Characterization, development and multiplexing of microsatellite markers in three commercially exploited reef fish and their application for stock identification

Laura Taillebois [Corresp., 1, 2] , Christine Dudgeon [3] , Safia Maher [3] , David A Crook [1] , Thor M Saunders [4] , Diane P Barton [1, 4] , Jonathan A Taylor [4] , David J Welch [5] , Stephen J Newman [6] , Michael J Travers [6] , Richard J Saunders [7, 8] , Jennifer Ovenden [3]

[1] Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, Northern Territory, Australia

[2] North Australia Marine Research Alliance, Arafura Timor Research Facility, Darwin, Northern Territory, Australia

[3] Molecular Fisheries Laboratory, School of Biomedical Sciences, University of Queensland, Brisbane, Queensland, Australia

[4] Fisheries Research, Northern Territory Department of Primary Industries and Fisheries, Berrimah, Northern Territory, Australia

[5] C2O Fisheries, Cairns, Queensland, Australia

[6] Western Australian Fisheries and Marine Research Laboratories, Department of Fisheries, Government of Western Australia, North Beach, Western Australia, Australia

[7] Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Douglas, Queensland, Australia

[8] Animal Science, Queensland Department of Agriculture and Fisheries, Brisbane, Queensland, Australia

Corresponding Author: Laura Taillebois
Email address: laura.taillebois@cdu.edu.au

Thirty-four microsatellite loci were isolated from three reef fish species; golden snapper *Lutjanusjohnii*, black jewfish *Protonibeadiacanthus* and grass emperor *Lethrinuslaticaudis* using a next generation sequencing approach. Both IonTorrent single reads and Illumina MiSeq paired-end reads were used, with the latter demonstrating a higher quality of reads than the IonTorrent. From the 1-1.5 million raw reads per species, we successfully obtained 10 to 13 polymorphic loci for each species, which satisfied stringent design criteria. We developed multiplex panels for the amplification of the golden snapper and the black jewfish loci, as well as post-amplification pooling panels for the grass emperor loci. The microsatellites characterized in this work were tested across three locations of northern Australia. The microsatellites we developed can detect population differentiation across northern Australia and may be used for genetic structure studies and stock identification.

1 **Characterization, development and multiplexing of microsatellite markers in three**

2 **commercially exploited reef fish and their application for stock identification**

4 Laura Taillebois[1,2], Christine Dudgeon[3], Safia Maher[3], David A. Crook[1], Thor Saunders[4], Diane

5 P. Barton[1,4], Jonathan Taylor[4], David J. Welch[5], Stephen J. Newman[6], Michael J. Travers[6],

6 Richard J. Saunders[7,8] & Jennifer Ovenden[3]

8 [1] Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin,

9 NT 0810, Australia

10 [2] North Australia Marine Research Alliance, Arafura Timor Research Facility, Darwin, NT 0810,

11 Australia

12 [3] Molecular Fisheries Laboratory, School of Biomedical Sciences, The University of Queensland,

13 St. Lucia, QLD 4072, Australia

14 [4] Fisheries Research, Northern Territory Department of Primary Industries and Fisheries,

15 Berrimah, NT 0810, Australia

16 [5] C_2O Fisheries, Cairns, QLD 4870, Australia

17 [6] Western Australian Fisheries and Marine Research Laboratories, Department of Fisheries,

18 Government of Western Australia, P.O. Box 20, North Beach, WA 6920, Australia

19 [7] Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Douglas,

20 QLD 4814, Australia

21 [8] Animal Science, Queensland Department of Agriculture and Fisheries, QLD, Australia

23 **Corresponding auhtor:** Laura Taillebois, Email: laura.taillebois@cdu.edu.au.

25 **Abstract**

26

27 Thirty-four microsatellite loci were isolated from three reef fish species; golden snapper

28 *Lutjanus johnii*, black jewfish *Protonibea diacanthus* and grass emperor *Lethrinus laticaudis*

29 using a next generation sequencing approach. Both IonTorrent single reads and Illumina MiSeq

30 paired-end reads were used, with the latter demonstrating a higher quality of reads than the

31 IonTorrent. From the 1-1.5 million raw reads per species, we successfully obtained 10 to 13

32 polymorphic loci for each species, which satisfied stringent design criteria. We developed

33 multiplex panels for the amplification of the golden snapper and the black jewfish loci, as well as

34 post-amplification pooling panels for the grass emperor loci. The microsatellites characterized in

35 this work were tested across three locations of northern Australia. The microsatellites we

36 developed can detect population differentiation across northern Australia and may be used for

37 genetic structure studies and stock identification.

38

39   **1. Introduction**

40

41   Microsatellites are hypervariable, nuclear-encoded and codominant-inherited markers used for a

42   variety of aquaculture and fisheries applications, including determining the spatial extent of

43   fisheries stocks and other important applications of population genetics. *De novo* discovery of

44   microsatellites is required for analyses in the laboratory with each non-model species studied,

45   however the costs are high and the procedure involving cloning is time-consuming (Peters et al.

46   2009). The alternative to *de novo* development is cross-species amplification where existing

47   microsatellite loci of related species are used on the target species; but this is often hampered by

48   the lack of conserved flanking sequences of microsatellites or the lack of data on related species.

49   The adoption of Next-Generation Sequencing (NGS) by researchers using microsatellite loci has

50   made the discovery of microsatellite markers easier (Gardner et al. 2011) and is becoming the

51   preferred method for developing microsatellites (Abdelkrim et al. 2009; Castoe et al. 2010;

52   Fernandez-Silva et al. 2013). Once the microsatellites are identified, major cost and time

53   reductions in the laboratory are achieved through polymerase chain reaction (PCR) multiplexing.

54   The challenge of PCR multiplexing is to combine several microsatellite primers into one PCR

55   cocktail to amplify several microsatellite loci at the same time.

56   Herein, we describe the discovery, characterization, development and multiplexing of

57   microsatellite loci of three reef fish species of commercial and recreational significance: golden

58   snapper (*Lutjanus johnii*, Lutjanidae), black jewfish (*Protonibea diacanthus*, Sciaenidae) and

59   grass emperor (*Lethrinus laticaudis*, Lethrinidae). *Lutjanus johnii,* is a highly prized sport and

60   food fish and is harvested in the commercial, recreational, charter and indigenous sectors of

61   northern Australia and many other fisheries worldwide (Allen, Swainston & Ruse 1997). The

62   catch of *L. johnii* has been declining in the Northern Territory since 1997 and this species is

63   considered overfished (Grubert et al. 2013; Saunders et al. 2014a). Sciaenids form the basis of

64   commercial and recreational fisheries of tropical and temperate regions worldwide (Lenanton &

65   Potter 1987; Rutherford et al. 1989) and several large species are considered threatened or

66   vulnerable due to over-fishing (Rao et al. 1992; Saunders et al. 2014b; True, Loera & Castro

67   1997). Among Sciaenid species, *P. diacanthus* is vulnerable to over-exploitation because of its

68   predictable   aggregating   behavior   (Bowtell   1995;   Bowtell   1998;   Phelan   2001).

69   *Lethrinus laticaudis* is considered an excellent eating fish and is targeted by commercial fishers

70 and recreational anglers across northern Australia (Coleman 2004). Although *L. laticaudis* is

71 considered robust to fishing pressure (Grubert, Kuhl & Penn 2010) due to its high reproductive

72 capacity (*i.e.* serial batch spawners, high spawning frequency, high batch fecundity) (Ayvazian,

73 Chatfield & Keay 2004), it is heavily exploited in some areas. These three fish species are of

74 high economic value and the sustainability of the fisheries they support is potentially threatened

75 by over harvesting and thus requires the development of suitable management programs. The

76 development of genetic tools is necessary to further investigate their population genetics and

77 assess stock structure.

78   In this study, we provide novel polymorphic microsatellite loci for the three species. We

79 also describe a fast and cost-effective protocol for species-specific microsatellite marker

80 discovery using genomic sequencing and multiplexing. Finally, we explore the relevance of the

81 described microsatellite markers for further population genetics by looking at the genetic

82 differences found between two locations in the Northern Territory for the three study species.

83 This will inform us on the potential to use these markers for the identification of stocks for

84 management purposes. This is the first report of the nuclear genomes of the three study species

85 and provides useful baseline information for future genetic studies of these important species.

86

87 **2. Materials and methods**

88

89   **2.1.   Sample and extraction**

90 Samples selected for the production of the microsatellite loci were derived from muscle tissue

91 collected by the Northern Territory Department of Primary Industries and Fisheries and the

92 Western Australian Department of Fisheries under Charles Darwin University Animal Ethics

93 permit A13014. The *L. johnii* sample was a 210 mm male caught at 6 m depth in Darwin

94 Harbour, Northern Territory, Australia (Middle Arm, 130°58'0.24"E, 12°39'0.97"S) in 2013. The

95 *P. diacanthus* sample was a 890 mm male caught in Fenton patches, Northern Territory,

96 Australia (130° 42.084'E, 12° 10.664'S) in 2013. The *L. laticaudis* sample (WAM16-001) was a

97 419 mm male collected from East of the Lacepede Islands, Western Australia, Australia in 2013.

98 Genomic DNA from *L. johnii* and *P. diacanthus* was extracted using Qiagen DNeasy Blood &

99 Tissue columns (Qiagen, Germantown, USA) following the manufacturer's instructions.

100 *Lethrinus laticaudis* genomic DNA was extracted using a salting-out method as described in

101 Broderick et al. (2011). Genomic DNA from all samples for testing the loci and further

102 genotyping was extracted using ISOLATE II Genomic DNA Kit (Bioline) following the

103 manufacturer's instructions. This resulted in 100μL of eluted DNA for each sample. All the

104 DNA extracts were quantified using the Qubit v3 (ThermoFisher) fluorometric method.

105

106 **2.2.   Next-Generation Sequencing and primer selection**

107 The purified genomic DNA of *L. johnii* and *P. diacanthus* was prepared for direct shotgun

108 sequencing using the Iron ExpressTM fragment library kit and sequenced on an IonTorrent

109 Personal Genome Machine using an Ion318 chip (Life Technologies Corporation, Grand Island,

110 NY). The purified genomic DNA of *L. laticaudis* was sequenced on an Illumina® MiSeq as part

111 of a 2x300bp run at the Australian Genome Research Facility. Because two different NGS

112 platforms were used to scan the genomes of the three species we were able to compare their

113 performance for microsatellite design and to assess whether equivalent results were obtainable

114 from each platform. *Lutjanus* and *Lethrinus* genera and Sciaenidae are known to have genome

115 size   comparable   to   other   fish   species   (average   size   for   *Lutjanus* = 1066 Mb,

116 *Lethrinus* = 1192 Mb,   Sciaenidae = 753 Mb,   Perciformes = 919 Mb;   Gregory   2001).   The

117 paired-end reads obtained with the MiSeq run were merged using FLASH source code (Magoč &

118 Salzberg 2011) and their quality was checked in FastQC (Andrews 2010); the first 10 bp were

119 trimmed in Geneious v 9.0 (Drummond et al. 2011).

120   From the NGS data we looked for sequences longer than 300 bp that contained a

121 microsatellite repeat that would be suitable for primer design. These sequences were checked for

122 microsatellite motifs and forward and reverse primers were designed using the software QDD2

123 beta (Meglécz et al. 2010). Sequences with target microsatellites and primers were then filtered

124 according to the following criteria: only pure repeats were selected; all dinucleotide repeats were

125 excluded; repeats greater than 8 were selected; loci with a predictive target sequence length

126 above 300 bp were selected; primers with a distance less than 20bp from the repeat sequence

127 were excluded; and the PCR primers with a PCR_PRIMER_ALIGNSCORE equal or above 6

128 were excluded to discard primers with high alignment scores to the amplicon. A unique pair of

129 primers was selected for each locus. The PCR predicted sequences for all the loci were imported

130 into Geneious v 9.0 and blasted (MEGABLAST) against the NCBI GenBank database to check

131 if the microsatellites fell into coding regions. Sequences that would be homologous to any other

132 NCBI sequence likely to be functional were excluded. All the primers were blasted against their
133 original genomic database built using the NGS reads. Only microsatellites with primers that had
134 one hit across the whole genome were kept for further steps to increase the chance of each
135 primer to amplify a unique sequence. As a final check point before wet laboratory work and to
136 make sure each pair of primers bound to the 5' and 3' ends of a unique sequence containing a
137 microsatellite we selected, containing microsatellites sequences and pairs of primers were
138 mapped *de novo*. For each species, we selected the 48 microsatellites that contained the best
139 quality repeats with the highest number of tri- tetra- or penta-nucleotide repeats possible and
140 with no small dinucleotide repeats between the primer and the microsatellite sequence to avoid
141 any noise that may interfere with scoring genotypes.

142   Forward primers were tagged on the 5' end with the universal CAG sequence (5'-
143 CAGTCGGGCGTCATCA-3'). Inclusion of the 5'-tail will allow use of a CAG-tagged universal
144 primer in the PCR that is fluorescently labeled for detection on the sequencing machine
145 (Schuelke 2000). Additionally, a pig-tail (5'-GTTTCTT-3') was added to the 5' end of the
146 reverse primers to increase the accuracy of genotyping and ensure the consistency of the
147 amplicon size (Brownstein, Carpten & Smith 1996). The resulting 48 pairs of primers were
148 synthetized by Integrated DNA Technologies (www.idtdna.com).

149

150   **2.3.   Loci and primers testing**

151 For each species, the 48 pairs of primers were tested over a set of genomic DNA extracted from
152 eight individuals of the target species. The fluorescent dye labelling of PCR fragments in one
153 reaction was performed with three primers. Amplification reactions were carried out in a 8.8 µL
154 volume comprising 1 µL of DNA, 4.84 µL of 2x Bioline Taq Mastermix, 4.4 pmol of locus-
155 specific forward CAG-tagged primer, 22 pmol of locus-specific reverse pig-tail-tagged primer
156 and 22 pmol of universal CAG primer labeled with a 6-FAM fluorescent dye. The heating cycle
157 parameters used for amplification were 95ºC for 3 min, 37 cycles of denaturation at 94ºC for 15s,
158 annealing for 15s at 57ºC and elongation at 72ºC for 60s. A final extension at 72ºC for 30s was
159 performed. Post-amplification, the PCR products were diluted with water 1:20. We added 2 µL
160 of these diluted PCR products to 10 µL of Hi-Di formamide (ABI) and 0.05 µL of GenScan-500
161 LIZ (ABI) size standard. Samples were denatured at 95ºC for 3 min and sized on the ABI 3730xl
162 capillary sequencer (Applied Biosystems, Carlsbad, California, USA) using the conditions set

163   down by the manufacturer. Chromatograms were analysed using Geneious v 9.0 (Drummond et

164   al. 2011).

165   Criteria used to select the best loci among the 48 tested for each species included the

166   amplification success rate, peak intensity, the presence or absent of stutter peaks, the

167   polymorphism of the loci, the number of alleles and heterozygosity. The best loci were

168   individually tested against a further 23 samples of the target species.

169

170   **2.4.   Multiplex optimization of PCR**

171   In order to reduce the cost and time of genotyping for further genetic studies, the newly designed

172   microsatellites were combined into multiplex panels of 2 to 4 loci. The panels were set up based

173   on the microsatellite allele-size range. The primers for all the loci of each panel were combined

174   in a single PCR to allow the amplification of several microsatellite loci at the same time. When

175   allele-size ranges overlapped, alternate dyes were employed to allow the discrimination of each

176   locus on the chromatograms. Each of the four ABI dyes FAM, VIC, NED and PET were tailed

177   with a unique M13 primer: M13F (5'-TTTCCCAGTCACGACGTTG-3'), M13V (5'-

178   GCGGATAACAATTTCACACAGG-3'), M13N (5'-TAAAACGACGGCCAGTGC-3') and

179   M13P (5'-CACAGGAAACAGCTATGACC-3'). The 5' end of the forward primer for the locus

180   was synthetized with the corresponding M13 tail to allow fluorescent labeling of PCR product

181   using a 3-primer PCR protocol as described above (Schuelke 2000). Several multiplex trials were

182   conducted to find the best combination of loci with the optimal concentration of primers and

183   PCR parameters. Primer pairs that failed to amplify within a multiplex were removed from the

184   panels and further optimization focused on the remaining primer pairs. For each species, the

185   multiplex trials were all evaluated against 8 samples that were the same for those used in the 23

186   samples above, allowing the consistency to be checked across the templates.

187

188   **2.5.   Genetic variability and population genetics**

189   In order to test if the herein developed microsatellites would be useful to discriminate fish stocks

190   across northern Australia we collected samples from two locations in the Northern Territory and

191   one location in Western Australia and assayed their population structure. For each of the three

192   species, samples were collected from Camden Sound (Western Australia, Australia), Wadeye

193   (Northern Territory, Australia) close to the Western Australia border and Vanderlin Islands in the

194   bottom of the Gulf of Carpentaria (Northern Territory, Australia) (Figure 1). Fourteen *L. johnii*,

195   eighteen *P. diacanthus* and twenty-eight *L. laticaudis* were collected from Camden Sound

196   (Western Australia, Australia); twenty-nine *L. johnii*, twenty-five *P. diacanthus* and twenty-

197   seven *L. laticaudis* were collected from Wadeye (Northern Territory, Australia); twenty-five

198   *L. johnii*, twenty-nine *P. diacanthus* and twenty-nine *L. laticaudis* were collected from Vanderlin

199   Islands (Northern Territory, Australia). All individuals were assayed as part of the multiplex

200   panels or PCR pooled. The multiplex PCR were comprised of 2-5 µL of DNA depending on the

201   samples (approximately 20 ng total), 8 µL of 2x Bioline Taq Mastermix and 4 µL of panel

202   primer mix containing various quantities of primers as described in Table 1. Concentrations of

203   the different primers were adjusted to obtain homogenous PCR products revealed by similar

204   intensity chromatogram peaks for each of the dyes within each panel. The heating cycle

205   parameters, sizing of the alleles and chromatogram analyses were conducted using the same

206   method as stated above.

207   The obtained datasets were statistically evaluated. The potential for null alleles, large

208   allele dropout and stuttering to interfere with scoring accuracy was evaluated for each

209   microsatellite locus in each sample using Microchecker v.2.2.3 (Van Oosterhout et al. 2004). The

210   software Arlequin 3.5.2.2 (Excoffier & Lischer 2010) was used to calculate the number of alleles

211   (A), expected (H_e) and observed (H_o) heterozygosity and conduct exact tests of conformance of

212   genotypic proportions to Hardy-Weinberg equilibrium expectations. Estimation of probability

213   values (P_{HW}) employed a Monte Carlo Markov Chain (MCMC) of 10^5 steps and 5.10^4

214   dememorization. Genotypic equilibrium between pairs of microsatellites (linkage disequilibrium)

215   was tested in Arlequin with 10,000 permutations. Fixation indices (F_{ST}) between pairs of sample

216   localities were estimated as implemented in Arlequin to identify possible spatial boundaries

217   among sample locations.

218

219   **3.   Results and Discussion**

220

221   The IonTorrent sequencing technology allowed the production of longer reads (range 8-620 bp)

222   compared to the Illumina MiSeq that produced 300 bp reads fixed by the method. The paired-end

223   reads of the MiSeq run were merged to increase their length to 300-575 bp and allow the

224   detection of at least 300 bp length sequences that may contain a microsatellite locus. This

225   resulted in 1,169,198 reads, which is less than what was used to select microsatellite loci in the

226   first two species. The MiSeq run produced higher quality sequences than the IonTorrent (Phred

227   score: 36 vs 29). Quality profiles along the IonTorrent reads showed that the quality of the

228   sequencing decreased with length meaning that the end of the longer reads (> 325 bp) had a

229   lower quality then at their start. QDD Pipe1 detected between 110,000 and 170,000 sequences

230   containing a microsatellite sequence depending on the species (Table 2). This number was

231   independent of the type of NGS platform used. From those sequences QDD Pipe2 removed the

232   low complexity sequences (no BLAST to itself), putative minisatellites (short sequences of

233   repeated nucleotides) and sequences that had BLAST hit to other sequences to only keep the

234   singletons and unique consensus sequences. QDD Pipe3 designed primers for all QDD Pipe2

235   output reads. The resulting number of sequences that contained a microsatellite sequence and the

236   corresponding primers were given in the final output of QDD pipeline, and varied between

237   20,000 and 30,000 depending on the species (Table 2). After applying the filtering criteria

238   described previously, 97 potentially amplifiable microsatellite reads were found for

239   *P. diacanthus*, 121 for *L. johnii* and 103 for *L. laticaudis*. From those microsatellite reads, we

240   selected the ones with the smallest number of repeats but greater than 8 and eliminated those

241   with small repeats between the primer and the microsatellite to reach 48 microsatellite loci per

242   species being ultimately tested in the laboratory.

243   The testing of 144 primer pairs resulted in the selection of 34 polymorphic loci that could

244   be reliably scored and showed consistent amplification success. We selected a final set of 10 loci

245   for *L. johnii*, 11 loci for *P. diacanthus* and 13 loci for *L. laticaudis* (Table 3). Multiplex panels of

246   microsatellites were developed for the two species *L. johnii* and *P. diacanthus*, and the

247   optimization of each panel resulted in the efficient assay and unambiguous scoring of

248   microsatellites in the two species. Although the M13 labeling system worked very well for

249   *L. johnii* and *P. diacanthus* it did not amplify successfully as part of PCR multiplexes for

250   *L. laticaudis*. The reasons why it did not work well for this species are still unclear as the quality

251   of the DNA was even across the three species and the same protocol was followed. However, in-

252   house experiments showed that lengthening the labeled forward primer might facilitate the PCR

253   reaction when multiplexing several loci. Direct labeling of the forward primer may also be

254   another option for multiplexing a large number of loci. For *L. laticaudis*, the loci were all

255 amplified in individual PCR with the CAG labeling system as described above. The resulting

256 PCR products were then pooled according to the panels described in Table 1 before the ABI run.

257   Genotypes from 10 microsatellites were obtained from multiplexed PCR for 68

258 individuals of *L. johnii*. There was 1.91 % missing data and the number of alleles for each locus

259 varied between 4 and 23 (Table 3). Microchecker indicated the possible occurrence of null

260 alleles at location Wa for locus *Luj012* and at CS for locus *Luj018* with possible stuttering or

261 scoring errors for the latter. There were only 3/45 significant tests for linkage disequilibrium

262 between pairs of loci (*Luj094* x *Luj051*, *Luj051* x *Luj091* and *Luj076* x *Luj082*) and there was no

263 deviation from Hardy-Weinberg equilibrium (HWE) detected. Heterozygosity was moderate to

264 high for all loci (mean overall loci 0.707 +/- 0.200) and generally similar to expectations (around

265 0.7 for marine fish, DeWoody & Avise 2000). Genotypes from 11 microsatellites were obtained

266 from multiplexed PCR for 72 individuals of *P. diacanthus*. There was 2.77 % missing data and

267 the number of alleles for each locus varied between 5 and 19 (Table 3). Microchecker indicated

268 the possible occurrence of null alleles at location CS for locus *Prd068* and at VI for locus

269 *Prd012*. There were only 3/55 significant tests for linkage disequilibrium between pairs of loci

270 (*Prd046* x *Prd018*, *Prd020* x *Prd018 and Prd018* x *Prd045*) and overall deviations from Hardy-

271 Weinberg equilibrium (HWE) were detected at two loci *Prd023* (p-value=0.034) and *Prd018* (p-

272 value=0.008). Heterozygosity was variable and with an overall mean lower than for *L. johnii*

273 (0.673 +/- 0.185). Genotypes from 10 microsatellites were obtained from pooled post PCR

274 products for 84 individuals of *L. laticaudis* as 3 of our developed microsatellites did not

275 amplified consistently in all the samples. There was 2.61 % missing data and the number of

276 alleles for each locus varied between 9 and 22 (Table 3). Microchecker indicated the possible

277 occurrence of null alleles at location CS for loci *Lel033* and *Lel012*. There were 6/78 significant

278 tests for linkage disequilibrium between pairs of loci (*Lel040* x *Lel012*, *Lel040* x *Lel032*, *Lel011*

279 x *Lel027*, *Lel012* x *Lel032, Lel012* x *Lel013* and *Lel028* x *Lel027*) and overall deviations from

280 Hardy-Weinberg equilibrium (HWE) were detected at a single locus *Lel012* (p-value=0.005).

281 Heterozygosity was high and with an overall mean higher than for the two other species (0.834

282 +/- 0.078).

283   A pattern of genetic differentiation with low but significant population-pairwise F_{ST}

284 (range 0.007-0.017) was observed in the three species (Table 4). These levels of differentiation

285 are similar in magnitude to those reported for other marine fish species with potentially high

286 gene flow (O'Reilly et al. 2004). The three species also presented different patterns of structure

287 between the three locations. *L. laticaudis* presented pairwise differentiation between Western

288 Australia location (CS) and the two other Northern Territory locations (VI and Wa). Similarly,

289 *P. diacanthus* presented a structure between the two most distant populations of CS and VI

290 whereas CS and Wa remained undifferentiated. By contrast, *L. johnii* did not present structure

291 between Western Australia and Northern Territory. However, VI and Wa in the Northern

292 Territory had a significant pairwise F_{ST}. Pairwise F_{ST} values were low in all comparisons for the

293 three species (range 0.001-0.017) as commonly reported in marine species and, 4 out of 9

294 pairwise comparisons were significant meaning the microsatellites we developed can accurately

295 detect population differentiation across Northern Australia and may be used for genetic structure

296 studies and stock identification.

297

298 **4. Conclusion**

299

300 In conclusion, we applied the direct sequencing of a genomic library approach to develop

301 microsatellite loci and it resulted in a significant reduction in laboratory effort compared to

302 traditional protocols for microsatellite discovery (cloning and Sanger sequencing). Merged

303 paired-end reads from the MiSeq platform demonstrated higher quality of reads than the

304 IonTorrent single reads. From the 1-1.5 million raw reads, we selected a reduced number of loci

305 to test (48) and successfully amplified satisfactory polymorphic loci for 10 to 13 of them

306 depending on the species. However, the NGS data revealed the potential for hundreds to

307 thousands of potentially amplifiable microsatellites to be discovered. The microsatellites

308 characterized in this work will be available to explore the population genetics and stock structure

309 of these highly valuable species.

310

311 **Acknowledgements**

312 The authors thank staff from the Western Australian Department of Fisheries and the Northern

313 Territory Department of Primary Industries as well as Indigenous rangers for sample collection.

314 We extend our thanks to Samuel Williams at the Molecular Fisheries laboratory and Sean Corley

315 at the Animal Genetics Laboratory for their assistance with laboratory work.

316

References

Abdelkrim J, Robertson BC, Stanton J-AL, and Gemmell NJ. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46:185.

Allen GR, Swainston R, and Ruse J. 1997. *Marine fishes of tropical Australia and south-east Asia*. Perth: Western Australian Museum.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ayvazian S, Chatfield B, and Keay I. 2004. The age, growth, reproductive biology and stock assessment of grass emperor, *Lethrinus laticaudis* in Shark Bay, Western Australia. Department of Fisheries, Research Division, Westrern Australian Marine Research Laboratories.

Bowtell B. 1995. Heed jewfish warnings. *Fish and Boat July*.

Bowtell B. 1998. Huge schools of black jew about. *Fish and Boat July*.

Broderick D, Ovenden J, Buckworth R, Newman S, Lester R, and Welch D. 2011. Genetic population structure of grey mackerel *Scomberomorus semifasciatus* in northern Australia. *Journal of Fish Biology* 79:633-661.

Brownstein MJ, Carpten JD, and Smith JR. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20:1004-1006.

Castoe TA, Poole AW, Gu W, Jason de Koning A, Daza JM, Smith EN, and Pollock DD. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources* 10:341-347.

Coleman APM. 2004. The national recreational fishing survey: the Northern Territory. Department of Business Industry and Resource Development. Fishery Report 72.

DeWoody J, and Avise J. 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* 56:461-473.

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, and Markowitz S. 2011. Geneious v5. 4.

Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* 10:564-567.

Fernandez-Silva I, Whitney J, Wainwright B, Andrews KR, Ylitalo-Ward H, Bowen BW, Toonen RJ, Goetze E, and Karl SA. 2013. Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *PLoS one* 8:e55990.

Gardner MG, Fitch AJ, Bertozzi T, and Lowe AJ. 2011. Rise of the machines–recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* 11:1093-1101.

Gregory TR. 2001. Animal genome size database. *Available at* http://www.genomesize.com.

Grubert M, Saunders T, Martin JM, Lee H, and Walters CJ. 2013. Stock Assessments of Selected Northern Territory Fishes. Northen Territory Government. Fishery Report 110.

Grubert MA, Kuhl PJ, and Penn JW. 2010. Ecological risk assessment. Northern Territory coastal line fishery. Norther Territory Department of Resources. Fishery Report 103.

Lenanton R, and Potter I. 1987. Contribution of estuaries to commercial fisheries in temperate Western Australia and the concept of estuarine dependence. *Estuaries* 10:28-35.

363 Magoč T, and Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve
364 genome assemblies. *Bioinformatics* 27:2957-2963.
365 Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, and Martin J-F. 2010. QDD: a
366 user-friendly program to select microsatellite markers and design primers from large sequencing
367 projects. *Bioinformatics* 26:403-404.
368 O'Reilly P, Canino M, Bailey K, and Bentzen P. 2004. Inverse relationship between F_{ST} and
369 microsatellite polymorphism in the marine fish, walleye pollock (Theragra chalcogramma):
370 implications for resolving weak population structure. *Molecular Ecology* 13:1799-1814.
371 Peters MB, Ovenden JR, Broderick D, Lance SL, Hagen C, and Glenn TC. 2009. Fifteen
372 microsatellite loci for the jungle perch, *Kuhlia rupestris*. *Molecular Ecology Resources* 9:1467-
373 1469.
374 Phelan MJ. 2001. Sciaenid aggregations in northern Australia: an example of successful
375 outcomes through collaborative research. *Fisheries Centre Research Reports*:100.
376 Rao TA, Mohan RL, Chakraborty S, Murty VS, Nair KS, Vivekanandan E, and Raje S. 1992.
377 Stock assessment of sciaenid resources of India. *Indian Journal of Fisheries* 39:85-103.
378 Rutherford ES, Tilmant JT, Thue EB, and Schmidt TW. 1989. Fishery harvest and population
379 dynamics of spotted seatrout, *Cynoscion nebulosus,* in Florida Bay and adjacent waters. *Bulletin
380 of Marine Science* 44:108-125.
381 Saunders T, Newman S, Keag M, and Errity C. 2014a. Golden Snapper *Lutjanus johnii*. Status of
382 key Australian fish stocks reports 2014. Fisheries Research and Development Corporation.
383 Canberra. 523-529.
384 Saunders T, Roelofs A, Newman S, and Errity C. 2014b. Black jewfish *Protonibea diacanthus*.
385 Status of key Australian fish stocks reports 2014. Fisheries Research and Development
386 Corporation. Canberra. 344-350.
387 Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature
388 Biotechnology* 18:233-234.
389 True CD, Loera AS, and Castro NC. 1997. Technical Notes: Acquisition of Broodstock of
390 Totoaba macdonaldi: Field Handling, Decompression, and Prophylaxis of an Endangered
391 Species. *The Progressive Fish-Culturist* 59:246-248.
392 Van Oosterhout C, Hutchinson WF, Wills DP, and Shipley P. 2004. MICRO-CHECKER:
393 software for identifying and correcting genotyping errors in microsatellite data. *Molecular
394 Ecology Notes* 4:535-538.

Technical details on the multiplex polymerase chain reaction (PCR) and post-PCR pooled products of microsatellite loci in *Lutjanus johnii*, *Protonibea diacanthus* and *Lethrinus laticaudis*.

Included in the table are the multiplex PCR panels for *L. johnii*, *P. diacanthus* and post-PCR pooled products panels for *L. laticaudis*, primer mix quantities per reaction (μL) within each multiplex and fluorescent dye labels used for each locus in the PCR reactions.

Species	Panel	Microsatellite	Quantity (µL)	ABI Dye
Lutjanus johnii	1	*Luj094*	0.2	FAM
		Luj068	0.6	PET
		Luj027	0.6	FAM
		Luj076	0.6	VIC
	2	*Luj051*	0.2	VIC
		Luj090	0.4	FAM
		Luj114	0.6	VIC
		Luj091	0.6	PET
	3	*Luj072*	0.7	FAM
		Luj082	0.7	VIC
Protonibea diacanthus	1	*Prd012*	0.2	PET
		Prd023	0.2	VIC
		Prd044	0.2	FAM
		Prd042	0.2	NED
	2	*Prd018*	0.4	NED
		Prd045	0.2	PET
		Prd046	0.2	FAM
		Prd020	0.2	PET
	3	*Prd036*	0.2	VIC
		Prd049	0.2	FAM
		Prd024	0.2	PET
Lethrinus laticaudis	1	*Lel033*	-	NED
		Lel012	-	PET
		Lel040	-	VIC
		Lel011	-	FAM
	2	*Lel032*	-	NED
		Lel028	-	PET
		Lel041	-	VIC
		Lel013	-	FAM
	3	*Lel036*	-	NED
		Lel039	-	PET
		Lel044	-	VIC
		Lel047	-	FAM
	4	*Lel027*	-	FAM

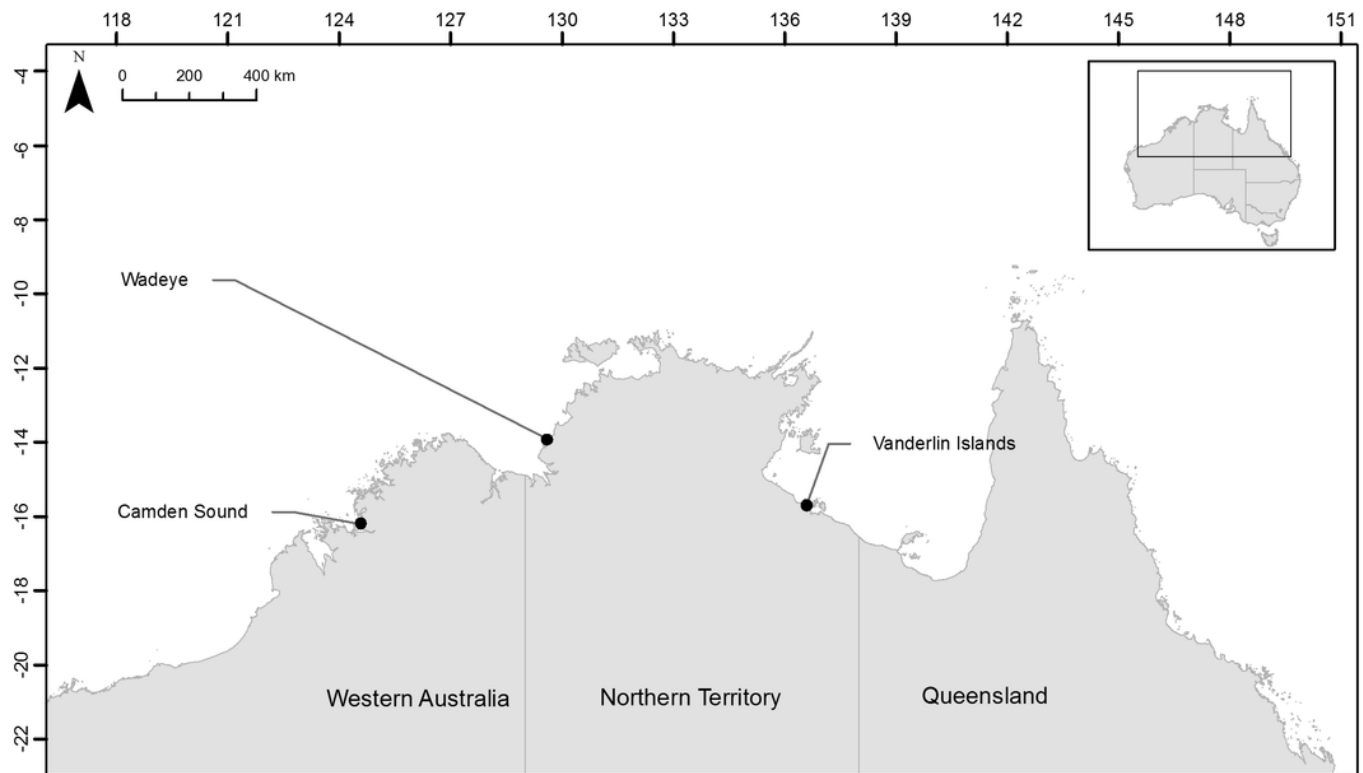
2   Primer mix for each locus is all initially made with 6 µL of forward (M13) primer (10µM),
3   30 µL of reverse primer (10µM), 30 µL of M13 labeled dye (10µM) and 84 µL of water. The

4   primer mixes of each panel are mixed in the proportion given in the table with water added to
5   reach 4 µL.

Next-generation sequencing and bioinformatics details obtained from FastQC software (Andrews, 2010) and QDD pipeline (Meglécz et al., 2010) for *Lutjanus johnii, Protonibea diacanthus* and *Lethrinus laticaudis*.

	Lutjanus johnii	*Protonibea diacanthus*	*Lethrinus laticaudis*
Genomic DNA extraction	Qiagen DNeasy	Qiagen DNeasy	Salting out
NGS Technology	IonTorrent | Ion318Chip	IonTorrent | Ion318Chip	Illumina® MiSeq
Library preparation	Iron ExpressTM	Iron ExpressTM	
Type of reads	Single reads	Single reads	Paired-end reads
Number of reads	1,374,891	1,587,789	2,800,640
Merged reads	-	-	1,169,198
Reads length	8-620	8-618	300
Merged reads length	-	-	300-575
FASTQC			
% GC	41	42	39
Sequence quality < Phred 20	yes at positions >325bp	yes at positions >350 bp	no
Per sequence quality - Phred score	29	29	36
Sequence length distribution	peak at 350 bp	peak at 350 bp	plateau at 520-540 bp
QDD2			
QDD2 pipe 1 - Sequence preparation and microsatellite detection			
Number sequence length ≥80bp	1,235,685	1,405,082	1,169,198
Number sequence with microsatellite	109,641 (8.9%)	167,702 (11.9%)	130,269 (11.1%)
QDD2 pipe 2 - Sequence similarity detection			
Total # input sequences	109,641	167,702	130,269
Numer of unique consensus sequences	18,978	N/A	N/A
Number of singleton sequences	49,122	N/A	N/A
Number of reads in output	68,100	63,789	69,714
QDD2 pipe 3 - Primer design			
Total number of input sequences	68,100	63,789	69,714
Total number of sequences with target MS	67,461	63,785	69,714
Total number of sequences with primers	29,485	20,233	19,867
Filtering QDD output			
Total # input sequences	29,485	20,233	19,867
Total # sequences after filtering criteria	121	97	103

Figure 1

Location of the three sampling sites across northern Australia.

Characteristics of the 34 microsatellite markers developed in *Lutjanus johnii*, *Protonibea diacanthus* and *Lethrinus laticaudis*.

n is the sample size, $\#A$ is the number of alleles at each loci, H_E is the expected heterozygosity, H_O is the observed heterozygosity and P_{HW} is the p-value of the exact tests of conformance of genotypic proportions to Hardy-Weinberg equilibrium expectations.

Species	Locus	Primer sequences (5'-3') (Fluorescent label)		Repeat motif	GenBank accession no.	n	Allele size range (bp)	#A	H_O	H_E	P_{HW}
Lutjanus johnii	*Luj027*	F:	CTGGGCCACACTGATAAAGC (FAM)	(AGC)9	KX387441	68	152-179	8	0.309	0.306	0.606
		R:	GGCTCTGAACCTGGGAGATT								
	Luj094	F:	TCTCAGAGGGTTTGATGCAG (FAM)	(AATC)9	KX387437	68	223-239	4	0.426	0.470	0.205
		R:	CTTTGGCGCTTTCTATCAGC								
	Luj076	F:	CGGGTCGAGTCTGTTTGTGT (VIC)	(AAG)15	KX387436	66	200-233	10	0.818	0.811	0.289
		R:	CTTCAGACGGATTAGCAGCA								
	Luj068	F:	CCTAGGGTGTCAGTGTCAGTCA (PET)	(AAAG)20	KX387435	68	174-258	18	0.882	0.936	0.212
		R:	TGCCTGTATGTTCTCTTGAGC								
	Luj090	F:	ATCCTAATGCATCGTGCTTG (FAM)	(AGC)17	KX387444	68	194-278	23	0.868	0.922	0.146
		R:	GGCATGTTCTATTGAGGTTGG								
	Luj051	F:	TGCAGAGCAACAGAACAACAC (VIC)	(ACTG)10	KX387440	67	172-192	6	0.687	0.596	0.377
		R:	CACCTTGCGTTTGCAGTCT								
	Luj114	F:	CCATAACTGCTGTTCTGTATCTGG (VIC)	(AGC)9	KX387442	68	276-314	11	0.794	0.739	0.725
		R:	AATACGGCAGATCTCGGGTT								
	Luj091	F:	TCATTCCCAGGAGCTCAAAT (PET)	(ACAG)12	KX387438	64	219-279	12	0.813	0.782	0.499
		R:	AATCGTCACTTTCGACCCAC								
	Luj072	F:	ACTCGAAGAACACAGCCCAC (FAM)	(AGC)9	KX387443	65	192-204	5	0.738	0.675	0.052
		R:	CACATTTGAATCCTTGCTGG								
	Luj082	F:	AAGTACATCGGAGGGCTGAG (VIC)	(ACGAT)12	KX387439	65	220-275	12	0.800	0.833	0.548
		R:	TGTTATCAAAGTTCACCGATACAAA								
Protonibea diacanthus	*Prd044*	F:	ACAAAGTTTCCTCCTCTGGC (FAM)	(AAG)13	KX387452	71	181-211	11	0.746	0.804	0.324
		R:	CACGTTCCATCTTTATTTATTTGC								
	Prd023	**F:**	**TCGTGTGAACACTTTGATGC (VIC)**	**(ATC)11**	**KX387448**	**72**	**292-316**	**9**	**0.875**	**0.847**	**0.034**
		R:	CTCGTCTCTGCTCTTGGTCC								
	Prd042	F:	TACCTTTGAGATGCGAGCG (NED)	(AGC)12	KX387451	72	230-248	7	0.694	0.698	0.526
		R:	GTCAAAGCCATCAATCCAGC								
	Prd012	F:	AGGCTGTTTGAACTGCAGGG (PET)	(AAAG)20	KX387445	64	195-271	19	0.828	0.898	0.324
		R:	CATGCTGAGCAATATGTGGG								
	Prd046	F:	TCATCCTGAGTTTGTGCTGG (FAM)	(AGC)9	KX387454	72	224-236	5	0.347	0.351	0.693
		R:	CATGAGTAAGCAGAGCGTGG								
	Prd018	**F:**	**ATGAACGGCATCAGTCAGC (NED)**	**(ACAG)9**	**KX387446**	**68**	**179-207**	**8**	**0.662**	**0.784**	**0.008**
		R:	CGTCTGATAAACAGCACTGCC								
	Prd020	F:	CAATGTTCTGCAAGAGCTGC (PET)	(ATC)11	KX387447	71	189-216	10	0.746	0.713	0.837
		R:	TCAAATGTCAAAGTCCAGTCC								
	Prd045	F:	GTCTATCCATGTTCCAGCCC (PET)	(ATCC)11	KX387453	64	279-309	8	0.672	0.639	0.400

		R:	TCATCCCAAAGTGACCAACC								
	Prd049	F:	CCTTGTCCTCCTTTCAGGC (FAM)	(ACC)9	KX387455	72	216-234	5	0.278	0.316	0.280
		R:	GGGTCATTAAACATGGCAGC								
	Prd036	F:	TCACGTGAAGCGTCTACAGC (VIC)	(AAG)12	KX387450	72	227-257	9	0.653	0.677	0.746
		R:	AAAGGAGGAAACACAGAGCC								
	Prd024	F:	AGAGTGTCCGAGTCCAGAGG (PET)	(AAG)11	KX387449	72	199-217	6	0.736	0.683	0.892
		R:	CAGTACCTGGTGATGGGAGC								
Lethrinus laticaudis	*Lel011*	F:	CTGTCGGAGGTAAAGTGCG (FAM)	(AGC)9	KX387422	84	237-285	11	0.655	0.660	0.251
		R:	CTCATGGTGTTGAGGATGGG								
	Lel040	F:	TGGTTGCAGACAACTGCC (VIC)	(AGC)9	KX387431	84	171-216	14	0.821	0.825	0.367
		R:	CTTAAGAGCAGTGATCCAGGC								
	Lel033	F:	AGTGCGACAAAGAAATGGC (NED)	(AGAT)16	KX387428	84	167-143	20	0.893	0.926	0.438
		R:	CATTTGTCAGTTATGAAACTTGGC								
	Lel012	**F:**	**GCGAGGGTCTGCTACTATAGGG (PET)**	**(AAT)9**	**KX387423**	**76**	**246-334**	**22**	**0.711**	**0.846**	**0.005**
		R:	TGTAAAGTGTAAACCACGTCCC								
	Lel013	F:	CCTGAACCTGGAGAACTCGG (FAM)	(ATC)12	KX387424	82	242-287	10	0.744	0.829	0.150
		R:	ACTGAGGGAGGAGATAAAGGG								
	Lel041	F:	CTGCTGTTCTGGGTTGCC (VIC)	(AAT)19	KX387432	-	-	-	-	-	-
		R:	CAACAAGCTGTTGGTGTCCC								
	Lel032	F:	AAATCTGCATTATGAAATTGGC (NED)	(AAAG)16	KX387427	83	173-233	16	0.867	0.883	0.209
		R:	CAGCTCCTTGAGTTTAGTCCC								
	Lel028	F:	CAGTAGCTTTAATAGTTAGGCACCC (PET)	(AAAG)13	KX387426	83	200-244	14	0.843	0.877	0.954
		R:	GGCTGTCCAGAGTGAGGC								
	Lel047	F:	AAAGAATGGGAAGAATGACCC (FAM)	(AGAT)11	KX387434	-	-	-	-	-	-
		R:	AAGCCAAGTGATTAAGAAACCC								
	Lel027	F:	CACTAAGGGTCCATGTTGCC (FAM)	(AAT)22	KX387425	79	196-238	15	0.911	0.914	0.480
		R:	TCTGTAATGAATGATCAAACCG								
	Lel044	F:	TTCTACTTGACCCTGGTAGGC (VIC)	(ATCC)11	KX387433	83	151-199	11	0.759	0.816	0.268
		R:	AATGTAATGCCATAAGCGGG								
	Lel036	F:	TCCAATTTACACCAAACTAGGC (NED)	(AAAG)15	KX387429	-	-	-	-	-	-
		R:	CCGGAATGATCTGCAGGC								
	Lel039	F:	CTTGTAGAGTGTCAACGAGGG (PET)	(AAT)11	KX387430	80	196-214	9	0.700	0.764	0.747
		R:	CATGATGCAATAACCATCCC								

Pairwise F_{ST} estimates for *Lutjanus johnii*, *Protonibea diacanthus,* and *Lethrinus laticaudis* for three locations.

Estimates are based on 10 microsatellite data from 68 individuals of *L. johnii*, 11 microsatellite data from 72 individuals of *P. diacanthus* and 10 microsatellite data from 84 individuals of *L. laticaudis* among the three sampling locations Camden Sound (CS), Wadeye (Wa) and Vanderlin Islands (VI). The comparisons that differed significantly from zero (p<0.05) are shaded in grey.

	Lutjanus johnii		*Protonibea diacanthus*		*Lethrinus laticaudis*	
	F_{ST}	p-value	F_{ST}	p-value	F_{ST}	p-value
Wa-CS	0.001	0.308	0.009	0.090	**0.017**	**0.000**
Wa-VI	**0.007**	**0.043**	0.007	0.069	0.007	0.069
CS-VI	0.006	0.148	**0.014**	**0.013**	**0.009**	**0.017**