Electronic Supplemental Material 1:

Genetic differences between male and female Antarctic krill (*Euphausia superba*).

Manuscript title: Sex identification from distinctive gene expression patterns in Antarctic krill (*Euphausia superba*)

Journal: Polar Biology

Authors: Leonie Suter, Andrea Maree Polanowski, Robert King, Chiara Romualdi, Gabriele Sales, So Kawaguchi, Simon Neil Jarman, Bruce Emerson Deagle

Corresponding author: Leonie Suter, leonie.suter@aad.gov.au

Introduction

The underlying mechanism of sex determination in Antarctic krill (*Euphausia superba*, hereafter called krill) is not known. Sex determination mechanisms could theoretically range from genetic to environmental to parasitic or any combination of these (Bachtrog et al. 2014; Carmichael et al. 2013; Duron et al. 2008; Kato et al. 2011). Furthermore, sex determination mechanisms can evolve rapidly, and may even differ between populations of the same species (Pen et al. 2010). Similar to other Euphausiacea species, krill probably do not have heteromorphic sex chromosomes (Thiriot-Quiévreux and Cuzin-Roudy 1995; Thiriot-Quiévreux et al. 1998), but a sex determination region may still exist within the krill genome. However, the krill genome is very large with an estimated size of 47 gigabases (Deagle et al. 2015; Jeffery 2011) and contains many repetitive elements (Deagle et al. 2015), which are the primary reasons that genome sequencing and assembly has not been carried out (Jarman and Deagle 2016).

One goal of our study was to design a molecular test to determine sex of krill. Ideally, such a test would be DNA based, as storage of samples for DNA extraction is easier than for RNA extraction (Wit et al. 2012). We pursued two separate avenues to develop a simple molecular DNA test, including restriction site-associated DNA sequencing (RAD-seq) (Fig. 1), with the aim to determine the sex of krill at any developmental stage, as well as to potentially determine sex-ratios in mixed bulk samples of krill containing both sexes.

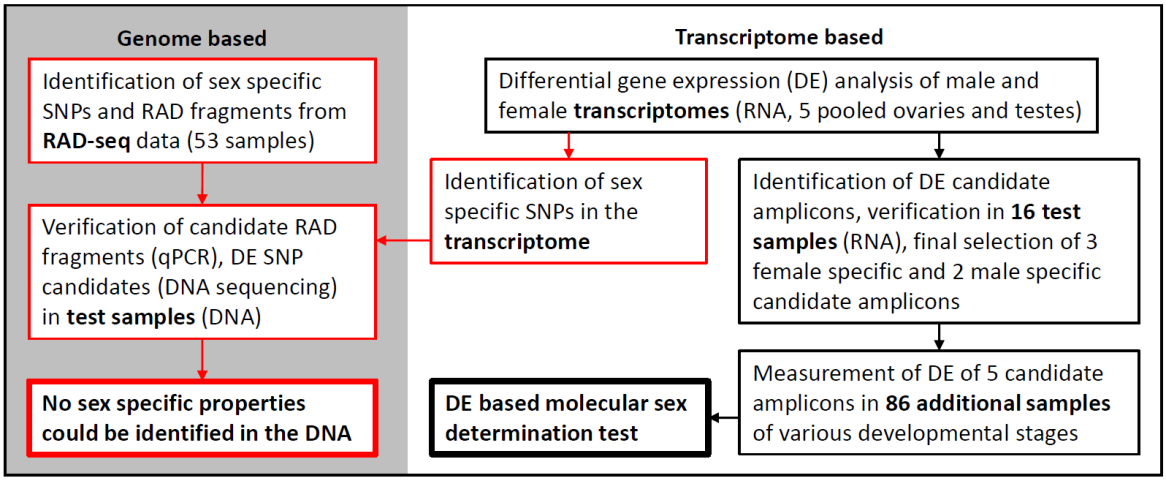

Figure 1: Flow chart of methods applied to identify genetic differences between male and female krill. Red outlines indicate methods and results described in this electronic supplemental material, whereas black outlines describe methods and results from the main manuscript.

Material and Methods

RAD-seq

To search for sex-specific differences in krill DNA, we re-analysed RAD-seq data originally generated for a population genomics study (Deagle et al. 2015). We used data from 53 samples of known sex (26 males, 27 females) originating from the "Mawson" and "Casey" populations to identify sex-specific DNA markers.

RAD-seq sequences of the male and of the female krill sample with the most reads were used to generate two separate assemblies (Deagle et al. 2015). To find sex-linked single nucleotide polymorphisms (SNPs) in the krill genome, RAD-seq sequences were mapped to these two assemblies. Reads mapping to more than one reference sequence were discarded, and the maximum number of mismatches allowed was three. SNPs were called using SAMtools (Li et al. 2009) under the following parameters: minor variant frequency of 0.075, minimum 6X coverage, minimum phred genotype quality score of 15 and minimum per cent of the samples genotyped of 80%. Two separate assemblies (one for each sex) were used to ensure sex-specific markers were not being excluded and the two resulting data sets were examined separately for all subsequent analyses.

Various parameters and filters were applied to exclude ambiguous markers. For example, SNPs with tri-allelic calls within single individuals were excluded, as the RAD fragment was likely to be present in more than one copy in the krill genome. Additionally, if the most common genotype variant of a particular SNP occurred at a higher frequency than 80% across all sexed individuals, it was excluded, as it was unlikely to be sex-linked in our sample with nearly equal numbers of males and females. To screen for sex linked SNPs, bi-allelic SNPs were assigned numeric values (0 for the first homozygous genotype, 0.5 for heterozygotes and 1 for the second homozygous genotype) and correlations between sex and numeric genotypes were calculated using R (R Core Team 2017).

Next we explored whether there were any RAD sequences unique to one sex by mapping all sequences to the male and female assembly and examining whether the sequences were present in both sexes. For example, a male specific sequence would only show up in males but not females when mapped to the male assembly, and would be completely absent in the female assembly. As variable sequencing coverage across samples meant that many samples had a lot of missing data, male specific candidate fragments were chosen if at least ten male samples had positive read numbers on the male assembly, and average female read number was less than one (to allow for minimal false positive reads in females but to exclude reads that were simply more common in males but also present in females). Candidate reads were then considered reads that had a *p*-value of 0.05 when

comparing male read numbers vs female read numbers. This way, 45 male specific candidates and 82 female specific candidates were obtained. These candidate sequences were then blasted against both the male and female RAD assemblies as well as the krill transcriptome, and any fragments that had multiple good hits were excluded from further analyses. For the remaining top ten male and top ten female candidate fragments, primer pairs were designed using primer3 (Koressaar and Remm 2007; Untergasser et al. 2012) where possible. RAD fragment length of 90 base pairs (bp) meant that for two male and two female candidates no suitable primer sequences could be found, even with primer design conditions less stringent than the standard settings.

To verify whether the candidate fragments were only present in one sex, 16 primer pairs (eight from the male and female assemblies each) were PCR amplified on a LightCycler 480 II using genomic DNA from three male and three female krill (a subsample of the "Test" samples, compare to main manuscript). PCR reaction mix contained 0.1 µL each of 10 µM forward and reverse primer; 0.1 µL of bovine serum albumin (BSA, 20 mg ml^{-1}), 0.5 µL EvaGreen dye, 20 x in water (Biotium, Hayward, CA, USA); 5 µL 2 x Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs, Ipswich, MA, USA) and 1 µL of DNA template (10 ng $µL^{-1}$) in a total reaction volume of 10 µL. PCRs were done on a LightCycler 480 II (Roche Diagnostics Australia Pty Limited, North Ryde, NSW, Australia) and thermal cycling conditions were 98 °C for 2 mins followed by 40 cycles of 98 °C for 5 secs, 58 °C for 20 secs and 72 °C for 20 secs, followed by a melt curve from 50 °C to 99 °C at a ramp rate of 2.2 °C/s with five acquisitions per degree.

Transcriptome

We screened newly sequenced transcriptomic data from krill gonads (compare to main manuscript) for sex-specific SNPs using the software discosnp++ (Uricaru et al. 2015), which allows SNP calling in the absence of a reference genome. Eight SNPs had a classification rank of > 0.9 (i.e. the particular SNP was to more than 90% unique to one sex), and primer pairs surrounding these SNPs were designed as described in the main manuscript to characterise the SNPs in additional samples. To avoid amplification of multiple similar fragments, primer pair sequences were blasted against the krill transciptome (compare to methods in main manuscript) and consequently three primer pairs with hits to multiple transcripts were excluded. Genomic DNA of eight males and eight females ("Test" samples, compare to main manuscript) was used to PCR amplify the remaining five fragments, followed by fragment purification with Ampure magnetic beads (Agencourt) and bi-directional sequencing by dye terminator v 3.1 chemistry on an ABI 3130 capillary sequencer (Applied Biosystems).

Results

RAD-seq sex-specific genomic differences

Applying stringent filters to the SNPs identified in the RAD-seq analysis reduced the number of SNPs from 55 703 to 2 855 bi-allelic SNPs mapped to the male assembly, and from 63 092 to 3 363 bi-allelic SNPs mapped to the female assembly. The different assemblies contained distinct markers, but this was due to different sequences recovered from each reference specimen. None of these sequence variants identified in markers from either assembly were consistently sex-linked across all 53 samples, therefore we did not pursue this part of our study further.

Based on the same RAD-seq data, 16 fragments that were suspected to be only present in one or the other sex were PCR amplified using genomic DNA of three additional male and female samples each. PCRs amplified the same fragments across all six samples, thus sex-bias of these RAD fragments could not be confirmed.

Transcriptome sequence analyses to identify genomic differences between sexes

We identified eight potentially sex-linked SNPs in the gonad transcriptomic data from five pooled males and females. Five of these SNPs were PCR amplified and sequenced using genomic DNA of the 16 “Test” samples, however, no sex-linkage of the SNPs could be confirmed.

Discussion

In this study we could not detect any sex-specific properties of the genome related to sex determination of krill, despite investigating thousands of genetic markers spread throughout the genome. This suggests that krill probably do not have a simple XX/XY or ZW/ZZ sex determination system with a large genomic region or a heteromorphic sex chromosome responsible for sex determination. It is possible that there is a smaller genetic region responsible for sex determination, or a more complex genetic sex determination system (Bachtrog et al. 2014) that is more difficult to disentangle from the data used in this study. It is also possible that several mechanisms may interact, for example a basic genetic sex determination system may occasionally be overwritten by environmental factors (Bachtrog et al. 2014; Palaiokostas et al. 2013). Detection of genetic markers in such a scenario would be challenging, as they would not always be strictly associated with one sex. Developing good genetic markers in large numbers is difficult in krill due to the enormous size of the genome (Jeffery 2011), with many regions existing in multiple copies (Deagle et al. 2015), and due to the high intraspecific genetic diversity (Deagle et al. 2015; Goodall-Copestake et al. 2010; Jarman and Deagle 2016). Although RAD sequencing has proven to be very effective in detecting markers associated with sex in some species (Carmichael et al. 2013; Gamble et al. 2015; Gamble and Zarkower 2014), it may not be the most appropriate method to find suitable genetic markers in krill, as the majority of SNPs in our analysis had to be discarded due to multiple copies of the same sequence in the krill genome. Thus other methods may have to be employed in the future to discover sex-specific markers in krill. Affordable long-read sequencing technologies such as nanopore sequencing (Branton et al. 2008) or PacBio sequencing (Rhoads and Au 2015) may enable better assembly of this large, complex genome, which would provide a valuable resource for studies of krill genetics. This would allow questions on fundamental biological characteristics of krill, such as sex determination, to be fully addressed.

References

Bachtrog D, Mank JE, Peichel CL et al. (2014) Sex determination: why so many ways of doing it? PLoS Biol 12:e1001899 doi:10.1371/journal.pbio.1001899

Branton D, Deamer DW, Marziali A et al. (2008) The potential and challenges of nanopore sequencing. Nat Biotechnol 26:1146-1153

Carmichael SN, Bekaert M, Taggart JB et al. (2013) Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*) using RAD sequencing. PLoS One 8 doi:10.1371/journal.pone.0077832

Deagle BE, Faux C, Kawaguchi S, Meyer B, Jarman SN (2015) Antarctic krill population genomics: apparent panmixia, but genome complexity and large population size muddy the water. Mol Ecol 24:4943-4959 doi:10.1111/mec.13370

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstadter J, Hurst GD (2008) The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol 6:27 doi:10.1186/1741-7007-6-27

Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D (2015) Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. Mol Biol Evol 32:1296-1309 doi:10.1093/molbev/msv023

Gamble T, Zarkower D (2014) Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Mol Ecol Resour 14:902-913 doi:10.1111/1755-0998.12237

Goodall-Copestake WP, Pérez-Espona S, Clark MS, Murphy EJ, Seear PJ, Tarling GA (2010) Swarms of diversity at the gene cox1 in Antarctic krill. Heredity 104:513 doi:10.1038/hdy.2009.188

Jarman SN, Deagle BE (2016) Genetics of Antarctic krill. In: Biology and Ecology of Antarctic Krill. Springer, pp 247-277

Jeffery NW (2011) The first genome size estimates for six species of krill (Malacostraca, Euphausiidae): large genomes at the north and south poles. Polar Biol 35:959-962 doi:10.1007/s00300-011-1137-4

Kato Y, Kobayashi K, Watanabe H, Iguchi T (2011) Environmental sex determination in the branchiopod crustacean *Daphnia magna*: deep conservation of a *Doublesex* gene in the sex-determining pathway. PLoS Genet 7:e1001345 doi:10.1371/journal.pgen.1001345

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289-1291

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079

Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ (2013) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. PLoS One 8:e68389 doi:10.1371/journal.pone.0068389

Pen I, Uller T, Feldmeyer B, Harts A, While GM, Wapstra E (2010) Climate-driven population divergence in sex-determining systems. Nature 468:436

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics 13:278-289 doi:10.1016/j.gpb.2015.08.002

Thiriot-Quiévreux C, Cuzin-Roudy J (1995) Karyological study of the Mediterranean krill *Meganyctiphanes norvegica* (Euphausiacea). J Crust Biol 15:79-85

Thiriot-Quiévreux C, Leitão A, Cuzin-Roudy J (1998) Chromosome diversity in Mediterranean and Antarctic euphausiid species (Euphausiacea). J Crust Biol 18:290-297

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115-e115

Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, Chikhi R, Lemaitre C, Peterlongo P (2015) Reference-free detection of isolated SNPs. Nucleic Acids Res 43:e11 doi:10.1093/nar/gku1187

Wit P, Pespeni MH, Ladner JT, Barshis DJ, Seneca F, Jaris H, Therkildsen NO, Morikawa M, Palumbi SR (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Mol Ecol Resour 12:1058-1067