A peer-reviewed version of this preprint was published in PeerJ on 17 May 2016.

View the peer-reviewed version (peerj.com/articles/2019), which is the preferred citable publication unless you specifically need to cite this preprint.

Ewers-Saucedo C, Zardus JD, Wares JP. 2016. Microsatellite loci discovery from next-generation sequencing data and loci characterization in the epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1758) PeerJ 4:e2019 https://doi.org/10.7717/peerj.2019

Microsatellite loci discovery from next-generation sequencing data and marker characterization in the epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1798)

Christine Ewers-Saucedo[1], John D. Zardus[2], John P. Wares[1]

[1] Department of Genetics, University of Georgia, Athens, GA, USA

[2] Department of Biology, The Citadel, Charleston, SC, USA

Corresponding author:

Christine Ewers-Saucedo

120 E. Green Street, Athens, GA 30602, USA

Email address: ewers.christine@gmail.com

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

10 **Abstract**

11 Microsatellite markers remain an important tool for ecological and evolutionary research, but are
12 unavailable for many non-model organisms. One such organism with rare ecological and
13 evolutionary features is the epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1758).
14 *Chelonibia testudinaria* appears to be a host generalist, and has an unusual sexual system,
15 androdioecy. Genetic studies on host specificity and mating behavior are impeded by the lack of
16 fine-scale, highly variable markers, such as microsatellite markers. In the present study, we
17 discovered thousands of new microsatellite loci from next-generation sequencing data, and
18 characterized 12 loci thoroughly. We conclude that 11 of these loci will be useful markers in
19 future ecological and evolutionary studies on *C. testudinaria*.

20 **Introduction**

21 Microsatellite loci are valuable tools in ecological and evolutionary studies (e.g. Jarne
22 and Lagoda, 1996; Vignal et al. 2002; Selkoe and Toonen, 2006). Next-generation sequencing
23 approaches have revolutionized microsatellite loci development, allowing the rapid discovery of
24 thousands of potential microsatellite loci in the genome of non-model organisms (Castoe et al.,
25 2012). However, a thorough characterization of potential microsatellite loci remains labor-
26 intensive. It is nonetheless necessary to use these markers successfully in evolutionary and
27 ecological studies.

28 A non-model organism for which genetic and genomic resources are lacking is the
29 epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1758). *Chelonibia testudinaria* uses diverse

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

marine animals as substratum, such as sea turtles, manatees, swimming crabs, and horseshoe crabs. Host-specific morphotypes were previously described as distinct species (Darwin, 1854; Hayashi, 2013). Recent molecular analyses indicate that *C. testudinaria* is a host generalist, and *C. patula* (Ranzani, 1818) and *C. manati* (Gruvel, 1903) are now considered synonyms of *C. testudinaria* (Cheang et al., 2013; Zardus et al., 2014). However, a fine-scale genetic assessment of host-specificity based on highly polymorphic nuclear markers, such as microsatellite markers, is still lacking. Instead of host-specific divergence, genetic lineages are delineated by geographic affinities. The three major lineages are restricted to the Indo-West Pacific, Tropical Eastern Pacific and Atlantic Ocean, respectively (Rawson et al., 2003). These lineages likely represent separate species based on their levels of genetic differentiation (Zardus et al., 2014).

All lineages of *C. testudinaria* exhibit a rare sexual system: androdioecy. Androdioecy is characterized by the co-existence of hermaphrodites and males in the same reproductive population. Understanding mating success and mating patterns of both sexes would greatly advance our understanding of this rare sexual system. This would be most easily achieved with genetic parentage assignment – but its prerequisite, highly variable genetic markers, are not yet developed.

In order to overcome these shortcomings, we used next-generation sequencing to discover microsatellite markers for *C. testudinaria.* We characterized 12 promising markers using individuals of *C. testudinaria* from the Atlantic coast of the United States and from the Northeastern coast of Australia.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

Material & Methods

Specimen collections

Specimen collections of the Atlantic lineage took place on Nannygoat Beach, Sapelo Island, GA, USA (31.48° N, 81.24° W) between 2012 and 2014 under the collection permit of the University of Georgia Marine Institute, and sanctioned by the Georgia DNR Wildlife Services. We chose to collect from the horseshoe crab *Limulus polyphemus* (Linnaeus, 1758) because it is relatively abundant and easy to sample: each spring and early summer, horseshoe crabs crawl onto beaches to mate and lay their eggs. During this process, we removed one individual of *C. testudinaria* per host individual with a sharp knife directly on the beach, and preserved it in 95% EtOH immediately after collection. We collected specimens of the Indo-West Pacific lineage in the vicinity of Townsville, Queensland, Australia (23° S, 143° E), in September 2012 from green turtles (*Chelonia mydas* Brongniart, 1800). Working with officials from the Queensland Department of Environment and Heritage Protection, turtles were captured in-water for routine tagging and release. During capture barnacles were collected and immediately preserved in 95% EtOH.

Microsatellite marker development

We extracted genomic DNA from the feeding appendage of a single large hermaphroditic *C. testudinaria* collected from a horseshoe crab with Gentra Puregene Tissue Kit (Qiagen), and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

measured DNA concentration with a Qubit 2.0 Fluorometer (Life Technologies). Genomic DNA was fragmented into approximately 700bp lengths (insert size) and shotgun-sequenced on an Illumina MiSeq sequencer (PE250). We quality-checked paired-end reads with FastQC (Andrews, 2015). The software FASTQMCF was used to trim adapters, cut low quality ends and remove low quality reads and their mate-pair read (Aronesty, 2011). We calculated the haploid genome size by mapping genomic reads to 52 nuclear single-copy gene fragments available from the acorn barnacle *Semibalanus balanoides* (Regier et al., 2010), and calculating the median coverage for all 52 gene fragments. We then took the grand median coverage of all gene fragments, and dividing the total number of amplified base pairs by the grand median coverage.

We executed the perl script PALFINDER to identify short sequence repeat regions (Castoe et al., 2012). The script also calls PRIMER3, version 2.0.0, to identify potential primer pairs that span the repeat region (Rozen & Skaletsy, 2000). The minimum number of repeat units was chosen as in Castoe et al. (2012). A repeat unit, also called kmer, is defined as the length of the short sequence repeat. For example, a dimer would be a repeat of two base pairs (e.g. GC), and a tetramer would be a repeat of four basepairs (e.g. AGGT). PRIMER3 parameters were the default values. The search resulted in a large number of potentially amplifiable loci (PALs), repeat regions for which primers were identified. We filtered the results by removing all PALs which occurred less than two times and more than the estimated genome coverage in the genomic reads based on the following reasoning: If the number of primer occurrences is low, the primer sequence may contain sequencing error. If the number of occurrences is higher than the expected genome coverage, the primer region may occur more than once in the genome, leading to amplification of multiple loci (genomic regions). Neither of these outcomes is desirable because

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

a good marker occurs only once in the genome, and has a primer sequence that matches the genomic sequence well. R scripts for screening PALFINDER output as well as calculating genome coverage are available as supplementary information.

Of the filtered PALs, we chose 48 PALs for trial amplification, which differed in kmer length, kmer motif (e.g. AG vs TG) and fragment size. We extracted and amplified DNA of 16 *C. testudinaria* individuals for trials. DNA was extracted from feeding appendages of barnacles with the Chelex method (Walsh et al., 1991). Trials used the method of Schuelke (2000) to amplify fragments and simultaneously tag forward primers with a fluorescent dye. Loci that amplified and scored consistently in all individuals were fluorescently labeled with 6-FAM, NED or HEX (Applied Biosystems, Custom Oligo Synthesis Center), and used on a larger number of individuals to characterize the microsatellite markers.

Microsatellite marker amplification

Genomic DNA was extracted from feeding appendages of barnacles with the Chelex method (Walsh et al., 1991). PCR amplifications were performed in 20ul volumes containing final concentrations of 1x PCR buffer (Bioline), 5% bovine serum albumin 10 mg/mL (Sigma), 200 mM each dNTP, 2 nM MgCl, 0.5 mM each primer, 0.5 units of Promega GoTaq DNA Polymerase, and 1 µl template DNA. PCR conditions were as follows: 4 min initial denaturation, followed by 40 cycles with 45 sec denaturing at 94°C, 60 sec annealing at 55°C, 60 sec extension at 72°C and a final extension time of 10 min. The PCR were carried out in a MJ Research PCR Engine. HiDi and ROX500 size standard were added to each sample, and fragment length analysis was carried out at the Georgia Genomics Facility on an ABI 3730xl. Peaks were called

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

and binned with the microsatellite plugin of Geneious version 8.1 (Kearse et al., 2012).

Microsatellite marker characterization

We inspected peak calls for fragment size consistency, using the R package MSATALLELE (Alberto, 2009). MSATALLELE plots peak calls of a locus in histogram form, facilitating visual binning of alleles. If bins could not be clearly assigned, the locus was excluded from the subsequent analysis.

We tested whether loci were in Hardy-Weinberg Equilibrium (HWE) by using 1999 Monte Carlo permutations, as implemented by the function HW.TEST in the R package PEGAS (Paradis, 2010). We recorded the number of alleles, range of fragment sizes, and allelic richness of each locus. The frequency of null alleles was computed based on the method of Brookfield (1996). Genotyping error rates were calculated by repeating genotyping for all individuals. In addition, we amplified the markers for 24 individuals from Queensland, Australia, to assess if the markers could be used in cross-lineage analysis. A R script detailing these analyses is available as supplementary information.

Results

Microsatellite marker development

The MiSeq run generated 15,324,079 paired-end reads (35-251 bp long) with 81.05% > Q30. Raw reads are available in NCBI's short read archive (SRA) under accession number XXX.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

After quality control, 13,498,280 paired-end reads (19-251 bp long) remained, for a total of 6.2 Gb. The median genome coverage was 8x (min = 3, max = 24) for 52 nuclear single-copy gene fragments, and the haploid genome size is therefore approximately 800 Mb (= 6.2 Gb / 8x). The PALFINDER script detected 629,990 microsatellite repeat regions, of which 29,627 (5.38%) were potentially amplifiable loci (PALs) with forward and reverse primer. A summary of detected microsatellite repeat regions is available as supplementary information. A list of all detected microsatellite regions (with primer sequences) is available on figshare (www.figshare.com; DOI: 10.6084/m9.figshare.2070070). After removing PALs with more than eight or less than two occurrences of either forward or reverse primer in the sequence read data, 17,265 PALs remained. We chose 48 loci for trial amplification, which differed in kmer length and repeat motif. Of those 48 loci, 12 loci amplified and scored consistently throughout the trials, and were tagged with fluorescently labeled dye (Table 1).

Microsatellite marker characterization

We genotyped 42 individuals successfully at more than half of the 12 consistently scoring loci. Visual inspection of peak call histograms revealed that peak calls of Ctest2 did not have clearly defined bins, and were excluded from subsequent analyses. The number of alleles of the 11 scorable loci ranged from six to 30 (Figure 1). Microsatellite genotype and collecting date for each individual are available as supplementary information. For the Atlantic population, six loci were not in HWE, and showed homozygote excess (Table 2). The estimated percentage of null alleles ranged from 0% to 20.3%. Allelic richness ranged from 3.6 to 20.7, and genotyping error rates ranged from 0 to 7.32% (Table 2).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

All 11 loci amplified in some of the 23 Australian individuals, and had at least two alleles. For the Pacific lineage of C. testudinaria, none of the loci showed significant deviations from HWE. Allelic richness was significantly lower than in the Atlantic individuals (1.8-6.8). The percentage of null alleles ranged from 0% to 30% (Table 3).

Discussion

The present study developed and characterized 11 microsatellite markers for the epizoic barnacle *Chelonibia testudinaria*. Several loci are not in HWE, probably due to null alleles. Their high allelic diversity and scoring consistency should nonetheless make them useful in ecological and evolutionary studies. In addition, we provide the resources to evaluate thousands of additional potentially amplifiable loci (PALs) for *C. testudinaria*.

Several loci were not in HWE, and displayed homozygote excess. Homozygote excess can have several causes: selection on these loci, the presence of null alleles, inbreeding, population substructure or large variance in reproductive success. Inbreeding is unlikely because most barnacles are obligate outcrossers and *C. testudinaria* has a widely-dispersing planktonic larval phase. Selection cannot be excluded as an explanation, but selection on several markers appears unlikely. Population substructure may be present, but if so, is neither host-induced nor geographical. Large variance in reproductive success can cause homozygote excess (Hedgecock, 1994), and has been invoked to explain homozygote excess in e.g. sea urchins (Addison & Hart, 2004). If variance in reproductive success is present, the effective population size of *C. testudinaria* should be low (Hedgecock, 1994). We estimated a Θ of 10 for the Atlantic *C.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

testudinaria population using Watterson's estimator (Watterson, 1975) on published COI data, which suggests a large effective population size (data not shown). These data do not support the variance-in-reproductive-success hypothesis. The most likely cause for homozygote excess is the presence of null alleles. Null alleles are ubiquitous in microsatellite loci, and are caused by mutations in the primer sequence, leading to the dropout of alleles. While the true genotype is heterozygous, the observed genotype is homozygous due to the non-amplification of one of the alleles. Null alleles become increasingly prevalent with increasing effective population size (Chapuis and Estoup, 2007). Chapuis and Estoup (2007) show that simulated null allele frequencies were larger than 0.2 for all loci when the population mutation rate (theta) was one, the largest value they simulated. We estimated null allele frequencies between zero and 0.3 for our microsatellite markers, well within the range of simulated data with large effective population size. Thus the observed homozygote excess can be explained by the presence of null alleles.

We were able to amplify all loci in both lineages, which was somewhat surprising given the combination of large effective population size and significant between-lineage divergence. Both factors increase the chance for primer sequences to differ between lineages. Further, results for the Indo-West Pacific lineage need to be evaluated with caution because primers were designed from an individual of the Atlantic lineage. Future studies should increase sample sizes for both lineages to compare and contrast genotypic diversity. In summary, we identified new genetic resources that can be used in future ecological and evolutionary studies on the epizoic, androdioecious barnacle *Chelonibia testudinaria*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

References

Alberto, F. 2009. Msatallele_1.0: An r package to visualize the binning of microsatellite alleles. *J. Hered.* **100**: 394-7.

Andrews, S. 2015. Fastqc a quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Aronesty, E. 2011. Fastq-mcf sequence quality filter, clipping and processor http://code.Google.com/p/ea-utils/wiki/fastqmcf.

Brookfield J (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**, 453-455.

Castoe, T. A., Poole, A. W., de Koning, A. P. J., Jones, K. L., Tomback, D. F., Oyler-McCance, S. J., Fike, J. A., Lance, S. L., Streicher, J. W., Smith, E. N. & Pollock, D. D. 2012. Rapid microsatellite identification from illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE* **7**: e30953.

Chapuis, M.-P. & Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* **24**: 621-631.

Cheang, C., Tsang, L., Chu, K., Cheng, I.-J. & Chan, B. 2013. Host-specific phenotypic plasticity of the turtle barnacle *Chelonibia testudinaria*: A widespread generalist rather than a specialist. *PLoS ONE* **8**: e57592.

Darwin, C. 1854. *A monograph on the sub-class Cirripedia, with figures of all the species. The Balanidae, the Verrucidae, etc*. Ray society, London.

Gruvel, A. 1903. Cirrhipédes opercules nouveaux ou peu connus de la collèction du muséum.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

Bull. Mus.: 23-25.

Hayashi, R. 2013. A checklist of turtle and whale barnacles (cirripedia: Thoracica: Coronuloidea). *J. Mar. Biol. Assoc. U.K.* **93**: 143-182.

Hedgecock, D. (1994) Does variance in reproductive success limit effective population sizes of marine organisms. In: *Genetics and the evolution of aquatic marine organisms*, (Beaumont, A. R., ed.). pp. 122-134. Chapman & Hall, London.

Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**, 424-429.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. & Drummond, A. 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data (version 8.0.3). *Bioinformatics* **28**: 1647-1649.

Linnaeus, C. 1758. *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis,* 10 ed. Tomus I. L. Salvii, Stockholm, Sweden.

Paradis, E. 2010. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*: btp696.

Ranzani, A. C. 1818. Osservazioni su i balanidae parte iii. *Opusculi Scientifici* **2**: 63-93.

Rawson, P. D., Macnamee, R., Frick, M. G. & Williams, K. L. 2003. Phylogeography of the coronulid barnacle, *Chelonibia testudinaria*, from loggerhead sea turtles, caretta caretta. *Mol. Ecol.* **12**: 2697-2706.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

234 Regier J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J. W.,
235 Cunningham, C. W. 2010. Arthropod relationships revealed by phylogenomic analysis of
236 nuclear protein-coding sequences. Nature. **463:** 1079-1083.
237 Rozen, S. & Skaletsky, H. (2000) Primer3 on the www for general users and for biologist
238 programmers. In: *Bioinformatics methods and protocols: Methods in molecular biology*,
239 (Krawetz, S. & Misener, S., eds.). pp. 365-386. Humana Press, Totowa, NJ.
240 Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: A practical guide to using and
241 evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.
242 Schuelke, M. 2000. An economic method for the fluorescent labeling of pcr fragments. *Nat*
243 *BioTech* **18**: 233-234.
244 Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on snp and other types of
245 molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34**, 275-
246 306.
247 Walsh, P. S., Metzger, D. A. & Higuchi, R. 1991. Chelex 100 as a medium for simple extraction
248 of DNA for pcr-based typing from forensic material. *BioTech.* **10**: 506-13.
249 Watterson, G. A. 1975. On the number of segregating sites in genetical models without
250 recombination. Theoretical Population Biology **7:** 256-276.
251 Zardus, J. D., Lake, D. T., Frick, M. G. & Rawson, P. D. 2014. Deconstructing an assemblage of
252 “turtle” barnacles: Species assignments and fickle fidelity in *Chelonibia*. *Mar. Biol.* **161**:
253 45-59.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

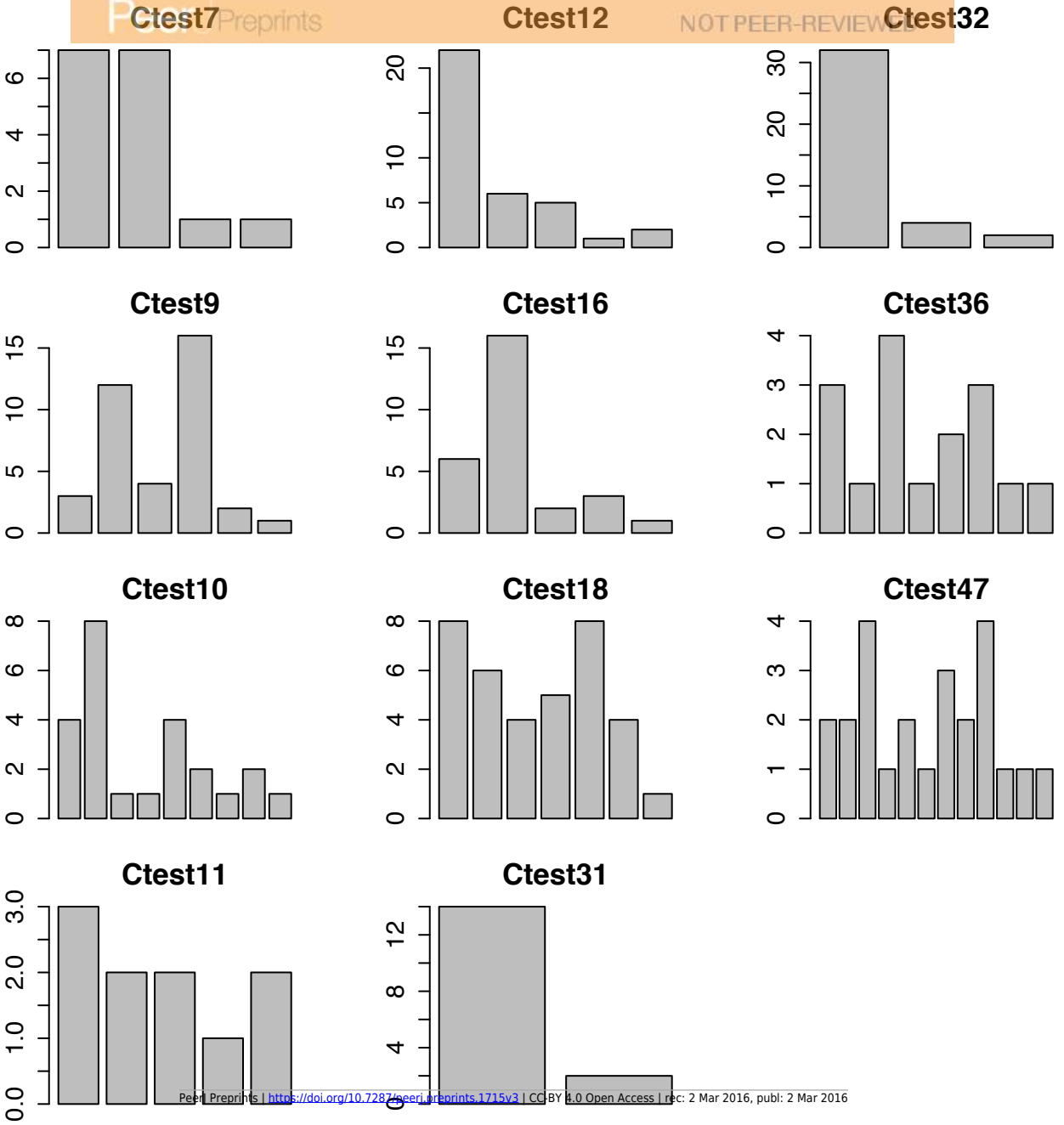
Captions

Figure 1. Allele frequencies for each microsatellite locus. Each barplot represents a locus, each bar an allele, and the height of each bar indicates the frequency of each allele in in the data. Sample sizes are indicated in table 2.

Table 1. Microsatellite loci amplification information. All loci were amplified at 55°C annealing temperature. "Dye" refers to the fluorescent color label for each forward primer. Ned is yellow, 6-FAM is blue and HEX is green. Labeling forward primers with different colors allows multiplexing several primer sets in the same reaction. "Multiplex reaction" refers to the multiplexing PCR scheme, e.g. all loci with the same multiplex code are amplified in the same reaction.

Table 2. Microsatellite loci characterization for *Chelonibia testudinaria* of the Atlantic lineage. Range refers to the smallest (min) and largest (max) allele observed. Frequency of null alleles was estimated after Brookfield (1996). Genotyping error rates were based on re-genotyping of the all 42 Atlantic individuals. Abbreviations: n = number of individuals; Obs het. = observed heterozygosity; Exp het. = expected heterozygosity.

Table 3. Microsatellite loci characterization for *Chelonibia testudinaria* of the Indo-West Pacific lineage. Abbreviations: Obs het. = observed heterozygosity; Exp het. = expected heterozygosity.



Peer J Preprints | https://doi.org/10.7287/peerj.preprints.1715v3 | CC-BY 4.0 Open Access | rec: 2 Mar 2016, publ: 2 Mar 2016

Table 1

Locus	Kmer	Forward primer sequence	Reverse primer sequence	Dye	Multiplex reaction
Ctest2		ACACACATCACTGGACTCG	CAGTAAGCAGCTCTGTTCG	NED	BB
Ctest7	4	GTTATCCGTCATTCCATCC	GACGTAACCACCTTGTCG	6-FAM	AA
Ctest9	4	AACAGATGTGACATTGATGC	TTGTACTGTCCTTGTAACGC	6-FAM	BB
Ctest10	2	ATACGCACAAACTCACACC	TGTCCTCTTACAGAGATCGG	HEX	BB
Ctest11	2	GTGTCCACCTTTATGTCTGG	AGTTGAAAATACGCACGC	HEX	CC
Ctest12	4	AACTGGTGGACAGTCTGG	CATCTTTATGAGTAGCGAGG	HEX	AA
Ctest16	4	TCAGGTACAGCATTATCGC	CAAGGACCATCAATTACCC	6-FAM	CC
Ctest18	5	TTCATGAATCACTTCCTGG	GTAATCAAATAAGGCGATGC	NED	AA
Ctest31	4	GTACGCCGAAAGTAAAGC	AGCTCTGACAAAGTTATGCC	6-FAM	DD
Ctest32	4	AGAAATCCATAATCGTCTGG	ATAACGACGTAATCAGCACC	NED	CC
Ctest36	4	AGATATTGGTGGAACGAGC	CACAACATACTCAACGAACG	HEX	DD
Ctest47	5	GTTGACACGATGACATAACG	ACAATTCCAGCTCTGTTAGC	NED	DD

Table 2

Locus	n	Range min	Range max	Number of alleles	Obs het.	Exp het.	HWE p-value	Allelic richness	Frequency null alleles	Genotyping error rate
Ctest7	34	206	314	18	0.559	0.817	0.007	16.113	0.142	0
Ctest9	34	388	432	8	0.618	0.69	0.022	7.449	0.043	0.021
Ctest10	34	264	278	8	0.647	0.766	0.141	7.453	0.067	0.032
Ctest11	38	140	318	27	0.868	0.93	0.559	22.678	0.032	0.073
Ctest12	34	388	476	23	0.735	0.917	0.006	20.261	0.095	0.024
Ctest16	35	355	367	4	0.371	0.39	0.137	3.703	0.014	0
Ctest18	33	450	485	7	0.606	0.776	0.033	6.984	0.096	0.061
Ctest31	36	292	336	8	0.556	0.664	0.001	6.542	0.065	0.068
Ctest32	35	316	464	4	0.571	0.518	0.493	3.647	0	0
Ctest36	33	336	524	23	0.455	0.891	0	20.072	0.231	0.025
Ctest47	37	255	385	10	0.622	0.662	0.259	8.356	0.024	0.017

Table 3

Locus	n	Range min	Range max	Number of alleles	Obs het.	Exp het.	HWE p-value	Allelic richness	Frequency null alleles
Ctest7	8	186	314	4	1	0.609	0.004	3.067	0
Ctest9	19	408	456	6	0.737	0.702	0.806	3.946	0
Ctest10	12	260	344	9	0.5	0.813	0.001	5.479	0.172
Ctest11	5	136	236	5	0.2	0.78	0.001	4.628	0.326
Ctest12	18	354	374	5	0.333	0.576	0	3.377	0.154
Ctest16	14	351	375	5	0.571	0.61	0.158	3.525	0.024
Ctest18	18	440	480	7	0.278	0.829	0	5.221	0.301
Ctest31	8	224	304	2	0.25	0.219	1	1.797	0
Ctest32	19	308	316	3	0	0.277	0	2.134	0.217
Ctest36	8	348	468	8	0.5	0.836	0.002	5.739	0.183
Ctest47	12	140	340	12	0.833	0.892	0.008	6.882	0.031