Marine Genomics 40 (2018) 58–63

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Method paper

Transcriptome of the bivalve *Limecola balthica* L. from Western Pacific: A new resource for studies of European populations

Andrey A. Yurchenko[a,b,*], Nataliia Katolikova[c,d], Dmitrii Polev[e], Irina Shcherbakova[e], Petr Strelkov[f,g]

[a] Theodosius Dobzhansky Center for Genome Bioinformatics, Saint Petersburg State University, 41A Sredniy Avenue, St. Petersburg 199004, Russia
[b] Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK
[c] Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia
[d] Institute of Translation Biomedicine, Saint-Petersburg State University, Universitetskaya nab., 7-9, St. Petersburg 199034, Russia
[e] Resource Centre "Biobank", Research Park, Saint Petersburg State University, 17 Botanicheskaya St., Peterhof, Saint-Petersburg 198504, Russia
[f] Department of Ichthyology and Hydrobiology, Saint Petersburg State University, 16 Line 29, Vasilevsky Island, St. Petersburg 199178, Russia
[g] Laboratory of Monitoring and Conservation of Natural Arctic Ecosystems, Murmansk Arctic State University, Kapitana Egorova 16, Murmansk 15183038, Russia

ARTICLE INFO

Keywords:
Macoma balthica
Limecola balthica
Marine biomarkers
Transcriptome
Hybridization
Assembly

ABSTRACT

The Baltic clam *Limecola balthica* L. (Tellinidae) is broadly used in ecophysiological, toxicological, evolutionary and environmental monitoring studies. However, it is poorly studied in respect of genome and gene functions. We obtained a transcriptome of *Limecola b. balthica* from Kamchatka (Western Pacific) generated with the use of Illumina high-throughput sequencing. We annotated 11,374 proteins, including 53 from the oxidative phosphorylation pathway and a number of pollution-stress biomarkers, recovered 254,540 single nucleotide variants within two annotated transcriptomes including 25,330 scorable in the previously published European data. Our results confirmed the available allozyme data indicating that nuclear genomes of the clams from the Baltic Sea were intermediate in their genetic composition between the Pacific (*L. b. balthica*) and the Atlantic (*L. b. rubra*) subspecies. At the same time, the mitochondrial genomes of *Limecola* from Kamchatka were nearly identical to the single published genome from the Baltic. The genomic diversity in *Limecola* was found to be high and comparable with that of other marine mollusks (0.0138 and 0.0142 heterozygous positions in the two studied transcriptomes). The data obtained in our study are a valuable resource for further development of genomic markers for evolutionary genetic and ecophysiological studies of L. *balthica complex*.

1. Introduction

Mollusks have a huge taxonomic diversity, deep evolutionary history, great ecological importance and considerable economic value. Nevertheless, they are among the most poorly studied organisms from the genomic point of view, with only a few genomes published to date. It is only recently that the amount of transcriptomic data began to grow (Romiguier et al., 2014; Gonzalez et al., 2015). The Baltic clam *Limecola balthica* L. (Tellinidae) is no exception in this respect. It is a very common littoral bivalve of boreal and sub-Arctic Seas and a key species in estuaries and the Baltic Sea (Väinölä, 2003; Gogina et al., 2016). Yet, despite the long-term use in ecophysiological, toxicological and environmental monitoring studies (e.g. Hummel et al., 1997; Wallace et al., 2003; Leiniö and Lehtonen, 2005), the genome and gene functions of this clam remain poorly known. This is all the more frustrating as data of this kind are also needed in evolutionary studies: the European L. *balthica* is a prominent object in studies of hybridization in the marine realm.

According to the allozyme and mitochondrial data, *L. balthica* is a species complex with two sub(species), *L. b. rubra* and L. *b. balthica*. These two taxa, separated since early Pleistocene (Simonarson et al., 1998; Nikula et al., 2007), are native to NE Atlantic and N Pacific, respectively. Recent natural trans-Arctic migrations have brought L. *b. balthica* into the northern edge of L. *b. rubra* range in Europe. Nowadays there is no pure L. *b. balthica* in Europe. The White, the Barents and the Baltic Seas are inhabited by diverse hybrid populations, which are mostly in local genetic equilibrium. These populations represent a unique phenomenon of the so-called hybrid swarms (Väinölä, 2003; Strelkov et al., 2007). The study of this model could provide fascinating insights into such issues as Quaternary biogeography (Luttikhuizen

* Corresponding author at: Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK.
E-mail address: andreyurch@gmail.com (A.A. Yurchenko).

https://doi.org/10.1016/j.margen.2018.03.007
Received 3 December 2017; Received in revised form 29 March 2018; Accepted 31 March 2018
Available online 12 April 2018

et al., 2003;Väinölä, 2003; Nikula et al., 2007), hybrid speciation (Riginos and Cunningham, 2007), cytonuclear incompatibility (Pante et al., 2012; Saunier et al., 2014) and architecture of hybrid zones (Nikula et al., 2008; Luttikhuizen et al., 2012). However, the genomic resources available for L. *balthica* are scarce: a panel of microsatellite loci (Becquet et al., 2009), 84 AFLP loci (Luttikhuizen et al., 2012), pooled low-coverage transcriptomic reads from 3 populations (Pante et al., 2012), and 6 nearly complete mitochondrial genomes (Saunier et al., 2014). Unfortunately, none of these studies was based on the material from the native range of L. *b. balthica* in the Pacific; only L. *b. rubra* from Western Europe and putatively hybrid clams from the Baltic Sea were examined. Studies of hybridization processes without considering the genomic resources of one of contributing species are likely to be ineffective as well as potentially misleading.

In this work, we present a transcriptome of *L balthica balthica* from the Western Pacific. It was generated using Illumina high-throughput sequencing and consists of the 11,248 annotated transcripts. In particular, we recovered nuclear genes from the oxidative phosphorylation (OXPHOS) pathway highlighted as the genomic outliers in previous population-based studies and potentially involved in mito-nuclear incompatibilities (Pante et al., 2012; Saunier et al., 2014). We also recovered markers commonly used in ecophysiological and toxicological studies of marine mollusks (Werner et al., 2004; Sarkar et al., 2006; Lehtonen et al., 2006; Jarque et al., 2014). Additionally, we compared the two assembled transcriptomes with the previously published transcriptomic data as well as mitochondrial genomes in the context of hybrid architecture, genomic diversity and divergence of the sub(species).

2. Materials and methods

2.1. Sample collection and processing

Two clams (hereinafter K_35 and K_36) were sampled in the Avacha Bay, Eastern Kamchatka (Fig. 1a; Table 1) and stored at −20 °C for 30 min, then at −80 °C in the freezer until the RNA extraction process (no specific permits were required for the collection). RNA was isolated from whole individuals using TRIzol reagent following the manufacturer's instructions (TRI Reagent, Sigma, T9424). DNA was obtained with Mint-2 cDNA synthesis kit according to the manufacturer's instructions (Evrogen, SK005). The samples were treated with a restriction enzyme *Sfi*I (NEB, R0123S) to remove parts of the Mint-2 adapters. The restricted DNA was cleaned up with Cleanup Standart (Evrogen, # BC022).

Table 1
Data descriptions of the samples in compliance with the MIGS standard.

Item	Description
Investigation_type	Eukaryote
Project_name	Transcriptome of the bivalve *Limecola balthica* L. from Western Pacific
Lat_lon	52°57′53″N 158°28′22″E
Geo_loc_name	Russia:Eastern Kamchatka:Avacha Bay
Collection_date	2010–08
Biome	Coastal sea water (ENVO:00002150)
Feature	Bay (ENVO:00000032)
Material	Sandy sediment (ENVO:01000118)
Depth	< 1.5 m
env_package	Water
seq_meth	Illumina HiSeq
Assembly method	SOAPdenovo-Trans
Ploidy	Diploid
BioProject	PRJNA384460
Short Read Archive (SRA) Experiments	SRR5758183-SRR5758186
Transcriptome Shotgun Assembly (TSA)	GGAT00000000

We used 50 ng of each double-stranded cDNA sample to prepare NGS-libraries with Nextera DNA Library Preparation Kit (Illumina, USA). All procedures were conducted in accordance with the protocol given in Nextera DNA Library Preparation Guide (2012). The libraries were sequenced with Illumina HiSeq 2500 system in paired-end mode (100 bp) using TruSeq Rapid SBS Kit HS (Illumina, USA) after the clusterization with TruSeq Rapid PE Cluster Kit HS (Illumina, USA) in accordance with Illumina HiSeq 2500 System User Guide (2014).

2.2. De novo assembly and sequence annotation

The raw reads were searched for the remains of adapters, other technical sequences and poly-A tails with cutadapt 1.12 (Martin, 2011)

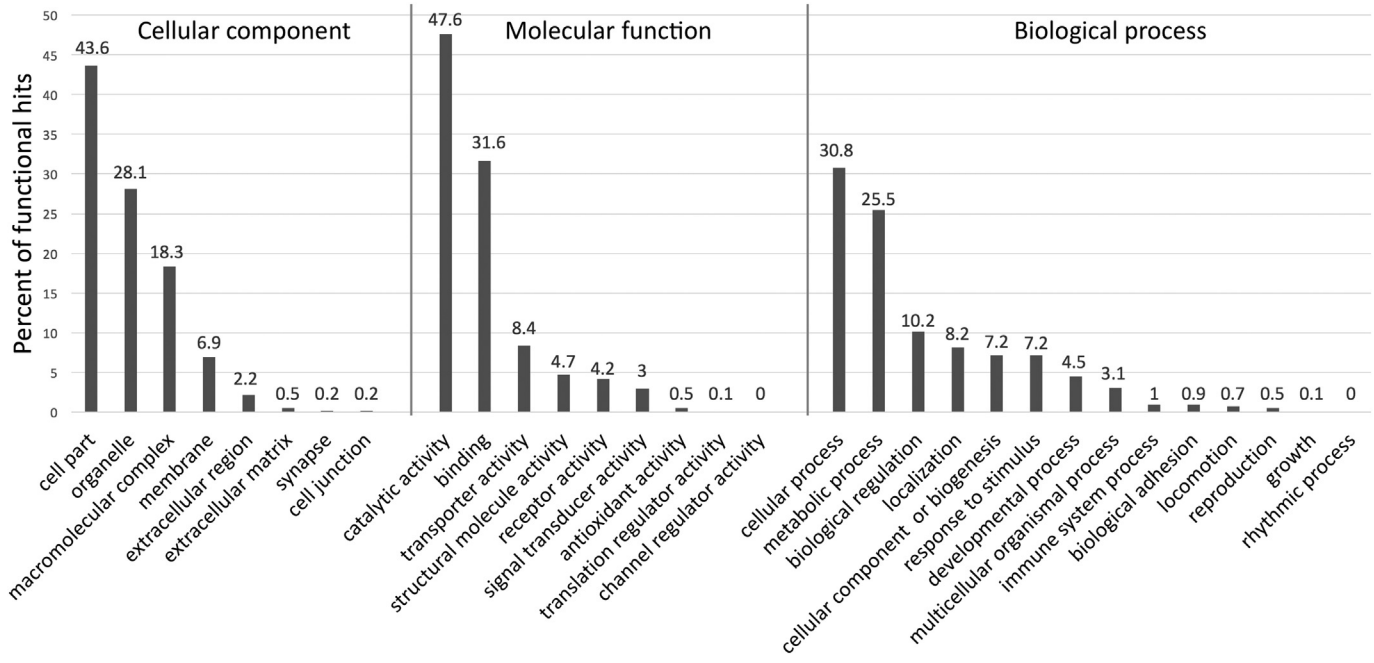

Fig. 1. GO Slim term distribution among the 5638 mapped to the PANTHER database proteins of L. *balthica*.