

# Carotenoid metabolic profiling and transcriptome-genome mining reveal functional equivalence among blue-pigmented copepods and appendicularia

NAZIA MOJIB,\* MAAN AMAD,† MANJULA THIMMA,‡ NAROA ALDANONDO,\*<sup>1</sup>  
MANDE KUMARAN‡ and XABIER IRIGOIEN\*

\*Red Sea Research Center, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal 23955-6900, Saudi Arabia, †Analytical Core Laboratory, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal 23955-6900, Saudi Arabia, ‡Biosciences Core Laboratory, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal 23955-6900, Saudi Arabia

## Abstract

The tropical oligotrophic oceanic areas are characterized by high water transparency and annual solar radiation. Under these conditions, a large number of phylogenetically diverse mesozooplankton species living in the surface waters (neuston) are found to be blue pigmented. In the present study, we focused on understanding the metabolic and genetic basis of the observed blue phenotype functional equivalence between the blue-pigmented organisms from the phylum Arthropoda, subclass Copepoda (*Acartia fossae*) and the phylum Chordata, class Appendicularia (*Oikopleura dioica*) in the Red Sea. Previous studies have shown that carotenoid–protein complexes are responsible for blue coloration in crustaceans. Therefore, we performed carotenoid metabolic profiling using both targeted and nontargeted (high-resolution mass spectrometry) approaches in four different blue-pigmented genera of copepods and one blue-pigmented species of appendicularia. Astaxanthin was found to be the principal carotenoid in all the species. The pathway analysis showed that all the species can synthesize astaxanthin from  $\beta$ -carotene, ingested from dietary sources, via 3-hydroxyechinenone, canthaxanthin, zeaxanthin, adonirubin or adonixanthin. Further, using de novo assembled transcriptome of blue *A. fossae* (subclass Copepoda), we identified highly expressed homologous  $\beta$ -carotene hydroxylase enzymes and putative carotenoid-binding proteins responsible for astaxanthin formation and the blue phenotype. In blue *O. dioica* (class Appendicularia), corresponding putative genes were identified from the reference genome. Collectively, our data provide molecular evidences for the bioconversion and accumulation of blue astaxanthin–protein complexes underpinning the observed ecological functional equivalence and adaptive convergence among neustonic mesozooplankton.

**Keywords:** astaxanthin, carotenoid-binding proteins, high-resolution mass spectrometry, mesozooplankton, transcriptome,  $\beta$ -carotene hydroxylase

Received 2 February 2014; revision received 24 April 2014; accepted 29 April 2014

## Introduction

Copepods and appendicularians are the two most abundant mesozooplankton grazers in the marine pelagic

environment. Due to the decline of primary production with depth, zooplankton needs to feed in the surface layers for at least a restricted period (Lampert 1989). They have to take the risk of staying on the surface for food where they are subjected to increased risk of biological stressors such as ultraviolet radiation (UVR) and predation (Johnsen & Jakobsen 1987). One of the physiological adaptation that may help zooplankton to attenuate the effects of UVR is the accumulation of either

Correspondence: Nazia Mojib,  
E-mail: Nazia.Mojib@kaust.edu.sa

<sup>1</sup>Present address: Marine Research Unit, AZTI-Tecnalia, Herrera Kaia Portualdea z/g, Pasaia, 20110, Basque Country, Spain

UVR absorbing compounds such as mycosporine-like amino acids or carotenoid pigments such as astaxanthin that may impart red or blue colour depending on the bound or free form (Hairston 1976; Persaud *et al.* 2007). The predominance of surface-dwelling blue-pigmented organisms was observed for the first time in living hauls sampled by special neuston nets in the Indian Ocean (Herring 1965). Among other wide varieties of organisms, the blue pigment in pontellid copepods was found to be a chromoprotein complex of carotenoid (astaxanthin) and protein. The blue carotenoproteins are widely distributed in more or less transparent neustonic zooplankton and may be involved in a wide range of functions including camouflage, photosensitivity, antioxidant activity and to some extent in the development (Cheeseman *et al.* 1967; Goodwin 1976). At individual level, studies have been carried out to characterize the carotenoproteins and their function in marine animals (Cheeseman *et al.* 1967; Maoka 2011). However, community-based study elucidating the underlying mechanism and functioning of blue pigment among zooplankton sharing the same niche and undergoing similar selection pressure has not been carried out yet.

Marine metazoans cannot synthesize carotenoids *de novo* and, therefore, accumulate them from food such as algae and modify them through metabolic reactions (Liaaen-Jensen 1998). Carotenoids and their metabolites found in these animals provide information about the food chain as well as metabolic pathways. Free astaxanthin and its esterified forms are the principal carotenoids present among crustaceans and specifically copepods (Fisher *et al.* 1964; Bandaranayake & Gentien 1982; Liaaen-Jensen 1998). It has long been known that copepods are capable of metabolically transforming  $\beta$ -carotene into astaxanthin via echinenone and canthaxanthin (Goodwin 1971), but the genes or enzymes involved in this metabolic conversion in copepods is not investigated yet. In several bacteria, algae and plants, two different enzymes exist: CrtZ ( $\beta$ -carotene, 3,3'-hydroxylase) that introduces a hydroxyl group at the 3 (and 3') positions of each of the two  $\beta$ -ionone rings of  $\beta$ -carotene, and CrtW ( $\beta$ -carotene ketolase) that introduces keto groups at carbons 4 and 4' of the  $\beta$ -ionone rings. These two enzymes sequentially convert  $\beta$ -carotene into astaxanthin (Martin *et al.* 2008). Another enzyme called CrtS ( $\beta$ -carotene hydroxylase of the P-450 monooxygenase family) from *Xanthophyllomyces dendrorhous* (yeast) encodes for both hydroxylase and ketolase activities (Ojima *et al.* 2006).

Ubiquitous nature of the highly hydrophobic and reactive carotenoids such as astaxanthin in aqueous cellular environment is due to their ability to form complexes with other molecules, such as proteins or lipids to improve stability and solubility. Carotenoids form

low molecular weight (between 20 and 100 kDa) water-soluble complexes with proteins, and this interaction is highly specific and stoichiometric (Britton & Halliwell 2008). In the course of evolution, nature has equipped marine organisms to manipulate and exploit this unique stable interaction between carotenoids and proteins to enhance the properties of carotenoids in order to adapt in specific environments (Pilbrow *et al.* 2012). Large multimeric protein complex  $\alpha$ -crustacyanin ( $\alpha$ -CRCN), found in the carapaces of crustaceans, where free astaxanthin colour is converted from the red to blue is a well-documented example that provides the molecular basis for camouflage against predation (Rao 1985).  $\alpha$ -CRCN is a large complex of eight heterodimeric  $\beta$ -CRCN units where the dimer is formed by subunits CRCN-A and CRCN-C in association with two astaxanthin molecules (Chayen *et al.* 2003). The formation of this complex results in the bathochromic shift in the emission spectrum of astaxanthin from red ( $\lambda_{\max} = 472$  nm in hexane) to purple as seen in  $\beta$ -CRCN ( $\lambda_{\max} = 580$ – $590$  nm) or blue in the case of  $\alpha$ -CRCN ( $\lambda_{\max} = 632$  nm) (Buchwald & Jencks 1968). CRCNs belong to the functionally diverse widespread lipocalin superfamily of proteins that are capable of binding small hydrophobic molecules (Flower 1996). The CRCN gene evolved from a duplicated ApoD (Apolipoprotein D) lipocalin ancestor and its occurrence is restricted to the class malacostraca of the subphylum crustacea (Wade *et al.* 2009), that subsequently led to the evolution of patterned shell colours.

The occurrence and chemistry of blue astaxanthin-protein complexes along with the genetic mechanisms had not been investigated and analysed in other invertebrate taxa specifically in copepods and appendicularia. It will be interesting to know if lipocalin proteins outside the malacostraca clade, that is, before the duplication event where they gained the neofunctional property of exhibiting arrays of shell colours, retained their property of binding and manipulating the physical and chemical properties of astaxanthin to show a blue phenotype. In the present study, we have used an integrative approach of high-resolution mass profiling and *in silico* analyses in neustonic mesozooplankton from the Red Sea to investigate the chemistry and genetics behind the exhibition of their blue phenotypes. We focused on two abundant surface grazers and phylogenetically well differentiated blue-pigmented organisms from the phylum Arthropoda (subclass Copepoda) and the phylum Chordata (class Appendicularia). The high-resolution mass profiling was used to examine the presence of astaxanthin and other carotenoid intermediates involved in the conversion from  $\beta$ -carotene to astaxanthin. Further, using bioinformatics based data mining from *de novo* assembled transcriptome and genome, putative  $\beta$ -carotene

hydroxylases and lipocalin class of carotenoid-binding proteins were also identified.

## Materials and methods

### Sample collection for metabolic profiling

Mesozooplankton were collected from the top 1 m of the water column using a Manta neuston net (150  $\mu\text{m}$  mesh) in 10–15 min long tows from the Red Sea (22.4575°N, 39.0305°E) in May 2012. The contents of the cod ends were poured onto a 150  $\mu\text{m}$  sieve, washed with filtered seawater and immediately snap-frozen in liquid nitrogen and subsequently stored at -80 °C.

### Species identification and sample preparation

In the laboratory, each blue-pigmented organism used in this study was morphologically identified (Fig. 1). A neighbor-joining, bootstrapped phylogenetic tree was constructed using 18S rRNA gene sequences available in National Center for Biotechnology Information (NCBI) database with MEGA 5.2.2 software (Tamura *et al.* 2011). The organisms were then sorted under the dissecting microscope (Zeiss Stemi 2000; Carl Zeiss Microscopy, Jena, Germany) from the frozen samples. The sorting process was carried out in dim light using micropipette capillary and forceps taking care to minimize the

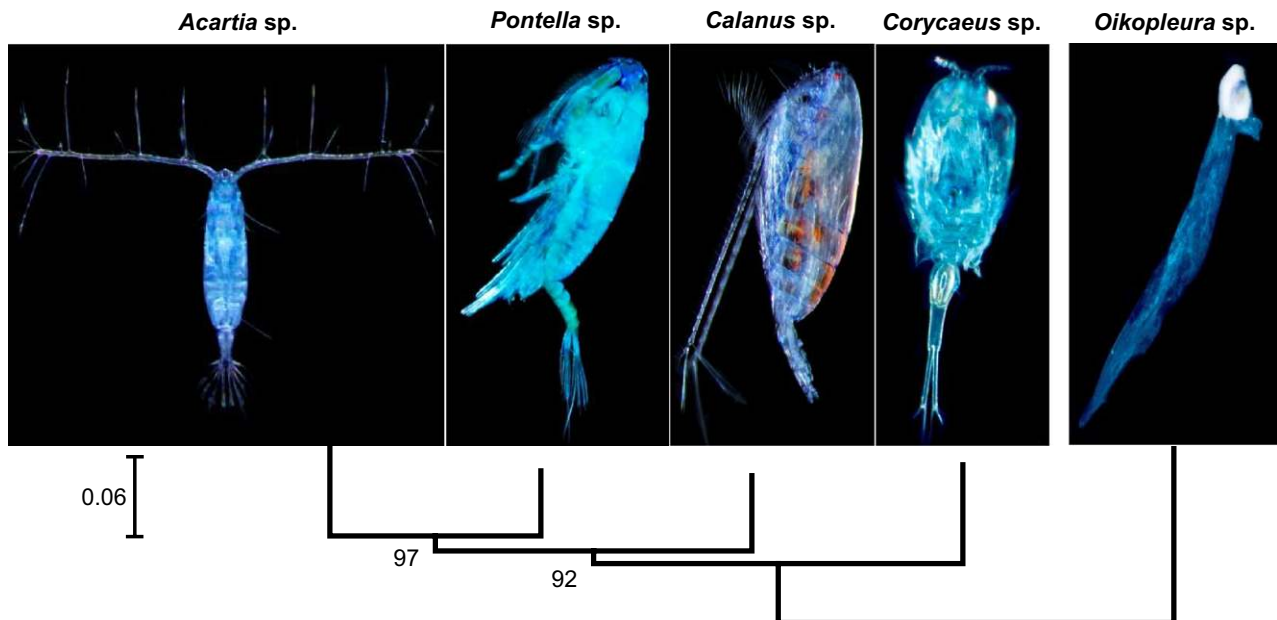
illumination and picking time as much as possible for downstream processing. For carotenoid extraction, approximately 50 individuals from each species were used per sample. The organisms were homogenized and carotenoids were extracted thrice in methanol. The blue-coloured methanol fraction containing dissolved carotenoids was then filtered through membrane (0.45  $\mu\text{m}$ ) and immediately analysed on high-performance liquid chromatography (HPLC).

### Chemicals and reagents

LC-MS-grade methanol and 2-propanol were purchased from Fisher Scientific (Suwanee, GA, USA), beta-carotene, astaxanthin, fucoxanthin, canthaxanthin from Sigma Aldrich (St. Louis, MO, USA) and zeaxanthin from VWR International (UK). The 1 mg/mL stock solutions for each standard were prepared in dichloromethane (DCM) (Fisher Scientific) separately in 2-mL glass vials. The 1 ng/ $\mu\text{L}$  working solution of each standard was prepared in 9:1 (*v/v*) methanol/water.

### HPLC-UV-MS/MS analysis

Carotenoid analysis was conducted using a Transcend series HPLC system (Thermo Scientific, Waltham, MA, USA) equipped with a photo diode array detector coupled to a TSQ Vantage triple quadrupole mass analyzer



**Fig. 1** The photomicrographs of the blue-pigmented copepods and appendicularia used in this study along with their phylogenetic relationship based on 18S rRNA sequences available in NCBI database (Table S2, Supporting information). The neighbor-joining tree is constructed with a sequence alignment of 1649 nt with bootstrap values (1000 replicates) using MEGA 5.2.2 software (Tamura *et al.* 2011). Photographs courtesy: Joan Costa Garcia.

equipped with atmospheric pressure chemical ionization (APCI) interface (Thermo Scientific). LC separation of the carotenoids was performed using an Eclipse XDB C18, 4.6 × 150 mm, 3.5 μm particle size column (Agilent Technologies, Santa Clara, CA, USA). The separation was achieved using an isocratic mobile phase composed of methanol/isopropanol (95:5, *v/v*). The column was maintained at 25 °C and the injection volume was 10 μL. The flow rate was set to 500 μL/min for the first 5 min of the run and then increased to 900 μL/min for another 15 min. The photodiode array (PDA) detector was set at 480 nm to record the peaks, and the UV-Vis spectra were recorded from 200 to 600 nm. Positive APCI ionization mode was used to detect the carotenoids, and instrumental parameter conditions were set as follows: sheath and auxiliary gasses at 60 and 30 arbitrary units, respectively, vaporizer temperature of 350 °C, source capillary temperature of 250 °C, corona discharge current of 5 μA and argon collision cell gas pressure of 1.5 mTorr. The mass spectrometer was operated in multiple reaction monitoring (MRM) mode tracing the specific transition of the specific precursor and product ions of each compound (Table S1, Supporting information). The TSQ Vantage triple quadrupole mass spectrometer was operated using the XCALIBUR software. Carotenoid characterization was performed by comparing the retention times (RTs) of the specific selected reaction monitoring chromatograms of carotenoids in reference to the available reference standards (β-carotene, astaxanthin, zeaxanthin, canthaxanthin and fucoxanthin). For each sample analysis, four biological replicates of each sample were run thrice (three technical replicates).

#### High-resolution mass spectrometry analysis

Accurate mass data were obtained using a Thermo LTQ Orbitrap mass spectrometer (Thermo Scientific) equipped with an APCI ion source, using the same HPLC and APCI source conditions as above. A full mass scan was acquired in the Orbitrap. The scan mass range was set to cover *m/z* of 200 to 1500 with a resolution of 100 000. Lock-mass option was enabled in all measurements and external mass calibration was used. The *m/z* calibration of the LTQ Orbitrap analyzer was performed with positive ESI ionization mode using a solution containing caffeine, MRFA (H-Met-Arg-Phe-Ala-OH peptide) and Ultramark 1621 as calibrators, according to the manufacturer's guidelines. Carotenoid identification in each sample was performed as discussed earlier using the XCALIBUR software. Tentative identification for other carotenoid compounds was based on accurate mass data, molecular formula assignment and isotopic distribution pattern matching as well as elution order of the identified carotenoids. The

specificity of the analysis was achieved by applying a mass window of 5 ppm to the theoretical mass of the analytes.

#### RNA isolation and sequencing

To generate the transcriptome of *A. fossae*, contents of the cod end of one of the plankton tows was brought back to the laboratory in seawater. After checking that the copepods were alive and active, blue *A. fossae* were sorted under a dissecting microscope, added directly to Qiazol lysis reagent (Qiagen, St. Louis, MO, USA) and homogenized using RNase-free pestle. Total RNAs were separated and precipitated with chloroform and isopropanol, respectively, followed by washing the RNA pellets with 80% ethanol and redissolved in nuclease-free water. Total RNAs were cleaned and concentrated with RNeasy Mini columns (Qiagen). The isolated RNAs were quality-checked using Bioanalyzer (Agilent Technologies) and NanoDrop (Thermo Scientific) to ensure that RNA quality index (RQI) values were >8.0 prior to library creation and sequencing by the KAUST Biosciences Core lab. One microgram of the total RNA was used to construct one RNA-Seq library with average insert sizes of 250–300 bp using Truseq™ RNA Sample Prep Kit (Illumina, San Diego, CA, USA) following manufacturer's recommendations. The transcriptome was pair-end sequenced generating 202 bp reads using the HiSeq™ 2000 sequencing system (Illumina) by following standard protocols. RNA-Seq data are available in NCBI Short read Archives (SRA) under the accession no. SRP036139.

#### Data processing, transcriptome assembly and annotation

Raw reads were quality trimmed using Trimmomatic (Lohse *et al.* 2012) and all the adapter sequences were removed. Ribosomal RNA (rRNA) reads were filtered out using the SortMeRNA pipeline (Kopylova *et al.* 2012). Trimmed and rRNA-filtered mRNA reads were de novo assembled using the TRINITY 2012-03-17-IU\_ZIH\_TUNED software (<http://trinityrnaseq.sourceforge.net/>) standard protocol (Haas *et al.* 2013). The obtained transcripts were then annotated using Trinotate workflow bundled with the TRINITY software package.

#### Homology search

For the identification of β-carotene hydroxylase homologues in *A. fossae*, known CrtZ, CrtW and CrtS protein sequences from different organisms were used as queries for a BLASTX (Altschul *et al.* 1990) search against *A. fossae* transcriptome assembly with E-value cut-off 10<sup>-3</sup> and low complexity filtered (Kuchibhatla *et al.*

2014). All hits were translated and checked manually for sequence similarity to the target query. Similarly, in *O. dioica*,  $\beta$ -carotene hydroxylase homologues were identified by BlastP search against 17 152 predicted proteins from its published genome (ASM20953v1) (Denoeud *et al.* 2010). Likewise, for the identification of carotenoid-binding proteins, protein sequences from subunits CRCN-C and CRCN-A from Class malacostraca were used as queries to Blast search against *A. fossae* transcriptome assembly and *O. dioica* genome using the same parameters. The abundance estimation or expression of all identified homologous transcripts in the blue *A. fossae* transcriptome assembly was carried out using the RSEM (RNA-Seq by Expectation Maximization) software (Li & Dewey 2011) bundled with the Trinity software package. The expression was measured as FPKM (fragments per kilobase of transcript per million mapped reads). The FASTA sequences for all the identified homologous proteins used for analysis in this study are provided in Table S2, Supporting information.

#### Phylogenetic placement and homology modelling

All the identified sequences were phylogenetically placed with the homologues from different taxa. Phylogenetic trees were constructed using Bayesian inference as method of analysis (Huelsenbeck & Ronquist 2001). For Bayesian analysis, the best-fixed evolutionary Whelan and Goldman (WAG) model (Whelan & Goldman 2001) was selected using the MEGA 5.2.2 software (Tamura *et al.* 2011). A gamma shape parameter with a fraction of invariant sites with four discrete values was used to calculate the log-likelihood of each site on the given tree. The final tree was visualized and edited with FIGTREE v1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>). The related data files are available on Dryad (<http://doi.org/10.5061/dryad.mm18b>).

The tertiary structures of the identified putative carotenoid-binding proteins were generated using the MODELLER software from HHPRED alignments on HHpred servers (Soding *et al.* 2005; Eswar *et al.* 2006). Verify3D was used where compatibility of an atom model (3D) is analysed with its own amino acid sequence (1D) to check the quality of the models generated (Luthy *et al.* 1992). The 3-D structures were visualized and superimposed using the Matchmaker application of the UCSF CHIMERA software (Pettersen *et al.* 2004).

## Results

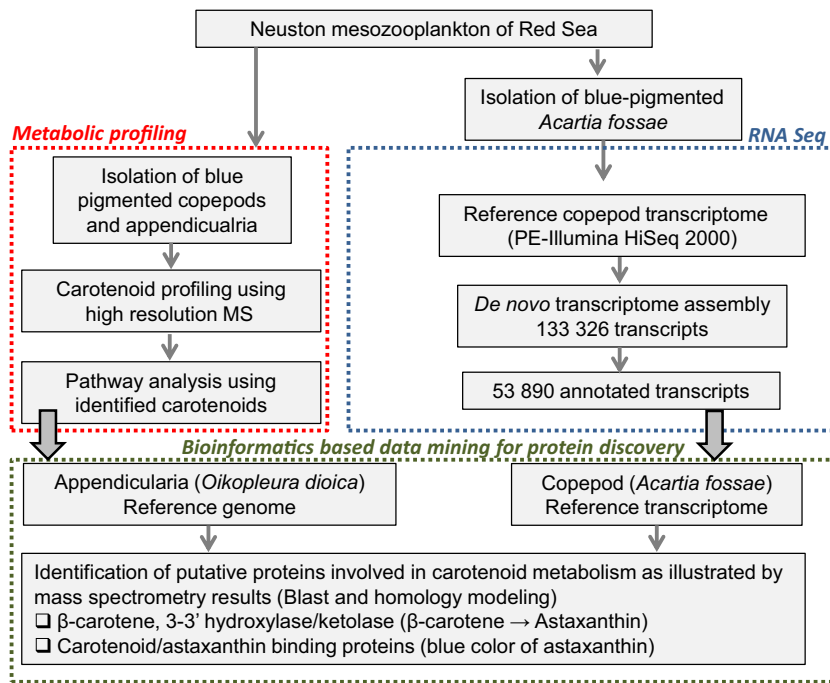
#### Overview of the samples and study design

Five different blue-pigmented organisms (four genera from subclass Copepoda and one genus from class

Appendicularia) were selected to incorporate phylogenetically and architecturally diverse species for carotenoid profiling (Fig. 1). The sample number assignment used throughout the study is as follows: 1, *Calanus* sp.; 2, *Pontella* sp.; 3, *Oikopleura* sp.; 4, *Acartia* sp. (blue); 5, *Acartia* sp. (Orange); 6, *Corycaeus* sp. The *A. fossae* showing only orange pigmentation were also included in the pigment analysis to compare with their blue counterpart. Overview of the experimental design, approaches and workflow is outlined in Fig. 2. For the gene identification, blue *A. fossae* and blue *O. dioica* were selected as representatives from the subclass Copepod and class Appendicularia, respectively. In *A. fossae*, homologous gene identification and expression values were inferred from its de novo assembled transcriptome. In *O. dioica*, respective homologous genes were mined from its reference genome (ASM20953v1) available in public database (NCBI). Our analysis workflow integrates metabolomics, transcriptomics and genomics data where we utilized the transcript sequences and available genome sequences in public databases to extract information about active pathways or gene functions related to blue pigmentation among mesozooplankton.

#### Carotenoid metabolic profiling

At first, we employed the LC-UV method to detect the carotenoids. Due to the low levels of intermediate carotenoids (except the final product astaxanthin) present in the samples, the UV-VIS spectrum could only be analysed for astaxanthin and its corresponding esters. Astaxanthin and its corresponding esters had UV absorbance at a wavelength of 480 nm. In order to detect the intermediate metabolites present at low concentrations, we employed mass spectrometry as described in the material and methods. Initially, chromatographic separations of carotenoid standards were optimized yielding the HPLC and APCI source conditions as mentioned in the materials and methods section. Table S1 (Supporting information) presents the retention times (RTs) of the selected precursors and product ion  $m/z$  values used to target carotenoid detection in all samples by comparing the specific RTs of the specific selected reaction monitoring chromatograms to the available standards ( $\beta$ -carotene, astaxanthin, zeaxanthin, canthaxanthin and fucoxanthin). The results showed that astaxanthin was the principal carotenoid present in all the six samples. Fucoxanthin was not identified in any of the samples suggesting absence of dietary algae in the samples used for analysis. For nontargeted carotenoid analysis, we shifted to Thermo LTQ Orbitrap mass spectrometer equipped with an APCI ion source using the optimized HPLC and APCI source conditions for further analysis. The nontargeted analysis screening identified additional



**Fig. 2** Experimental design and work flow for the illustration of the carotenoid metabolism in the mesozooplankton of Red Sea by integrating metabolic profiling results with transcriptomics (RNA-Seq) and genomics data.

carotenoids in the samples. The list of positively identified carotenoids across all six samples is shown in Table 1. The HPLC total ion chromatograms obtained from all the six samples clearly show that astaxanthin and its esters are the principal components among all the identified carotenoids (Fig. 3A, B). Chromatographic peaks of other identified carotenoids were present in trace amount and co-eluted between 0 and 2 min. These identified carotenoids were intermediate metabolites that occur transiently when conversion from  $\beta$ -carotene to astaxanthin takes place.

#### Astaxanthin pathway analysis

As stated earlier, detection of carotenoids and their metabolites can provide information about the food chain as well as metabolic pathways. Therefore, our strategy was to use the data obtained from carotenoid profiles to trace the metabolic pathway of astaxanthin conversion from  $\beta$ -carotene in all the six samples (Fig. 4). For this, we also used previously published results (Fraser *et al.* 1998) relating different enzymatic conversion from  $\beta$ -carotene to astaxanthin as template. The results of the analysis show that all the studied species follow the same metabolic pathways via almost the same intermediate metabolite formation. Only two exceptions were observed, canthaxanthin was not detected in *Pontella* sp. (sample 2) and  $\beta$ -cryptoxanthin was not detected in *A. fossae* (Orange, sample 5). Echinenone, one of the intermediate metabolite was not detected in any of the samples but its hydroxylated form, that is, 3-hydroxy-

chinenone was detected in all the six samples. Lutein was also detected in all the six samples.

#### Sequencing and de novo assembly of blue *Acartia fossae*

The transcriptome of the blue *A. fossae* assembled using Trinity produced a large number of contigs (or isotigs) clustered into a number of components (or genes) as outlined in Table 2. The assembly for *A. fossae* consisted of 133 326 contigs with 87 635 components. The maximum contig length was 10 538 bp with N50 contig length of 755 bp. The frequency distribution of contig lengths for the assembly is shown in Fig. S1 (Supporting information). The top 30 taxa generating hits to the *Acartia* contigs are illustrated in Fig. S2 (Supporting information). Around 50% matched to Phylum Arthropoda, with *Daphnia pulex* being the top matched species (16.6%). Transcripts from recently sequenced copepod *Calanus finmarchicus* transcriptome (Christie *et al.* 2013) were not available in the *nr* database and were therefore not used for annotation.

#### Identification of $\beta$ -hydroxylase

Putative  $\beta$ -carotene hydroxylase genes belonging to cytochrome P450 family were identified in both *A. fossae* and *O. dioica*. Thirty putative cytochrome 450s were identified in the entire *A. fossae* transcriptome when a Blast search was performed against known and experimentally characterized CrtS, CrtZ and CrtW enzymes

**Table 1** List of identified carotenoids across the six samples showing positive ion LTQ Orbitrap (APCI) MS data

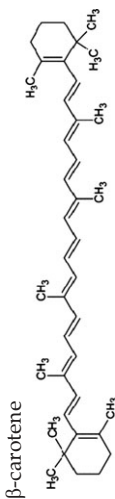
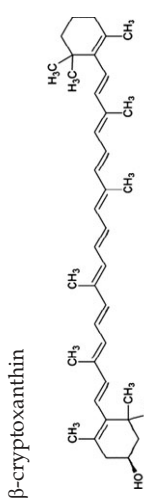
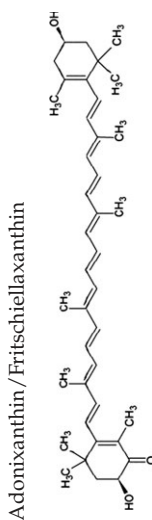
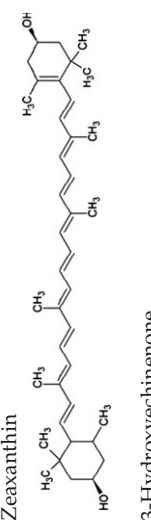
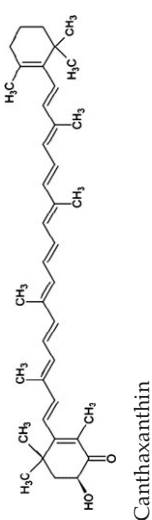
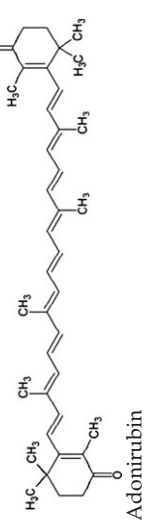
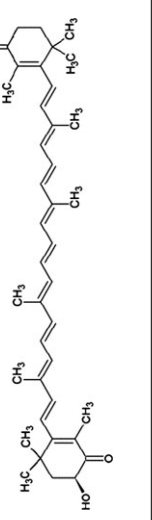
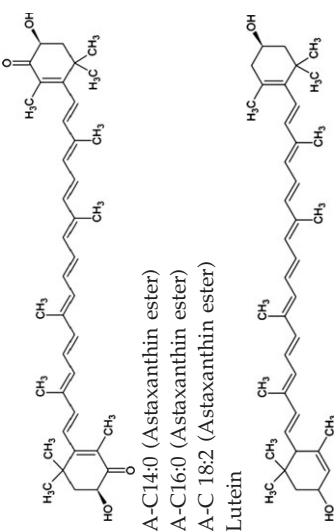
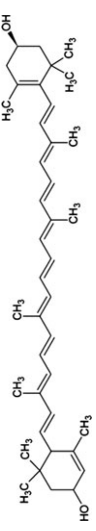
Carotenoid	Molecular formula/observed ion	Theoretical mass	Measured mass	Mass error (ppm)	Identified in samples
	C <sub>40</sub> H <sub>57</sub> /[M+H] <sup>+</sup>	537.44548	537.44751	3.7	1–6
	C <sub>40</sub> H <sub>57</sub> O/[M+H] <sup>+</sup> C <sub>40</sub> H <sub>55</sub> /[MH-H <sub>2</sub> O] <sup>+</sup>	553.44039 535.42983	553.44189 535.43170	2.7 3.4	1–4, 6
	C <sub>40</sub> H <sub>55</sub> O <sub>3</sub> /[M+H] <sup>+</sup> C <sub>40</sub> H <sub>53</sub> O <sub>2</sub> /[MH-H <sub>2</sub> O] <sup>+</sup>	583.41457 565.40401	583.4150 565.4048	1.1 1.1	1–6
	C <sub>40</sub> H <sub>55</sub> O/[M+H] <sup>+</sup>	551.42474	551.42651	3.2	1–6
	C <sub>40</sub> H <sub>55</sub> O <sub>3</sub> /[M+H] <sup>+</sup> C <sub>40</sub> H <sub>53</sub> O/[MH-H <sub>2</sub> O] <sup>+</sup>	567.41966 549.42909	567.42096 549.41098	2.2 3.2	1–6
	C <sub>40</sub> H <sub>53</sub> O <sub>2</sub> /[M+H] <sup>+</sup>	565.4041	565.4053	2.7	1, 3–6
	C <sub>40</sub> H <sub>53</sub> O <sub>3</sub> /[M+H] <sup>+</sup> C <sub>40</sub> H <sub>51</sub> O <sub>2</sub> /[MH-H <sub>2</sub> O] <sup>+</sup>	581.39892 563.38836	581.40015 563.38959	2.1 2.1	1–6

Table 1 Continued

Carotenoid	Molecular formula/observed ion	Theoretical mass	Measured mass	Mass error (ppm)	Identified in samples
Astaxanthin (A)		597.39384	597.39442	0.9	1–6
A-C14:0 (Astaxanthin ester)		807.59275	807.59433	2.6	1–6
A-C16:0 (Astaxanthin ester)		835.62405	835.62553	2.4	1, 2, 4, 6
A-C 18:2 (Astaxanthin ester)		859.62405	859.62566	2.5	1–3, 4, 6
Lutein		569.43531	569.43671	2.4	1–6
		551.42474	551.42609	2.4	

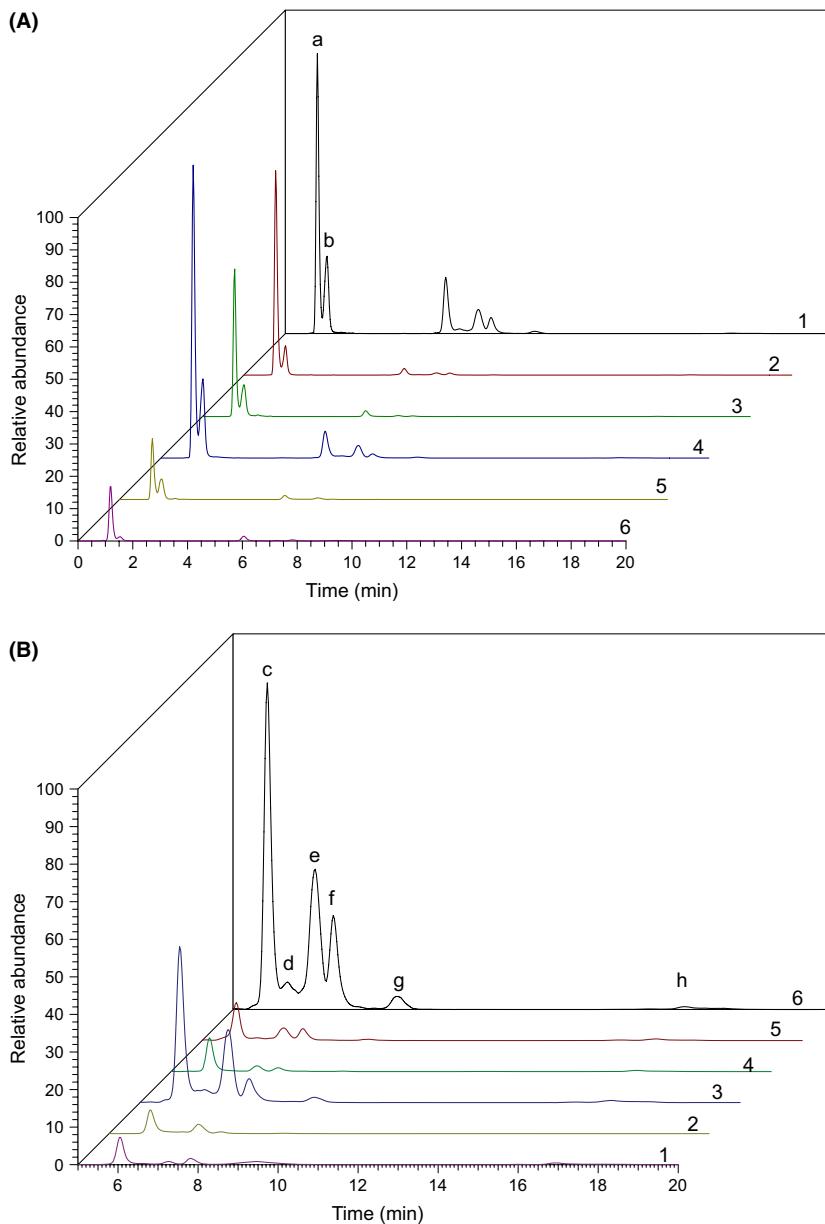
Samples: 1- *Calanus* sp.; 2- *Pontella* sp.; 3- *Oikopleura* sp.; 4- *Acartia* sp. (blue); 5- *Acartia* sp. (orange); 6- *Corycaeus* sp.

(Table S3, Supporting information). Eight cytochrome P450 transcripts with expression values >10 FPKM with their BLASTX E-values in blue *A. fossae* are presented in Table 3. In *O. dioica*, two putative  $\beta$ -carotene hydroxylase genes were identified from its genome and their protein IDs are as follows: CBY21750 and CBY13480. Alignment of CrtS (cytochrome-P450 hydroxylase of *X. dendrorhous*) conserved motifs with those of the cytochrome-P450 hydroxylases from *Acartia* and *Oikopleura* showed conservation of residues common to all families of P450 proteins such as oxygen binding site, the haeme-binding and K-helix domains (Fig. S3, Supporting information). Phylogenetic placement of the five full-length  $\beta$ -carotene hydroxylase enzymes from *A. fossae* and two from *O. dioica* shows distinct clustering outside the well-established groups of  $\beta$ -hydroxylases from bacteria, cyanobacteria, archaea, green algae, plants and fungi (Fig. 5). The three  $\beta$ -hydroxylases from *A. fossae* (representing arthropods) clustered with enzymes from mammals and amphibians (protein identity ranging between 20–42%; Table S4, Supporting information). Furthermore, they lack identity to  $\beta$ -carotene ketolase (CrtW) enzymes existing in databases.

#### Identification of carotenoid-binding proteins of the lipocalin family

Six and three putative proteins from the lipocalin family were identified in both *A. fossae* and *O. dioica*, respectively, after the BLAST search. Then, these sequences were aligned with the known CRCNs from the class malacostraca and ApoD from different taxa and the phylogenetic relationship was established (Fig. 6). The identified gene sequences did not cluster with and were clearly divergent from the recognized CRCNs. In addition to the target species, two putative carotenoid-binding proteins from *Daphnia* also clustered among the copepod proteins. The identified sequences from both *A. fossae* and *O. dioica* did not appear to be CRCN homologues when primary structures were aligned against the sequences in the database using BLAST. Therefore, we used a method based on sequence 'profiles', that is, sequence-profile comparison (PSI-Blast) (Altschul *et al.* 1997) and profile-profile comparison (HHpred) (Soding *et al.* 2005) to characterize the identified distant homologues of CRCNs in *Acartia* and *Oikopleura*. Interestingly, when this method was used to search for homology, the first hit was ApoD (PDB ID: 2hzq\_A) with a homology probability of 100% and the second hit was CRCN A2 subunit (1gka\_B), also with 100% homology probability (Table 4). After this, the 3D structures of two protein sequences from *A. fossae* (comp\_54867 and comp\_54972) and one (CBY\_33375) from *O. dioica* with the highest HHpred scores were





**Fig. 3** Positive ion LTQ-Orbitrap atmospheric pressure chemical ionization (APCI)-MS total ion chromatograms of extracted carotenoids (A) in six samples (1, *Calanus* sp.; 2, *Pontella* sp.; 3, *Oikopleura* sp.; 4, *Acartia* sp. (blue); 5, *Acartia* sp. (Orange); 6, *Corycaeus* sp.) with peak identity as follows: Astaxanthin (a,b); (B) zoomed in total ion chromatograms showing astaxanthin esters in all six samples- Astaxanthin-C14:0 (c,e); Astaxanthin-C18:2 (d); Astaxanthin-C16:0 (f,g); and  $\beta$ -carotene (h).

generated using the MODELLER software (Eswar *et al.* 2006) with the CRCN A2 subunit (1gka\_B) as template and visualized using the UCSF CHIMERA software (Pettersen *et al.* 2004) (Fig. 7A) (The generated PDB files are provided as Appendices S2-S4 in the Supporting information). The overall score of the model was good for more than 90% of amino acid residues (Fig. S4, Supporting information). Upon superimposing the predicted structures using the Matchmaker tool of Chimera, it was observed that these proteins fold similar to CRCN, forming hydrophobic calyx region rich in aromatic amino acid residues suggesting similar putative function of binding to astaxanthin (highlighted in blue; Fig. 7B). Further, detailed comparison between the

multiple alignment using primary structures alone and primary and secondary structures together provided evidence of functional homology among the identified lipocalin proteins in *Acartia*, *Oikopleura* and CRCNs in *Homarus gammarus*, whose astaxanthin-binding function has been studied extensively (Chayen *et al.* 2003; Fig. 7C). Both alignments showed strong conservation of the three lipocalin structurally conserved regions (SCRs) within all the sequences (shaded in blue; Fig. 7C). Outside the core SCRs, specific Y51 (in CRCN-A), previously thought to bind carotenoids (Keen *et al.* 1991), aligned perfectly with the Y72 (comp\_54867), Y64 (comp\_54972) and Y61 (CBY\_33375) residues when secondary structures were also used for alignment. Absolute conser-

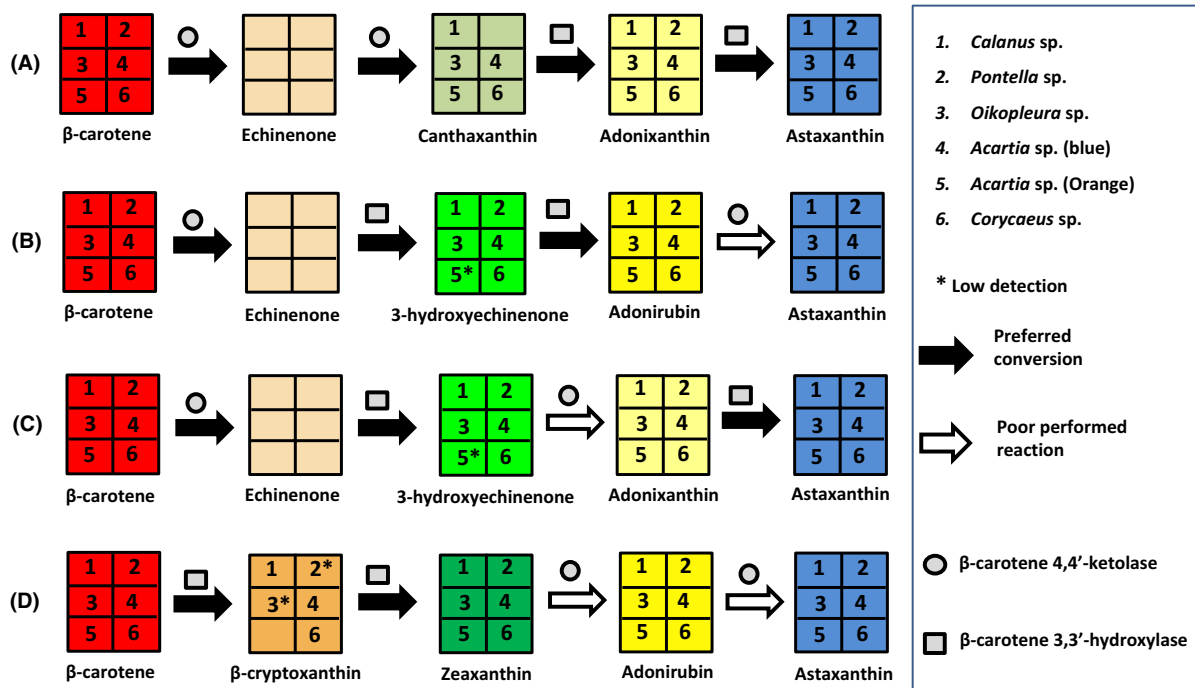


Fig. 4 Enzymatic pathways to astaxanthin formation in all the six organisms showing the functional equivalence among blue mesozooplankton community and predominance of  $\beta$ -carotene 3,3'-hydroxylase activity. (A–D) Shows the enzymatic routes to astaxanthin formation (modified after Fraser *et al.* 1998). Respective carotenoids are drawn as coloured boxes containing the sample numbers in which they were detected.

Table 2 Overview of mRNA sequencing, de novo assembly and annotation statistics of *Acartia fossae* transcriptome

Blue <i>Acartia fossae</i> transcriptome	
Total raw reads (2 × 100 bp)	34 385 018
Quality Filtered mRNA (rRNA filtered)	27 905 504
Total no. of Trinity transcripts/ isoforms (N50)	133 326 (755)
Total no. of Trinity components/genes	87 635
No. of protein coding transcripts	53 890
No. of 53 890 genes expressed, FPKM > 0 (Median FPKM)	30 616 (3.6)
No. of 53 890 genes expressed, FPKM ≥ 10 (Median FPKM)	8383 (28.9)
Blast annotation (%)	33 299 (61.8)
Pfam annotation (%)	31 197 (57.9)
GO annotation (%)	32 038 (59.4)
Full-length transcript (min. 20% alignment coverage)	10 695
Full-length transcript (min. 60% alignment coverage)	5204
Full-length transcript (min. 90% alignment coverage)	2909

vation was also observed for cysteine residues known to form two disulphide bridges in CRCN-A when secondary structures were used for alignment.

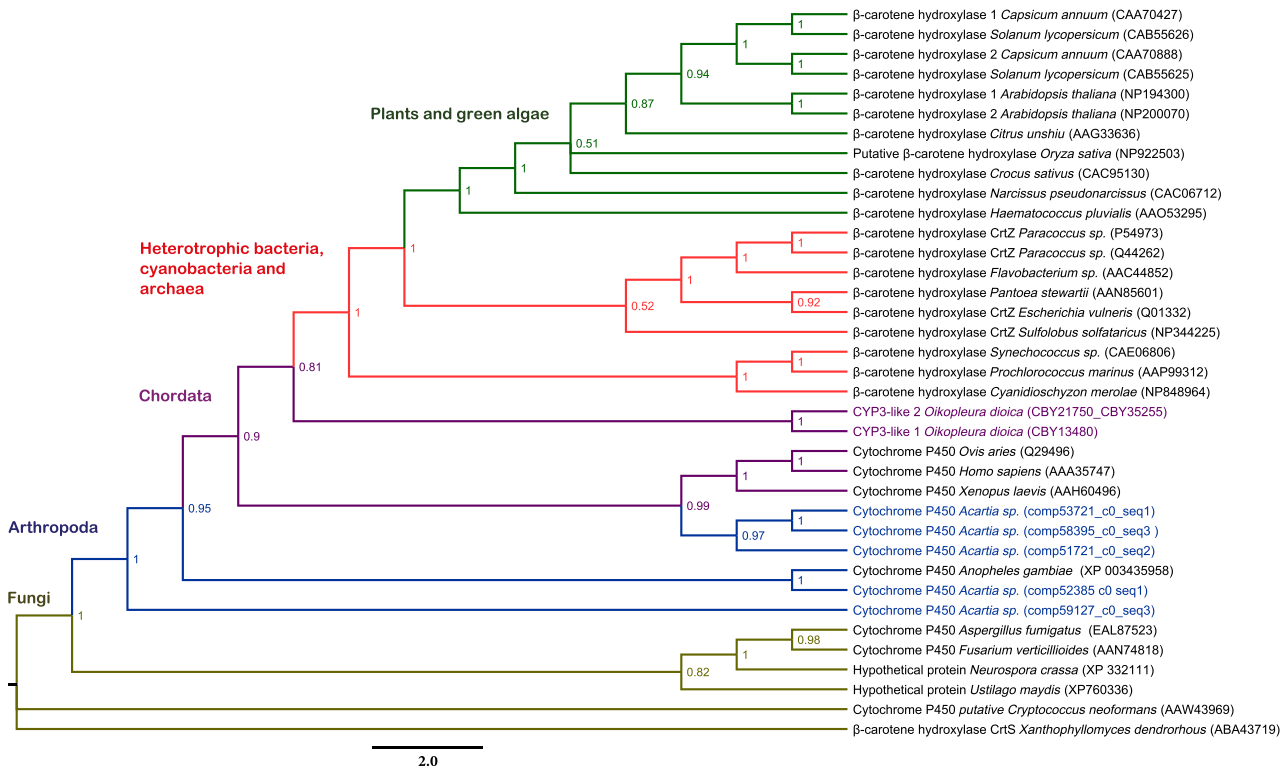
Table 3 Putative  $\beta$ -carotene hydroxylase of P450 family coding transcripts with FPKM > 10 identified in blue *Acartia fossae* by in silico transcriptome mining

Putative $\beta$ -carotene hydroxylase Blue copepod ( <i>Acartia fossae</i> )			
Transcript_ID		E-value (BLASTX)	FPKM <i>Acartia</i>
1	comp59127_c0_seq3	1.66E-009	41.11
2	comp56035_c0_seq7	3.69E-008	40.98
3	comp59141_c0_seq3	1.81E-018	37.01
4	comp58395_c0_seq3	3.69E-013	25.47
5	comp52385_c0_seq1	3.54E-008	12.87
6	comp51721_c0_seq2	2.07E-005	11.13
7	comp53721_c0_seq1	1.73E-011	10.86
8	comp51607_c0_seq1	1.85E-005	10.42

FPKM, Fragments Per Kb of transcript per Million mapped reads.

### Discussion

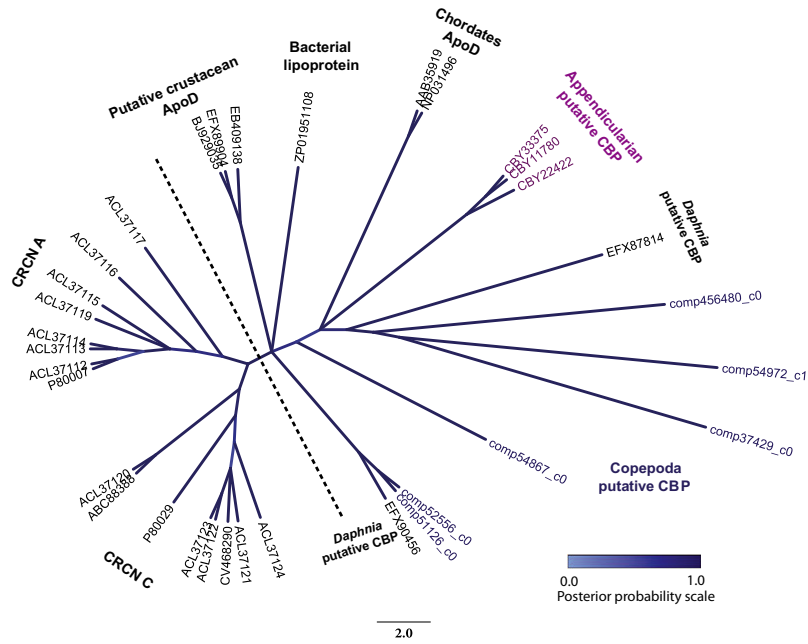
Blue pigmentation is ubiquitous in phylogenetically diverse neustonic organisms, from Phylum Cnidaria (*Vellela* sp. and *Physalia* sp.) to Chordata, through Class Crustacea and Phylum Mollusca (*Glaucus* sp., *Janthina* sp.). It can, therefore, be safely assumed that the blue colour has an important adaptive value for organisms



**Fig. 5** Phylogenetic relationships of  $\beta$ -carotene hydroxylases from various organisms with their accession numbers in parentheses. Phylogenies were estimated using the program MRBAYES version 3.2.2 (Huelsenbeck & Ronquist 2001). Metropolis-coupled Markov chain Monte Carlo algorithm was used for sampling of trees from the posterior probability distribution that allowed running of multiple Markov chains. A run with four chains was performed for 400 000 generations, under a fixed WAG model with set parameters as described in the materials and methods section. Numbers at the nodes indicate Bayesian posterior probability values.

living in the surface of the ocean. This is confirmed by the persistence of both the metabolic pathways to produce astaxanthin as well as the protein structures binding to it as stated in our results. We found that astaxanthin is the principal carotenoid present in both copepods and appendicularia. They follow the same metabolic pathways of astaxanthin conversion from  $\beta$ -carotene to astaxanthin via almost the same intermediate metabolite formation (Fig. 4). The absence of carotenoid intermediate, echinenone (a 4-keto-  $\beta$ -carotene) in the second step of this biochemical conversion in all samples suggests absence of  $\beta$ -carotene 4, 4'-ketolase activity. This agrees with our finding where we were unable to find any transcripts related to ketolase enzyme. Another possibility could be that the transient nature of these intermediates and their immediate conversion to 3-hydroxyechinenone limits the possibility to detect them. Astaxanthin presence has been observed in many copepods, but actual sources are likely to be dietary phytoplankton or bacterioplankton. Our results suggest that both copepods and appendicularia obtain the main precursor, that is,  $\beta$ -carotene from the dietary algae and oxidize it into astaxanthin. We have also looked into the genetic basis of this metabolic

conversion by identifying putative genes or enzymes in both copepods and appendicularia. The identified cytochrome-P450 hydroxylases from *Acartia* and *Oikopleura* encode polypeptides similar to that of chordates (including mammals and amphibians) and fungi. They showed low sequence similarities with  $\beta$ -hydroxylases (CrtZ) from bacteria, algae and plants. The phylogenetic placement of cytochrome P450 hydroxylases from *Acartia* and *Oikopleura* suggests that their evolution origin is different from the enzymes catalysing similar reactions in micro-organisms or plants and they are more closely related to the homologous proteins of each other. As CrtS from fungus *X. dendrorhous* is a well-characterized enzyme known for astaxanthin conversion from  $\beta$ -carotene and closely related to the identified cytochrome P450 hydroxylases from *Acartia* and *Oikopleura*, we further compared the identified proteins with CrtS to explore their putative functionality in relation to astaxanthin conversion. All the encoded proteins contain the oxygen-binding site including the invariant Glycine<sup>340</sup> and Threonine<sup>343</sup> (in CrtS of *X. dendrorhous*) (Teutsch *et al.* 1993), the haeme-binding domain that is involved in the folding of the haeme-binding pocket (Poulos *et al.* 1987), <sup>394</sup>ESLR<sup>397</sup> domain (in CrtS of



**Fig. 6** Phylogenetic analysis of lipocalin superfamily proteins from various organisms. They are represented by their accession numbers or transcript IDs. comp54867\_c0, comp54972\_c1, comp52556\_c0, comp456480\_c0, comp37429\_c0, comp51126\_c0 (*Acartia fossae* transcripts, putative carotenoid-binding proteins, CBP); P80029 (CRCN-C1, *H. gammarus*); ACL37124 (CRCN-C, *Panulirus versicolor*); ACL37121 (CRCN-C, *Cherax quadricarinatus*); CV468290 (CRCN-C, *Litopenaeus vannamei*); ACL37122 (CRCN-C, *Marsupenaeus japonicus*); ACL37123 (CRCN-C, *Penaeus monodon*); ABC88388 (CRCN-like lipocalin, *Macrobrachium rosenbergii*); ACL37120 (CRCN-C, *Alpheus* sp.); P80007 (CRCN-A2, *H. gammarus*); ACL37112 (CRCN-A1, *Panulirus cygnus*); ACL37117 (CRCN-A, *P. monodon*); ACL37116 (CRCN-A, *M. japonicus*); ACL37119 (CRCN-A, *Gonodactylus smithii*); ACL37115 (CRCN-A, *Dardanus megistos*); ACL37113 (CRCN-A2, *P. cygnus*); ACL37114 (CRCN-A, *C. quadricarinatus*); BJ929035 [putative Apolipoprotein D (ApoD), *Daphnia magna*]; EB409138 (putative ApoD, *P.monodon*); AAB35919 (ApoD, *Homo sapiens*); NP031496 (ApoD, *Mus musculus*); ZP01951108 (lipoprotein, *Vibrio cholera*); EFX87814, EFX89904 & EFX90456 (hypothetical proteins, *Daphnia pulex*); CBY33375, CBY11780 & CBY22422 (putative ApoD, *O. dioica*). Phylogenies were estimated using the program MRBAYES, version 3.2.2 (Huelsenbeck & Ronquist 2001). Branches are gradient coloured by their posterior probability support values.

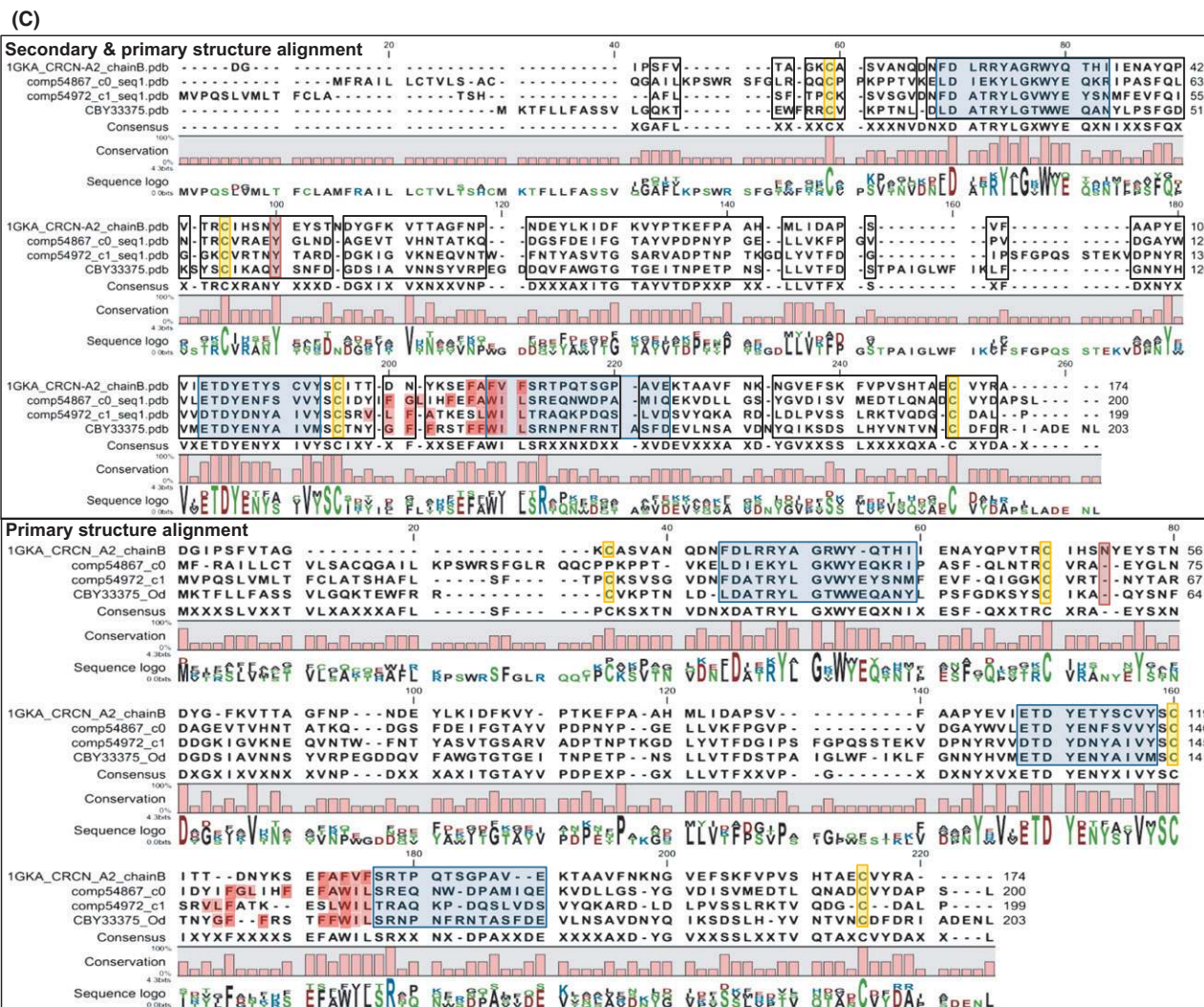
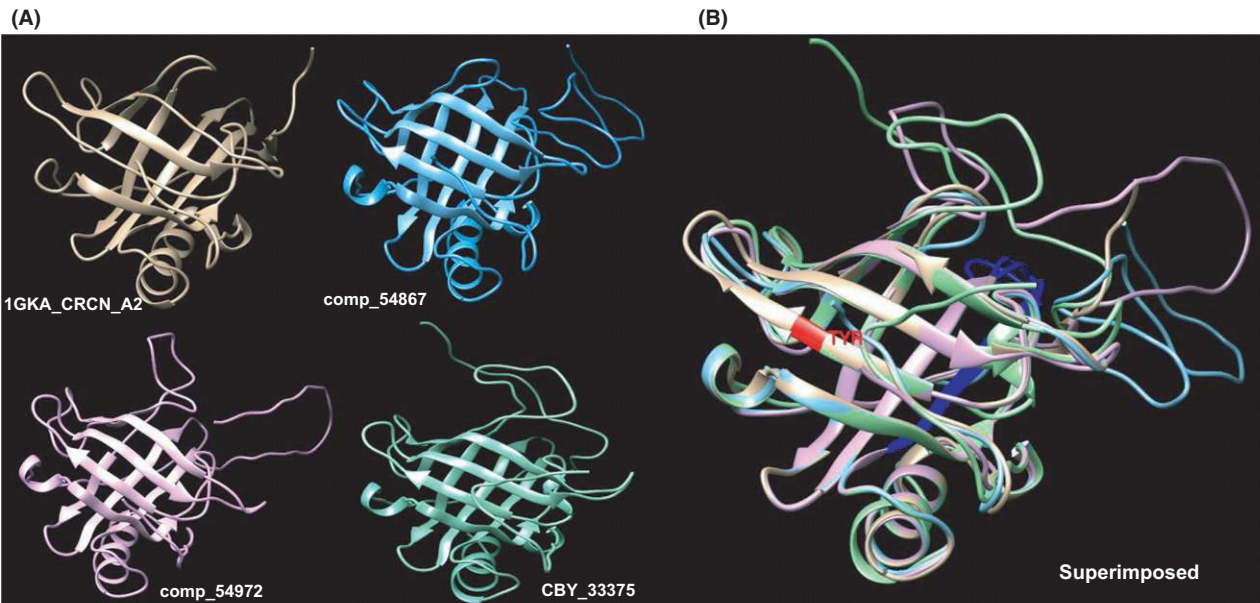
**Table 4** Putative carotenoid-binding proteins after transcriptome/genome mining showing 100% homology to Apolipoprotein D (ApoD) and crustacyanin as predicted by HHpred

Carotenoid-binding proteins Transcript/Protein_ID	1st hit—ApoD (2hzq_A)				2nd hit—Crustacyanin A2 subunit (1gka_B)			
	Prob	E-value	P-value	Score	Prob	E-value	P-value	Score
comp54972_c1_seq1*	100	3.2E-37	9.9E-42	229.9	100	2.8E-36	8.8E-41	224.7
comp54867_c0_seq1*	100	2.9E-37	8.9E-42	228.5	100	7.6E-37	2.4E-41	226.1
comp51126_c0_seq2*	100	1.7E-31	5.4E-36	212.8	100	2.7E-31	8.4E-36	211.3
comp52556_c0_seq3*	100	3.4E-29	1E-33	196.2	100	5.9E-29	1.8E-33	194.2
CBY33375†	100	3.8E-35	1.2E-39	223.1	100	4.2E-35	1.3E-39	222.6

\**Acartia fossae*; †*Oikopleura dioica*; Prob-% probability of homology.

*X. dendrorhous*) involved in the maintenance of the three-dimensional structure and found in all cytochrome P450 proteins (Yu & Strobel 1997) and conserved cysteine<sup>495</sup> (in CrtS of *X. dendrorhous*) residue that provides the thiolate ligand to the haeme group (Shimizu *et al.* 1988). It has been known that in different organisms, the conversion from  $\beta$ -carotene to astaxan-

thin takes place by the combined action of a  $\beta$ -hydroxylase and a  $\beta$ -ketolase enzyme that work successively on different carotenoid intermediates (Martin *et al.* 2008). Both the hydroxylase and ketolase activities have been reported to be found in CrtS of *X. dendrorhous* (Ojima *et al.* 2006), but further biochemical evidence is required to elucidate this bifunctional activity (Martin *et al.*



2008). As mentioned earlier, we did not find any transcript related to  $\beta$ -ketolase enzyme in the *A. fossae* transcriptome or the corresponding protein in the *Oikopleura* genome. Even the identified cytochrome P450 hydroxylases from both organisms showed no sequence similarities with the  $\beta$ -carotene ketolase enzymes. This suggests that the identified cytochrome P450 proteins might be bifunctional with both hydroxylase and ketolase activities in *Acartia* and *Oikopleura*, resulting in the formation and accumulation of astaxanthin as confirmed by our LCMS analyses (Figs 3 and 4).

Astaxanthin appears bright red in colour in free, unbound form or when esterified to fatty acids and blue when bound to proteins (Kobayashi & Sakamoto 1999). Our model ecosystem in this study consists of surface mesozooplankton specifically copepods and appendicularia with predominantly astaxanthin in protein bound blue form. After exploring the metabolic and genetic mechanisms of astaxanthin formation, we searched for the proteins responsible for imparting blue colour to astaxanthin. Using a combination of methods like BLAST homology, sequence-profile (PSI-BLAST) (Altschul *et al.* 1997) and profile-profile comparison (HHpred) (Soding *et al.* 2005), we identified six and three putative carotenoid-binding proteins from the lipocalin family in both *Acartia* and *Oikopleura*, respectively. Our results also support the previous observations of lipocalin family proteins displaying remarkable structural similarities despite low sequence similarities at the protein level. No specific gene or protein sequences have been identified in a wide range of invertebrate phyla with similar biochemical properties as malacostracan crustacean CRCNs (Wade *et al.* 2009). Our study provides evidences for the existence of blue astaxanthin-protein complexes and corresponding putative carotenoid-binding proteins in two distant phylogenetic groups, copepod and appendicularia. In addition, two putative proteins from *Daphnia* also clustered among the copepod proteins and support the observation of carotenoproteins in the blood and eggs of *Daphnia* (Green 1957). Phylogenetic analyses of the malacostracan CRCNs that act as astaxanthin modifiers

to produce a blue phenotype indicated that they originated and evolved from an ApoD-like ancestor through duplication and divergence (Wade *et al.* 2009). Wade *et al.* (2009) further hypothesized that there was neofunctionalized duplication that subsequently led CRCNs to evolve their capacity to bind astaxanthin in forming patterned shell colours. It was unclear whether related crustacean specific clade of ApoD-like or EST proteins have the capacity of binding chromophores. Our study strongly suggests that before the duplication event among malacostracan clade, ApoD-like proteins in crustacean specific clade, that is, copepod (outside the malacostraca clade) and even appendicularia (outside crustacean) were capable of binding astaxanthin and show the blue phenotype. It is logical to hypothesize that there may have been a rapid evolution of the sequence within Crustacea and that the decapod CRCN is derived from a shared ancestor with the crustacean gene. But there is not enough information from other species to suggest that this also occurred between crustaceans and appendicularians. The putative carotenoid-binding proteins in both *Acartia* and *Oikopleura* were clearly distinguished from the crustacean CRCNs; they are examples of independent convergent evolution and functional equivalence. In order to confirm that the sequences identified are truly the functional carotenoid-binding proteins, we made an attempt to purify the blue carotenoid protein from *A. fossae* as described in Appendix S1 (Supporting information). We could determine the size of the blue pigment to be around 20–30 kDa after size-exclusion chromatography, which agrees with the predicted sizes of the identified transcripts (e.g. comp54867\_c0\_seq1, predicted MW 22.6 kDa). Similarly, the spectral maxima absorbance of the same corresponded to the carotenoproteins observed in pontellids in earlier study (Herring 1965).

The blue pigmentation in neustonic organisms, in particular in subtropical areas, has generally been associated with the protection against UVR and/or predation. In oligotrophic areas, low nutrient availability owing to low primary productivity and dissolved organic carbon results in greater water transparency (Sommaruga 2001). This increased transparency not

---

**Fig. 7** Three-dimensional structure models of (A) astaxanthin-binding proteins from *A. fossae* (comp\_54867, comp\_54972) and *O. dioica* (CBY\_33375) predicted by HHpred alignment and MODELLER using Crustacyanin A2 subunit (PDB ID: 1GKA) as template. (B) Superimposition of all three models on the template showing aligned Tyrosine residue Y51 in CRCN-A, Y72 in comp\_54867, Y64 in comp\_54972 and Y61 in CBY\_33375 (highlighted in red), the hydrophobic calyx region rich in aromatic amino acid residues (highlighted in blue) and (C) Structure-based multiple sequence alignment using UCSF Chimera (Pettersen *et al.* 2004). Alignment using primary and secondary structures (black box) together and primary structure alone of the identified lipocalin proteins in *Acartia* and *Oikopleura* with CRCN in *H. gammarus* showed three lipocalin structurally conserved regions (SCRs) within all the sequences (shaded in blue). The aligned cysteine residues involved in disulphide bridges are indicated in yellow boxes. The hydrophobic calyx region rich in aromatic amino acid residues are indicated in red.

only allows a high penetration of UVR but also leads to increased visibility by the visual predators such as fish (Stambler 2005). This is particularly relevant for species that do not perform diel vertical migration because of the size and limited swimming capacity (i.e. small copepods and appendicularia) or species specifically adapted to live in the neuston. UVR causes damage either by generation of reactive oxygen species (ROS) that may react with the biological tissues or by directly inducing DNA damage resulting in the formation of thymine dimers that might have heritable mutagenic effects depending on the genes affected (Zagarese *et al.* 1997; Hansson *et al.* 2007). Further, UVR synergistically increases the tissue damaging property of phototoxic pollutants such as polycyclic-aromatic hydrocarbon (PAH) if present in the water (Arfsten *et al.* 1996). Astaxanthin is a potent antioxidant with higher antioxidant properties than  $\beta$ -carotene and  $\alpha$ -tocopherol (Miki 1991). The high levels of carotenoids impart photoprotection in copepods and the concentration of pigmentation directly influences their behaviour and vertical migration (Hairston 1976). Direct photoprotective mechanism via reflectance or absorbance of UVR does not hold true for astaxanthin because it does not have UVR absorbance (Cianci *et al.* 2002). Rather indirect photoprotection of astaxanthin via scavenging reactive oxygen radicals formed by UVR stress could result in increased survival when exposed to high levels of UVR (Sommer *et al.* 2006). Furthermore, astaxanthin may also impart protection against other oxidative stresses inducing environmental factors such as high salinity or xenobiotics.

All the above-mentioned conclusions are related to studies where astaxanthin existed in free or esterified form imparting red coloration to copepods. Occurrence of astaxanthin in blue colour, when bound to protein transforming the whole organisms into blue has been mentioned in very few studies earlier in 1960s and 70s (Herring 1965; Hairston 1976). But very little research has been carried out on comparing advantages and disadvantages of red and blue coloration and implications of the change from red to blue and vice versa. In one study, it was found that red copepods can tolerate 2-fold greater levels of UVB than blue or mixed copepods ( $P < 0.001$ ), where total astaxanthin content was much higher in red copepods than blue or mixed copepods (Hudelson *et al.* 2011). Both *A. fossae* and *O. dioica* are known to present red coloration as well (as observed in this study). Although astaxanthin can provide photoprotection against UVR through its antioxidant properties, its presence in high concentrations also exposes copepods to an increased risk of predation by visual predators such as fish (or birds) leading to the exhibition of phenotypic plasticity (Vestheim & Kaartvedt 2006). The fact that two phylogenetically very distant

neustonic organisms present evolutionary convergence to blue pigmentation suggests that visual predation in surface waters is a strong evolutionary driving force.

## Acknowledgements

We thank captain and crew of the Coastal and Marine Resources Core lab (CMOR) at King Abdullah University of Science and Technology (KAUST) for their technical support during sampling. We thank Dr. Mohsen El-Sherbiny (King Abdulaziz University, Jeddah) for his assistance with the identification of *A. fossae*. We are also thankful to Analytical Core and Bioscience Core labs at KAUST for their support and services. The research was supported by baseline funding provided by KAUST to Prof. Xabier Irigoien.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Altschul SF, Madden TL, Schaffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.
- Arfsten DP, Schaeffer DJ, Mulveny DC (1996) The effects of near ultraviolet radiation on the toxic effects of polycyclic aromatic hydrocarbons in animals and plants: a review. *Ecotoxicology and Environmental Safety*, **33**, 1–24.
- Bandaranayake WM, Gentien P (1982) Carotenoids of *Temora*-*Turbinata*, *Centropages-Furcatus*, *Undinula-Vulgaris* and *Euchaeta-Russelli*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, **72**, 409–414.
- Britton G, Halliwell JR (2008) Carotenoid-protein interactions. In: *Carotenoids* (eds Britton G, Liaaen-Jensen S, Pfander H), pp. 99–118. Birkhäuser Verlag, Basel, Switzerland.
- Buchwald M, Jencks WP (1968) Properties of the crustacyanins and the yellow lobster shell pigment. *Biochemistry*, **7**, 844–859.
- Chayen NE, Cianci M, Grossmann JG *et al.* (2003) Unravelling the structural chemistry of the colouration mechanism in lobster shell. *Acta Crystallographica: Section D, Biological Crystallography*, **59**, 2072–2082.
- Cheeseman DF, Lee WE, Zagalsky PF (1967) Carotenoproteins in invertebrates. *Biological Reviews*, **42**, 132–160.
- Christie AE, Roncalli V, Wu LS *et al.* (2013) Peptidergic signaling in *Calanus finmarchicus* (Crustacea, Copepoda): in silico identification of putative peptide hormones and their receptors using a de novo assembled transcriptome. *General and Comparative Endocrinology*, **187**, 117–135.
- Cianci M, Rizkallah PJ, Olczak A *et al.* (2002) The molecular basis of the coloration mechanism in lobster shell: beta-crustacyanin at 3.2-Å resolution. *Proceedings of the National Academy of Sciences of the USA*, **99**, 9795–9800.

- Denoeud F, Henriot S, Mungpakdee S *et al.* (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science*, **330**, 1381–1385.
- Eswar N, Webb B, Marti-Renom MA *et al.* (2006) Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics* **Chapter 5**, Unit 5.6.
- Fisher LR, Kon SK, Thompson SY (1964) Vitamin A and carotenoids in certain invertebrates 7. Crustacea: Copepoda. *Journal of the Marine Biological Association of the UK*, **44**, 685–692.
- Flower DR (1996) The lipocalin protein family: structure and function. *Biochemical Journal*, **318**, 1–14.
- Fraser PD, Shimada H, Misawa N (1998) Enzymic confirmation of reactions involved in routes to astaxanthin formation, elucidated using a direct substrate in vitro assay. *European Journal of Biochemistry*, **252**, 229–236.
- Goodwin TW (1971) Pigments – Arthropoda. In: *Chemical Zoology* (eds Florkin M, Scheer BT), pp. 279–288. Academic Press, New York.
- Goodwin TW (1976) *Chemistry and Biochemistry of Plant Pigments*, 2nd edn. Academic Press, New York.
- Green J (1957) Carotenoids in *Daphnia*. *Proceedings of the Royal Society of London: Series B, Biological Sciences*, **147**, 392–401.
- Haas BJ, Papanicolaou A, Yassour M *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494–1512.
- Hairston NG (1976) Photoprotection by carotenoid pigments in copepod *Diaptomus-Nevadensis*. *Proceedings of the National Academy of Sciences of the USA*, **73**, 971–974.
- Hansson LA, Hylander S, Sommaruga R (2007) Escape from UV threats in zooplankton: a cocktail of behavior and protective pigmentation. *Ecology*, **88**, 1932–1939.
- Herring PJ (1965) Blue pigment of a surface-living oceanic copepod. *Nature*, **205**, 103–104.
- Hudelson KE, Barst BD, Smith JD, Roberts AP (2011) Effects of carotenoprotein expression on UV tolerance in high elevation copepods. In: *96th ESA Annual Meeting*, Austin, Texas.
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.
- Johnsen GH, Jakobsen PJ (1987) The effect of food limitation on vertical migration in *Daphnia-Longispina*. *Limnology and Oceanography*, **32**, 873–880.
- Keen JN, Caceres I, Eliopoulos EE, Zagalsky PF, Findlay JB (1991) Complete sequence and model for the A2 subunit of the carotenoid pigment complex, crustacyanin. *European Journal of Biochemistry*, **197**, 407–417.
- Kobayashi M, Sakamoto Y (1999) Singlet oxygen quenching ability of astaxanthin esters from the green alga *Haematococcus pluvialis*. *Biotechnology Letters*, **21**, 265–269.
- Kopylova E, Noe L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.
- Kuchibhatla DB, Sherman WA, Chung BY *et al.* (2014) Powerful sequence similarity search methods and in-depth manual analyses can identify remote homologs in many apparently “orphan” viral proteins. *Journal of Virology*, **88**, 10–20.
- Lampert W (1989) The adaptive significance of diel vertical migration of zooplankton. *Functional Ecology*, **3**, 21–27.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
- Liaaen-Jensen S (1998) Carotenoids in food chain. In: *Carotenoids: Biosynthesis and Metabolism* (eds Britton G, Liaaen-Jensen S, Pfander H), pp. 359–371. Birkhäuser, Basel, Switzerland.
- Lohse M, Bolger AM, Nagel A *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**, W622–W627.
- Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.
- Maoka T (2011) Carotenoids in marine animals. *Marine Drugs*, **9**, 278–293.
- Martin JF, Gudina E, Barredo JL (2008) Conversion of beta-carotene into astaxanthin: two separate enzymes or a bifunctional hydroxylase-ketolase protein? *Microbial Cell Factories*, **7**, 3.
- Miki W (1991) Biological functions and activities of animal carotenoids. *Pure and Applied Chemistry*, **63**, 141–146.
- Ojima K, Breitenbach J, Visser H *et al.* (2006) Cloning of the astaxanthin synthase gene from *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*) and its assignment as a beta-carotene 3-hydroxylase/4-ketolase. *Molecular Genetics and Genomics*, **275**, 148–158.
- Persaud AD, Moeller RE, Williamson CE, Burns CW (2007) Photoprotective compounds in weakly and strongly pigmented copepods and co-occurring cladocerans. *Freshwater Biology*, **52**, 2121–2133.
- Pettersen EF, Goddard TD, Huang CC *et al.* (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, 1605–1612.
- Pilbrow J, Garama D, Carne A (2012) Carotenoid-binding proteins; accessories to carotenoid function. *Acta Biochimica Polonica*, **59**, 163–165.
- Poulos TL, Finzel BC, Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. *Journal of Molecular Biology*, **195**, 687–700.
- Rao KR (1985) Pigmentary effectors. In: *Integuments, Pigments and Hormonal Processes* (eds Bliss DE, Mantel LH). Academic Press, New York.
- Shimizu T, Hirano K, Takahashi M, Hatano M, Fujii-Kuriyama Y (1988) Site-directed mutagenesis of rat liver cytochrome P-450d: axial ligand and heme incorporation. *Biochemistry*, **27**, 4138–4141.
- Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, **33**, W244–W248.
- Sommaruga R (2001) The role of solar UV radiation in the ecology of alpine lakes. *Journal of Photochemistry and Photobiology B: Biology*, **62**, 35–42.
- Sommer F, Agurto C, Henriksen P, Kiorboe T (2006) Astaxanthin in the calanoid copepod *Calanus helgolandicus*: dynamics of esterification and vertical distribution in the German Bight, North Sea. *Marine Ecology Progress Series*, **319**, 167–173.
- Stambler N (2005) Bio-optical properties of the northern Red Sea and the Gulf of Eilat (Aqaba) during winter 1999. *Journal of Sea Research*, **54**, 186–203.
- Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Teutsch HG, Hasenfratz MP, Lesot A *et al.* (1993) Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinn-



- mate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. *Proceedings of the National Academy of Sciences of the USA*, **90**, 4102–4106.
- Vestheim H, Kaartvedt S (2006) Plasticity in coloration as an antipredator strategy among zooplankton. *Limnology and Oceanography*, **51**, 1931–1934.
- Wade NM, Tollenaere A, Hall MR, Degnan BM (2009) Evolution of a novel carotenoid-binding protein responsible for crustacean shell color. *Molecular Biology and Evolution*, **26**, 1851–1864.
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*, **18**, 691–699.
- Yu XC, Strobel HW (1997) Hydroperoxide-mediated cytochrome P450-dependent 8-anilino-1-naphthalenesulfonic acid destruction, product formation and P450 modification. *Molecular and Cellular Biochemistry*, **167**, 159–168.
- Zagarese HE, Feldman M, Williamson CE (1997) UV-B-induced damage and photoreactivation in three species of *Boeckella* (Copepoda, Calanoida). *Journal of Plankton Research*, **19**, 357–367.

---

X.I. and N.M. conceived and designed the research. N.M. and N.A. performed the sampling and species identification. N.M. and M.A. performed the carotenoid profiling experiments and analysed the data. N.M., M.T. and M.K. contributed in preparing RNA-Seq library and de novo assembly of the transcriptome. N.M. performed the rest of the analyses. N.M. and X.I. wrote the paper.

---

### Data accessibility

RNASeq raw sequences generated using Illumina have been submitted to SRA at NCBI under the accession no. SRP036139. All FASTA sequences and PDB files used in

the analysis are provided as supporting information. The multiple sequence alignments and phylogenetic tree data files are deposited in the Dryad repository: <http://doi.org/10.5061/dryad.mm18b>.

### Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Frequency distribution of the de novo assembled *Acartia fossae* transcriptome using Trinity.

**Fig. S2** Top 30 taxa generating BLASTX hit to *Acartia* transcripts.

**Fig. S3** Alignment of CrtS (cytochrome-P450 hydroxylase of *X. dendrorhous*) conserved motifs with those of the cytochrome-P450 hydroxylases from *Acartia* and *Oikopleura*.

**Fig. S4** Compatibility analysis of atomic model (3D) with its own amino acid sequence (1D) by Verify3D tool.

**Table S1** Precursor and product ion *m/z* and instrument parameter values for carotenoid detection.

**Table S2** FASTA sequences analysed in the study.

**Table S3** List of identified putative cytochrome P450 proteins in *Acartia fossae* with FPKM value and annotation.

**Table S4** Pairwise percent identity matrix for the identified beta hydroxylases.

**Appendix S1** Purification of blue pigment from *Acartia fossae*.

**Appendix S2** PDB\_CBY33375\_Oikopleura\_dioica.

**Appendix S3** PDB\_comp54867\_c0\_Acartia\_fossae.

**Appendix S4** PDB\_comp54972\_c1\_Acartia\_fossae.