



Data in Brief

Microbiome analyses of pacific white shrimp (*Litopenaeus vannamei*) collected from disparate geographical locations



Arun S. Seetharam^a, Emily Kawaler^b, Zhi-Qiang Du^c, Max F. Rothschild^b, Andrew J. Severin^{a,*}

^a Genome Informatics Facility, Office of Biotechnology, Iowa State University, Ames, IA 50011, USA

^b Department of Animal Sciences, Iowa State University, Ames, IA 50011, USA

^c College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China

ARTICLE INFO

Article history:

Received 3 August 2015

Received in revised form 4 August 2015

Accepted 7 August 2015

Available online 12 August 2015

Keywords:

Shrimp

Litopenaeus vannamei

Microbiome

Next-generation sequencing

Geographical diversity

ABSTRACT

In this study, the tail muscle microbiota of pacific white shrimp (*Litopenaeus vannamei*) sourced from five countries across Central and South America and Southeast Asia were determined and compared. The genomic DNA was sequenced at around 10× coverage for each geographical location and was assembled *de novo* for comparative analysis. The assembled sequences for all the lines were classified based on their similarity to the sequences in the public database. We found that there is high correlation among the microbiota of shrimp from disparate regions, as well as the presence of some DNA from bacteria known to cause food poisoning in humans. Sequencing data has been deposited at NCBI-SRA database and can be found under the BioProject ID PRJNA282154.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications	
Organism/cell line/tissue	<i>Litopenaeus vannamei</i> , muscle tissue, genomic DNA
Sex	N/A
Sequencer or array type	Illumina HiSeq 2000
Data format	Raw
Experimental factors	Frozen packaged shrimp imported from Indonesia, Vietnam, Thailand, Venezuela, and Honduras were acquired from US supermarkets. From each package, 6–10 shrimp are used to isolate genomic DNA and subsequent sequencing.
Experimental features	Reads were assembled into contigs after pooling all the isolates. The presence/absence of contigs in each isolate followed by diversity analyses was performed to characterize the microbiota.
Consent	N/A
Sample source location	N/A

1. Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/bioproject/PRJNA282154>.

* Corresponding author.

E-mail address: severin@iastate.edu (A.J. Severin).

2. Experimental design, materials and methods

2.1. Sample preparation

Frozen *L. vannamei* samples were purchased at the local grocery. Each bag had been packaged and imported from different countries: Indonesia, Vietnam, Thailand, Venezuela, and Honduras. Genomic DNA was isolated from 6 to 10 shrimp of each location and was sequenced separately. Tissue was sampled after peeling the shell and dissecting the tail muscle from the shrimp. DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) was used for genomic DNA isolation. Single-end libraries were constructed using the Illumina TruSeq DNA Sample Preparation Kit (Illumina, Inc., San Diego, CA, USA), as per instructions. Each library was ligated to a different index-tag adapter and sequencing was performed using the TruSeq Unique and Universal Adaptors on a HiSeq 2000 sequencer, which produced 100 bp single-end reads [1].

2.2. Bioinformatics analyses

Sequence quality was assessed using FastQC (v 0.10.1) [2]. Digital normalization was performed on the raw reads using Khmer (v 1.01) [3] before assembling the reads. Khmer was run with the default settings except for the cutoff value (*c*), which was set for 10. About

Table 1
Total number of reads used for the assembly. The raw number indicates the actual number of reads obtained from the sequencing machine, whereas normalized read count indicates the number of reads retained after performing digital normalization. Only normalized reads were used for the assembly.

Origin	File names	Raw	Normalized	Retained (%)
Honduras	H12_TAGGCATG_L006_R1_001.fastq	2,330,575	1,381,635	59.28%
	H1_TAAGGCGA_L006_R1_001.fastq	14,712,575	8,115,248	55.16%
	H3_CGTAAGTAG_L006_R1_001.fastq	1,453,462	859,454	59.13%
	H5_AGGCAGAA_L006_R1_001.fastq	2,221,334	1,367,269	61.55%
	H7_TCCTGAGC_L006_R1_001.fastq	5,865,487	3,773,662	64.34%
	H8_GGACTCCT_L006_R1_001.fastq	6,583,439	3,665,752	55.68%
Indonesia	IO2_AGGCAGAA_L007_R1_001.fastq	9,177,514	5,357,801	58.38%
	IO3_TCCTGAGC_L007_R1_001.fastq	3,197,185	2,168,825	67.84%
	IO4_GGACTCCT_L007_R1_001.fastq	14,167,640	8,210,733	57.95%
	IO5_TAGGCATG_L007_R1_001.fastq	11,396,545	6,795,882	59.63%
	IO6_CTCTCTAC_L007_R1_001.fastq	10,435,103	6,227,204	59.68%
	IO7_CAGAGAGG_L007_R1_001.fastq	7,342,116	4,524,518	61.62%
	IO8_GCTACGCT_L007_R1_001.fastq	4,276,929	2,587,323	60.49%
	IO9_CGAGGCTG_L007_R1_001.fastq	4,233,317	2,611,674	61.69%
	T10_CTCTCTAC_L008_R1_001.fastq	11,028,257	6,402,355	58.05%
Thailand	T12_CAGAGAGG_L008_R1_001.fastq	13,584,934	8,121,686	59.78%
	T1_TAAGGCGA_L008_R1_001.fastq	15,171,966	8,789,103	57.93%
	T3_CGTAAGTAG_L008_R1_001.fastq	2,504,357	1,720,428	68.70%
	T4_AGGCAGAA_L008_R1_001.fastq	7,341,971	4,321,487	58.86%
	T5_TCCTGAGC_L008_R1_001.fastq	5,802,738	3,669,116	63.23%
	T7_GGACTCCT_L008_R1_001.fastq	5,869,011	3,614,750	61.59%
	T9_TAGGCATG_L008_R1_001.fastq	23,687,158	13,906,904	58.71%
	V10_CAGAGAGG_L005_R1_001.fastq	6,584,510	4,175,631	63.42%
	V11_GCTACGCT_L005_R1_001.fastq	14,586,243	7,767,055	53.25%
Venezuela	V12_CGAGGCTG_L005_R1_001.fastq	15,782,536	8,900,226	56.39%
	V1_TAAGGCGA_L005_R1_001.fastq	21,065,170	11,223,993	53.28%
	V2_CGTAAGTAG_L005_R1_001.fastq	11,940,498	7,142,047	59.81%
	V3_AGGCAGAA_L005_R1_001.fastq	17,026,545	9,454,844	55.53%
	V4_TCCTGAGC_L005_R1_001.fastq	10,803,314	6,567,535	60.79%
	V5_GGACTCCT_L005_R1_001.fastq	7,554,075	4,433,658	58.69%
	V8_TAGGCATG_L005_R1_001.fastq	10,224,862	6,104,682	59.70%
	V9_CTCTCTAC_L005_R1_001.fastq	5,872,950	3,047,919	51.90%
	VN12_GTAGAGGA_L008_R1_001.fastq	9,564,954	4,872,018	50.93%
Vietnam	VN1_GCTACGCT_L008_R1_001.fastq	7,092,963	3,912,680	55.16%
	VN2_CGAGGCTG_L008_R1_001.fastq	14,190,209	8,738,927	61.58%
	VN3_AAGAGGCA_L008_R1_001.fastq	3,430,308	2,069,722	60.34%
	VN4_CGAGGCTG_L006_R1_001.fastq	35,679,513	20,000,119	56.05%
	VN5_AAGAGGCA_L006_R1_001.fastq	11,274,952	6,484,876	57.52%
	VN8_AAGAGGCA_L007_R1_001.fastq	5,376,761	2,709,392	50.39%
	VN9_GTAGAGGA_L007_R1_001.fastq	6,887,031	3,555,985	51.63%

50–60% of the total reads retained after normalization (see Table 1) were pooled and used for assembling the genome. The Ray assembler (v 2.3.1) [4] was used to generate the initial assembly using the default k-mer size (33). Obtained scaffolds were classified as prokaryotic or non-prokaryotic based on NCBI-BLAST (v 2.2.30+) [5] searches against the NR database (see Table 2). All scaffolds of prokaryotic origin were used for the diversity analyses. Bowtie2 (v 2.2.0) [6] was used to map the reads back to the indexed draft assembly and classify them as present or absent in an isolate (Table 3).

Conflict of interest

The authors declare that there are no conflicts of interest.

Table 2

Number of scaffolds with matches to NR database and its classification (prokaryotes/eukaryotes).

Classification	≥5000 K nt	1000–5000 K nt	≥1000 nt
Eukaryotes	163 (24.33%)	4661 (5.04%)	4824 (5.14%)
Non-eukaryotes	378 (56.42%)	5534 (5.98%)	5912 (6.30%)
Scaffolds with hits (NR)	541 (80.75%)	10,195 (11.02%)	10,736 (11.43%)
Scaffolds without hits	129 (19.25%)	82,360 (88.98%)	82,489 (87.85%)
Total scaffolds	670	92,555	93,895

Acknowledgments

This project was supported by funding from the USDA NIFA, the NRSP8 Aquaculture Genome Coordination Program and the College of Agriculture and Life Sciences, the State of Iowa and Hatch funds. The advice and assistance provided by Dr. James Dickson is appreciated.

References

- [1] E. Kawaler, A.S. Seetharam, Z.-Q. Du, A.J. Severin, M.F. Rothschild, A comparison of the microbiomes of *Litopenaeus vannamei* from disparate geographical regions (submitted for publication) 2015.
- [2] FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- [3] C.T. Brown, A. Howe, Q. Zhang, A.B. Pyrkosz, T.H. Brom, A reference-free algorithm for computational normalization of shotgun sequencing data (eprint arXiv:1203.4802) 2012.
- [4] S. Boisvert, F. Laviolette, J. Corbeil, Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* 17 (11) (2010) 1519–1533.
- [5] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* 215 (3) (1990) 403–410.
- [6] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4) (2012) 357–359.

Table 3

Mapping percent for various isolates to the draft assembly.

Origin	Isolate	Normalized	Aligned	Percent
Honduras	H12_TAGGCATG_L006_R1_001.fastq	2,330,575	135,668	5.82%
	H1_TAAGGCGA_L006_R1_001.fastq	14,712,575	816,820	5.55%
	H3_CGTACTAG_L006_R1_001.fastq	1,453,462	72,543	4.99%
	H5_AGGCAGAA_L006_R1_001.fastq	2,221,334	125,013	5.63%
	H7_TCCTGAGC_L006_R1_001.fastq	5,865,487	406,069	6.92%
	H8_GGACTCCT_L006_R1_001.fastq	6,583,439	352,841	5.36%
	I02_AGGCAGAA_L007_R1_001.fastq	9,177,514	584,184	6.37%
	I03_TCCTGAGC_L007_R1_001.fastq	3,197,185	199,306	6.23%
Indonesia	I04_GGACTCCT_L007_R1_001.fastq	14,167,640	970,809	6.85%
	I05_TAGGCATG_L007_R1_001.fastq	11,396,545	791,905	6.95%
	I06_CTCTCTAC_L007_R1_001.fastq	10,435,103	717,542	6.88%
	I07_CAGAGAGG_L007_R1_001.fastq	7,342,116	517,977	7.05%
	I08_GCTACGCT_L007_R1_001.fastq	4,276,929	289,595	6.77%
	I09_CGAGGCTG_L007_R1_001.fastq	4,233,317	303,002	7.16%
	T10_CTCTCTAC_L008_R1_001.fastq	11,028,257	677,696	6.15%
	T12_CAGAGAGG_L008_R1_001.fastq	13,584,934	915,014	6.74%
Thailand	T1_TAAGGCGA_L008_R1_001.fastq	15,171,966	877,129	5.78%
	T3_CGTACTAG_L008_R1_001.fastq	2,504,357	177,428	7.08%
	T4_AGGCAGAA_L008_R1_001.fastq	7,341,971	452,370	6.16%
	T5_TCCTGAGC_L008_R1_001.fastq	5,802,738	389,395	6.71%
	T7_GGACTCCT_L008_R1_001.fastq	5,869,011	380,675	6.49%
	T9_TAGGCATG_L008_R1_001.fastq	23,687,158	1,531,907	6.47%
	V10_CAGAGAGG_L005_R1_001.fastq	6,584,510	460,195	6.99%
	V11_GCTACGCT_L005_R1_001.fastq	14,586,243	810,484	5.56%
Venezuela	V12_CGAGGCTG_L005_R1_001.fastq	15,782,536	964,607	6.11%
	V1_TAAGGCGA_L005_R1_001.fastq	21,065,170	1,127,598	5.35%
	V2_CGTACTAG_L005_R1_001.fastq	11,940,498	789,086	6.61%
	V3_AGGCAGAA_L005_R1_001.fastq	17,026,545	1,025,844	6.02%
	V4_TCCTGAGC_L005_R1_001.fastq	10,803,314	722,363	6.69%
	V5_GGACTCCT_L005_R1_001.fastq	7,554,075	452,362	5.99%
	V8_TAGGCATG_L005_R1_001.fastq	10,224,862	653,237	6.39%
	V9_CTCTCTAC_L005_R1_001.fastq	5,872,950	278,656	4.74%
Vietnam	VN12_GTAGAGGA_L008_R1_001.fastq	9,564,954	177,669	1.86%
	VN1_GCTACGCT_L008_R1_001.fastq	7,092,963	149,412	2.11%
	VN2_CGAGGCTG_L008_R1_001.fastq	14,190,209	332,937	2.35%
	VN3_AAGAGGCA_L008_R1_001.fastq	3,430,308	69,846	2.04%
	VN4_CGAGGCTG_L006_R1_001.fastq	35,679,513	896,728	2.51%
	VN5_AAGAGGCA_L006_R1_001.fastq	11,274,952	294,998	2.62%
	VN8_AAGAGGCA_L007_R1_001.fastq	5,376,761	85,758	1.59%
	VN9_GTAGAGGA_L007_R1_001.fastq	6,887,031	132,293	1.92%