The data archived here includes:

(1) List of samples used for RAD-seq, Floragenex RAD-seq project overview document, report on Illumina sequencing and raw sequence data from all 148 krill in FASTQ format.

1a_Full_RAD_info_Dryad.pdf

1b_FGX_Project_Readme_v1.3.pdf

1c_FGX_Sequencing_QC_Report.pdf

Raw_Illumina_Sequence available at the Australian Antarctic Data Centre:
http://dx.doi.org/10.4225/15/556FAB354BE19

(2) The unique RAD tags (unitags) identified using the Floragenex bioinformatic pipeline. There are two assemblies given here in FASTA format – each using sequence data from a single individual krill (S1CasM03 or S2MawF01b). The RAD tags from sample Krill_S1CasM03 were used to call SNPs in downstream analysis.

2a_Deagle_Krill_S1CasM03_sequence_1_RAD_unitags.fasta
(239,441 total sequences)

2b_Deagle_Krill_S2MawF01b_sequence_1_RAD_unitags.fasta
(231,791 total sequences)

2c_FGX_unitag_assembler_log.txt

(3) Log giving parameters used to produce sorted binary alignment map (BAM) files generated from Bowtie alignment of raw Illumina data to the specified reference genome. SAMtools pileups files are used in downstream population genetics analysis.

3a_BSP_pipeline_log.txt

SAM_BAM_Pileups_vs_S1CasM03 and SAM_BAM_Pileups_vs_S2MawF01b available at the Australian Antarctic Data Centre:
http://dx.doi.org/10.4225/15/556FAB354BE19)

(4) PopGen Directory and log giving the parameters used to produce these summaries. Files are in VariantCall Format (VCF) 4.1. They summarise the genotype information for each krill. There are several datasets here produced using different reference RAD tags and varying stringency of genotype calls (std = standard; ultrstr = ultrastringent and rel =relaxed). Data from Krill_S1CasM03_vs_All_std.vcf were used in the paper (12114 SNPs).

4a_20131119_BDeagle_Krill_S1CasM03_vs_All_std.vcf ** Data used in paper**

4b_Full_RAD_info.csv ** Data on where each sample was collected**

4c_FGX_popgen_log.txt

Alternative PopGen datasets

4d_20131009_BDeagle_Krill_S1CasM03_vs_All_ultrstr.vcf
4e_20131009_BDeagle_Krill_S1CasM03_vs_All_rel_Rename.vcf
4f_20131010_BDeagle_Krill_S2MawF01b_vs_all_ultrastr.vcf
4g_20131010_BDeagle_Krill_S2MawF01b_vs_all_std.vcf
4h_20131010_BDeagle_Krill_S2MawF01b_vs_all_rel.vcf

(5) Components of the .vcf file parsed out and put into separate .csv files for convenience. Data from: 20131119_BDeagle_Krill_S1CasM03_vs_All_std.vcf. The separate datasets include Genotypes, Sequencing Depth and Alternate Allele counts for 12114 SNPs in 148 krill. The first 11 columns have information on the SNP (e.g. which position on each RAD tag the SNPs were called, the reference nucleotide, etc.).

5a_Dryad_GenoFGX_20131119_Sup_S1CasM03_vs_All_std.csv

5b_Dryad_SeqDepth_20131119_Sup_S1CasM03_vs_All_std.csv

5c_Dryad_AltCount_20131119_Sup_S1CasM03_vs_All_std_MAF_point1per.csv

(6) Data and R code for principal component analysis (PCA) of the count data from the 'Core Dataset' (a few invariant SNPs removed ~ 12000 markers). Four .csv data files are required to run the R code. The R code creates a dataset of counts of each nucleotide at each variable SNP. PCA is also done on re-sampled data with a maximum of 25 sequences for each marker. Finally, PCA is done on only the samples run in batch one to avoid the batch effect.

6a_Rcode_Count_PCA.r **(.txt file)**

6b_Rcount_Dryad_Geno.csv

6c_Rcount_Dryad_Depth.csv

6d_Rcount_Dryad_Alt.csv

6e_Rcount_Dryad_RunOneTwo.csv

(7) MtDNA sequence data for 142 krill. Two fragments were sequenced: 655 bp Cytochrome c Oxidase subunit I (COI) and 569 bp NADH Dehydrogenase subunit 1 (ND1). 136 COI and 139 ND1 sequences were obtained (140 krill had at least one mtDNA sequence). Sequences were trimmed to 593 bp (COI) and 494 bp (ND1) to standardise length. Files archived here include sample information (krill not in RAD-seq labelled _Extra), edited and raw sequences.

7a_Krill_Samples_mtDNA.pdf

7b_COI_For_Rev_Aligned_Trim_Edit_F2_NArm.fas

7c_NADH_ForRev_Align_Trim_edit_NRm.fas

7d_Raw mtDNA chromatogram files (Folder)