Next Generation Sequencing to Discover Genetic Markers for Pacific White Shrimp

A.S. Leaflet R2594

Zhi-Qiang Du, postdoctoral research associate;
Max F. Rothschild, distinguished professor

Summary and Implications

Recently, a new method of sequencing called next generation sequencing has been widely applied due to its lower cost per sequencing output compared to traditional sequencing technologies. Genomic regions of Pacific white shrimp have been selected to construct materials for sequencing using the Illumina Solexa technology. We then used *de novo* techniques to assemble the sequences into 8,007 contigs, and identified ~ 256K genetic markers. Due to the potential complex genome structure of Pacific white shrimp, we had a success rate of validating about 2/3 of the genetic markers chosen. This important resource could be used for the improvement of growth rate and disease resistance by the shrimp industry.

Introduction

Massively parallel sequencing technology has been widely applied in the search for genes underlying important phenotypes, such as those associated with human diseases or agriculture production. Compared to conventional methods, these new methods are fast and orders of magnitude more cost-effective. Pacific white shrimp, *Litopenaeus vannamei*, occupy an important position in the aquaculture industry. In order to discover more genetic variants, we employed the reduced representation sequencing method to further interrogate the shrimp genome,.

Materials and Methods

We selected eight animals from four F2 families (two per each family) previously created to construct a SNP genetic map for Pacific white shrimp. Standard DNA isolation methods were used. Massive parallel DNA sequencing was performed on an Illumina/Solexa Genome Analyzer II (DNA Facility, Iowa State University) with the option of single read and 75 cycles of sequencing on six lanes.

Results and Discussion

In total, 8,007 contigs were assembled and five of them matched to mitochondrial sequences (Table 1). The partitioning of the shrimp nuclear genome had ~46.37 million (99.99%) short reads aligned to 8,002 contigs from the nuclear genome, with an average length of 215 bp and an average coverage depth of 52 fold. However, a total of 3,937 short reads were aligned to the mitochondrial genome, but with a relatively longer average length (344 bp), and deeper average coverage depth (131 fold).

The distribution of the lengths of the assembled contigs revealed that few contigs were longer than 3 kb, and most of which were clustered around 150 bp. The short length of the assembled contigs and the deep coverage depth suggested that this library could be over-sampled, and no extra sequence information could be used to further extend the assembled contigs. Individual short reads were aligned back to the contigs and ~ 256K potential SNPs were recorded. After the comparison of the assembled contigs to our previously discovered shrimp genomic sequences, 16 of 24 potential SNPs appeared to be real as seen by sequencing and genotyping data, respectively. We have demonstrated here for the first time that Pacific White shrimp could have a complicated genome structure, and we propose that a more conservative method should be adopted when sequencing and assembling the shrimp genome.

Acknowledgments

We appreciate the funding support from CP Indonesia, Shrimp Improvement Systems, Golden Dragon Research, and State of Iowa and Hatch Funding. Technical help from Drs. Michael Baker and Gary Polking in the DNA facility at Iowa State University is also appreciated.

Table 1. Statistics for *de novo* assembly of short reads for shrimp.

	Contigs	Contigs matched to mitochondrion
Contigs assembled	8,002	5
Reads matched to contigs	46,374,552	3,937
Average length (bp)	215	344
Average coverage	52 ×	131×
Potential SNPs	256,160	