Marine Genomics 37 (2018) 114–119

Contents lists available at ScienceDirect

Marine Genomics

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

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
Available online 14 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)

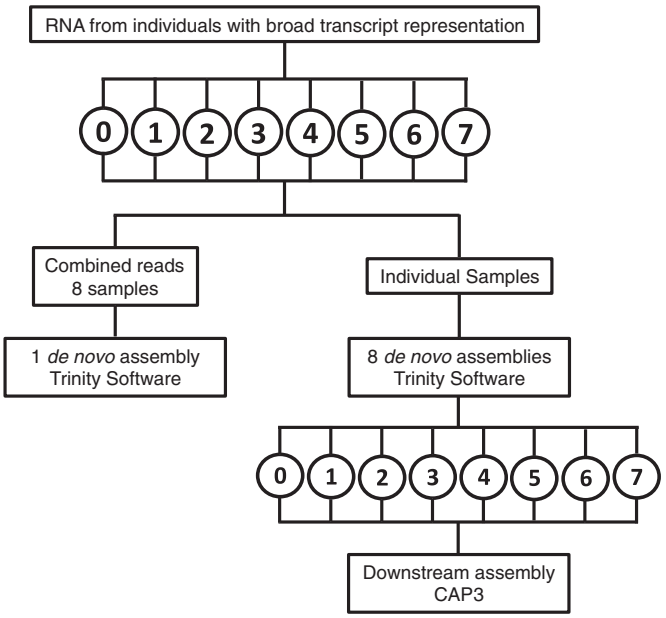

Fig. 2. Diagram of the two workflows used for generating *de novo* transcriptomes from Illumina generated RNA-Seq data for eight *Neocalanus flemingeri* females that represent a range of physiological states starting from dormant to reproductively active. The numbers 0 to 7 gives the time in weeks post collection from 400 to 700 m depth starting with the day of collection ("0").

were combined and assembled using Trinity (Haas et al., 2013) ("Full-Trinity"); and 2) quality-filtered reads for each individual were assembled independently using Trinity (Fig. 2). The second strategy generated eight *de novo* assemblies that were clustered and further assembled using CAP3 Assembly program (Huang and Madan, 1999). For the Trinity assemblies we used version 2.0.6 with initial parameters set to: –seqType fq–CPU 32–max_memory 200G –min_contig_length 300 –normalize_max_read_cov50. Clustering using CAP3 Assembly program, was performed with default settings with an overall percent identity cutoff of 85% (setting: -p $N > 85$). Trinity and the CAP3 Assembly program were run on the National Center for Genome Analysis Support's (NCGAS; Indiana University, Bloomington, IN, USA) Mason Linux cluster. For each individual assembly, mapping (Bowtie2; v2.1.0) (Langmead et al., 2009) and "Bench-marking universal single-copy orthologs" (BUSCO) analysis (v1.22) (Simão et al., 2015) were performed. The software OrthoVenn (Wang et al., 2015) was used to compare the number of annotated orthologous genes across three individual assemblies representing early, mid and late stage females.

2.4. Functional annotation

Transdecoder analysis (default settings; v. 3.0.0; open reading frame > 100 amino acid) (Haas et al., 2013) was performed on the CAP3 assembly. Transcripts with coding regions (CDS) were annotated using BLASTx against the NCBI SwissProt protein database (downloaded on 18th September, 2015) on a local Beowulf Linux computer cluster (Altschul et al., 1997). The resulting BLAST annotations using a maximum E-value of 10^{-3} were then used to retrieve Gene Ontology (GO) terms with UniProt (http://www.uniprot.org/help/uniprotkb).

The assembly and annotation statistics of the *N. flemingeri de novo* transcriptome were compared with those of other calanoid copepods for *C. finmarchicus, C. glacialis, Labidocera madurae* and *Eurytemora affinis* (Lenz et al., 2014; Tarrant et al., 2014; Almada and Tarrant, 2016; Bailey et al., 2017; Roncalli et al., 2017, in press).

Table 2
Summary of assembly statistics and BUSCO analysis for Trinity *de novo* assemblies. For the "Full-Trinity" quality filtered RNA-Seq reads from eight females collected from week 0 (collection day) to week 7 were combined prior to assembly and analysis. Number of quality filtered RNA-Seq reads, Trinity assembly and mapping statistics and BUSCO results for three samples ("Week 0", "Week 3" and "Week 7") of the eight samples are also given. Assembly statistics for the remaining five samples are provided as supplementary material (Supplementary file 2).

		Full-Trinity	Week 0	Week 3	Week 7
Quality cleaned reads (#)		149,805,914	18,798,468	16,605,228	15,637,432
Transcripts (#)	Trinity	245,135	57,804	54,192	58,038
Minimum length (bp)		301	301	301	301
Maximum length (bp)		26,066	24,980	24,663	13,915
N50		557	1089	1105	1086
N25		983	2024	2074	1929
N75		405	595	599	604
Overall mapping (%)	Bowtie	86[a]	89	89	88
Mapping > 1 time (%)		37	33	32	38
Core eukaryotic genes (#)[b]	BUSCO				
Complete (%)		41	62	60	63
Duplicated (%)		8	11	10	11
Fragmented (%)		26	12	13	10
Missing (%)		32	26	27	26

[a] Mapping percentage is the average of the eight samples mapped back to the assembly transcriptome. Mapping statistics for individual samples against this assembly are presented in Supplementary file 2.

[b] "Complete" is defined as a gene with a predicted length that is within two standard deviations of the BUSCO group mean length that get annotation against the "Arthropoda databases" (Simão et al., 2015). "Duplicated" indicates that multiple transcripts annotated to the same core gene such as transcripts with predicted isoforms. "Fragmented genes" refers to transcripts that encode partial proteins.

3. Results and discussion

3.1. De novo assemblies

Comparison between the assemblies showed that in spite of the greater number of transcripts (240 K *vs* 54 K) and the presence of long transcripts (> 20,000), the "Full-Trinity" assembly showed evidence of significant fragmentation compared with the individual assemblies (Table 2, Supplementary file 2). Assembly statistics for N50 and N25 lengths were below 1000 bp, which was approximately half of the equivalent lengths obtained for the single individual assemblies. "Bench-marking universal single-copy orthologs" (BUSCO) analysis confirmed high fragmentation in the "Full-Trinity" assembly with only 40% of complete CORE genes, and the number of fragmented genes being a factor of two higher than that of the individual assemblies (Table 2). Although fragmentation in the "Full-Assembly" is suggested by the N50 and BUSCO values, it is not a result of the poor quality of the sequencing since overall mapping of the reads for this assembly was > 85% (80% defined high quality; Haas personal communication) and was comparable to the single assemblies (Table 2, Supplementary file 2).

Although the eight assemblies were of high quality, each was generated from RNA-Seq data obtained from single individuals that represented a range of physiological states in the transition from dormant to reproductively active. Thus, any single individual assembly was unlikely to have the depth of coverage needed for a reference transcriptome to investigate the physiology of diapause. Comparison of the number of annotated orthologous genes across three individual assemblies showed that while > 14,000 orthologous genes were shared

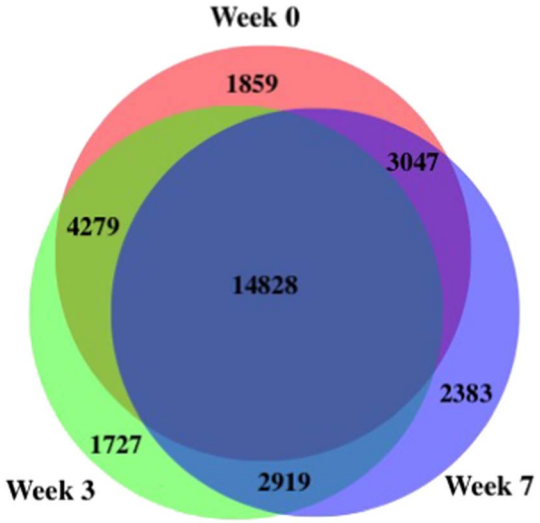

Fig. 3. Proportional Venn diagram showing the degree of overlap of orthologous genes among three *de novo* assemblies generated for individual females collected at three different time points. *De novo* assemblies were compared and orthologous genes were identified using OrthoVenn (Wang et al., 2015). Color coding – red: "Week 0"; green: "Week 3" and purple: "Week 7". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
De novo assembly and annotation statistics for *N. flemingeri* reference transcriptome. The reference was generated by clustering 8 individual *de novo* assemblies ("Week 0" to "Week 7") using the cluster CAP3 Assembly program.

Assembly		
Trinity transcripts (from 8 individual assemblies) (#)		469,806
Clustered contig (#)		140,841
Minimum transcripts length (bp)		301
Longest contig length (bp)		24,981
Total length of all sequence in assembly (bp)		142,589,362
GC Content (%)		44.13
N50 (bp)		1452
Overall mapping (%)		92
Mapping > 1 time (%)		59
Annotation		
Transcripts with coding regions (CDS) (#)	Transdecoder	108,092
Transcripts with BLAST hits (#)	SwissProt	62,126
Transcripts with GO terms (#)	UniProt	59,544
Transcripts with KEGG terms (#)	KEGG	52,555
Core eukaryotic genes (#)	BUSCO	
Complete genes (%)[a]		79
Complete duplicated (%)		33
Fragmented genes (%)		6
Missing genes (%)		15

[a] "Complete" is defined as a gene with a predicted length that is within two standard deviations of the BUSCO group mean length that get annotation against the "Arthropoda databases" (Simão et al., 2015). "Duplicated" indicates that multiple transcripts annotated to the same core gene such as transcripts with predicted isoforms. "Fragmented genes" refers to transcripts that encode partial proteins.

among the "Week 0", "Week 3" and "Week 7" assemblies, a large number of genes were missing from either one or two assemblies (Fig. 3). However, the negative effect of combining RNA-Seq data from multiple individuals was already apparent with two samples with a 10% decline in N50 length. This decline increased to > 30% when four samples were assembled together. In order to remove redundancy while preserving sequencing depth across the eight assemblies an additional clustering step using the CAP3 Assembly program was added to the workflow (Fig. 2, Table 3). The resulting clustered transcriptome had > 140 K transcripts with an N50 length > 1400 bp, and the longest transcripts being > 14,000 nucleotides long (Table 3). > 90% of the reads mapped back to the transcriptome although the number of ambiguous reads was high with > 50% of the reads mapped more than once (Table 3).

3.2. Functional annotation

Transdecoder analysis on the CAP3 assembly identified 77% of the transcripts with coding regions (CDS) (Table 3). More than 50% of these transcripts were successfully annotated against SwissProt, and over 95% of those were annotated with GO terms. Within the "biological process" (BP) category, *N. flemingeri* transcripts covered broad conserved eukaryotic processes with "cellular process", "metabolic process" and "single-organism process" representing the three top gene ontology (GO) terms (Supplementary file 3). Approximately 80% (2102) of the BUSCO complete orthologs genes were identified in the reference, with only 6% being fragmented (Table 3), an outcome much better than the results for either the individual or the combined read "Full-Trinity" assembly.

3.3. Comparison with other calanoid copepods

The number of assembled transcripts and longest transcript are comparable across *de novo* transcriptomes with the *N. flemingeri* being most similar to the *C. finmarchicus* GOM transcriptome (Table 4). Based on the BUSCO analysis, the number of predicted core proteins was similar across the transcriptomes of all species with an approximate coverage of 72 to 79% of complete genes (Table 4). The annotation statistics suggests that *N. flemingeri* transcriptome is of at least similar quality and depth as these others.

It should be noted, however, that the high-quality assembly was not obtained using the more common workflow of assembling reads from multiple samples (Haas et al., 2013). The version of Trinity used here is designed to preserve isoforms derived from either polymorphisms or splice variants (Haas et al., 2013; https://github.com/trinityrnaseq/trinityrnaseq/wiki). The geographic distribution of *N. flemingeri* spans the subarctic North Pacific with an estimated population abundance of 10^9 individuals per km^2 (Miller and Clemons, 1988; Yoshiki et al., 2013), and sub-populations are connected *via* ocean currents (Mackas and Coyle, 2005; Gibson et al., 2013; Coyle et al., 2013). Thus, even individuals collected from a single station may have different origins and be genetically diverse, which could contribute to the fragmentation of the assembly. Variant calling analysis of the RNASeq data for the individuals used in the reference transcriptome (using the workflow described in: http://samtools.sourceforge.net/mpileup.shtml) identified > 300 K SNPs, that corresponds to a frequency of 2.7 SNPs per 1000 bp. This number is higher than what was reported for the human genome where the expected frequency is 1 SNP per 1000 bp (1000 Genomes Project Consortium, 2010). The genetic diversity among the *N. flemingeri* females may have contributed to the fragmentation of the assembly.

4. Conclusions

The transcriptomic resource described here is the first for this species, and it is the first resource generated to specifically elucidate the developmental changes associated with diapause and emergence from diapause. Obtaining a high-quality transcriptome required a modified workflow involving assemblies of RNA-Seq data obtain from individuals followed by the clustering of eight assemblies using the CAP3 Assembly program. This resource was developed to provide a reference for high-throughput gene expression profiling. It has the potential to be used not only to elucidate the basic physiology of the diapause program in *N. flemingeri*, but also to record physiological differences between individuals collected in different years or from different locations.

Supplementary data to this article can be found online at http://dx.

Table 4
Comparison of *de novo* assemblies for calanoid copepods.

	N. flemingeri[a]	*C. finmarchicus*		*C. glacialis*[d]	*L. madurae*[e]	*Eurytemora affinis*[f]
		GOM[b]	NOR[c]			
Sequencing platform	Illumina NextSeq	Illumina HiSeq	Illumina HiSeq	Illumina HiSeq	Illumina NextSeq	Illumina HiSeq
Assembly software	Trinity & CAP3	Trinity	Trinity	CLC	Trinity	Trinity
Transcripts (#)	140,841	206,041	241,140	59,353	211,002	138,581
Minimum Length (bp)	301	301	200	200	301	201
Maximum Length (bp)	24,981	23,068	25,048	13,363	23,836	23,627
N50	1452	1418	988	1019	1184	2087
Transcripts with BLAST hits (#)	62,126	81,554		*np*	62,980	*np*
Transcripts with GO terms (#)	59,544	78,467	*np*	19,790	60,097	*np*
KEGG	52,555	67,673	*np*	*np*	57,192	
BUSCO						
Complete (%)	79	79	72	75	76	
Duplicated (%)	34	20	14	4	0.2	
Fragmented (%)	6	8	13	18	11	
Missing (%)	15	12	14	6	12	

np: Not provided in the study.
[a] Current study.
[b] Data from (Lenz et al., 2014). BLAST and BUSCO analysis were performed in 2017 using publicly accessible NCBI "transcriptome shotgun assembly" (GAXK00000000).
[c] Data from (Tarrant et al., 2014). BLAST and BUSCO analysis were performed in 2017 using publicly accessible NCBI "transcriptome shotgun assembly" (GBFB01000000).
[d] Data from (Bailey et al., 2017).
[e] Data from: Roncalli et al., 2017, in press.
[f] Data from (Almada and Tarrant, 2016). BUSCO analysis was performed in 2017 using publicly accessible NCBI "transcriptome shotgun assembly" (GBGO00000000).

doi.org/10.1016/j.margen.2017.09.002.

Availability of supporting data

N. flemingeri sequence data have been submitted to the National Center of Biotechnology Information under the Bioproject PRJNA324453. The raw sequence reads (SRA) for the eight biological replicates has been deposited at DDBJ/EMBL/GenBank under the accession Nos. SRR5873554, SRR5873555, SRR5873556, SRR5873557, SRR5873558, SRR5873559, SRR5873560, SRR5873561. This Transcriptome Shotgun Assembly (TSA) has been deposited at DDBJ/EMBL/GenBank under the accession GFUD00000000.

Authors' contribution

VR, PHL and RRH conceived and performed the experiments; VR, MCC, SAS analyzed the data and evaluated the conclusions; VR and PHL wrote the manuscript. All authors approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We would like to thank A. E. Christie, D. K. Hartline and L. Hata from the University of Hawaii at Manoa, M. Belanger and R. Nilsen from the Georgia Genomics Facility at the University of Georgia, C. Clark and Catilin Smoot from University of Alaska Fairbanks. We would like to thank the crew of the USFWS R/V Tiglax for at sea support. This research was supported by National Science Foundation Grant OCE-1459235 to PHL and Andrew E Christie and National Science Foundation Grant OCE-1459826 to RRH. Additional support was provided by the National Science Foundation NCGAS under Grants DBI-1458641 and ABI-1062432 to Indiana University. The views expressed herein are those of the authors and do not reflect the views of the funding agencies. This is the University of Hawaii at Manoa School of Ocean and Earth Science and Technology contribution Number XXXX.

References

1000 Genomes Project Consortium, 2010. A map of human genome variation from population scale sequencing. Nature 467 (7319), 1061.
Almada, A.A., Tarrant, A.M., 2016. Vibrio elicits targeted transcriptional responses from copepod hosts. FEMS Microbiol. Ecol. 92 (6) (fiw0).
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17), 3389–3402. http://dx.doi.org/10.1093/Nar/25.17.3389.
Bailey, A., De Wit, P., Thor, P., Browman, H.I., Bjelland, R., Shema, S., Fields, D.M., Runge, J.A., Thompson, C., Hop, H., 2017. Regulation of gene expression is associated with tolerance of Artic copepod *Calanus glacialis* to CO_2-acidified water. Ecol. Evol. http://dx.doi.org/10.1002/ece3.3063.
Baumgartner, M.F., Tarrant, A.M., 2017. The physiology and ecology of diapause in marine copepods. Annu. Rev. Mar. Sci. 9, 387–411.
Coyle, K.O., Gibson, G.A., Hedstrom, K., Hermann, A.J., Hopcroft, R.R., 2013. Zooplankton biomass, advection and production on the northern Gulf of Alaska shelf from simulations and field observations. J. Mar. Syst. 128, 185–207.
Denlinger, D.L., 2002. Regulation of diapause. Annu. Rev. Entomol. 47 (1), 93–122.
Denlinger, D.L., Armbruster, P.A., 2014. Mosquito diapause. Annu. Rev. Entomol. 59, 73–93.
Gewin, V., 2015. North Pacific 'blob' stirs up fisheries management. Nature 524 (7566), 396.
Gibson, G.A., Coyle, K.O., Hedstrom, K., Curchitser, E.N., 2013. A modeling study to explore on-shelf transport of oceanic zooplankton in the Eastern Bering Sea. J. Mar. Syst. 121, 47–64.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (8), 1494–1512. http://dx.doi.org/10.1038/Nprot.2013.084.
Hirche, H.J., 1996. Diapause in the marine copepod, *Calanus finmarchicus*—a review. Ophelia 44 (1–3), 129–143.
Hu, Z.Z., Kumar, A., Jha, B., Zhu, J., Huang, B., 2017. Persistence and predictions of the remarkable warm anomaly in the Northeastern Pacific Ocean during 2014–16. J. Clim. 30 (2), 689–702.
Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9 (9), 868–877.
Huang, X., Poelchau, M.F., Armbruster, P.A., 2015. Global transcriptional dynamics of diapause induction in non-blood-fed and blood-fed *Aedes albopictus*. PLoS Negl. Trop. Dis. 9 (4), e0003724.
Kintisch, E., 2015. 'The Blob' invades Pacific, flummoxing climate experts. Science 348 (6230), 17–18.
Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, 3. http://dx.doi.org/10.1186/Gb-2009-10-3-R25.
Lenz, P.H., Roncalli, V., Hassett, R.P., Wu, L.S., Cieslak, M.C., Hartline, D.K., et al., 2014. *De novo* assembly of a transcriptome for *Calanus finmarchicus* (Crustacea, Copepoda) — the dominant zooplankter of the North Atlantic Ocean. PLoS One 9 (2), e88589. http://dx.doi.org/10.1371/journal.pone.0088589.
Mackas, D.L., Coyle, K.O., 2005. Shelf–offshore exchange processes, and their effects on

mesozooplankton biomass and community composition patterns in the northeast Pacific. Deep Sea Res. II 52 (5), 707–725.
Mackas, D.L., Tsuda, A., 1999. Mesozooplankton in the eastern and western subarctic Pacific: community structure, seasonal life histories, and interannual variability. Prog. Oceanogr. 43 (2), 335–363.
Miller, C.B., Clemons, M.J., 1988. Revised life history analysis for large grazing copepods in the subarctic Pacific Ocean. Prog. Oceanogr. 20, 293–313.
Peterson, W.T., Fisher, J.L., Strub, P.T., Du, X., Risien, C., Peterson, J., Shaw, C.T., 2017. The pelagic ecosystem in the northern California current off Oregon during the 2014–2016 warm anomalies within the context of the past 20 years. J. Geophys. Res. http://dx.doi.org/10.1002/2017JC012952.
Poelchau, M.F., Reynolds, J.A., Elsik, C.G., Denlinger, D.L., Armbruster, P.A., 2013a. RNA-Seq reveals early distinctions and late convergence of gene expression between diapause and quiescence in the Asian tiger mosquito, *Aedes albopictus*. J. Exp. Biol (jeb-089508).
Poelchau, M.F., Reynolds, J.A., Elsik, C.G., Denlinger, D.L., Armbruster, P.A., 2013b. Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, *Aedes albopictus*. Proc. R. Soc. Lond. Ser. B Biol. Sci. 280 (1759), 20130143.
Sim, C., Denlinger, D.L., 2009. A shut-down in expression of an insulin-like peptide, ILP-1, halts ovarian maturation during the overwintering diapause of the mosquito *Culex pipiens*. Insect Mol. Biol. 18 (3), 325–332.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31 (19), 3210–3212. http://dx.doi.org/10.1093/bioinformatics/btv351.
Strom, S.L., Fredrickson, K.A., Bright, K.J., 2016. Spring phytoplankton in the eastern coastal Gulf of Alaska: photosynthesis and production during high and low bloom years. Deep Sea Res. 132, 107–121.
Tarrant, A.M., Baumgartner, M.F., Hansen, B.H., Altin, D., Nordtug, T., Olsen, A.J., 2014. Transcriptional profiling of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod *Calanus finmarchicus*. Front. Zool. 11 (1). http://dx.doi.org/10.1186/s12983-014-0091-8.
Wang, Y., Coleman-Derr, D., Chen, G., Gu, Y.Q., 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43, W78–W84. Web Server issue. http://dx.doi.org/10.1093/nar/gkv487.
Willette, T.M., Cooney, R.T., Hyer, K., 1999. Predator foraging mode shifts affecting mortality of juvenile fishes during the subarctic spring bloom. Can. J. Fish. Aquat. Sci. 56 (3), 364–376.
Willette, T.M., Cooney, R.T., Patrick, V., Mason, D.M., Thomas, G.L., Scheel, D., 2001. Ecological processes influencing mortality of juvenile pink salmon (*Oncorhynchus gorbuscha*) in Prince William Sound, Alaska. Fish Oceanogr. 10 (s1), 14–41.
Yoshiki, T.M., Chiba, S., Sugisaki, H., Sasaoka, K., Ono, T., Batten, S., 2013. Interannual and regional variations in abundance patterns and developmental timing in mesozooplankton of the western North Pacific Ocean based on Continuous Plankton Recorder during 2001–2009. J. Plankton Res. 35 (5), 993–1008.