nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0482-x

In the format provided by the authors and unedited.

OPEN

Genomics of habitat choice and adaptive evolution in a deep-sea fish

Michelle R. Gaither[1,7]*, Georgios A. Gkafas[1,2], Menno de Jong[1], Fatih Sarigol[1,3], Francis Neat[4], Thomas Regnier[4], Daniel Moore[1], Darren R. Gröcke[5], Neil Hall[6,8], Xuan Liu[6], John Kenny[6], Anita Lucaci[6], Margaret Hughes[6], Sam Haldenby[6] and A. Rus Hoelzel[1]*

[1]Department of Biosciences, Durham University, Durham, UK. [2]Department of Ichthyology and Aquatic Environment, School of Agricultural Sciences, University of Thessaly, Volos, Greece. [3]Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. [4]Marine Scotland Science, Aberdeen, UK. [5]Department of Earth Sciences, Durham University, Durham, UK. [6]Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool, UK. [7]Present address: University of Central Florida, Genomics and Bioinformatics Cluster, Department of Biology, Orlando, FL, USA. [8]Present address: Earlham Institute, Norwich, UK. *e-mail: michellergaither@gmail.com; a.r.hoelzel@dur.ac.uk

Supplementary Methods

Sampling

Draft genome: A single specimen (Crup98) was used for Illumina sequencing while two other samples were required for PacBio sequencing (Scotia54 and Scotia90). See below for details on library preparation and sequencing. All samples were collected from the same haul at 1025m (57.3813ºN, -9.5972ºW) in 2006 during a Marine Scotland Science deep-water survey (Supplementary Table 6).

Genome re-sequencing: Samples for re-sequencing were collected during the Marine Scotland Science deep-water survey on September 11th 2015 from 750m (56.2665ºN; -9.2503ºW), 1000m (56.2523ºN; -9.3558ºN), 1500m (56.3443ºN; -9.6465ºW) and 1800m (56.3392ºN; -9.9092ºW) along a single transect off the west coast of Scotland (Supplementary Fig. 1; Supplementary Table 6).

Juveniles: Samples were collected during the Marine Scotland Science deep-water survey on September 8th 2015 from 1000m (57.6400ºN; -9.7223ºW) along transect 2 in Supplementary Fig. 1. Adults were also collected from 1000m and 1800m along the same transect.

Adults screened at outlier loci: Samples for outlier screening were collected during the Marine Scotland Science deep-water survey between September 5th-14th 2015 and as part of the ECOMAR project between 2007 and 2011 (ref 1) (Supplementary Fig. 1; Supplementary Table 6)

RNASeq[2]: Tissues of *Coryphaenoides rupestris* for RNASeq were collected from Skagerrak region in Norway in February 2015 with the help of Odd Aksel Bergstad and Søvik Guldborg at the Institute of Marine Research in Arendal, Norway (Supplementary Fig. 1; Supplementary Table 6).

RADSeq: Samples used for RADSeq were used in a previous study[1] with the exception of the individuals from Skagerrak (Norway). We sequenced a total of 207 samples from nine sampling sites across the North Atlantic including along the Mid-Atlantic Ridge and four locations in the Hebrides region (Supplementary Fig. 1; Supplementary Table 6).

Trawl survey and data used in modelling ontogenetic strategies: Marine Scotland Science's *RV Scotia* was used to make a deep water bottom trawl survey of the continental slope west of Scotland between the years 1998-2015. The survey took place in September of each year with trawling carried out along and down the slope from 55 to 59°N at approximately 9°W and at depths from 500 to 2000m. The same fishing net was used throughout with the cod-end fitted with an internal liner (20 mm mesh size) which ensured retention of fishes as small as a few cm in length. Due to the tendency for the elongate fragile tails of Macrourid species to break, length was measured as the distance (cm) from the snout to the first ray of the anal fin (the pre-anal fin length – PAFL). The total number of individuals was determined from a subsample (by weight) of the total catch with the final numbers per cm size class raised according to the proportion sampled. Efforts were made to ensure the sub-sample was representative of the entire catch, i.e. baskets were sampled randomly with respect to the sequence by which they were processed. The data set comprised of 120 hauls yielding a total of 176,733 fish. Not all fish were sexed, and any possible bias by depth

would need to be assessed by investigating possible sampling biases for size, however both sexes were represented at all depths sampled.

Genome sequence

Illumina library preparation and sequencing: High molecular weight DNA was extracted from a single individual, following a phenol chloroform protocol[3], and used to make a 350 bp and 550 bp insert fragment libraries using the Illumina TruSeq Nano method, starting with 100 ng and 200 ng DNA, respectively. DNA was sheared to the relevant size using a Covaris M220 Focused-ultrasonicator (Covaris). The resulting fragmented DNA was purified using Agencourt AMPure XP beads (Beckman Coulter) and checked for size distribution on an Agilent 2100 Bioanalyzer (Agilent Genomics). Following the manufacturer's protocol, fragments were end repaired, A-tailed and adapter ligated. Enrichment was achieved with eight rounds of PCR. Libraries were checked using a Qubit Fluorometer (Invitrogen) and a high sensitively chip on the Agilent Bioanalyzer.

Mate pair libraries were made using the Nextera Mate Pair Sample Preparation Kit (Illumina) using the gel plus option. Following manufacturer's protocol 4 µg of genomic DNA was tagmented. During this step, by use of a specially formulated mate pair transposome, the genomic DNA sample is simultaneously fragmented and tagged with a biotinylated mate pair junction adapter. In later steps this biotin adapter serves to facilitate purification of mate pair fragments. The DNA was cleaned on a Zymo DNA clean column (Zymo Research). The range of fragment sizes produced was checked using a DNA 12000 Bioanalyzer chip. Subsequently, the strand displacement reaction was performed and the DNA purified with AMPure XP beads. The resulting DNA fragments were size selected

using a 0.75% cassette for the BluePippin (Sage Science) using two ranges: 2-4 Kbp and 9-11 Kbp. DNA was recovered and ligated overnight at 30°C following the manufacturer's protocol. After incubation and heat inactivation, exonuclease was added to remove DNA which hadn't circularised. This DNA was then sheared using the Covaris M220 and purified using Streptavidin Magnetic Beads. The samples were subsequently end repaired, A-tailed and adapter ligated according to manufacturer's instructions and enrichment was achieved with 10 rounds of PCR. Libraries were checked for quantity and quality using a Qubit assay and a high sensitivity Bioanalyzer chip.

The Qubit and Bioanalyzer QC information was used to pool the fragment and mate pair libraries into a single pool. The quality and quantity of the pool was assessed using a Qubit assay and the Agilent Bioanalyzer and subsequently by qPCR using the Illumina Library Quantification Kit (Kapa Biosystems) on a Roche Light Cycler LC480II (Roche Molecular Systems) according to manufacturer's instructions. The template DNA was denatured according to the protocol described in the Illumina cBot user guide and loaded at 10 pM and later 12 pM concentration. The sequencing was carried out on two lanes of an Illumina HiSeq 2000 at 2 x 100 bp paired-end sequencing with v3 chemistry and later on two lanes of a HiSeq 2500 at 2 x 125 bp paired-end sequencing with v4 chemistry.

PacBio libraries preparation and sequencing: 7 µg of DNA was purified using 1x AMPure XP beads and the quantity and quality was assessed using Nanodrop and Qubit assay. In addition, the Fragment Analyser (Advanced Analytical), using a high sensitivity genomic DNA kit, was used to determine the average size of the DNA and the extent of degradation. This procedure was also used at the steps indicated below to determine the

average fragment size of the DNA. DNA was sheared using a Covaris g-TUBE at 5200 rpm and purified with AMPure XP beads. The DNA fragments were size selected on a 0.75% BluePippin cassette in the range 7-20 kb. DNA was treated with Exonuclease VII at 37°C for 15 minutes. For DNA end repair, the sample was incubated for 20 minutes at 37°C with DNA damage repair mix supplied in the SMRTbell library kit (PacBio). This was followed by five minutes of incubation at 25°C with end repair mix. DNA was cleaned using 0.5x AMPure XP beads and 70% ethanol washes.

DNA was ligated to the SMRTbell adapters overnight at 25°C. Ligation was terminated by incubation at 65°C for 10 minutes followed by exonuclease treatment for 1 hour at 37°C. The SMRTbell library was then purified using 0.5x AMPure XP beads. The quantity of the library was determined by Qubit assay and the average fragment size determined using a Fragment Analyser. The SMRTbell library was annealed to the SMRT sequencing primers at values predetermined by the Binding Calculator (PacBio) and a complex made with the DNA Polymerase (P6/C4 chemistry). The complex was bound to Magbeads which were used for 32 SMRT cells of sequencing. Sequencing was carried out using 240 minute movie times on the Pacific Biosciences RSII instrument.

Initial processing and QC sequence data: Basecalling and de-multiplexing of indexed Illumina reads was performed by CASAVA v. 1.8.2 (Illumina) to produce sequence data, in FASTQ format. The raw FASTQ files were trimmed to remove Illumina adapter sequences using CUTADAPT v. 1.2.1 (ref. 4). The reads were further trimmed to remove low quality bases, using SICKLE v. 1.200 (ref. 5) with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed. Singleton reads, i.e. those that

did not have a corresponding paired read, were subsequently discarded (Supplementary Table 7).

Filtered sub-reads from PacBio sequencing runs were generated from raw sequences generated from 32 SMRT cells. These were generated using the Pacific Biosciences SMRT Portal software v. 2.3 using the RS_Subreads.1 protocol with default parameters (minimum subread length=50 bp, minimum polymerase read quality=75, minimum polymerase read length=50 bp).

Genome assembly: Illumina read-pairs were assembled using ABySS version 1.6 (ref. 6), using multiple k-mer sizes. The optimal assembly was selected (k=81 bp) and the metrics are shown in Supplementary Table 8. This set of scaffolds was then carried forward, filtering those with a size of <1000 bp. Sequences were subsequently scaffolded using mate pair reads with SSPACE (version 3.0)[7], using 'bowtie' as the alignment tool, and allowing for a 50% divergence from the expected insert size. Scaffolded sequences were then further scaffolded and gap-filled using PacBio filtered subreads with PBJelly[8]. The new gap-filled scaffolds were subsequently subjected to another round of first scaffolding with SSPACE and then with PBJelly to yield a final assembly. All assembly metrics are presented in Supplementary Table 8. To assess coverage, all paired end reads were aligned to the assembled scaffolds using Bowtie2. Only one alignment per read was accepted and in the case of equally best alignments, one was selected at random. PCR and optical duplicate reads were marked and removed using the MarkDuplicates tool from picard-tools. Mean depth of coverage was calculated from the generated BAM file using genomeCoverageBed from the bedtools package. The mean coverage was 104x

based on this method, and this drops to 61x when the stringency is increased to mappings of quality 10 or greater (so that there is less than 1 in 10 chance that the read actually maps to a different location on the genome). Although the PacBio reads were based on two samples, this was not a critical issue since the primary function of these data was to aid the assembly of scaffolds, and neither long-read assemblers nor PBJelly require 100% match to the reference.

Genome Completeness: To assess completeness of the final genome assembly, the scaffolds were analysed using BUSCO v. 2.0.1 (ref. 42; vertebrata_odb9 database). BUSCO searches for core single-copy orthologues within an assembled genome and reports how many are completely or partially present to provide an estimate of genome completeness. Additionally, it reports the number of duplicated orthologues, which can be indicative of multiple contigs representing the same region, i.e. assembly artifacts, or multiple contigs representing multiple alleles of the same region, e.g. both copies of a region in a diploid genome. Supplementary Table 9 shows the BUSCO results for the final assembly.

Supplementary Table 9 shows an estimated genome completeness of 82.9% based on complete (single copy and duplicated) orthologues. Including partially complete (fragmented) sequences, 94.3% of the single-copy orthologues were detected, suggesting a high, but not complete, inclusion of all genetic content. 3.5% of orthologues were present in multiple copies. This figure is low enough to suggest that (a) in the majority of cases, both alleles of the diploid genome have been collapsed into a single copy, i.e. the assembly is likely to represent the haploid genome, and (b) the ancient genome duplication that occurred prior to the Teleost radiation (320-350 Ma) may either be small/partial, or perhaps not

represented in this assembly, i.e. gene copies may have been collapsed into a single copy. Alternatively, the duplicated genes may have diverged significantly that they represent different orthologous groups entirely.

Gene prediction and genome annotation: Final PacBio scaffolds were used as input to MAKER v. 2.31.8 (ref. 10), an annotation pipeline which can utilize various gene prediction tools, including SNAP[11] which was utilized here. Both transcript and protein evidence are provided to SNAP within MAKER to assist with predicting gene models.

Prior to annotation with MAKER, transcripts were assembled from RNAseq Illumina data derived from multiple tissues from *Corphanoides rupestris* which were obtained as follows. We extracted total RNA, including small RNA, from *C. rupestris* tissues using the Total RNA Purification Plus Kit (Norgen BioTek Corp) following the manufacturer's protocol. Five tissue types were extracted from a single individual collected from the Skagerrak region in Norway including muscle, liver, heart, brain, and gills. From a second individual, also from Skagerrak, an additional brain tissue was extracted for a total of six tissues. Extracted RNA was quantified and quality was assessed using gel electrophoreses and a NanoDrop Lite Spectrophotometer (Thermo Scientific). The KAPA Stranded mRNA-Seq Kit for Illumina platforms (Kapa Biosystems) was used to prepare libraries with a starting input of 2.0-4.0 µg of total RNA. Libraries were prepared following the manufacture's protocol for 200-300 bp insert size. Additionally, libraries for small RNA were prepared using the TruSeq Small RNA Library Prep Kit (Illumina) using an initial input of 1 µg of total RNA and following manufacturer's protocol until the cDNA purification step. Here we quantified each library using a High Sensitivity DNA Kit on a 2200 Tape Station (Agilent

Technologies) and pooled the six libraries in equal molar amounts and used a Pippin Prep (Sage Science**)** to select fragments between 135-170 bp. Each final library (six mRNA libraries and one small RNA library) was quantified using qPCR and a commercial library quantification kit (Kapa Biosystems) and were pooled in equal amounts and sequenced on a single Illumina HiSeq 2500 lane (125 bp paired end reads) at the DBS Genomics facility at Durham University. As with the other Illumina sequences utilized in this study, raw FASTQ files were adaptor and quality trimmed using CUTADAPT and SICKLE.

Transciptome assembly was performed using TRINITY v. r2013_08_14 (ref. 12). To improve the accuracy of gene predictors, it is important to first identify and mask repeats within the genome that is to be annotated. Repeats were detected within the final assembly using RepeatModeler v. 1.0.4 (ref. 13). The resulting set of repeats was provided to MAKER during gene prediction.

Gene Prediction: A first MAKER run was carried out to initially train SNAP using the assembled transcripts as RNA evidence and uniprot/swissprot proteins as protein evidence. Following this initial run, two additional MAKER runs were carried out to iteratively train and refine the SNAP gene predictions. Subsequently a final iteration of MAKER was carried out, keeping all gene predictions (regardless of whether the prediction lacked direct transcript or protein mapping evidence). This yielded a set of 102,302 predicted genes. All predicted protein sequences were annotated with Pfam and GO term, using InterProScan 5 (ref. 14), and were subsequently filtered to retain only genes/proteins that either (a) were supported by evidence provided to MAKER, whether protein or RNA, or (b) contained a

Pfam domain, as determined by InterProScan annotation. This was carried out using MAKER accessory scripts and ultimately yielded 33,252 protein coding genes.

In addition to GO and Pfam domain annotation, genes/proteins were also annotated with KOG (eukaryotic orthologous groups) and CDD (conserved domain database) information and descriptions by means of rpsBLAST searches of proteins against the respective databases (ftp://ftp.ncbi.nih.gov/pub/COG/KOG and https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd_help.shtml). Of the 33,252 protein coding genes, 47.0% (15,630) were annotated with GO terms [45.5% (15,114) with gene names, gene descriptions and GO terms], 33.7% (11,209) were annotated with conserved domain information, and 56.3% (18,730) were annotated with KOG information. In total, 70.0% (23,295) of genes were annotated at least once. This information is presented in Supplementary Figure 7.

Whole genome re-sequencing and QC

We extracted DNA using an E.Z.N.A Genomic Tissue DNA Kit (Omega Bio-Tek, USA) from 60 individuals (Supplementary Table 6) following the manufacturer's protocol and subsequently stored DNA at -20°C. DNA quality was assessed using gel electrophoresis to ensure high molecular weight and quantified using a Qubit 2.0 Fluorometer (Invitrogen). A single microgram of DNA was used for each sample and libraries were prepared using the TruSeq DNA PCR-Free Library Prep (Illumina). Libraries were prepared following manufactures protocol for a 350 bp insert. Final libraries were assessed using a High Sensitivity DNA Kit on a 2200 TapeStation (Agilent Technologies) and each library was quantified using qPCR and a commercial library quantification kit (Kapa Biosystems). We

pooled six individuals for sequencing on each of 10 lanes on an Illumina HiSeq 2500 (125 bp paired end reads) at the DBS Genomics facility at Durham University.

Variant discovery and genotyping: We trimmed Illumina adaptors and filtered the raw reads with Trimmomatic v. 0.33 (ref. 15). Adaptors were trimmed allowing for a maximum of three mismatches (ILLUMINACLIP: TruSeq3-PE-2.fa:3:30:10:8:true). Leading and trailing bases of quality below 3 were trimmed (LEADING:3 TRAILING:3) and a sliding window four bases wide was used to cut sequences once average quality per base dropped below 20 (SLIDINGWINDOW:4:20). Resulting reads less than 50 bases long were dropped. We used BWA v. 0.7.12 (ref. 16) (bwa mem command with -aM flags) to align the filtered reads onto the *C. rupestris* draft genome. After the alignment we used Picard[17] to add read groups and to remove duplicates.

Restriction-site Associated DNA Sequencing (RADSeq)

We prepared RADSeq libraries for individuals from nine populations across the North Atlantic (Supplementary Fig. 1) using a double digest protocol, with few modifications[3]. We digested genomic DNA using the SphI and MluCl restriction endonucleases (New England Biolabs) and employed a Pippin Prep (Sage Science) to select all fragments between 375-450 bp. The unique Illumina indices were incorporated onto the P2 adaptor end of DNA fragments using 10-12 rounds of PCR and a commercial library amplification kit (Kapa Biosystems). Library fragment size was estimated using a High Sensitivity DNA Kit on a 2200 TapeStation (Agilent Technologies) and each library was quantified using qPCR and a commercial library quantification kit (Kapa Biosystems). Libraries were standardized and pooled in equal molar amounts. Final pools were quantified as above and

sequenced on two lanes of an Illumina HiSeq 2500 (100 bp paired end reads) at the DBS Genomics facility at Durham University.

Raw data QC and filtering using STACKS: Raw paired end sequences were de-multiplexed, trimmed to a common length (92 bp), and reads with Phred scores below 20 were discarded using the process_radtags command in STACKS v. 1.35 (ref. 18). Individuals with less than 1 million reads were discarded, leaving 207 individuals across the nine populations (Supplementary Fig. 1; Supplementary Table 6). Paired reads were aligned to the *Coryphaenoides rupestris* draft genome using BWA v. 0.7.12 (bwa mem -aM). Each resulting sam file was converted to bam format using SAMTOOLS v. 1.3. (samtools view -bSh -q 10; ref. 19). Loci were assembled using the STACKS ref_map.pl (-m 3 -S) pipeline and its component programs. Loci were generated by the merging of three or more similar "stacks." Stacks in which the numbers of reads were more than two standard deviations above the mean were assumed to be repetitive elements and removed from analysis. Using the "populations" component in STACKS we filtered the dataset in two ways. First we filtered all 207 individuals retaining only those loci with ≥8x coverage and that aligned in ≥80% of individuals in seven of the nine populations. We used the minimum minor allele frequencies flag of 0.05. The resulting dataset consisted of 4989 loci. Next we ran the "populations" component in STACKS on only the four populations in the Hebrides region (Supplementary Fig. 1a: H1, 1188m; H2, 2023m; H3, 1934m; H4, 1031m). We retained only those loci with ≥10x coverage and that aligned in ≥80% of individuals in all four populations. We used the minimum minor allele frequencies flag of 0.05 and resolved 12,113 loci. From these datasets, GENEPOP (ref. 20) input files were produced using the "populations" component in STACKS and implementing the "write_single_snp" option. The

resulting files were converted into the appropriate formats using the program PGDSPIDER v. 2.1.0.3 (ref. 21) and used for population level analyses, including the detection of outliers putatively under selection (e.g. ref. 22).

PCR Screening

For each of the 4 outlier loci we designed a set of 3 oligonucleotide primers. The forward and reverse primers were designed to amplify a ~400 bp fragment that included the SNP while a third internal primer was designed such that the 3' end of the primer was homologous to one of the residues of the heterozygous SNP pair (Supplementary Table 10a). A single band resulted when the residue included in the third primer was not present, indicating a homozygote for the other allele. Candidate primers were aligned against the reference genome using Blast. Only those primers with just a single blast hit were considered further. To confirm primer specificity primers were screened using the web-based program Beacon Designer (http://www.premierbiosoft.com/molecular_beacons/) for length, melting temperature, GC content, GC clamp and primer secondary structure. For a subset of samples (and all juvenile samples), a further PCR screen was undertaken to resolve heterozygotes from homozygotes from double-banded results (Supplementary Table 10b).

PCR amplifications: PCR conditions were as shown in Supplementary Table 10. The PCR cycling profile was: 95^{o}C for 15 minutes; 30 cycles of 95^{o}C for 1 minute, annealing temperature for 30 seconds and 72^{o}C for 30 seconds; and final extension at 72^{o}C for 15 minutes. PCR products were verified by 1% agarose gel electrophoresis.

References

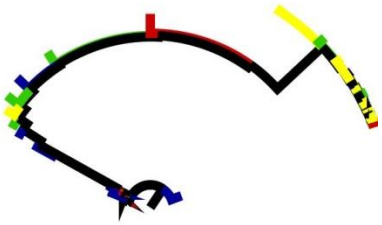
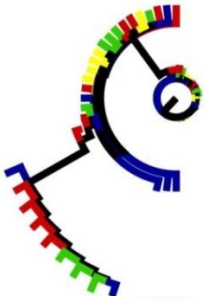
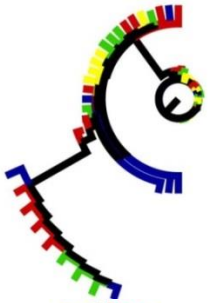
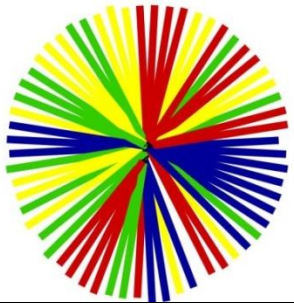
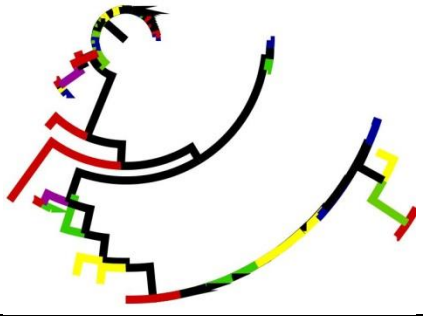
1. Priede, I. G. *et al.* The ecosystem of the Mid-Atlantic Ridge at the sub-polar front and Charlie–Gibbs Fracture Zone; ECO-MAR project strategy and description of the sampling programme 2007–2010. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **98, Part B,** 220–230 (2013).
2. Peterson, P. K., Weber, J. N., Kay, E. H., Fisher, H. S., Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135 (2012)
3. Hoelzel, A. R. (Ed) *Molecular Genetic Analysis of Populations: A Practical Approach*. (Oxford University Press, 1998).
4. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
5. Joshi N. A. & Fass J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. (2011)
6. Birol, I. *et al.* De novo transcriptome assembly with ABySS. *Bioinformatics* **25,** 2872–2877 (2009).
7. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27,** 578–579 (2011).
8. English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLOS ONE* **7,** e47768 (2012).
9. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).
10. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12,** 491 (2011).
11. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5,** 59 (2004).
12. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29,** 644–652 (2011).
13. Smit, A. F. A. & Hubley, R. RepeatModeler Open-1.0. (2010).
14. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30,** 1236–1240 (2014).
15. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
16. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
17. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).
18. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes Genomes Genet.* **1,** 171–182 (2011).
19. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
20. Rousset, F. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **8,** 103–106 (2008).

21. Lischer, H. E. L. & Excoffier, L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28,** 298–299 (2012).
22. Beaumont, M. A. & Nichols, R. A. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. B* **263,** 1619–1626 (1996).

Supplementary Table 1. F_{ST} comparisons among depths at 44,650 neutral loci. F_{ST} values are in the lower diagonal, while p-values are in the upper diagonal (none are significant).

	750m	**1000m**	**1500m**	**1800m**
750m		0.999	0.999	0.999
1000m	-0.0091		0.999	0.999
1500m	-0.0104	-0.0095		0.999
1800m	-0.0087	-0.0074	-0.009	

Supplementary Table 2. Results from SAGUARO. Colour coding is the same as for Fig.1.

Cactus	regions >300bp	Cumulative length	% genome	Illustration
5 (shows strongest separation for 1800m – yellow bars)	77	1,287,729	0.155	
0	54	395,791	0.06	
3	212	2,275,822	0.306	
4	21,162	436,433,952	52.63	
9	13,681	171,227,263	20.64	

Supplementary Table 3. A) Results of Hardy-Weinberg equilibrium tests for SNPs identified from 60 re-sequenced genomes. **B)** Nine non-synonymous outlier loci associated with depth. Loci were identified from the 60 re-sequenced genomes. Observed and expected frequencies compared to Hardy-Weinberg expectations are shown with the results of a chi-square test and associated *p*-value.

A)

Dataset	# SNPs	# SNPS out of HWE	Excess heterozygotes		Heterozygote Deficiency	
			# SNPs	F_{IS}	# SNPs	F_{IS}
Neutral SNPs	5,928,719	188,688 (3.18%)	183,858	-0.891 ± 0.094	4,830	0.774 ± 0.042
Outlier SNPs	346	34 (9.83%)	0	-	34	0.741 ± 0.131

B)

Locus	Genotype	Observed	Expected	Chi-Square	*p*-value
1041_33310	TT	24	13		
	TC	8	30		
	CC	28	17	18.70	< 0.001
1041_33313	GG	25	14		
	GA	7	30		
	AA	28	16	20.67	< 0.001
1041_33314	TT	25	14		
	TC	8	30		
	CC	27	16	18.65	< 0.001
1041_367049	CC	21	10		
	TC	5	28		
	TT	31	19	22.81	< 0.001
1041_367059	TT	21	10		
	TC	6	28		
	CC	29	18	20.71	< 0.001
4264_107730	TT	7	5		
	TC	19	24		
	CC	34	31	1.05	> 0.1
3193_255831	GG	5	14		
	GA	19	30		
	AA	35	15	14.73	< 0.001
5367_301064	GG	14	7		
	GA	13	27		
	AA	33	26	8.06	< 0.005
9875_168788	AA	21	11		
	AC	10	29		
	CC	29	20	14.04	< 0.001

Supplementary Table 4. Results from tests for Hardy-Weinberg equilibrium. Tests were conducted in **A)** juvenile samples from 1000 m and **B)** comparing juvenile allele frequencies against allele frequencies for adults from the same depth. **C)** Frequency of the 'deep water' alleles on contig 1041 loci identified by RADseq screening. Total number of alleles are in parentheses. **D)** Genotype frequencies for RADseq loci on contig 1041 associated with depth. Observed and expected frequencies tested against Hardy-Weinberg expectations are shown with the results of a chi-square test and associated *p*-value.

A)

Locus	Genotype	Observed	Expected	Chi-Square	*p*-value
1041	TT	35	27		
	TC	18	34		
	CC	18	10	8.24	< 0.005
3193_255831	GG	51	40		
	GA	18	39		
	AA	20	10	12.40	< 0.001
5367_301064	GG	10	5		
	GA	18	29		
	AA	46	40	4.66	< 0.05
9875_168788	AA	24	17		
	AC	30	44		
	CC	35	28	4.62	< 0.05

B)

	Allele frequencies			
Locus	Juveniles	Adults	Chi-Square	*p*-value
1041	T (0.62)	T (0.75)		
	C (0.38)	C (0.25)	5.91	< 0.02
3193_255831	G (0.76)	G (0.41)		
	A (0.24)	A (0.59)	44.45	< 0.00001
5367_301064	G (0.28)	G (0.41)		
	A (0.72)	A (0.59)	5.98	< 0.02
9875_168788	A (0.44)	A (0.75)		
	C (0.56)	C (0.25)	36.57	< 0.00001

C)

Population	Locus				
	4914_85[a]	4916_53[b]	4957_15[c]	5067_6[d]	5092_85[e]
shallow (<1800m)	0.268 (224)	0.279 (222)	0.304 (240)	0.570 (214)	0.291 (292)
deep (>1800m)	0.909 (88)	0.909 (88)	0.976 (82)	0.811 (90)	0.922 (90)

a:ATGCATTGGCCGGCTCGCGGCTGCATAAAGGAGATATTTCCCAGCCAATCAGGAACATCTAAA ATGTTGGGGAATGCACCACCAC/GCAGCAG

b:ATTGGACACGCTTGCAACTCTCACGAGCGCGCGCAACTCGCGTATGCACCGCG/ATGACGCAG GGGGAGGCGGCGTGGTGGTATGCACTTGGN

c:ATTCCCCCGCNCTCA/GCCCCCTCATAAAGGGACTATTGACAGCCCTTCCTGAGCCCACCGCCA ATTCGGCTGAGATGGCATTTTGCCCCNNN

d:ATTCCGG/ATTGTCCTTAAAGTTACACGGTACACCATCTCAGTGACGCTGTTATGGATAACGAT ATCGCAATGAAGACGTGTTGCCCTGGTGT

e:ATTACTTTTATTTTGAAATATTCTACCTTGTATGCCTTTGACAGAGAAGTAAGGGCCCAACTCC ATGGCCACATGCTGTGGATTT/GACACAC

D)

Locus	Genotype	Observed	Expected	Chi-Square	*p*-value
4914_85	AA	74	46		
	AB	22	79		
	BB	62	33	47.6	<0.00001
4916_53	AA	60	33		
	AB	22	77		
	BB	73	45	45.0	<0.00001
4957_15	AA	68	37		
	AB	17	80		
	BB	76	44	58.6	<0.00001
5067_6	AA	70	63		
	AB	55	71		
	BB	28	19	4.1	<0.05
5092_85	AA	72	37		
	AB	24	94		
	BB	94	59	60.8	<0.00001

Supplementary Table 5. A) The genes associated with depth-correlated outlier SNPs. Analysis of the 346 outlier SNPs detected from the Manhattan plot analyses of the 60 re-sequenced genomes revealed associations with 69 genes (one SNP was within 30 Kb of the identified locus, 15 within 10 Kb and the rest were within introns or exons). Genome reference and the proteins for which they code (http://www.uniprot.org/) are listed. **B)** Results of Gene Ontology (GO) analysis. GO anaylsis revealed twenty-nine terms that were significantly over represented in the outlier gene list. Sixteen of the 29 terms were associated with development or morphogenesis. *P*-values are after correction for false discovery. Source refers to Biological Process Propagated (BPP) or Slim lists. Numbers in the first column correspond to entries in table S16. **C)** Genes associated with GO terms**.** Numbers in the first column refer to numbers in column one of Table S15. Odds ratio for support, percent of list associated with a given GO term, and the list of associated loci are provided.

A)

Genome reference	Locus	Genome reference	Locus
CRUP_019093-RA	ACSL3	CRUP_012018-RA	LPHN2
CRUP_003594-RA	ADIP	CRUP_021651-RA	LPHN2
CRUP_006544-RA	AGAP3	CRUP_019574-RA	MARE2
CRUP_006545-RA	AGAP3	CRUP_019710-RA	MBNL1
CRUP_003591-RA	ALAT2	CRUP_019569-RA	MCM4B
CRUP_019091-RA	AP1S3	CRUP_003578-RA	MIB1
CRUP_017717-RA	AQPA	CRUP_003581-RA	MIB1
CRUP_006542-RA	ASB10	CRUP_020048-RA	MLRP2
CRUP_006540-RA	ASI1C	CRUP_003587-RA	MMP16
CRUP_006541-RA	ASI1C	CRUP_019703-RA	MTEF4
CRUP_006537-RA	ATG9A	CRUP_023136-RA	MUL1A
CRUP_021650-RA	B3GA2	CRUP_003584-RA	NPC1
CRUP_016848-RA	BRNP3	CRUP_010770-RA	OBSCN
CRUP_003583-RA	CABL1	CRUP_010774-RA	OBSCN
CRUP_012025-RA	CAC1E	CRUP_010771-RA	OBSL1
CRUP_012026-RA	CAC1E	CRUP_003598-RA	ODFP2
CRUP_012027-RA	CAC1E	CRUP_012288-RA	PARL
CRUP_012029-RA	CAC1E	CRUP_006550-RA	PCMD1
CRUP_012030-RA	CAC1E	CRUP_003589-RA	PTER
CRUP_019570-RA	CEBPD	CRUP_032792-RA	PTER
CRUP_021468-RA	CLCN7	CRUP_011818-RA	PTH1R
CRUP_012281-RA	CO4A2	CRUP_019709-RA	RAP2A
CRUP_011822-RA	CP8B1	CRUP_006546-RA	RBCC1
CRUP_017716-RA	CRFR2	CRUP_003567-RA	ROCK1
CRUP_019089-RA	CUL3	CRUP_019092-RA	SCG2
CRUP_014581-RA	D19L3	CRUP_012285-RA	SMS2
CRUP_019486-RA	ECE2	CRUP_024137-RA	SNAI2
CRUP_003586-RA	EGFR	CRUP_006551-RA	SNTG1
CRUP_019643-RA	EPHB1	CRUP_019571-RA	SPIDR
CRUP_019644-RA	EPHB1	CRUP_019572-RA	SPIDR
CRUP_025695-RA	EPHB1	CRUP_020046-RA	TECTA
CRUP_003582-RA	GATA6	CRUP_019568-RA	UB2V2
CRUP_029896-RA	GPC1	CRUP_003588-RA	UBC3
CRUP_012278-RA	GPC5	CRUP_012295-RA	UBC4
CRUP_003600-RA	GPR54	CRUP_027378-RA	VATO
CRUP_003576-RA	GRB1L	CRUP_011830-RA	WNT3A
CRUP_019828-RA	HES5	CRUP_011829-RA	WNT9A
CRUP_021652-RA	KAPCB	CRUP_019492-RA	YETS2
CRUP_011833-RA	KGUA	CRUP_012290-RA	YOF5
CRUP_012031-RA	KLH14	CRUP_012019-RA	ZN830

B)

	Gene Ontology Term	*p*-value	Source List
1	cellular protein localization(GO:0034613)	2.72E-03	BPP
2	anatomical structure formation involved in morphogenesis(GO:0048646)	2.72E-03	BPP
3	cellular macromolecule localization(GO:0070727)	2.72E-03	BPP
4	protein localization to organelle(GO:0033365)	3.58E-03	BPP
5	embryo development ending in birth or egg hatching(GO:0009792)	3.92E-03	BPP
6	chordate embryonic development(GO:0043009)	3.92E-03	BPP
7	positive regulation of cellular component organization(GO:0051130)	9.99E-03	BPP
8	embryonic morphogenesis(GO:0048598)	9.99E-03	BPP
9	catenin import into nucleus(GO:0035411)	1.92E-02	BPP
10	regulation of catenin import into nucleus(GO:0035412)	1.92E-02	BPP
11	regulation of multicellular organismal development(GO:2000026)	2.17E-02	BPP
12	regulation of cell death(GO:0010941)	2.17E-02	BPP
13	cardioblast differentiation(GO:0010002)	2.33E-02	BPP
14	cardiovascular system development(GO:0072358)	2.95E-02	BPP
15	circulatory system development(GO:0072359)	2.95E-02	BPP
16	in utero embryonic development(GO:0001701)	2.96E-02	BPP
17	regulation of cellular response to stress(GO:0080135)	2.96E-02	BPP
18	embryo development(GO:0009790)	2.96E-02	BPP
19	cell morphogenesis involved in differentiation(GO:0000904)	2.96E-02	BPP
20	regulation of intracellular protein transport(GO:0033157)	3.01E-02	BPP
21	regulation of alcohol biosynthetic process(GO:1902930)	3.02E-02	BPP
22	cardiocyte differentiation(GO:0035051)	3.39E-02	BPP
23	regulation of cell differentiation(GO:0045595)	3.89E-02	BPP
24	neuron differentiation(GO:0030182)	4.40E-02	BPP
25	regulation of establishment of protein localization(GO:0070201)	4.40E-02	BPP
26	cellular component morphogenesis(GO:0032989)	4.44E-02	BPP
27	cell differentiation(GO:0030154)	5.81E-04	Slim
28	proteinaceous extracellular matrix(GO:0005578)	8.94E-04	Slim
29	anatomical structure development(GO:0048856)	3.41E-02	Slim

C)

	Odds Ratio	List %	Loci Associated with GO Terms
1	1.77E+00	15.85	ATG9A,PARL,SPIDR,NPC1,RAP2A,ACSL3,AP1S3,WNT3A,SNAI2,EGFR,OBSL1,AGAP3,OBSCN
2	1.96E+00	14.63	EPHB1,GATA6,ROCK1,GPC1,WNT3A,MIB1,SNAI2,CUL3,OBSL1,HES5,OBSCN,SCG2
3	1.77E+00	15.85	ATG9A,PARL,SPIDR,NPC1,RAP2A,ACSL3,AP1S3,WNT3A,SNAI2,EGFR,OBSL1,AGAP3,OBSCN
4	2.15E+00	10.98	ATG9A,PARL,SPIDR,WNT3A,SNAI2,EGFR,OBSL1,AGAP3,OBSCN
5	2.09E+00	10.98	WNT9A,GATA6,WNT3A,MMP16,MIB1,EGFR,CUL3,HES5,MBNL1
6	2.10E+00	10.98	WNT9A,GATA6,WNT3A,MMP16,MIB1,EGFR,CUL3,HES5,MBNL1
7	1.93E+00	10.98	SPIDR,ACSL3,EPHB1,ROCK1,WNT3A,MIB1,SNAI2,CUL3,OBSL1
8	2.09E+00	9.76	WNT9A,GATA6,WNT3A,MMP16,MIB1,CUL3,HES5,MBNL1
9	4.05E+00	3.66	WNT3A,SNAI2,EGFR
10	4.05E+00	3.66	WNT3A,SNAI2,EGFR
11	1.57E+00	13.41	RAP2A,WNT9A,EPHB1,GATA6,ROCK1,GPC1,WNT3A,MIB1,SNAI2,OBSL1,HES5
12	1.57E+00	13.41	PARL,NPC1,WNT9A,EPHB1,GATA6,ROCK1,WNT3A,SNAI2,EGFR,OBSCN,SCG2
13	3.87E+00	3.66	ECE2,GATA6,WNT3A
14	1.69E+00	10.98	ECE2,EPHB1,GATA6,ROCK1,WNT3A,MIB1,SNAI2,OBSL1,SCG2
15	1.69E+00	10.98	ECE2,EPHB1,GATA6,ROCK1,WNT3A,MIB1,SNAI2,OBSL1,SCG2
16	2.16E+00	7.32	GATA6,WNT3A,MIB1,EGFR,CUL3,MBNL1
17	2.16E+00	7.32	SPIDR,NPC1,RAP2A,EPHB1,SNAI2,EGFR
18	1.68E+00	10.98	WNT9A,GATA6,WNT3A,MMP16,MIB1,EGFR,CUL3,HES5,MBNL1
19	1.66E+00	10.98	RAP2A,EPHB1,ROCK1,GPC1,WNT3A,SNAI2,EGFR,CUL3,OBSL1
20	2.43E+00	6.1	PARL,ACSL3,WNT3A,SNAI2,EGFR
21	3.58E+00	3.66	PTH1R,ACSL3,SNAI2
22	2.80E+00	4.88	ECE2,GATA6,WNT3A,OBSL1
23	1.50E+00	12.2	RAP2A,WNT9A,GATA6,ROCK1,GPC1,WNT3A,MIB1,SNAI2,OBSL1,HES5
24	1.47E+00	12.2	RAP2A,WNT9A,EPHB1,ROCK1,GPC1,WNT3A,MIB1,EGFR,OBSL1,HES5
25	2.03E+00	7.32	PARL,SPIDR,ACSL3,WNT3A,SNAI2,EGFR
26	1.46E+00	12.2	RAP2A,EPHB1,ROCK1,GPC1,WNT3A,SNAI2,EGFR,CUL3,OBSL1,OBSCN
27	1.95E+00	12.2	ECE2,WNT9A,EPHB1,GATA6,GPC1,WNT3A,PTER,SNAI2,HES5,MBNL1
28	2.55E+00	7.32	WNT9A,GPC5,GPC1,WNT3A,TECTA,MMP16
29	1.38E+00	10.98	PTH1R,ECE2,WNT9A,EPHB1,GATA6,WNT3A,SNAI2,HES5,MBNL1

Supplementary Table 6. Samples used in this study. Level of sequencing is listed with sample name, region of collection, depth in meters, latitude (Lat), longitude (Long) and month and year of collection (Date).

Sequencing	Sample	Region	Depth	Lat	Long	Date
Draft genome	Crup98	Hebrides	1025	57.3813	-9.5972	Sep-06
Draft genome	Scotia54	Hebrides	1025	57.3813	-9.5972	Sep-06
Draft genome	Scotia90	Hebrides	1025	57.3813	-9.5972	Sep-06
re-sequencing	Crup_Cru10	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru13	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru14	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru15	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru17	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru19	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru21	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru22	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru24	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru26	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru28	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru7	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru8	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_Cru9	Hebrides	750	56.2347	-9.2568	Nov-15
re-sequencing	Crup_J100	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_J66	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J67	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J68	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J70	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J71	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J75	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J76	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J77	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J78	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J80	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J81	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J82	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J83	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J85	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J86	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J87	Hebrides	1800	56.2927	-9.8812	Nov-15
re-sequencing	Crup_J99	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L01	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L011	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L013	Hebrides	1500	56.297	-9.6332	Nov-15

re-sequencing	Crup_L017	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L019	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L02	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L021	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L022	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L023	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L025	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L05	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L06	Hebrides	1500	56.297	-9.6332	Nov-15
re-sequencing	Crup_L26	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L27	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L29	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L30	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L31	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L32	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L35	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L36	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L42	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L43	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L44	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L45	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L46	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L48	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L49	Hebrides	1000	56.2062	-9.3827	Nov-15
re-sequencing	Crup_L50	Hebrides	1000	56.2062	-9.3827	Nov-15
RADSeq	Crup_SKG_55F04	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55F05	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55F06	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55F07	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55F08	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55F09	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55G03	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55G07	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55G08	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55G09	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55G10	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H01	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H02	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H03	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H04	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H05	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H06	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H07	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_SKG_55H08	Skagerrak	492	58.3897	9.9266	Jan-15

RADSeq	Crup_SKG_55H09	Skagerrak	492	58.3897	9.9266	Jan-15
RADSeq	Crup_FS_32I10	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_32I8	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_32J1	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_32J4	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_32J6	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_32J7	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_32J9	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33A10	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33A8	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33A9	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33B10	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33B4	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33B6	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33B7	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33B8	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33B9	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33C10	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33C1	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33C7	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33D3	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33D4	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33D5	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33D8	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33D9	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33F10	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33G1	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33G2	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33G4	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33H10	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33I6	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_FS_33J3	Faraday	990	49.8714	-29.6383	Nov-08
RADSeq	Crup_CGS_33E4	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33E6	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33E7	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33E8	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33E9	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F1	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F3	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F5	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F6	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F7	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F8	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_33F9	CG south	1263	51.5639	-30.3183	Nov-08

RADSeq	Crup_CGS_35I10	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_35I7	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_35J2	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_35J7	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_35J8	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36A5	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36A6	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36A7	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36A8	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36A9	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36B10	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36B5	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36B6	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36B9	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36C1	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36C3	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36C4	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGS_36C6	CG south	1263	51.5639	-30.3183	Nov-08
RADSeq	Crup_CGN_36C10	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36C8	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36C9	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D10	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D1	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D2	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D3	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D4	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D5	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D6	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D7	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D8	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36D9	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36E1	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36E2	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36E3	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36E4	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36E5	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_CGN_36E6	CG north	1650	52.9817	-34.8708	Nov-08
RADSeq	Crup_Biscay_14G10	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G2	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G3	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G4	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G5	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G6	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G7	Bay of Biscay	1339	47.9042	-8.36	Apr-99

RADSeq	Crup_Biscay_14G8	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14G9	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H10	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H2	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H4	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H5	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H6	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H7	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H8	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14H9	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14I2	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Biscay_14I3	Bay of Biscay	1339	47.9042	-8.36	Apr-99
RADSeq	Crup_Heb4_26G10	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26G7	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26G9	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26H1	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26H2	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26H3	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26H4	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26H6	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26I3	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26I5	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26I7	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26J10	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26J3	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26J7	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_26J9	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27A1	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27A2	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27A3	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27A6	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27A8	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27A9	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb4_27B6	Hebrides	1030	56.8333	-9.1667	Sep-06
RADSeq	Crup_Heb1_15A10	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15A8	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15A9	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15B2	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15B3	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15B6	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15B7	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15B8	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15B9	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15C10	Hebrides	1188	55.2861	-10.185	Apr-99

RADSeq	Crup_Heb1_15C1	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15C3	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15C4	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15C6	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15C9	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15D2	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15D9	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15E3	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15E4	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15E6	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15E7	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb1_15E8	Hebrides	1188	55.2861	-10.185	Apr-99
RADSeq	Crup_Heb3_16A10	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16A9	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B10	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B1	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B2	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B3	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B4	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B5	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B6	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B7	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B8	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16B9	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C1	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C2	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C3	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C4	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C5	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C6	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C7	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb3_16C8	Hebrides	1934	56.1908	-10.0947	Apr-99
RADSeq	Crup_Heb2_15H10	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I10	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I1	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I2	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I3	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I4	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I5	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I6	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I7	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I8	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15I9	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15J10	Hebrides	2023	56.275	-10.2733	Apr-99

RADSeq	Crup_Heb2_15J1	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15J2	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15J4	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15J5	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_15J9	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A1	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A2	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A3	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A4	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A5	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A6	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A7	Hebrides	2023	56.275	-10.2733	Apr-99
RADSeq	Crup_Heb2_16A8	Hebrides	2023	56.275	-10.2733	Apr-99
RNASeq	Crup_RNASeq_Skag_001	Skagerrak	492	58.3897	9.9266	Jan-17
RNASeq	Crup_RNASeq_Skag_002	Skagerrak	492	58.3897	9.9266	Jan-17

Supplementary Table 7. Summary of read counts from Illumina sequencing before and after adapter and quality trimming.

Sample	Untrimmed reads	Trimmed reads	Trimmed read pairs	Singleton reads
Sample_3kb_part1	133,959,150	132,355,431	65,391,174	1,573,083
Sample_3kb_part2	144,237,682	143,375,457	71,275,274	824,909
Sample_10kb_part1	117,507,416	115,868,767	57,291,035	1,286,697
Sample_10kb_part2	128,330,640	127,262,653	63,292,435	677,783
Sample_200-300bp_part1	329,792,724	326,004,707	161,176,637	3,651,433
Sample_200-300bp_part2	376,126,200	374,100,199	186,119,850	1,860,499
Sample_400-500bp_part1	243,944,250	238,814,862	116,959,301	4,896,260
Sample_400-500bp_part2	350,068,778	346,718,739	171,860,070	2,998,599

Supplementary Table 8. Genome assembly metrics. A is the initial assembly with ABySS, accepting scaffolds larger than 1000 bp in length. **B** represents **A** scaffolded with mate-pair reads using SSPACE. **C** is **B** scaffolded and gap-filled with PacBio reads using PBJelly. **D** is **C** scaffolded with mate-pair reads using SSPACE. **E** represents **D** scaffolded and gap-filled with PacBio reads using PBJelly. Lengths are reported in base pairs (bp).

Metric	**A**	**B**	**C**	**D**	**E**
Number of Scaffolds	146,377	87,682	61,854	48,293	39,765
Total length	494,686,465	595,280,705	743,702,681	774,768,614	829,209,412
N-bases	38,232,632	140,939,398	30,962,710	62,376,549	39,129,839
Non-N bases	456,453,833	454,341,307	712,739,971	712,392,065	790,079,573
N50 Number	16,743	2,394	2,348	1,162	1,171
N50	5,270	50,057	65,578	149,434	159,738
Longest Scaffold	320,319	1,264,117	1,336,384	2,200,949	2,284,816
Shortest Scaffold	1,000	1,000	888	913	914

Supplementary Table 9. Genome completeness. Percentage of complete, duplicated, fragmented, and missing core orthologues in the final genome assembly based on 2586 single-copy orthologues.

Category	Orthologues (%)
Complete – single copy	79.4
Complete - duplicated	3.5
Fragmented	11.4
Missing	5.7

Supplementary Table 10. Primer sequences. a) Primers used for initial gene screening. b) Primers used for second screening (with yellow highlighting showing changes). Optimal melting temperature (Tm) and GC content are listed. Experimental annealing temperatures and $MgCl_2$ concentrations used in PCR reactions are shown.

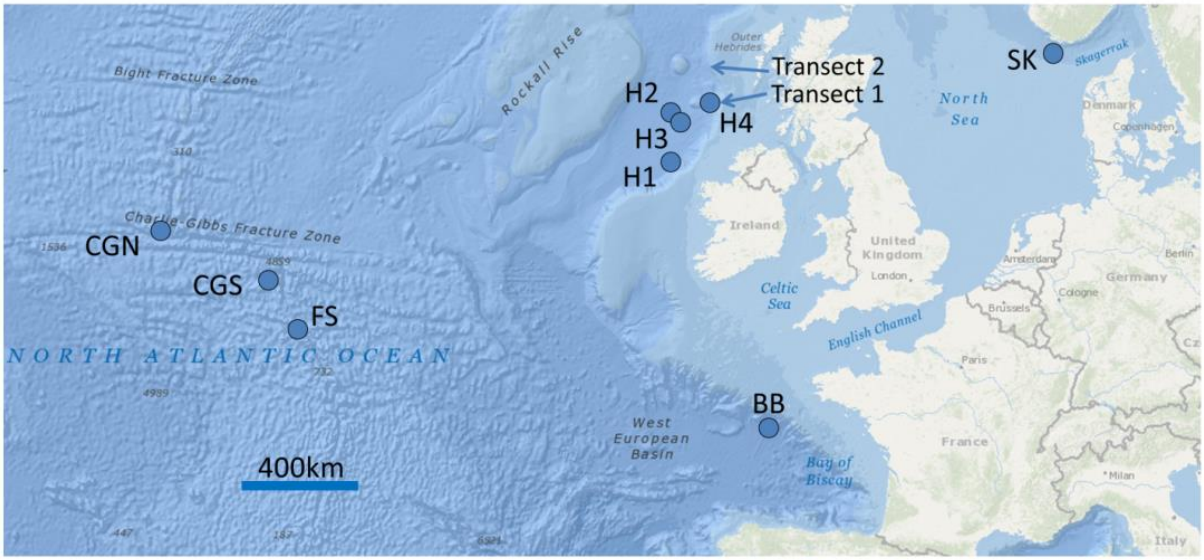
a)

Gene	Primer Sequence	Tm (°C)	GC%	Annealing temp (°C)	$MgCl_2$(mM)
Rock 1	F: TCCTGCCCCTCACCTCTCC	60.25	68.42		
	R: CCCCAGCCCTAGCCCAACC	62.29	73.68		
	R_I: TTCCCCTCTCGCCGCCCAC	64.48	73.68	63.7	1.0
YOF5	F: GCCTTTCTAATAATGTGC	45.96	38.89		
	R: TGGCAATTCTCCTCTATC	47.87	44.44		
	R_I: AGAGTTTGTCCCTTGGCT	52.89	50	48.0	1.5
OBSL1	F: TAAGCAGCAAGCTCAAAG	50.36	44.44		
	R: AAAGCCTCAGAAGCAATC	49.94	44.44		
	F_I: CTCAAGTGCAAAACTTAG	46.14	38.89	52.0	1.0
CAC1E	F: TACTAAGTGTCAGCTAAG	44.85	38.89		
	R: GTCAGTAAGACAACACAG	47.45	44.44		
	F_I: TTAGCAAAGAGCAAAGAG	47.24	38.89	52.0	1.0

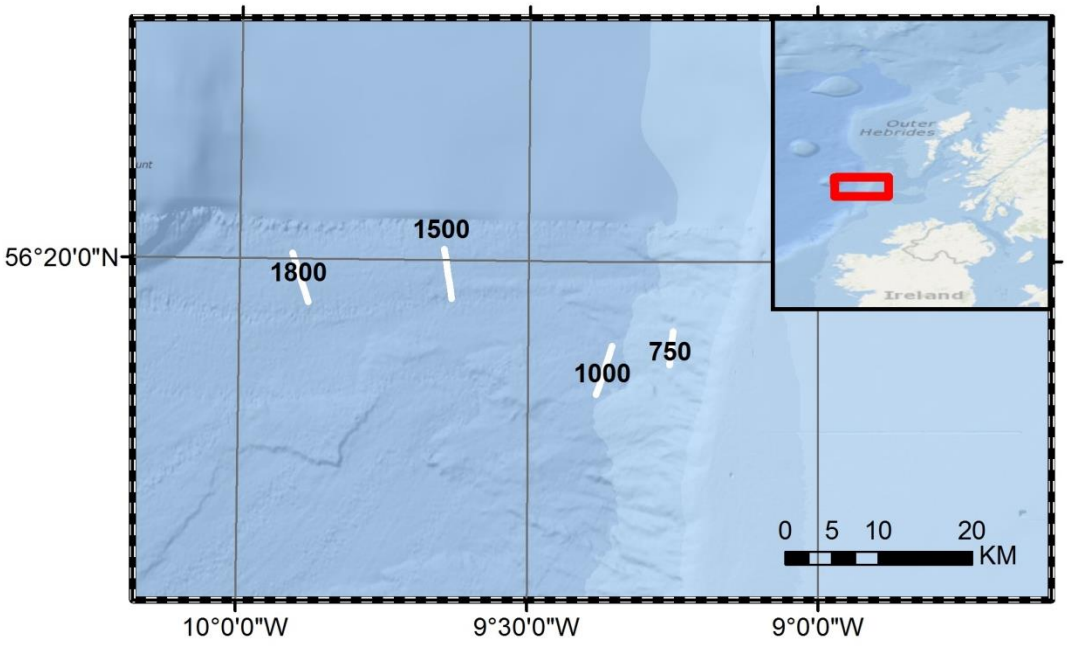
b)

Gene	Primer Sequence	Tm (°C)	GC%	Annealing temp (°C)	$MgCl_2$(mM)
Rock 1	F: TCCTGCCCCTCACCTCTCC	60.25	68.42		
	R: CCCCAGCCCTAGCCCAACC	62.29	73.68		
	R_I: TTCCCCTCTCGCCG**T**CC**GT**	**63.16**	**68.42**	63.0	1.0
YOF5	F: CCAGAGCTTGACCCCCAC	45.96	38.89		
	R: CGTGCCAAACTTCCCCAG	47.87	44.44		
	R_I: TTCCCGCTTCCTGTTGG**C**	**54.11**	**55.56**	48.0	1.5
OBSL1	F: TAAGCAGCAAGCTCAAAG	50.36	44.44		
	R: AAAGCCTCAGAAGCAATC	49.94	44.44		
	F_I: CTCAAGTGCAAAACTTA**A**	**45.48**	**33.33**	52.0	1.0
CAC1E	F: TACTAAGTGTCAGCTAAG	44.85	38.89		
	R: GTCAGTAAGACAACACAG	47.45	44.44		
	F_I: TTAGCAAAGAGCAAAGA**T**	**46.29**	**33.33**	52.0	1.0

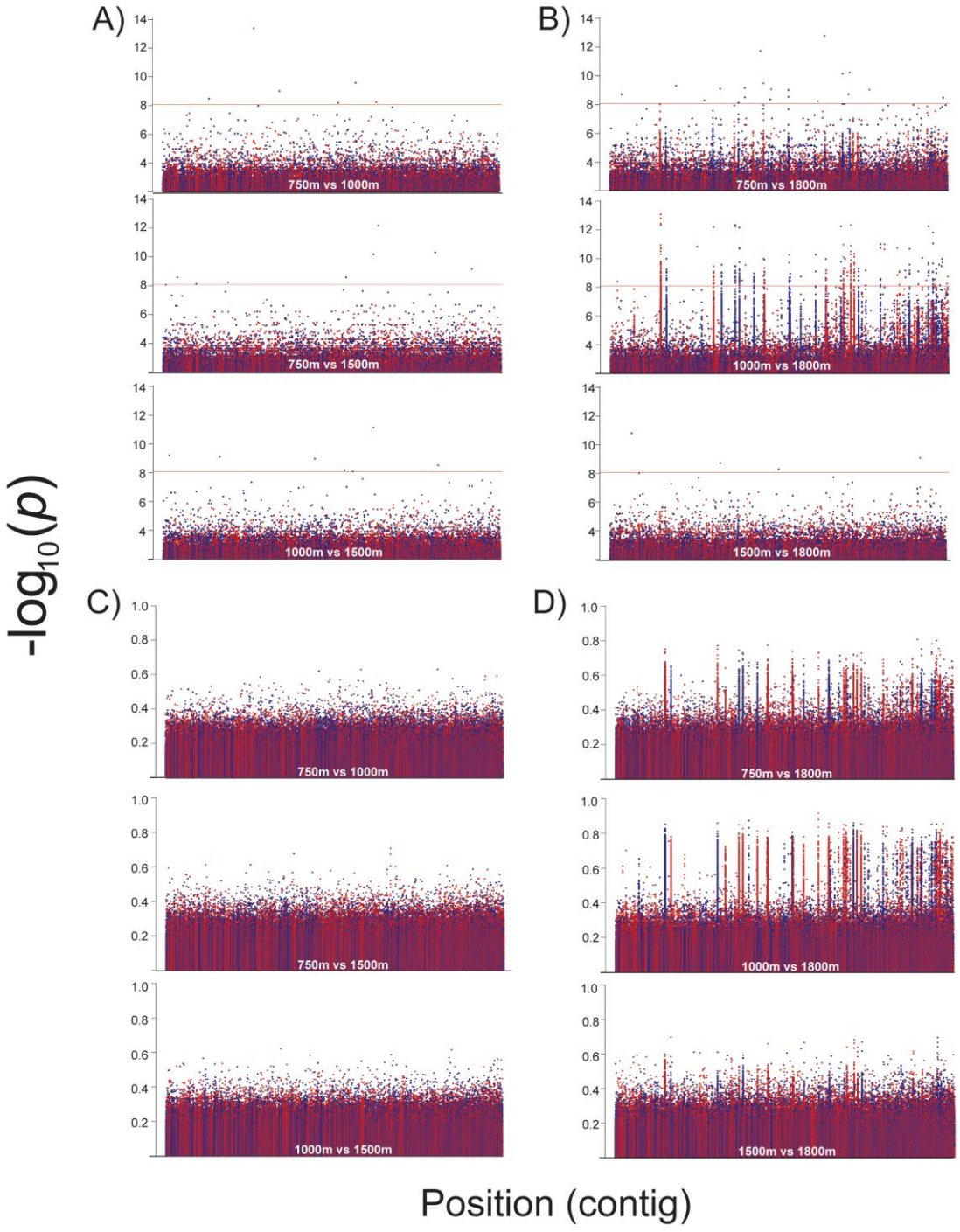
A)



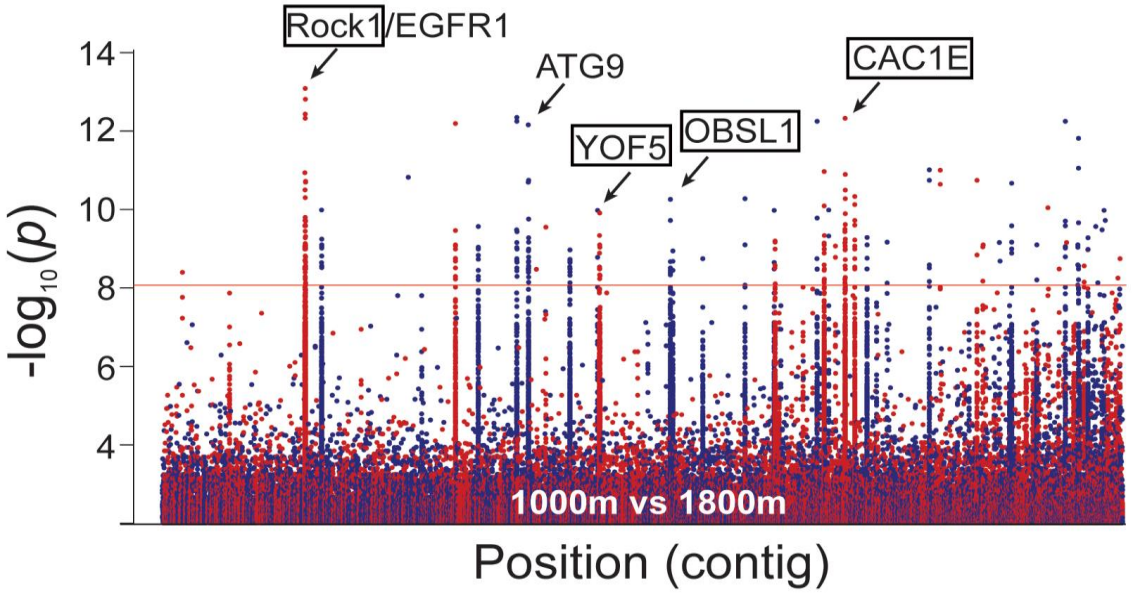
B)



Supplementary Figure 1 | Sampling sites. (a) Samples used in the RADseq analyses from north (CGN) and south (CGS) of the Charlie Gibbs Fracture Zone, from Faraday Seamounts (FS), the Bay of Biscay (BB), Skagerrak (SK), and four sites in the Hebrides (H1-H4). More precise location details are given in Table S1. The locations of two transect sampling sites are indicated. **(b)** The 60 individuals for which whole genomes were re-sequenced were collected along this transect during the Marine Scotland Science deep-water survey on September 11th 2015. This is indicated as 'transect 1' on the map in part a. Juveniles were collected at 'transect 2'.

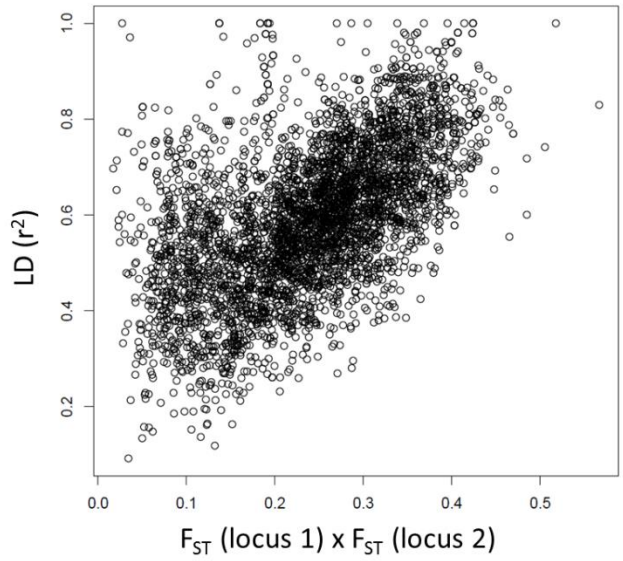

A)
B)
C)
D)
750m vs 1000m
750m vs 1500m
1000m vs 1500m
750m vs 1800m
1000m vs 1800m
1500m vs 1800m
-log10(p)
Position (contig)

E)

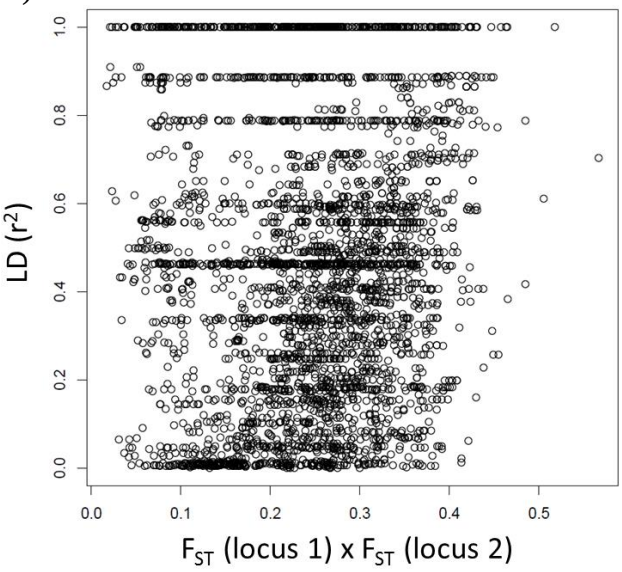


Supplementary Figure 2 | Results from the generalized linear model (GLM) and F_{ST} pairwise analysis. GLM was used to examine the correlation between genotype and depth across 60 re-sequenced genomes of *Coryphaenoides rupestris*. Negative log 10 of *p*-values based on the GLM are plotted in **(a, b)** while **(c, d)** show results based on F_{ST}. Comparisons in **(a, c)** involve shallower populations while those in **(b, d)** show comparisons with the populations at 1800m. **e**) Shows the position of the genes showing non-synonymous changes and those screened at the population level (in boxes).

A)

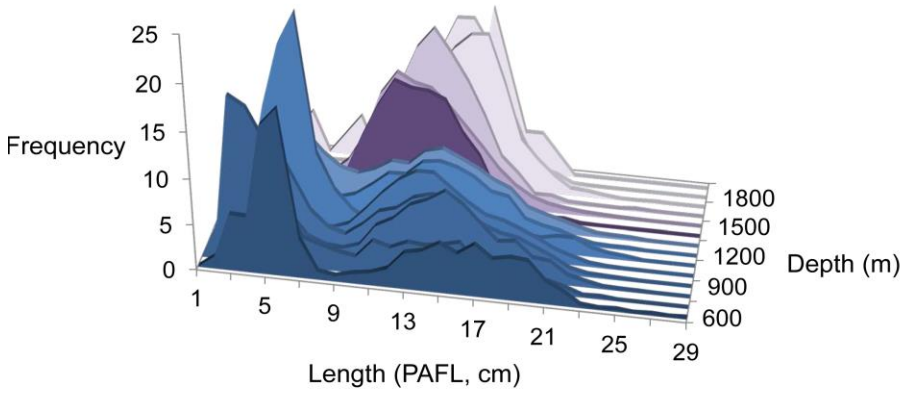


B)

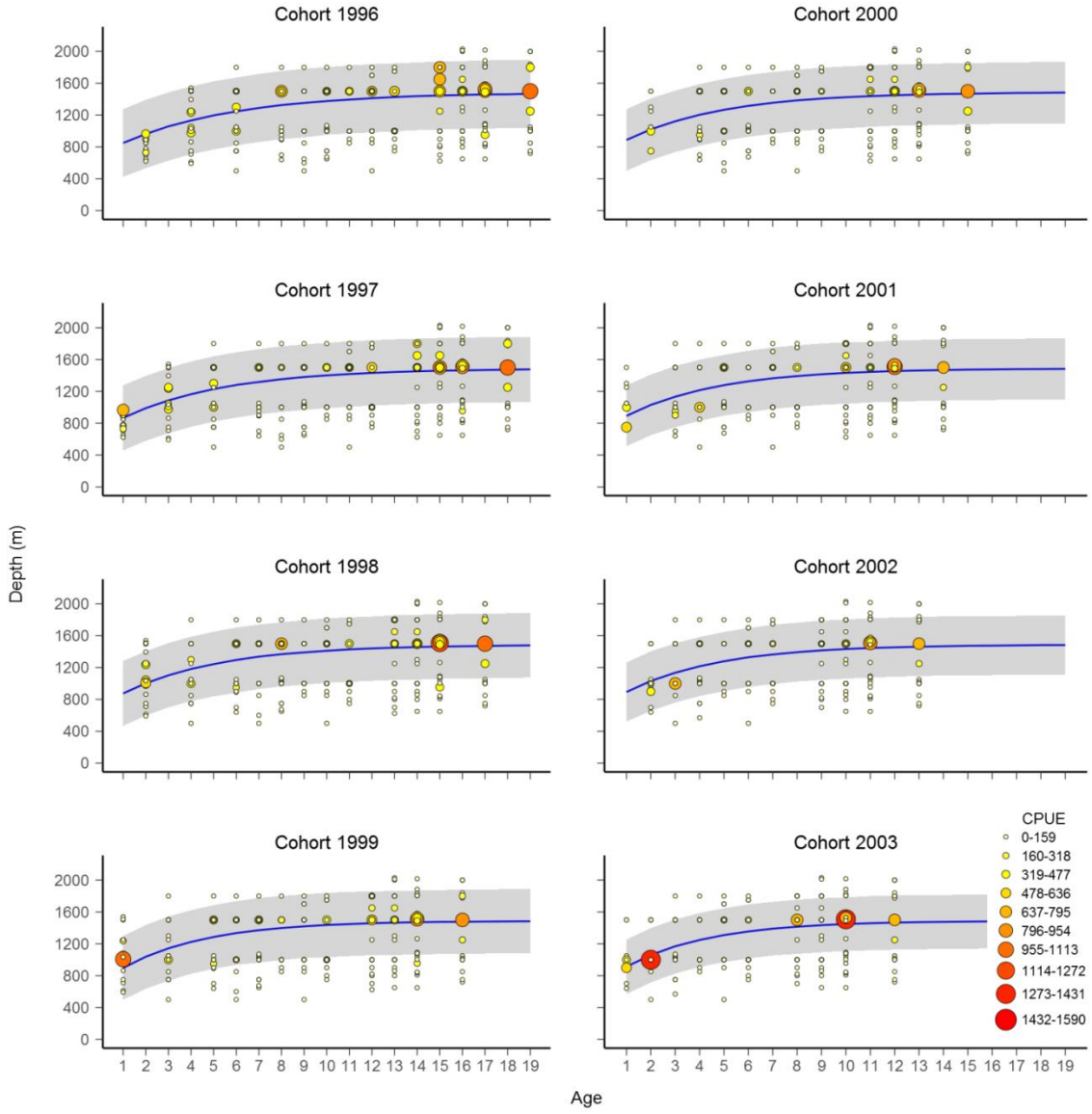


Supplementary Figure 3 | Plots of F_{ST} product against r^2 to assess for evidence of a 2-locus Wahlund effect (where a linear relationship is expected), based on all 346 outlier loci. (a) comparison between 1800m and the combined shallower sample sets (750m, 1000m and 1500m). **(b)** 1000m only.

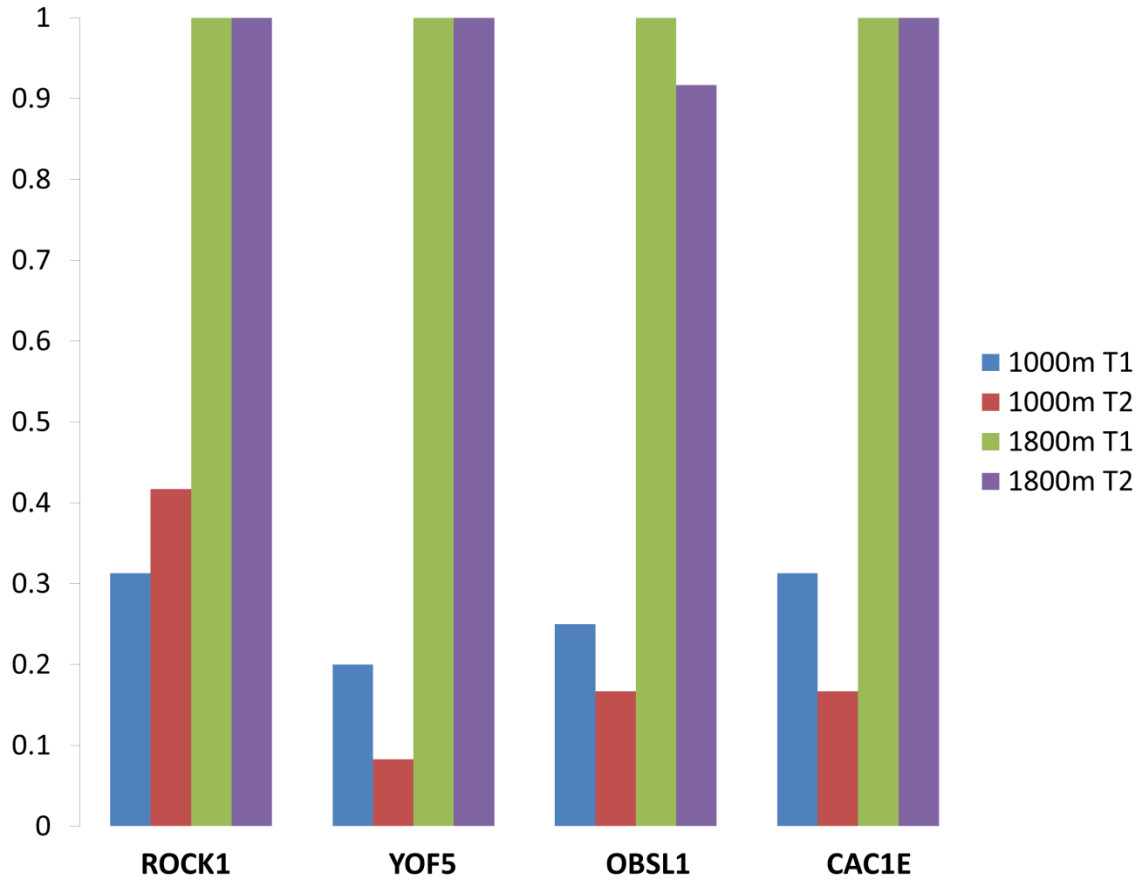
A)


Frequency
25
20
15
10
5
0
1
5
9
13
17
21
25
29
Length (PAFL, cm)
1800
1500
1200
900
600
Depth (m)

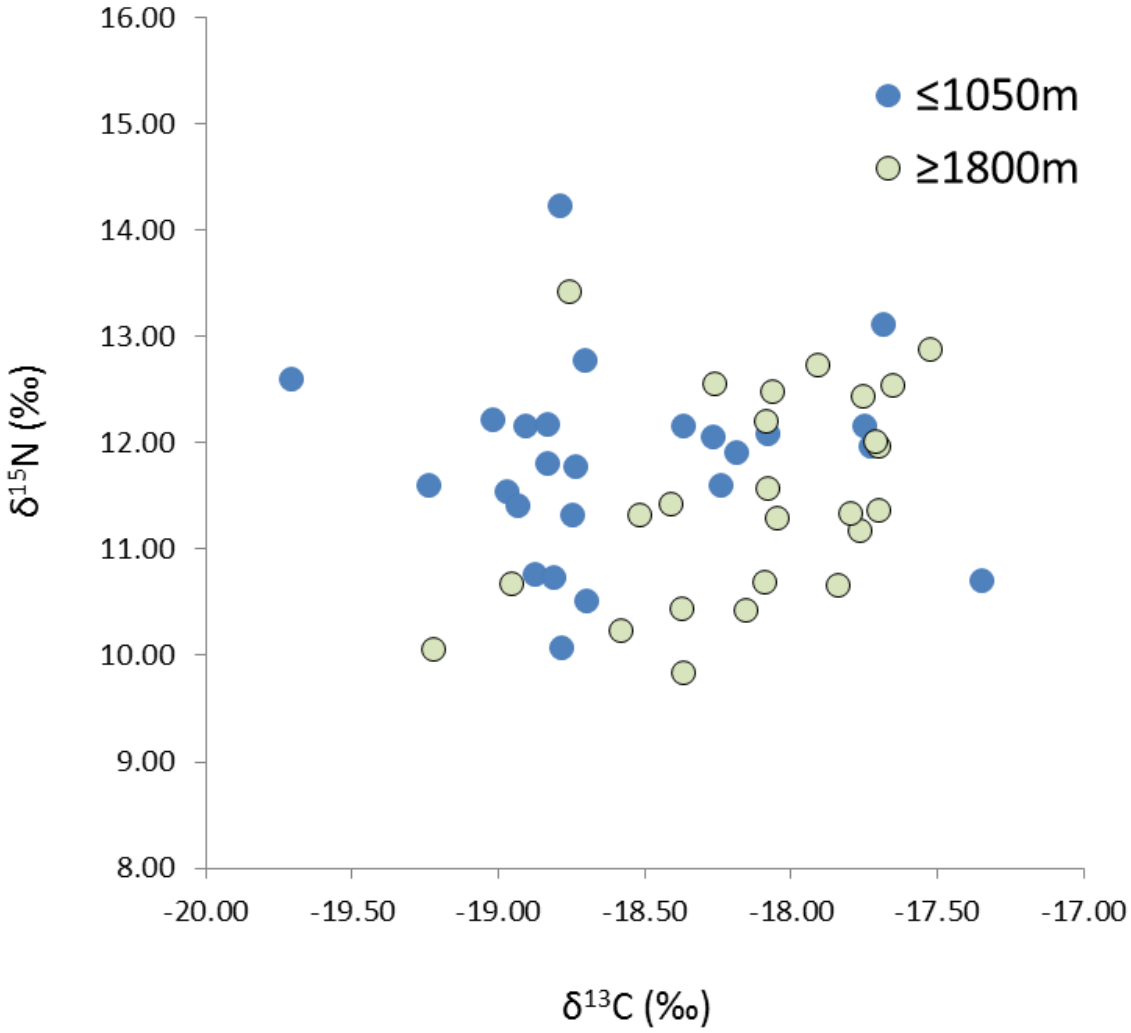
B)


Cohort 1996
Cohort 2000
Cohort 1997
Cohort 2001
Cohort 1998
Cohort 2002
Cohort 1999
Cohort 2003
Depth (m)
2000
1600
1200
800
400
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Age
CPUE
0-159
160-318
319-477
478-636
637-795
796-954
955-1113
1114-1272
1273-1431
1432-1590

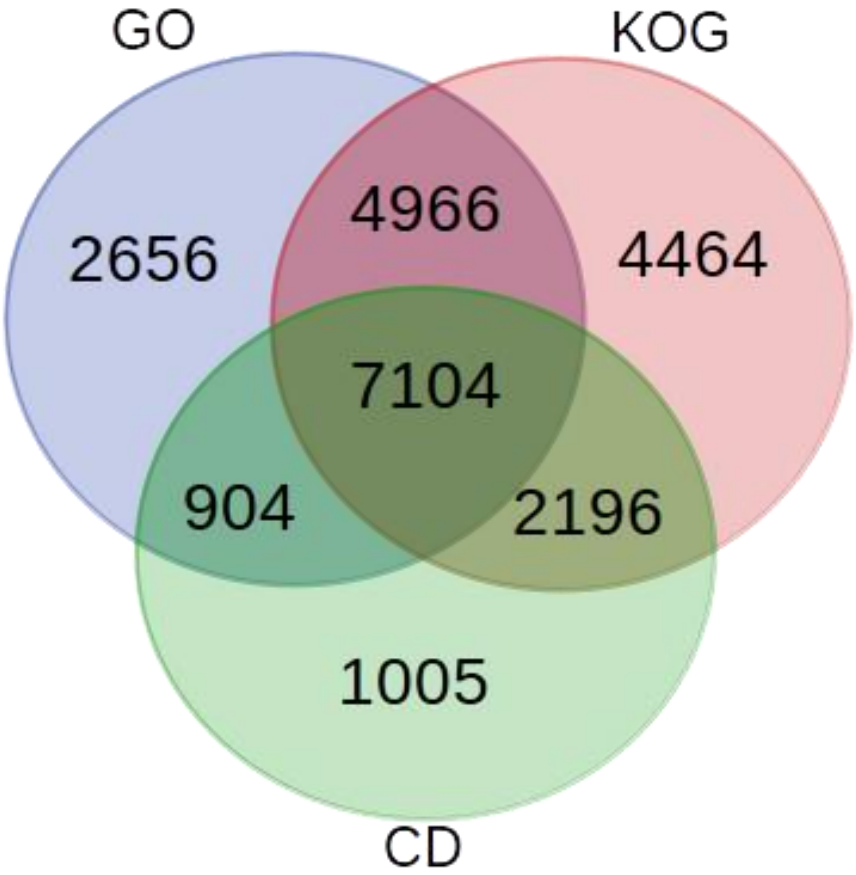
Supplementary Figure 4 | Cohort data on ontogenic migrations. (a) Frequency distribution of *C. rupestris* lengths caught at different depths. Juvenile size fish are predominantly caught in waters of 1000m or shallower, while adult age classes are caught at all depths. **(b)** Modelled ontogenetic depth migration from eight cohorts. Blue line is median trajectory for all cohorts with 95% confidence intervals (grey shaded area). Point size and colour reflects density of individuals (red and large = high, small and yellow = low). This model output was based on all available data. Due to the negligible variations in cohort-specific trajectories, the overall trajectory is presented. The key reflects catch per unit effort (CPUE).



Supplementary Figure 5 | Proportion of 'deep adapted' alleles at four loci along two transects (T1 & T2). See Supplementary Fig. 1 for transect locations.



Supplementary Figure 6 | Stable isotope profiles for *Coryphaenoides rupestris*. Individual fish were sampled at 1000m or 1050m (≤1050 m), or at 1800m or 2000m (≥1800m).



Supplementary Figure 7 | Venn diagram showing the number of genes annotated with either GO, KOG, or CD (conserved domain) information. 9,957 genes were un-annotated.