Isolation and characterisation of microsatellite DNA markers in the deep-sea amphipod *Paralicella tenuipes* by Illumina Miseq sequencing

Heather Ritchie1, Alan J. Jamieson2 and Stuart B. Piertney1

1 Institute of Biological and Environmental Sciences, University of Aberdeen, Zoology Building, Aberdeen, AB24 2TZ, UK

2 Oceanlab, University of Aberdeen, Newburgh, Aberdeenshire, AB41 6AA, UK

Address correspondence to Heather Ritchie, Institute of Biological and Environmental Sciences, University of Aberdeen, Zoology Building, Aberdeen, AB24 2TZ, UK or email: heather.ritchie.07@aberdeen.ac.uk

**Abstract**

Here we describe the development of 16 polymorphic microsatellite markers using an Illumina MiSeq sequencing approach in the deep-sea amphipod *Paralicella tenuipes*. A total of 25,577,844 DNA sequences were filtered for microsatellite motifs of which 197,873 sequences were identified. From these sequences 64 had sufficient flanking regions for primer design and 16 of these loci were polymorphic. Between five and 30 alleles were detected per locus, with an average of 13.63 alleles per locus, across a total of 120 individuals from five separate deep sea trenches from the Pacific Ocean. For the 16 loci, observed and expected heterozygosity values ranged from 0.116 to 0.414 and 0.422 to 0.820, respectively, with one locus displaying significant deviation from Hardy-Weinberg equilibrium. The microsatellite loci that have been isolated and described here are the first molecular markers developed for deep sea amphipods and will be invaluable for elucidating the genetic population structure and the extent of connectivity between deep ocean trenches.

Subject Areas: Population structure and phylogeography; Conservation genetics and biodiversity

Key words: deep-sea amphipod, hadal zone, population structure, *Paralicella*, microsatellite, high-throughput sequencing

**Introduction**

The Abyssal and Hadal zones are the two deepest marine biozones (3000-6000 m, and 6000 m – c. 11,000 m respectively). These are perhaps the most poorly explored and least understood ecosystems on Earth, particularly the hadal trenches (Jamieson and Fujii, 2011; Jamieson, 2015). The abyssal and hadal zones differ in that the former comprises large geographically expansive plains whilst the latter comprises deep trenches formed by tectonic convergence, the majority of which are located around the Pacific Rim (Jamieson et al., 2010). These trenches disrupt the continental shelf-slope-rise to abyssal plain continuum in that they form an array of spatially disjunct ultra-deep habitats which are often isolated by extremely large geographical distances. As a result of this isolation it was conventionally believed that hadal trenches lacked any level of connectivity through dispersal and gene flow, which would ultimately promote speciation and high levels of local species endemism (Wolff, 1970, 1960; Beliaev, 1989).

An excellent model system to test such supposition and examine patterns of connectivity, or conversely isolation, between trenches is the amphipods of the superfamily Lysianassoidea. They are a diverse, ubiquitous and abundant component of the abyssal and hadal fauna that can be readily sampled using baited traps (Blankenship et al., 2006; Jamieson et al., 2010; Fujii et al., 2013). The presence of seemingly cosmopolitan species of lyssianassoid amphipod, such as *Eurythenes gryllus*,within different trenches and across the adjoining abyssal plains would suggest that dispersal is possible between these habitats, despite the large geographical distance, and argues against any level of isolation among trenches. In contrast, however, apparent morphological and phylogeographic differences between populations of the same putative species suggests that populations are not panmictic and trenches do represent reproductively isolated habitats where gene flow is limited (Havermans et al., 2013).

The use of neutral molecular markers, such as microsatellite DNA length polymorphisms, can provide more direct estimates of the extent of population genetic structure between populations, and as such can be utilised to assess the extent to which hadal trenches represent demographically and evolutionarily independent units. Here we develop a suite of microsatellite markers for the lysianassoid amphipod *Paralicella tenuipes*, which is an abundant, cosmopolitan species found on the abyssal plains and hadal trenches around the entire Pacific Rim. We isolate and characterise multiple microsatellite loci by using Illumina MiSeq high-throughput DNA sequencing which is a useful tool for isolating new genetic markers in species with poor genomic resources and large genome sizes (Davey et al., 2011), while also circumventing the technical challenges of more traditional microsatellite enrichment protocols (Ekblom and Galindo, 2011).

**Materials and Methods**

**Sample Collection**

Specimens were collected from the Kermadec, Japan, Mariana, Peru-Chile and New Hebrides trenches around the Pacific Rim across multiple cruises (for details see Ritchie et al., 2015). Samples were collected using an autonomous lander vehicle (Jamieson et al., 2009) which was deployed with funnel traps secured to the feet and baited with locally sourced fish. Species were identified from species-specific mitochondrial COI barcodes according to Ritchie et al., (2015).

**DNA Extraction and Sequencing**

Genomic DNA (gDNA) was extracted from four whole individuals (~3-4mm) from the Japan trench using a conventional phenol-chloroform approach. Extractions were treated with 20 mg/mL RNAse A (Sigma-Aldrich). DNA was visualised for high molecular weight on a 1% TBE agarose gel then quantified using a Qubit 2.0 Fluorometer (Life Technologies) with subsequent dilution to ~20ng/µl. After purification a total of 5 µg of DNA pooled from four individuals was used for library preparation and shotgun sequencing by fragmentation at MWG Eurofins (Germany) following manufacturer’s protocols (Illumina, San Diego, CA).

**Microsatellite Discovery**

One lane of Illumina MiSeq v3, using a 2x350 bp paired-end length configuration, yielded 15,346 Mbp of data that passed Illumina quality control. Paired fragment libraries were aligned before *de novo* assembly using ABySS v1.5.2 (Simpson et al., 2009). Sequences had an approximate k-mer length of 25 bp (k=25) and a minimum number of five pairs were required to join ends (n=5). A total of 25,577,844 sequences were filtered for microsatellite arrays containing di-, tri- or tetra-nucleotide motifs with a minimum of five repeats using the software mreps v2.6 (Kolpakov et al., 2003). This returned 197, 873 microsatellite loci. Of these loci, 64 had over 50 base pairs of flanking sequence both upstream and downstream of the array to allow for primer design using PRIMER 3 (Rozen and Skaletsky, 2000).

**Primer Testing**

A total of 64 primer pairs were synthesised for amplification testing. Initial testing for PCR amplification success was undertaken using a total of 16 individuals from three different trench locations in the Pacific Ocean. PCR amplification reaction mixes contained 0.2 mM each dNTPs, 2.5 mM MgCl2, 0.5 µM each primer, 0.5U of BioTaq DNA polymerase (Bioline), 5-20 ng DNA template in 1x NH4 buffer (Bioline: 16 mM (NH4)2SO4, 67 mM Tris-HCl) in a total reaction volume of 10µL. PCR amplification was performed using a G-storm thermal cycler (G-storm Ltd, Surrey, UK) with the following touch-down conditions: initial denaturation at 94°C for 2 min, followed by; 20 cycles of denaturation at 94°C for 30 s, annealing at 65°C – 55°C (-0.5°C/cycle) for 30 s, extension at 72°C for 30 s; 15 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 30 s before a final elongation step at 72°C for 1 min.

After initial screening, 40 loci were excluded either due to PCR failure, lack of polymorphism or ambiguous profiles when genotyped. The remaining 16 loci were selected for genotyping across a total of 120 individuals from five different trenches to examine the level of polymorphism (Table 1). The selected loci are composed of one tetranucleotide repeat, three trinucleotide repeats and 12 dinucleotide repeats. Forward primers were labelled with one of four fluorescently labelled universal tails A: 6-Fam GCCTCCCTCGCGCCA, B: HEX GCCTTGCCAGCCCGC, C: ATTO 550 CAGGACCAGGCTACCGTG or D: ATTO 565 CGGAGAGCCGAGAGGTG (MWG Eurofins, Germany) following the procedures outlined in Blacket et al., (2012).

PCRs were performed in singleplex with a reaction mix containing 0.2mM each dNTPs, 2.5 mM MgCl2, 0.5 µM of reverse primer, 0.3 µM of forward primer with universal tail, 0.2 µM of flurochorome-labelled forward primer, 0.5U of BioTaq DNA polymerase (Bioline), 5-20 ng DNA template in 1x NH4 buffer (Bioline) in a total reaction volume of 10uL. PCR amplification was performed using a G-storm thermal cycler (G-storm Ltd, Surrey, UK) with the following touch-down conditions: initial denaturation at 94°C for 2 min, followed by; 20 cycles of denaturation at 94°C for 30 s, annealing at 65°C – 55°C (-0.5°C/cycle) for 30 s, extension at 72°C for 30 s; 15 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 30 s before a final elongation step at 72°C for 15 min.

Genotypes were resolved on an automatic ABI 3730 DNA Capillary DNA sequencer (Dundee DNA Sequencing Services Ltd) using a GS-500 (LIZ) size standard. Individual electropherograms were scored by eye using GeneMarker v 1.4 (SoftGenetics, 2010).

**Microsatellite loci analysis**

Micro-Checker 2.2.3 (van Oosterhout et al., 2004) was used to establish whether any observed heterozygote deficiencies were attributable to null alleles, scoring errors or large allelic dropout. Tests for linkage disequlibrium between all possible loci combinations were conducted using Genepop 4.1 (Rousset, 2011). Genetic diversity statistics were calculated using a variety of statistics: observed (HO) and expected (HE) heterozygosity, number of alleles per locus (NA) and the fixation index (FIS) were calculated alongside a test for deviation from Hardy-Weinberg equilibrium (PHWE), where p-values were subsequently adjusted after Bonferroni correction, using the diveRsity package in R (Keenan et al., 2013) (Table 2).

**Data Availability**

In accordance with the Journal of Heredity data archiving policy (Baker, 2013), we have deposited the primary data underlying these as analyses as follows: DNA sequences as Genbank accessions KT587291 – KT587306.

**Results and Discussion**

Here we demonstrate the utility of high throughput sequencing for isolating microsatellite loci in deep sea amphipods. In this case 197,873 microsatellite containing sequences were identified and from these a total of 64 microsatellite loci primers were designed for testing. Although there is a high attrition rate due to the generation of short reads with insufficient DNA sequence for primer design the number of primer sets that can be developed is still much greater than what would be expected from traditional microsatellite enrichment protocols. Next-generation sequencing produces a much larger, cost-effective database for microsatellite discovery more rapidly than traditional cloning-based approaches (Yu et al., 2011). It also has the additional benefit of limiting sequencing errors due to greater sequence coverage and quality control filters. While other next-generation sequencing approaches, such as RAD-seq, can provide a considerably greater number of markers with 1000s of single nucleotide polymorphisms (SNPs) for inferring population structure the relative amount of high-quality DNA required for each individual can be too high for samples with inherently poor DNA quality. Furthermore, these methods can also result in data with considerable noise where it is difficult to distinguish between true heterozygotes and artificially assembled variants of non-homologous but similar RAD-tag loci which is a major shortcoming for study systems without a reference genome (Mesak et al., 2014). This Illumina Miseq approach provides a valuable resource for a whole spectrum of taxa regardless of genome size and availability of genomic resources.

Of the 64 microsatellite loci investigated 16 were polymorphic. The number of alleles (NA) per locus varied from five to 30, with an average of 13.63 which is comparable to other amphipod microsatellite loci (Baird et al., 2012). Observed (HO) and expected (HE) heterozygosities ranged from 0.116 and 0.414, and 0.438 and 0.820, respectively. The expected heterozygosities are comparable to microsatellite loci obtained from other amphipod species but the observed heterozygosities are slightly lower (Baird et al., 2012).

Tests for linkage disequilibrium identified no significant interactions between any of the pairwise comparisons indicating that none of the loci are physically linked. Three markers showed significant deviations from HWE at the p<0.05 level (Paraten14, Paraten38 and Paraten54) due to a deficiency of heterozygotes but this was reduced to a single locus after Bonferroni correction (Paraten38). This lack of statistical significance is probably caused by low power due to small sample sizes and missing data which are both factors associated with deep-sea samples. Notwithstanding, the substantial heterozygote deficiency is probably biologically meaningful (Waples, 2015), and might be caused by population subdivision (Wahlund effect) or conflation of multiple cryptic species. These intriguing hypotheses will need to be addressed in a large-scale population study. Micro-Checker provided no support for this excess of homozygosity being due to null alleles, though this is somewhat difficult to reconcile given that only this single locus was not in Hardy Weinberg equilibrium. An alternative explanation is that the locus is linked to a gene under selection (Slatkin, 1995).

Amphipod samples retrieved from the deep sea frequently yield poor quality DNA due to the extreme change in hydrostatic pressure experienced by the cells which often results in cell death and DNA damage (Dixon et al., 2004). This is further compounded by a time delay between the death of the animal and its preservation on deck which allows endogenous nucleases to begin degrading the DNA (Hofreiter et al., 2001). As a result of this degradation the microsatellite primer sets were amplified in singleplex to maximise amplification success but with sufficient DNA quality these could be combined and the PCRs could be performed in multiplex.

Here we have produced novel polymorphic microsatellite markers from the first time in an abyssal and hadal amphipod species. These microsatellite loci will provide essential data for inferring the true patterns of movement across abyssal plains and between hadal trenches in an important amphipod species. This will allow us to address the long standing question of whether hadal trenches truly represent geographically isolated habitats which in turn have resulted in distinct, isolated populations of species.

**Funding**

This work was supported by the HADEEP projects, funded by the Nippon Foundation, Japan (2009765188), the Natural Environmental Research Council, UK (NE/E007171/1) and the Total Foundation, France. We acknowledge additional support from the Marine Alliance for Science and Technology for Scotland (MASTS) funded by the Scottish Funding Council (Ref: HR09011) and contributing institutions. We also acknowledge support from the Leverhulme Trust to SBP. Additional sea time was supported by NIWA’s ‘Impact of Resource Use on Vulnerable Deep-Sea Communities’ project (CO1\_0906).

**Acknowledgements**

We thank the chief scientists, crew and company of the Japanese RV *Hakuho-Maru* (KH0703 and KH0803), the RV *Tansei-Maru* (KT-09-03), the RV *Kairei* (KR0716), the German FS *Sonne* (SO197 and SO 209) and the New Zealand RV *Kaharoa* (KAH0190, KAH1109, KAH1202, KAH1301 and KAH1310). From NIWA we thank Malcolm Clark, Ashley Rowden, Kareen Schnabel, Sadie Mills for logistical support at the NIWA Invertebrate Collection. We also thank Fredrik Søreide from Promare, USA, for supply of the Puerto-Rico samples, and Niamh Kilgallen for amphipod identification.

**References**

Baird, H.P., Miller, K.J., Stark, J.S., 2012. Genetic population structure in the Antarctic benthos: insights from the widespread amphipod, Orchomenella franklini. *PLoS One* 7, e34363. doi:10.1371/journal.pone.0034363

Baker, C. S., 2013. Journal of Heredity adopts joint data archiving policy. *J. Hered.* 104, 1.

Beliaev, G.M., 1989. Deep sea ocean trenches and their fauna. *Scripps Inst. Oceanogr.*

Blacket, M.J., Robin, C., Good, R.T., 2012. Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Mol. Ecol. Resour.* 12, 456–463.

Blankenship, L.E., Yayanos, A.A., Cadien, D.B., Levin, L.A., 2006. Vertical zonation patterns of scavenging amphipods from the Hadal zone of the Tonga and Kermadec Trenches. *Deep Sea Res. Part I.* 53, 48–61. doi:10.1016/j.dsr.2005.09.006

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi:10.1038/nrg3012

Dixon, D., Pruski, A., Dixon, L., 2004. The effects of hydrostatic pressure change on DNA integrity in the hydrothermal-vent mussel Bathymodiolus azoricus: implications for future deep-sea. *Mutat. Res. Mol. Mech. Mutagen.* 552, 235–246.

Ekblom, R., Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. 107, 1–15. doi:10.1038/hdy.2010.152

Fujii, T., Kilgallen, N.M., Rowden, A.A., Jamieson, A.J., 2013. Deep-sea amphipod community structure across abyssal to hadal depths in the Peru-Chile and Kermadec trenches. *Mar. Ecol. Prog. Ser.* 492, 125–138. doi:10.3354/meps10489

Havermans, C., Sonet, G., d'Udekem d'Acoz, C., Nagy, Z.T., Martin, P., Brix, S., Riehl, T., Agrawal, S., Held, C., 2013. Genetic and morpholigcal divergences in the cosmopolitan deep-sea amphipod *Eurythenes gryllus* reveal a diverse abyss and bipolar species. *PLoS* *One*. 8, e74218. doi:10.1371/journal.pone.0074218

Hofreiter, M., Serre, D., Poinar, H.N., 2001. Ancient DNA. *Nat. Rev. Genet.* 2, 353–359.

Jamieson, A., 2015. The Hadal Zone Life in the Deepest Oceans. *Cambridge University Press.*

Jamieson, A., Fujii, T., 2011. Trench Connection. *Biol. Lett.* 7, 641–643.

Jamieson, A.J., Fujii, T., Mayor, D.J., Solan, M., Priede, I.G., 2010. Hadal trenches: the ecology of the deepest places on Earth. *Trends Ecol. Evol.* 25, 190–7. doi:10.1016/j.tree.2009.09.009

Jamieson, A.J., Fujii, T., Solan, M., Matsumoto, A.K., Bagley, P.M., Priede, I.G., 2009. Liparid and macrourid fishes of the hadal zone: in situ observations of activity and feeding behaviour. *Proc. Biol. Sci.* 276, 1037–45. doi:10.1098/rspb.2008.1670

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. Prodöhl, P.A., 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* 4, 782–788.

Kolpakov, R., Bana, G., Kucherov, G., 2003. mreps: efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res.* 31, 3672–3678.

Mesak, F., Tatarenkov, A., Earley, R. L., Avise, J. C., 2014. Hundreds of SNPs vs. dozens of SSRs: which dataset better characterizes natural clonal lineages in a self-fertilizing fish? *Front. Ecol. Evol.* 2(74), <http://dx.doi.org/10.3389/fevo.2014.00074>

Ritchie, H., Jamieson, A.J., Piertney, S.B., 2015. Phylogenetic relationships among hadal amphipods of the Superfamily Lysianassoidea: Implications for taxonomy and biogeography. *Deep Sea Res. Part I*. doi:10.1016/j.dsr.2015.08.014

Rousset, F., 2011. Genepop 4.1 for Windows/Linux/Mac OS X.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for General Users and for Biologist Programmers. *Bioinforma. Methods Protoc.* 132, 365–386.

Simpson, J., Wong, K., Jackman, S., 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 19, 1117–1123.

Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139, 457–462.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Mol. Ecol. Notes* 4, 535–538. doi:10.1111/j.1471-8286.2004.00684.x

Waples, R.S., 2015. Testing for Hardy-Weinberg proportions: Have we lost the plot? *J Hered* 106, 1-19.

Wolff, T., 1960. The hadal community, an introduction. *Deep Sea Res*. 6, 95–124. doi:10.1016/0146-6313(59)90063-2

Wolff, T., 1970. The concept of the hadal or ultra-abyssal fauna. *Deep Sea Res*. 17, 983–1003. doi:10.1016/0011-7471(70)90049-5

Yu, J.N., Won, C., Jun, J., Lim, Y.W., Kwak, M., 2011. Fast and cost-effective mining of microsatellite markers using NGS technology: an example of a Korean water deer *Hydropotes* *inermis* *argyopus*. *PLoS* *ONE* 6, e26933.