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

Christine Ewers-Saucedo, John D Zardus, John P Wares

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

1 **Microsatellite loci discovery from next-generation sequencing data and loci**

2 **characterization in the epizoic barnacle *Chelonibia testudinaria* (Linnaeus,**

3 **1798)**

5 Christine Ewers-Saucedo[1,2], John D. Zardus[3], John P. Wares[1,4]

7 [1] Department of Genetics, University of Georgia, Athens, GA, USA

8 [2] Evolution and Ecology, University of California, Davis, CA, USA

9 [3] Department of Biology, The Citadel, Charleston, SC, USA

10 [4] Odum School of Ecology, University of Georgia, Athens, GA, USA

12 Corresponding author:

13 Christine Ewers-Saucedo

14 Storer Hall, Davis, CA 95616, USA

15 Email address: ewers.christine@gmail.com

17 **Abstract**

18 Microsatellite markers remain an important tool for ecological and evolutionary research, but are

19 unavailable for many non-model organisms. One such organism with rare ecological and

20 evolutionary features is the epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1758).

21 *Chelonibia testudinaria* appears to be a host generalist, and has an unusual sexual system,

22 androdioecy. Genetic studies on host specificity and mating behavior are impeded by the lack of

23 fine-scale, highly variable markers, such as microsatellite markers. In the present study, we

24 discovered thousands of new microsatellite loci from next-generation sequencing data, and

25 characterized 12 loci thoroughly. We conclude that 11 of these loci will be useful markers in

26 future ecological and evolutionary studies on *C. testudinaria*.

27

28

29

30 **Introduction**

31   Microsatellite loci are valuable tools in ecological and evolutionary studies (e.g. Jarne

32 and Lagoda, 1996; Vignal et al. 2002; Selkoe and Toonen, 2006). Next-generation sequencing

33 approaches have revolutionized microsatellite loci development, allowing the rapid discovery of

34 thousands of potential microsatellite loci in the genome of non-model organisms (Castoe et al.,

35 2012). However, a thorough characterization of potential microsatellite loci remains labor-

36 intensive. It is nonetheless necessary if we want to use these markers successfully in

37 evolutionary and ecological studies.

38   A non-model organism for which genetic and genomic resources are lacking is the

39 epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1758). *Chelonibia testudinaria* uses diverse

40 marine animals as substratum, such as sea turtles, manatees, swimming crabs, and horseshoe

41 crabs. Host-specific morphotypes were previously described as distinct species (Darwin, 1854;

42 Hayashi, 2013). Recent molecular analyses indicate that *C. testudinaria* is a host generalist, and

43 *C. patula* (Ranzani, 1818) and *C. manati* (Gruvel, 1903) are now considered synonyms of *C.*

44 *testudinaria* (Cheang et al., 2013; Zardus et al., 2014). However, a fine-scale genetic assessment

45 of host-specificity based on highly polymorphic nuclear markers, such as microsatellite markers,

46 is still lacking.

47   While host species do not seem to provide barriers to gene flow, biogeography does:

48 Three major genetic lineages are restricted to the Indo-West Pacific, Tropical Eastern Pacific and

49 Atlantic Ocean, respectively (Rawson et al., 2003, Cheang et al., 2013; Zardus et al., 2014).

50 These lineages likely represent separate species based on their levels of genetic differentiation

51 (Zardus et al., 2014).

52   All lineages of *C. testudinaria* exhibit a rare sexual system: androdioecy. Androdioecy is

53   characterized by the co-existence of hermaphrodites and males in the same reproductive

54   population. Understanding mating success and mating patterns of both sexes would greatly

55   advance our understanding of this rare sexual system. This would be most easily achieved with

56   genetic parentage assignment – but its prerequisite, highly variable genetic markers, are not

57   available.

58   In order to overcome these shortcomings, we used next-generation sequencing to discover

59   microsatellite markers for *C. testudinaria*. We characterized 12 promising markers in

60   populations from the Atlantic coast of the United States and from the Northeastern coast of

61   Australia.

62

63

64   **Material & Methods**

65

66   Specimen collections

67   Specimen collections of the Atlantic lineage took place on Nannygoat Beach, Sapelo

68   Island, GA, USA (31.48° N, 81.24° W) between 2012 and 2014 under the collection permit of

69   the University of Georgia Marine Institute, and sanctioned by the Georgia DNR Wildlife

70   Services. We chose to collect from the horseshoe crab *Limulus polyphemus* (Linnaeus, 1758)

71   because it is relatively abundant and easy to sample: each spring and early summer, horseshoe

72   crabs crawl onto beaches to mate and lay their eggs. During this process, we removed one

73   individual of *C. testudinaria* per host individual with a sharp knife directly on the beach, and

74   preserved it in 95% EtOH immediately after collection. We collected specimens of the Indo-

75   West Pacific lineage in the vicinity of Townsville, Queensland, Australia (23° S, 143° E), in

76 September 2012 from green turtles (*Chelonia mydas* Brongniart, 1800). Working with officials

77 from the Queensland Department of Environment and Heritage Protection, turtles were captured

78 in-water for routine tagging and release. During capture barnacles were collected and

79 immediately preserved in 95% EtOH.

80

81 Microsatellite loci discovery

82   We extracted genomic DNA from the feeding appendage of a single large hermaphroditic

83 *C. testudinaria* collected from a horseshoe crab with Gentra Puregene Tissue Kit (Qiagen), and

84 measured DNA concentration with a Qubit 2.0 Fluorometer (Life Technologies). Genomic DNA

85 was fragmented into approximately 700bp lengths (insert size) and shotgun-sequenced on an

86 Illumina MiSeq sequencer (PE250). We quality-checked paired-end reads with FastQC

87 (Andrews, 2015). The software FASTQMCF was used to trim adapters, cut low quality ends and

88 remove low quality reads and their mate-pair read (Aronesty, 2011).

89   We executed the perl script PALFINDER to identify short sequence repeat regions (Castoe

90 et al., 2012). The script also calls PRIMER3, version 2.0.0, to identify potential primer pairs that

91 span the repeat region (Rozen & Skaletsy, 2000). The minimum number of repeat units was

92 chosen as in Castoe et al. (2012). A repeat unit, also called kmer, is defined as the length of the

93 short sequence repeat. For example, a dimer would be a repeat of two base pairs (e.g. GC), and a

94 tetramer would be a repeat of four basepairs (e.g. AGGT). PRIMER3 parameters were the default

95 values. The search resulted in a large number of potentially amplifiable loci (PALs), repeat

96 regions for which primers were identified. We filtered the results by removing all PALs which

97 occurred less than two times and more than the estimated genome coverage in the genomic reads

98 based on the following reasoning: If the number of primer occurrences is low, the primer

99   sequence may contain sequencing error. If the number of occurrences is higher than the expected

100 genome coverage, the primer region may occur more than once in the genome, leading to

101 amplification of multiple loci (genomic regions). Neither of these outcomes is desirable because

102 a good marker occurs only once in the genome, and has a primer sequence that matches the

103 genomic sequence well. We estimated genome coverage by mapping our genomic reads to 52

104 nuclear single-copy gene fragments available from the acorn barnacle *Semibalanus balanoides*

105 (Regier et al., 2010) with the Geneious version 6.0.3 Read Mapper, using the default settings

106 (Kearse et al., 2012), calculated the median coverage for each of the 52 gene fragments, and the

107 grand median of all genes. R scripts for screening PALFINDER output as well as calculating

108 genome coverage are available as supplementary information.

109   Of the filtered PALs, we chose 48 PALs for trial amplification, which differed in kmer

110 length, kmer motif (e.g. AG vs TG) and fragment size. We extracted and amplified DNA of 16

111 *C. testudinaria* individuals for trials. DNA was extracted from feeding appendages of barnacles

112 with the Chelex method (Walsh et al., 1991). Trials used the method of Schuelke (2000) to

113 amplify fragments and simultaneously tag forward primers with a fluorescent dye. Loci that

114 amplified and scored consistently in all individuals were fluorescently labeled with 6-FAM,

115 NED or HEX (Applied Biosystems, Custom Oligo Synthesis Center), and used on a larger

116 number of individuals to characterize the microsatellite loci.

117

118 Microsatellite loci amplification

119   Genomic DNA was extracted from feeding appendages of barnacles with the Chelex

120 method (Walsh et al., 1991). PCR amplifications were performed in 20ul volumes containing

121 final concentrations of 1x PCR buffer (Bioline), 5% bovine serum albumin 10 mg/mL (Sigma),

122   200 mM each dNTP, 2 nM MgCl, 0.5 mM each primer, 0.5 units of Promega GoTaq DNA

123   Polymerase, and 1 µl template DNA. PCR conditions were as follows: 4 min initial denaturation,

124   followed by 40 cycles with 45 sec denaturing at 94°C, 60 sec annealing at 55°C, 60 sec extension

125   at 72°C and a final extension time of 10 min. The PCR were carried out in a MJ Research PCR

126   Engine. HiDi and ROX500 size standard were added to each sample, and fragment length

127   analysis was carried out at the Georgia Genomics Facility on an ABI 3730xl. Peaks were called

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

129

130   Microsatellite loci characterization

131   We first characterized the microsatellite loci on 42 individuals from the Atlantic

132   population. We inspected peak calls for fragment size consistency, using the R package

133   MSATALLELE (Alberto, 2009). MSATALLELE plots peak calls of a locus in histogram form,

134   facilitating visual binning of alleles. If bins could not be clearly assigned, the locus was excluded

135   from the subsequent analysis. We tested whether loci were in Hardy-Weinberg Equilibrium

136   (HWE) by using 1999 Monte Carlo permutations, as implemented by the function HW.TEST in the

137   R package PEGAS (Paradis, 2010). Significance values were adjusted for multiple comparisons

138   based on the method of Holm (1979). We recorded the number of alleles, range of fragment

139   sizes, and allelic richness of each locus. The frequency of null alleles was computed based on the

140   method of Brookfield (1996). Genotyping error rates were calculated by repeating genotyping

141   for all individuals. After characterizing the loci in the Atlantic population, we amplified the loci

142   for 24 individuals from Queensland, Australia, to assess if the loci could be used in cross-lineage

143   analysis. Characterization on individuals of the Indo-West Pacific population were largely the

144   same as for the Atlantic population, but we did not assess genotyping error rates. A R script

145   detailing these analyses is available as supplementary information.

146

147

148   **Results**

149

150   Microsatellite marker development

151   The MiSeq run generated 15,324,079 paired-end reads (35-251 bp long) with 81.05% >

152   Q30. Raw reads are available in NCBI's short read archive (Study accession: SRP069253, run

153   accession: SRR3144544). After quality control, 13,498,280 paired-end reads (19-251 bp long)

154   remained, for a total of 6.2 Gb. The median genome coverage was 8x (min = 3, max = 24) for 52

155   nuclear single-copy gene fragments. The PALFINDER script detected 629,990 microsatellite repeat

156   regions, of which 29,627 (5.38%) were potentially amplifiable loci (PALs) with forward and

157   reverse primer. A summary of detected microsatellite repeat regions is available as

158   supplementary information. A list of all detected microsatellite regions (with primer sequences)

159   is available on figshare (www.figshare.com; DOI: 10.6084/m9.figshare.2070070). After

160   removing PALs with more than eight or less than two occurrences of either forward or reverse

161   primer in the sequence read data, 17,265 PALs remained. We chose 48 perfect repeat loci for

162   trial amplification, which differed in kmer length and repeat motif, but were otherwise chosen at

163   random. Of those 48 loci, 12 loci amplified and scored consistently throughout the trials, and

164   were tagged with fluorescently labeled dye (Table 1).

165

166   Microsatellite marker characterization

167   We genotyped 42 individuals successfully at more than half of the 12 consistently scoring

168   loci. Visual inspection of peak call histograms revealed that peak calls of Ctest2 did not have

169   clearly defined bins, and were excluded from subsequent analyses. The number of alleles of the

170   11 scorable loci ranged from six to 30 (Figure 1). Microsatellite genotype and collecting date for

171   each individual are available as supplementary information. For the Atlantic population, four loci

172   were not in HWE, and showed homozygote excess (Table 2). The estimated percentage of null

173   alleles ranged from 0% to 23%. Allelic richness ranged from 3.65 to 20.7, and genotyping error

174   rates ranged from 0 to 7% (Table 2).

175   All 11 loci amplified in some of the 23 Australian individuals, and had at least two alleles

176   (Table 3). No loci amplified in all individuals, and no individual failed to amplify all loci. For the

177   Pacific lineage of *C. testudinaria*, two loci deviated significantly from HWE expectations. The

178   locus Ctest7 showed heterozygote excess, even though not at a significant level (p-value = 0.13) .

179   Allelic richness of the Australian population was overall lower than in the Atlantic population.

180   The percentage of null alleles ranged from 0% to 33%.

181

182

183   **Discussion**

184   The present study developed and characterized 11 microsatellite markers for the epizoic

185   barnacle *Chelonibia testudinaria*. Several loci are not in HWE, probably due to null alleles.

186   Their high allelic diversity and scoring consistency should nonetheless make them useful in

187   ecological and evolutionary studies. In addition, we provide the resources to evaluate thousands

188   of additional potentially amplifiable loci (PALs) for *C. testudinaria*.

189   Several loci were not in HWE, and displayed homozygote excess. Homozygote excess

190   can have several causes: selection on these loci, the presence of null alleles, inbreeding,

191   population substructure or large variance in reproductive success. Inbreeding is unlikely because

192   most barnacles are obligate outcrossers and *C. testudinaria* has a widely-dispersing planktonic

193   larval phase. Selection cannot be excluded as an explanation, but selection on or near several

194   markers appears unlikely. Population substructure may be present, but if so, is neither host-

195   induced nor geographical. Large variance in reproductive success can cause homozygote excess

196   (Hedgecock, 1994), and has been invoked to explain homozygote excess in e.g. sea urchins

197   (Addison & Hart, 2004). If variance in reproductive success is present, the effective population

198   size of *C. testudinaria* should be low (Hedgecock, 1994). We estimated a population mutation

199   rate (called theta, Θ) of 10 for the Atlantic *C. testudinaria* population using Watterson's

200   estimator (Watterson, 1975) on published COI data, which suggests a large effective population

201   size (data not shown). These data do not support the variance-in-reproductive-success

202   hypothesis. The most likely cause for homozygote excess is the presence of null alleles. Null

203   alleles are ubiquitous in microsatellite loci, and are caused by mutations in the primer sequence,

204   leading to the dropout of alleles. While the true genotype is heterozygous, the observed genotype

205   is homozygous due to the non-amplification of one of the alleles. Null alleles become

206   increasingly prevalent with increasing effective population size (Chapuis and Estoup, 2007).

207   Chapuis and Estoup (2007) show that simulated null allele frequencies were larger than 0.2 for

208   all loci when Θ was one, the largest value they simulated. We estimated null allele frequencies

209   between zero and 0.3 for our microsatellite loci, well within the range of simulated data with

210   large effective population size. Thus the observed homozygote excess can be explained by the

211   presence of null alleles.

212   We were able to amplify all loci in both the Atlantic and Indo-West Pacific lineage,

213   which was somewhat surprising given large effective population size and significant between-

214 lineage divergence. Both factors increase the chance for mutation accumulation in the primer

215 sequences between lineages. Further, results for the Indo-West Pacific lineage need to be

216 evaluated with caution because primers were designed from an individual of the Atlantic lineage.

217 As expected by the concept of ascertainment bias (Li and Kimmel, 2013), the Atlantic lineage

218 had higher allelic richness than the Indo-West Pacific lineage. However, fewer loci violated the

219 expectations of HWE in the Indo-West Pacific lineage, likely an artifact of analyzing fewer

220 individuals, resulting in low power to detect significant deviations from HWE. Future studies

221 should increase sample sizes for both lineages to compare and contrast genotypic diversity.

222   Microsatellite markers have been developed successfully in a number of other barnacle

223 species (Dufresne et al. 1999; 2002; Dawson et al., 2010; Plough and Marko, 2014), providing

224 tools to address interesting ecological and evolutionary questions. We expect the loci developed

225 for *C. testudinaria* will be similarly useful.

226

227

228 **References**

229 Alberto, F. 2009. Msatallele_1.0: An r package to visualize the binning of microsatellite alleles.

230   *J. Hered.* **100**: 394-7.

231 Andrews, S. 2015. Fastqc a quality control tool for high throughput sequence data.

232   *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.*

233 Aronesty, E. 2011. Fastq-mcf sequence quality filter, clipping and processor

234   http://code.Google.com/p/ea-utils/wiki/fastqmcf.

235 Brookfield J. 1996. A simple new method for estimating null allele frequency from heterozygote

236   deficiency. *Molecular Ecology* **5**, 453-455.

237 Castoe, T. A., Poole, A. W., de Koning, A. P. J., Jones, K. L., Tomback, D. F., Oyler-McCance,

238   S. J., Fike, J. A., Lance, S. L., Streicher, J. W., Smith, E. N. & Pollock, D. D. 2012.

239   Rapid microsatellite identification from illumina paired-end genomic sequencing in two

240   birds and a snake. *PLoS ONE* **7**: e30953.

241 Chapuis, M.-P. & Estoup, A. 2007. Microsatellite null alleles and estimation of population

242   differentiation. *Mol. Biol. Evol.* **24**: 621-631.

243 Cheang, C., Tsang, L., Chu, K., Cheng, I.-J. & Chan, B. 2013. Host-specific phenotypic

244   plasticity of the turtle barnacle *Chelonibia testudinaria*: A widespread generalist rather

245   than a specialist. *PLoS ONE* **8**: e57592.

246 Darwin, C. 1854. *A monograph on the sub-class Cirripedia, with figures of all the species. The*

247   *Balanidae, the Verrucidae, etc*. Ray society, London.

248 Dawson MN, Grosberg RK, Yoel E S, Eric S. 2010. Population genetic analysis of a recent range

249   expansion: mechanisms regulating the poleward range limit in the volcano barnacle

250   *Tetraclita rubescens*. Molecular Ecology 19: 1585-1605.

251 Dufresne F, Bourget E, Bernatchez L. 2002. Differential patterns of spatial divergence in

252   microsatellite and allozyme alleles: further evidence for locus-specific selection in the

253   acorn barnacle, *Semibalanus balanoides*? Molecular Ecology 11:113-123.

254 Dufresne F, Parent M, Bernatchez L. 1999. Isolation and characterization of microsatellite

255   markers in the acorn barnacle *Semibalanus balanoides*. Molecular Ecology 8:1558.

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

257   *Bull. Mus.*: 23-25.

258 Hayashi, R. 2013. A checklist of turtle and whale barnacles (cirripedia: Thoracica:

259   Coronuloidea). *J. Mar. Biol. Assoc. U.K.* **93**: 143-182.

260   Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of

261   marine organisms. In: *Genetics and the evolution of aquatic marine organisms*,

262   (Beaumont, A. R., ed.). pp. 122-134. Chapman & Hall, London.

263   Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of

264   Statistics **6**:65-70.

265   Jarne P, Lagoda PJ. 1996. Microsatellites, from molecules to populations and back. *Trends in*

266   *Ecology & Evolution* **11**, 424-429.

267   Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,

268   Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. &

269   Drummond, A. 2012. Geneious basic: An integrated and extendable desktop software

270   platform for the organization and analysis of sequence data (version 8.0.3).

271   *Bioinformatics* **28**: 1647-1649.

272   Li, B., Kimmel, M. 2013. Factors Influencing Ascertainment Bias of Microsatellite Allele Sizes:

273   Impact on Estimates of Mutation Rates. Genetics **195**: 563-572.

274   Linnaeus, C. 1758. *Systema naturae per regna tria naturae, secundum classes, ordines, genera,*

275   *species, cum characteribus, differentiis, synonymis, locis,* 10 ed. Tomus I. L. Salvii,

276   Stockholm, Sweden.

277   Paradis, E. 2010. Pegas: An R package for population genetics with an integrated-modular

278   approach. *Bioinformatics*: btp696.

279   Plough LV, Marko PB. 2013. Characterization of microsatellite loci and repeat density in the

280   gooseneck barnacle, *Pollicipes elegans*, using next generation sequencing. Journal of

281   Heredity:est064.

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

283   Rawson, P. D., Macnamee, R., Frick, M. G. & Williams, K. L. 2003. Phylogeography of the

284   coronulid barnacle, *Chelonibia testudinaria*, from loggerhead sea turtles, caretta caretta.

285   *Mol. Ecol.* **12**: 2697-2706.

286   Regier J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J. W.,

287   Cunningham, C. W. 2010. Arthropod relationships revealed by phylogenomic analysis of

288   nuclear protein-coding sequences. Nature. **463:** 1079-1083.

289   Rozen, S. & Skaletsky, H. (2000) Primer3 on the www for general users and for biologist

290   programmers. In: *Bioinformatics methods and protocols: Methods in molecular biology*,

291   (Krawetz, S. & Misener, S., eds.). pp. 365-386. Humana Press, Totowa, NJ.

292   Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: A practical guide to using and

293   evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

294   Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat

295   BioTech* **18**: 233-234.

296   Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of

297   molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34**, 275-

298   306.

299   Walsh, P. S., Metzger, D. A. & Higuchi, R. 1991. Chelex 100 as a medium for simple extraction

300   of DNA for PCR-based typing from forensic material. *BioTech.* **10**: 506-13.

301   Watterson, G. A. 1975. On the number of segregating sites in genetical models without

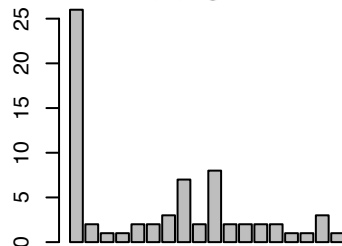
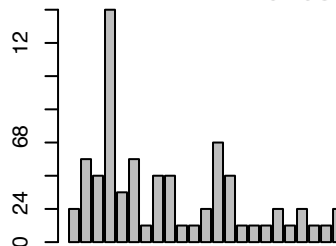
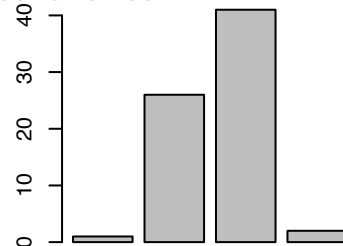
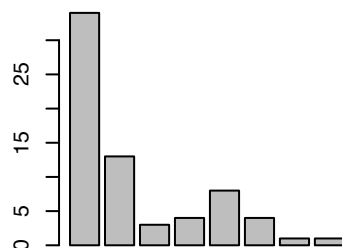
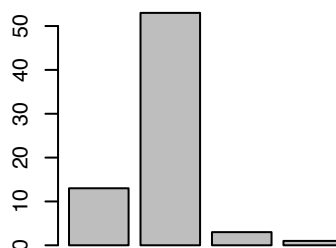
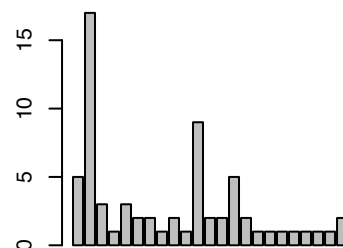
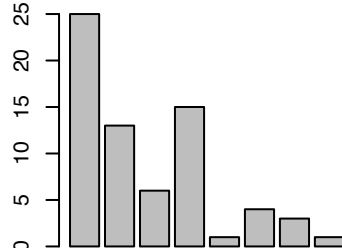
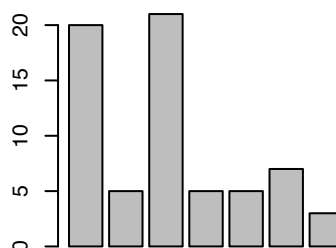
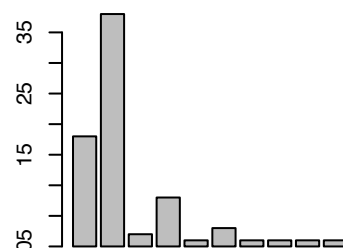
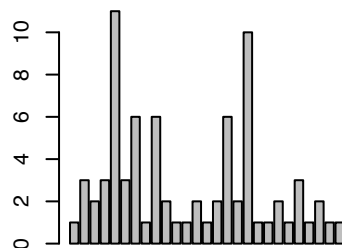
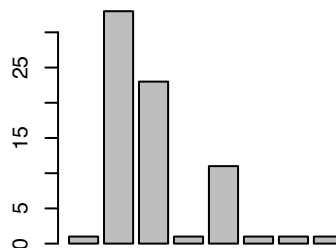
302   recombination. Theoretical Population Biology **7:** 256-276.

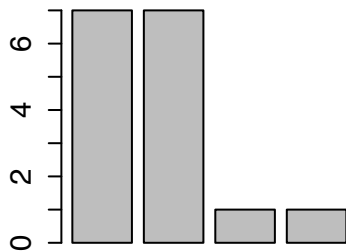
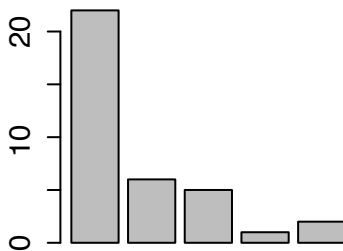
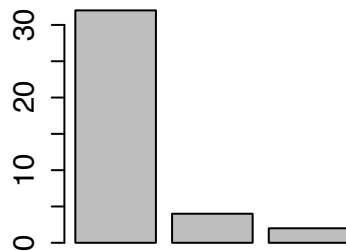
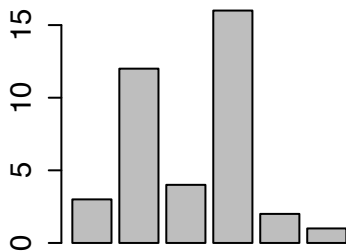
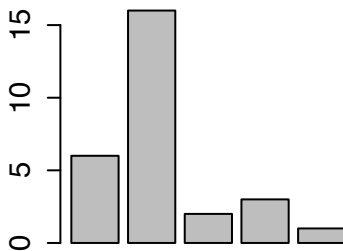
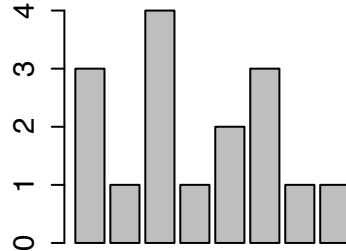
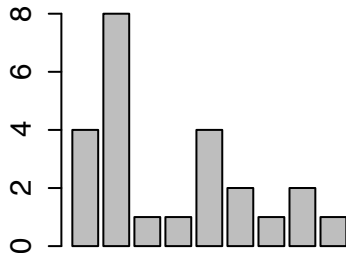
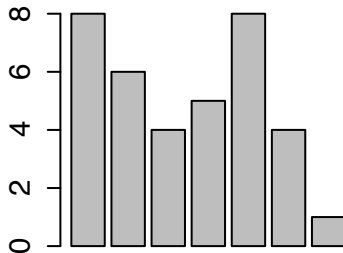
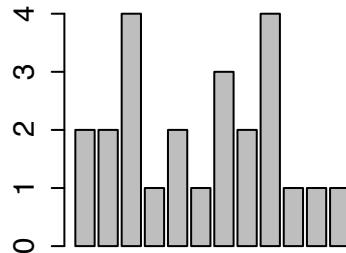
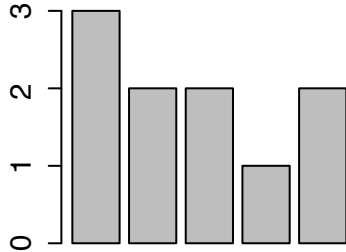
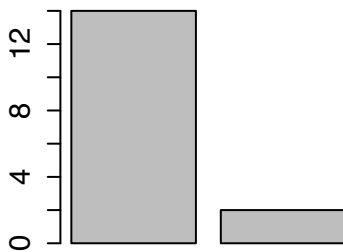
303   Zardus, J. D., Lake, D. T., Frick, M. G. & Rawson, P. D. 2014. Deconstructing an assemblage of

304   "turtle" barnacles: Species assignments and fickle fidelity in *Chelonibia. Mar. Biol.* **161**:

305   45-59.

Allele frequencies of each microsatellite locus amplified from the Atlantic lineage of *Chelonibia testudinaria*

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.

Allele frequencies of each microsatellite locus amplified from the Indo-West Pacific lineage of *Chelonibia testudinaria*

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 3.

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 were amplified in the same reaction.

Table 1.

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

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). Significance values of HWE test were adjusted for multiple comparisons (Holm, 1979). Genotyping error rates were based on re-genotyping of all Atlantic individuals. Abbreviations: n = number of individuals; Obs het. = observed heterozygosity; Exp het. = expected heterozygosity.

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	0.56	0.82	0.09	16.11	0.14	0
Ctest9	34	388	432	8	0.62	0.69	**0.03**	7.45	0.04	0.02
Ctest10	34	264	278	8	0.65	0.77	0.13	7.45	0.07	0.03
Ctest11	38	140	318	27	0.87	0.93	0.52	22.68	0.03	0.07
Ctest12	34	388	476	23	0.74	0.92	**0.01**	20.26	0.1	0.02
Ctest16	35	355	367	4	0.37	0.39	0.64	3.7	0.01	0
Ctest18	33	450	485	7	0.61	0.78	0.18	6.98	0.1	0.06
Ctest31	36	292	336	8	0.56	0.66	**<0.0001**	6.54	0.07	0.07
Ctest32	35	316	464	4	0.57	0.52	0.92	3.65	0	0
Ctest36	33	336	524	23	0.46	0.89	**<0.0001**	20.07	0.23	0.03
Ctest47	37	255	385	10	0.62	0.66	0.78	8.36	0.02	0.02

Microsatellite loci characterization for *Chelonibia testudinaria* of the Indo-West Pacific lineage

Range refers to the smallest (min) and largest (max) allele observed. Frequency of null alleles was estimated after Brookfield (1996). Significance values of HWE test were adjusted for multiple comparisons (Holm, 1979). Abbreviations: n = number of individuals; Obs het. = observed heterozygosity; Exp het. = expected heterozygosity.

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	0.61	0.13	3.07	0
Ctest9	19	408	456	6	0.74	0.7	1	3.95	0
Ctest10	12	260	344	9	0.5	0.81	0.08	5.48	0.17
Ctest11	5	136	236	5	0.2	0.78	0.11	4.63	0.33
Ctest12	18	354	374	5	0.33	0.58	**<0.0001**	3.38	0.15
Ctest16	14	351	375	5	0.57	0.61	1	3.53	0.02
Ctest18	18	440	480	7	0.28	0.83	**<0.0001**	5.22	0.3
Ctest31	8	224	304	2	0.25	0.22	1	1.8	0
Ctest32	19	308	316	3	0	0.28	0.08	2.13	0.22
Ctest36	8	348	468	8	0.5	0.84	0.1	5.74	0.18
Ctest47	12	140	340	12	0.83	0.89	0.14	6.88	0.03