Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

De novo transcriptome assembly of the calanoid copepod *Neocalanus flemingeri*: A new resource for emergence from diapause

Vittoria Roncalli[a,*], Matthew C. Cieslak[a], Stephanie A. Sommer[a,1], Russell R. Hopcroft[b], Petra H. Lenz[a]

[a] Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA
[b] College of Fisheries and Ocean Sciences, Institute of Marine Science, University of Alaska Fairbanks, 120 O'Neill, Fairbanks, AK 99775-7220, USA

ARTICLE INFO

Keywords:
Trinity
Illumina
CAP3 assembly program
Dormancy
Sub Artic Pacific Ocean

ABSTRACT

Copepods, small planktonic crustaceans, are key links between primary producers and upper trophic levels, including many economically important fishes. In the subarctic North Pacific, the life cycle of copepods like *Neocalanus flemingeri* includes an ontogenetic migration to depth followed by a period of diapause (a type of dormancy) characterized by arrested development and low metabolic activity. The end of diapause is marked by the production of the first brood of eggs. Recent temperature anomalies in the North Pacific have raised concerns about potential negative effects on *N. flemingeri*. Since diapause is a developmental program, its progress can be tracked using through global gene expression. Thus, a reference transcriptome was developed as a first step towards physiological profiling of diapausing females using high-throughput Illumina sequencing. The *de novo* transcriptome, the first for this species was designed to investigate the diapause period. RNA-Seq reads were obtained for dormant to reproductive *N. flemingeri* females. A high quality *de novo* transcriptome was obtained by first assembling reads from each individual using Trinity software followed by clustering with CAP3 Assembly Program. This assembly consisted of 140,841transcripts (contigs). Bench-marking universal single-copy orthologs analysis identified 85% of core eukaryotic genes, with 79% predicted to be complete. Comparison with other calanoid transcriptomes confirmed its quality and degree of completeness. Trinity assembly of reads originating from multiple individuals led to fragmentation. Thus, the workflow applied here differed from the one recommended by Trinity, but was required to obtain a good assembly.

1. Introduction

Planktonic copepods in the family Calanidae like *Calanus* and *Neocalanus* play a key role in the trophodynamics of the subpolar and polar ecosystems by transferring energy from primary producers to higher consumers. Three species of *Neocalanus* (*Neocalanus plumchrus, N. flemingeri, N. cristatus*) dominate the mesozooplankton across the subarctic Pacific Ocean and its marginal seas (Mackas and Tsuda, 1999). *N. flemingeri*, which is abundant in the Gulf of Alaska and Prince William Sound, serves as food source for commercially valuable fishes, such as haddock, pollock, cod, flounder and even salmon for at least part of their life cycle (Willette et al., 1999, 2001). Thus, the success of the Gulf of Alaska fishery depends on the abundance and nutritional quality (lipid content) of copepods such as *N. flemingeri* (Fig. 1).

The life cycle of many marine and terrestrial arthropods includes a period of dormancy, also called diapause, which is a mechanism for survival during an extended period of adverse environmental conditions (Mackas and Tsuda, 1999). Calanid copepods undergo post-embryonic diapause, which is similar to the diapause of mosquitoes (Hirche, 1996; Baumgartner & Tarrant, 2017; Denlinger, 2002). The life cycle of *N. flemingeri* is annual and involves a spring to early summer growth period in the epipelagic zone (0–100 m) followed by ontogenetic migration to deep waters (400–2000 m). At depth, *N. flemingeri* pre-adults undergo a terminal molt into non-feeding adult males and females, which then mate. After mating, the adult males die off and the females enter diapause, a state of arrested development and low metabolic activity. By September, all adult females are in diapause at depth and no males remain. Some time during winter/early spring, females produce multiple batches of eggs, which seed the population during the following spring (Miller and Clemons, 1988; Mackas and Tsuda, 1999). Not much is known about developmental progress from the time *N. flemingeri* females initiate diapause until eggs are released

* Corresponding author at: Békésy Laboratory of Neurobiology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA.
E-mail address: roncalli@hawaii.edu (V. Roncalli).
[1] Current address: Scripps Institution of Oceanography, University of California San Diego, 8622 Kennel Way, La Jolla, CA 92037, USA.

http://dx.doi.org/10.1016/j.margen.2017.09.002
Received 15 July 2017; Received in revised form 6 September 2017; Accepted 7 September 2017

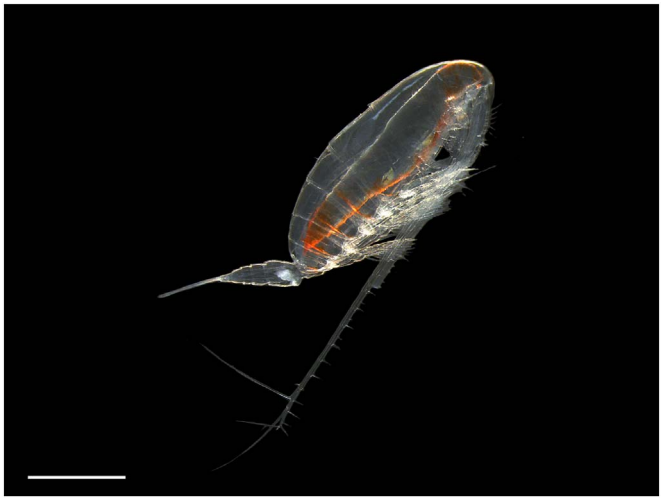

Fig. 1. Lateral view of *Neocalanus flemingeri* adult female (CVI) in diapause. Scale bar: 1 cm.

into the water column.

Diapause is a developmental program that involves changes in physiology as an organism prepares, enters, maintains and terminates diapause (Hirche, 1996). Much of what is known about the control and physiology of post-embryonic diapause has been elucidated in mosquitoes (review: Denlinger, 2002; Sim and Denlinger, 2009; Denlinger and Armbruster, 2014). Recently, these studies have included RNA-Seq approaches to quantify global changes in gene expression associated with the diapause program (*e.g.*, Poelchau et al., 2013a, 2013b; Huang et al., 2015). However, despite the ecological importance of calanid copepods, much less is known about their diapause program (reviews: Hirche, 1996; Baumgartner and Tarrant, 2017). Furthermore, compared with the insects and other copepods, there are several unusual features that characterize the diapause of *N. flemingeri*. Progress in the field has been limited by the challenges of working with marine organisms that inhabit difficult-to-access regions and/or are not amenable to experimental manipulation. However, transcriptomic tools open new opportunities to not only characterize the physiological transitions associated with the annual cycle, but to also recognize potential changes in the developmental program caused by environmental variability.

A recent multi-year temperature anomaly in the North Pacific ("Pacific blob"), detected in late 2013 through early 2016, caused major disruptions to marine communities throughout the eastern North Pacific, including the Gulf of Alaska and Prince William Sound (Peterson et al., 2017; Hu et al., 2017; Kintisch, 2015; Gewin, 2015). Considering that the life cycle of *N. flemingeri* is annual and involves a short growing season followed by a long diapause, questions have arisen on how environmental changes might affect this species. For example, a change in the food supply during the growing season could lead to inadequate lipid stores that fuel metabolic requirements during dormancy and egg production in the spring. In addition, changes in temperature could affect developmental rates in such a way that timing of life history transitions no longer coincides with the production cycle, which is driven by both light and nutrients (Strom et al., 2016).

Here, we developed a new transcriptomic resource for *N. flemingeri* that serves as a base for the investigation of physiological changes related to diapause in adult females. The *de novo* transcriptome, the first for any member of the *Neocalanus* genus, is compared with other publicly available transcriptomes of calanoid copepods, including *Calanus finmarchicus* and *C. glacialis* that undergo a pre-adult diapause, and two species with a facultative embryonic diapause.

Table 1
Summary of *Neocalanus flemingeri* collection details.

NCBI Accession No.	PRJNA324453
Collector	Russell R. Hopcroft; Vittoria Roncalli; Petra H. Lenz
Date of collection	September 2015
Time of collection	8.00 pm
Location	Latitude 60° 32.1′N; Longitude 147° 48.2′W
Depth	700–500 and 500–400 m
Temperature	5 °C
Salinity	33 PSU
Environment	Ocean deep water
Biotic relationship	Free living
Phenotype	*Neocalanus flemingeri* adult female (CVI)

2. Materials and methods

2.1. Sampling strategy

Neocalanus flemingeri adult females were collected during an oceanographic cruise in September 2015 part of a long-term observation program (LTOP) (http://www.sfos.uaf.edu/sewardline/) in Prince Williams Sound (station "PWS2"; Latitude 60° 32.1′N; Longitude 147° 48.2′W). Samples were collected between 700 and 500 and 500–400 m with an opening and closing multiple plankton sampler (0.5 m^2 cross-sectional area; 153 μm mesh nets; Multinet, Hydro-Bios) towed vertically from 700 m depth (Table 1). Plankton collections were immediately diluted with deep seawater, and stored in the dark at 5 °C prior to sorting. Healthy *N. flemingeri* adult females were sorted and either preserved immediately in RNA-Later ("Week 0") or into carboys for transport to University of Alaska Fairbanks. Females were then transferred into 750 mL Falcon tissue-culture flasks containing seawater that had been collected at 600 m depth. All animals were maintained in the dark at 5 °C. Three flasks with four females each were removed from the experiment each week to check for survival, any sign of egg production and preservation in RNA-Later. Weekly collections were made until week 7, which marked the beginning of egg release.

2.2. RNA extraction, cDNA library construction and sequencing

Total RNA was extracted from individual adult females (1 female per time point) using QIAGEN RNeasy Plus Mini Kit (catalog # 74134) in combination with a Qiashredder column (catalog # 79654) following the instructions of the manufacturer and stored at − 80 °C. Total RNA samples were shipped on dry ice to the University of Georgia Genomics Facility (dna.uga.edu) for library preparation and sequencing. Double-stranded cDNA libraries were prepared from total RNA extracted using the Kapa Stranded mRNA-seq kit (KK8420) following manufacturer's instructions. Briefly, RNA samples were first purified with two oligo-dT selection (polyA enrichment using oligodT beds), and then fragmented and reverse transcribed into double-stranded complementary cDNA. Each sample was tagged with an indexed adapter and paired-end sequenced (PE150 bp) using an Illumina NextSeq 500 instrument on a single lane.

2.3. Quality control and de novo assembly

Prior to the assembly, the quality of each RNASeq library (n = 8) was assessed using FASTQC (v1.0.0; Illumina Basespace Labs). For all libraries, FASTQ Toolkit (v.2.0.0; Illumina Basespace Labs) was used to: 1) trim the first 9 bp to remove Illumina adapters (TruSeqLT universal primer); 2) remove low quality reads ("Phred" cutoff score ≥ 30); and 3) set the minimum read length to 50 bp. An average of 5% of low quality reads were removed from each sample, with 15 to 22 million reads remaining per sample (Supplementary file 1).

Two strategies were used for the initial transcriptome assembly: 1) quality-filtered reads obtained for all 8 individuals (Week 0 to Week 7)