. vannamei* is one per 476 bp, which is lower than European hake (1/137 bp) [22], Eastern oyster (1/60 bp) [41] and higher than Atlantic cod (1/516 bp) [11], Atlantic salmon (1/614 bp) and Salt Marsh Beetle (1/898 bp) [42]. The SNP frequency was moderately higher in the reported species. The transition/transversion ratio (2.0) was also higher than Pacific oyster (1.3) [43] and Drosophila(1.5) [44].

Read depth is a key parameter affecting the predicting accuracy of SNPs [30]. One advantage of Illumina sequencing platform is the higher read depth comparing to 454 sequencing platform. It ensures the detection of true SNPs. In this study, average read depth of SNP position was 44 which was enough to guarantee the accuracy of discovered SNPs. It also could ensure that most of expected SNPs in the sequenced population could be detected [33]. SNPs with much higher read depth should be excluded since too high read depth might be caused by paralogous sequence variants [18].

Another advantage of SNP discovery using transcriptome data is to find the SNPs directly associated with interested traits, such as disease resistance or growth advantages. Researchers primarily focused on non-synonymous coding SNPs (nsSNPs) as those SNPs might influence the protein activity directly. Reports on human genome wide association studies (GWAS) showed that the synonymous SNPs might play the same role as those nsSNPs [45]. SNPs in 3′ or 5′ un-translated regions were also very important since some of them might lead to changes in mRNA binding sites [46,47]. Most of the annotated SNPs predicted in the present study were located in UTR region, and only 6,869 SNPs were non-synonymous SNPs. These SNPs are possibly to be used in the further genome wide association studies and genome selection breeding program of *L. vannamei*.

The SNP frequency in each unigene was calculated in this study. The SNP frequencies ranged from one per 1.4 kbp to one per 26 bp. The high polymorphic unigenes detected in this study could be used in the population diversity or population differentiation analysis. We arbitrarily categorized unigenes with SNP frequency over 0.014 as higher polymorphic genes. Gene Ontology analysis showed that some components tend to be less polymorphic.

Among 50 primers used for SNP validation, 34 primer pairs could amplified target sequence. As we used genome DNA for validation, primers may locate in boundary of exon and intron which resulted in failed amplification. Another reason for failed amplification may be large size intron insert between primer pair. We also found that the quality score of SNPs influenced the validation result. The accuracy for SNPs with quality score over 90 was higher than that with quality score over 20. It was reported previously that when selecting SNPs in genetic studies, we should consider the allele frequency of the SNPs [21]. Considering these together, both MAF and quality score should be in consideration in further genetic research.

Conclusion

In this study, next generation sequencing reads from two transcriptomes of *L. vannmei* were used for SNP discovery. A total of 96,040 high quality SNPs were predicted using the reads of the two transcriptomes together. Within those SNPs, approximately half could be annotated. The read depth and MAF analysis showed these predicted SNPs were accurate and common in this species. Besides, a moderately high genetic diversity and high heterozygosity in *L. vannmei* were found through characteristic analysis of SNPs. Overall, the SNPs predicted in this study will be useful in the genome wide association studies and whole genome selection studies.

Supporting Information

Table S1 KEGG pathway of annotated SNPs.
(XLSX)

Table S2 The SNPs validated by Sanger sequencing.
(XLSX)

Acknowledgments

We are grateful to Mr. Hao Huang from Guangtai Marine breeding company in Hainan province, China for the help in sample collection.

Author Contributions

Conceived and designed the experiments: FL JX. Performed the experiments: YY JW XZ JL. Analyzed the data: YY JW CL. Contributed reagents/materials/analysis tools: JW CL. Wrote the paper: YY FL.

References

1. Zhang L, Yang C, Zhang Y, Li L, Zhang X, et al. (2006) A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. Genetica 131: 37–49.
2. Alcivar-Warren A, Meehan-Meola D, Park SW, Xu Z, Delaney M, et al. (2007) ShrimpMap: A low-density, microsatellite-based linkage map of the pacific whiteleg shrimp, *Litopenaeus vannamei*: Identification of sex-linked markers in linkage group 4. Jounal of Shellfish Research 26: 1259–1277.
3. Du ZQ, Ciobanu DC, Onteru SK, Gorbach D, Mileham AJ, et al. (2010) A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. Animal Genetics 41: 286–294.
4. Perez F, Erazo C, Zhinaula M, Volckaert F, Calderon J (2004) A sex-specific linkage map of the white shrimp Penaeus (*Litopenaeus*) *vannamei* based on AFLP markers. Aquaculture 242: 105–118.
5. Glenn KL, Grapes L, Suwanasopee T, Harris DL, Li Y, et al. (2005) SNP analysis of AMY2 and CTSL genes in *Litopenaeus vannamei* and *Penaeus monodon* shrimp. Animal Genetics 36: 235–236.
6. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, et al. (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307: 1072–1079.
7. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087–1093.
8. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881–885.
9. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465: 627–631.
10. Matsunami H, Yu H, Xie W, Wang J, Xing Y, et al. (2011) Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers. PloS ONE 6: e17595.
11. Hubert S, Higgins B, Borza T, Bowman S (2010) Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). BMC Genomics 11: 191.
12. Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, et al. (2008) Identification and characterisation of novel SNP markers in Atlantic cod: Evidence for directional selection. BMC Genetics 9: 18.

13. Prudence M, Moal J, Boudry P, Daniel JY, Quere C, et al. (2006) An amylase gene polymorphism is associated with growth differences in the Pacific cupped oyster *Crassostrea gigas*. Animal Genetics 37: 348–351.
14. Nguyen MT, Barnes AC, Mather PB, Li YT, Lyons RE (2010) Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. Aquaculture 301: 7–15.
15. Gorbach DM, Hu ZL, Du ZQ, Rothschild MF (2009) SNP discovery in *Litopenaeus vannamei* with a new computational pipeline. Animal Genetics 40: 106–109.
16. Ciobanu DC, Bastiaansen JW, Magrin J, Rocha JL, Jiang DH, et al. (2010) A major SNP resource for dissection of phenotypic and genetic variation in Pacific white shrimp (*Litopenaeus vannamei*). Animal Genetics 41: 39–47.
17. Liu CZ, Wang X, Xiang JH, Li FH (2012) EST-derived SNP discovery and selective pressure analysis in Pacific white shrimp (*Litopenaeus vannamei*). Chinese Journal of Oceanology and Limnology 30: 713–723.
18. Helyar SJ, Limborg MT, Bekkevold D, Babbucci M, van Houdt J, et al. (2012) SNP Discovery Using Next Generation Transcriptomic Sequencing in Atlantic Herring (*Clupea harengus*). PloS ONE 7(8): e42089.
19. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. The Plant Journal 51: 910–918.
20. Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. Molecular Ecology Resources 10: 915–934.
21. Liu S, Zhou Z, Lu J, Sun F, Wang S, et al. (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. BMC Genomics 12: 53.
22. An HS, Byun SG, Kim YC, Lee JW, Myeong JI (2011) Wild and hatchery populations of Korean starry flounder (Platichthys stellatus) compared using microsatellite DNA markers. International Journal of Molecular Sciences 12: 9189–9202.
23. Sha Z, Wang S, Zhuang Z, Wang Q, Wang Q, et al. (2010) Generation and analysis of 10 000 ESTs from the half-smooth tongue sole Cynoglossus semilaevis and identification of microsatellite and SNP markers. Journal of Fish Biology 76: 1190–1204.
24. Fu BD, He SP (2012) Transcriptome analysis of silver carp (*Hypophthalmichthys molitrix*) by paired-end RNA sequencing. DNA Research 19: 131–142.
25. Xu J, Ji PF, Zhao ZX, Zhang Y, Feng JX, et al. (2012) Genome-wide SNP discovery from transcriptome of four common carp strains. PloS ONE 7(10): e48140.
26. Moen T, Hayes B, Baranski M, Berg PR, Kjoglum S, et al. (2008) A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. BMC Genomics 9: 223.
27. Zeng DG, Chen XL, Xie DX, Zhao YZ, Yang CL, et al. (2013) Transcriptome analysis of pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura Syndrome Virus (TSV) experimental infection. PloS ONE 8(3): e58627.
28. Li CZ, Weng SP, Chen YG, Yu XQ, Lu L, et al. (2012) Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. PloS ONE 7(10): e47442.
29. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59.
30. Li SH, Zhang XJ, Sun Z, Li FH, Xiang JH (2013) Transcriptome analysis on chinese shrimp *Fenneropenaeus chinensis* during WSSV acute infection. PloS ONE 8(3): e58627.
31. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644–652.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
33. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, et al. (2013) Development of strategies for SNP detection in RNA-Seq data: application to lymphoblastoid cell lines and evaluation using 1000 genomes data. PloS ONE 8(3): e58815.
34. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36: 3420–3435.
35. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35: W182–W185.
36. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644–652.
37. Zakas C, Schult N, McHugh D, Jones KL, Wares JP (2012) Transcriptome analysis and SNP development can resolve population differentiation of *Streblospio benedicti*, a developmentally dimorphic marine annelid. PloS ONE 7 (2): e31613.
38. Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, et al. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods 5: 247–252.
39. Miller NJ, Sun J, Sappington TW (2012) High-throughput transcriptome sequencing for SNP and gene discovery in a moth. Environmental Entomology 41: 997–1007.
40. Zhang X, Zhang T, Zhao C, Zhang B, Liu C, et al. Research progress in sequencing of *Litopenaeus vannamei*. Seventh International Crustacean Congress, Qingdao, China Jun 20–25, 2010(Abstract), P361.
41. Zhang LS, Guo XM (2010) Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. Aquaculture 302: 124–129.
42. Leu JH, Chen SH, Wang YB, Chen YC, Su SY, et al. (2011) A review of the major penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four penaeid shrimp. Marine Biotechnology 13: 608–621.
43. Tseng MC (2012) Evolution of microsatellite Loci of tropical and temperate *anguilla eels*. International Journal of Molecular Sciences 13: 4281–4294.
44. Vera M, Alvarez-Dios JA, Fernandez C, Bouza C, Vilas R, et al. (2013) Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. International Journal of Molecular Sciences 14: 5694–5711.
45. Chen R, Davydov EV, Sirota M, Butte AJ (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. PloS ONE 5(10): e13574.
46. Miyamoto Y, Mabuchi A, Shi DQ, Kubo T, Takatori Y, et al. (2007) A functional polymorphism in the 5′ UTR of GDF5 is associated with susceptibility to osteoarthritis. Nature Genetics 39: 529–533.
47. Saunders MA, Liang H, Li WH (2007) Human polymorphism at microRNAs and microRNA target sites. Proceedings of the National Academy of Sciences of the United States of America 104: 3300–3305.