Characterization, development and multiplexing of microsatellite markers using a next-generation sequencing approach in three commercially exploited reef fish (#11466)

First submission

Please read the **Important notes** below, and the **Review guidance** on the next page.
When ready **submit online**. The manuscript starts on page 3.

Important notes

Editor and deadline

Robert Toonen / 16 Jul 2016

Files	3 Table file(s) Please visit the overview page to **download and review** the files not included in this review pdf.
Declarations	**One or more DNA sequences were reported.** **Involves vertebrate animals.** **Involves a field study on animals or plants.**

Review guidelines

Please in full read before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this **pdf** and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standard**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (See **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusion well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

Characterization, development and multiplexing of microsatellite markers using a next-generation sequencing approach in three commercially exploited reef fish

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

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

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

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

[4] C2O Fisheries, Cairns, Queensland, Australia

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

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

[7] 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 will be available to study the population genetics and stock structure of these commercially exploited species.

Characterization, development and multiplexing of microsatellite markers using a next-generation sequencing approach in three commercially exploited reef fish

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

[1] Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT 0810, Australia

[2] Molecular Fisheries Laboratory, School of Biomedical Sciences, The University of Queensland, St. Lucia, QLD 4072, Australia

[3] Fisheries Research, Northern Territory Department of Primary Industries and Fisheries, Berrimah, NT 0810, Australia

[4] C_2O Fisheries, Cairns, QLD 4870, Australia

[5] Western Australian Fisheries and Marine Research Laboratories, Department of Fisheries, Government of Western Australia, P.O. Box 20, North Beach, WA 6920, Australia

[6] Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Douglas, QLD 4814, Australia

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

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

Abstract

Thirty-four microsatellite loci were isolated from three reef fish species; golden snapper *Lutjanus johnii*, black jewfish *Protonibea diacanthus* and grass emperor *Lethrinus laticaudis* 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 will be available to study the population genetics and stock structure of these commercially exploited species.

1. Introduction

Microsatellites are hypervariable, nuclear-encoded and codominant-inherited markers used for a variety of aquaculture and fisheries applications, including determining the spatial extent of fisheries stocks and other important applications of population genetics. *De novo* discovery of microsatellites ~~are~~ required for analyses in the laboratory with each non-model species studied, however the costs are high and the regular procedure involving cloning is time-consuming (Peters et al. 2009). The alternative to *de novo* development is cross-species amplification where existing microsatellite loci of related species are used on the target species; but this is often hampered by the lack of conserved flanking sequences of microsatellites or the lack of data on related species. The adoption of Next-Generation Sequencing (NGS) by researchers using microsatellite loci has made the discovery of microsatellite markers easier (Gardner et al. 2011) and is becoming the preferred method for developing microsatellites (Abdelkrim et al. 2009; Castoe et al. 2010). Once the microsatellites are identified, major cost and time reductions in the laboratory are achieved through polymerase chain reaction (PCR) multiplexing. The challenge of PCR multiplexing is to combine several microsatellite primers into one PCR cocktail to amplify several microsatellite loci at the same time.

Herein, we describe the discovery, characterization, development and multiplexing of microsatellite loci of three reef fish species of commercial and recreational significance: golden snapper (*Lutjanus johnii*, Lutjanidae), black jewfish (*Protonibea diacanthus*, Sciaenidae) and grass emperor (*Lethrinus laticaudis*, Lethrinidae). *Lutjanus johnii,* is a highly prized sport and food fish and is harvested in the commercial, recreational, charter and indigenous sectors of northern Australia and many other fisheries worldwide (Allen, Swainston & Ruse 1997). The catch of *L. johnii* has been declining in the Northern Territory since 1997 and this species is considered overfished and that overfishing is occurring (Grubert et al. 2013; Saunders et al. 2014a). Sciaenids form the basis of commercial and recreational fisheries of tropical and temperate regions worldwide (Lenanton & Potter 1987; Rutherford et al. 1989) and several large species are considered threatened or vulnerable due to over-fishing (Rao et al. 1992; Saunders et al. 2014b; True, Loera & Castro 1997). Among Sciaenid species, *P. diacanthus* is vulnerable to over-exploitation because of its predictable aggregating behavior (Bowtell 1995; Bowtell 1998; Phelan 2001). *Lethrinus laticaudis* is considered an excellent eating fish and is targeted by

commercial fishers and recreational anglers across northern Australia (Coleman 2004). Although *Le. laticaudis* is considered robust to fishing pressure (Grubert, Kuhl & Penn 2010) due to its high reproductive capacity (*i.e.* serial batch spawners, high spawning frequency, high batch fecundity) (Ayvazian, Chatfield & Keay 2004), it is heavily exploited in some areas. These three fish species are of high economic value and the sustainability of the fisheries they support is potentially threatened by over harvesting and thus requires the development of suitable management programs. The development of genetic tools is necessary to further investigate their population genetics and assess stock structure.

In this study, we provide novel polymorphic microsatellite loci for the three species. We also describe a fast and cost-effective protocol for species-specific microsatellite marker discovery using genomic sequencing and multiplexing. Finally, we discuss the different NGS approaches that were used and the differences found among the three study species. This is the first report of the nuclear genomes of the three study species and provides useful baseline information for future genetic studies of these important species.

2. Materials and methods

2.1. Sample and extraction

Samples selected for the production of the microsatellite loci were derived from muscle tissue collected by the Northern Territory Department of Primary Industries and Fisheries and the Western Australian Department of Fisheries under Charles Darwin University Animal Ethics permit A13014. The *L. johnii* sample was a 210 mm male caught at 6 m depth in Darwin Harbour, Northern Territory, Australia (Middle Arm, 130°58'0.24"E, 12°39'0.97"S) in 2013. The *P. diacanthus* sample was a 890 mm male caught in Fenton patches, Northern Territory, Australia (130° 42.084'E, 12° 10.664'S) in 2013. The *Le. laticaudis* sample (WAM16-001) was a 419 mm male collected from East of the Lacepede Islands, Western Australia, Australia in 2013. Genomic DNA from *L. johnii* and *P. diacanthus* was extracted using Qiagen DNeasy Blood & Tissue columns (Qiagen, Germantown, USA) following the manufacturer's instructions. *Lethrinus laticaudis* genomic DNA was extracted using a salting-out method as described in Broderick et al. (2011). Genomic DNA from all samples for testing the loci and further genotyping was extracted using ISOLATE II Genomic DNA Kit (Bioline) following the manufacturer's instructions. This resulted in 100µL of eluted DNA for each sample. All the DNA extracts were quantified using the Qubit v3 (ThermoFisher) fluorometric ~~machine~~.

2.2. Next-Generation Sequencing and primer selection

The purified genomic DNA of *L. johnii* and *P. diacanthus* was prepared for direct shotgun sequencing using the Iron ExpressTM fragment library kit and sequenced on an IonTorrent Personal Genome Machine using an Ion318 chip (Life Technologies Corporation, Grand Island, NY). The purified genomic DNA of *Le. laticaudis* was sequenced on an Illumina® MiSeq as part of a 2x300bp run at the Australian Genome Research Facility. Because two different NGS platforms were used to scan the genomes of the three species we were able to compare their performance for microsatellite design and to assess whether equivalent results were obtainable from each platform. *Lutjanus* and *Lethrinus* genera and Sciaenidae are known to have genome size comparable to other fish species (average size for *Lutjanus* = 1066 Mb, *Lethrinus* = 1192 Mb, Sciaenidae = 753 Mb, Perciformes = 919 Mb; Gregory 2001). The paired-end reads obtained with the MiSeq run were merged using FLASH source code (Magoč &

Salzberg 2011) and their quality was checked in FastQC (Andrews 2010); the first 10 bp were trimmed in Geneious v 9.0 (Drummond et al. 2011).

From the NGS data we looked for sequences longer than 300 bp that contained a microsatellite repeat that would be suitable for primer design. These sequences were checked for microsatellite motifs and forward and reverse primers were designed using the software QDD2 beta (Meglécz et al. 2010). Sequences with target microsatellites and primers were then filtered according to the following criteria: only pure repeats were selected; all dinucleotide repeats were excluded; repeats greater than 8 were selected; loci with a predictive target sequence length above 300 bp were selected; primers with a distance less than 20bp from the repeat sequence were excluded; and the PCR primers with an ALIGN_SCORE equal or above 6 were excluded. A unique pair of primers was selected for each locus. The PCR predicted sequences for all the loci were imported into Geneious v 9.0 and blasted (MEGABLAST) against the NCBI GenBank database to check if the microsatellites fell into coding regions. Sequences that would be homologous to any other NCBI sequence likely to be functional were excluded. All the primers were blasted against their original genomic database built using the NGS reads. Only microsatellites with primers that had one hit across the whole genome were kept for further steps to ~~ensure that~~ each primer would amplify a unique sequence. The resulting microsatellites and pairs of primers were assembled *de novo* to check that each pair of primer bound to the 5' and 3' ends of a unique sequence containing a microsatellite repeat. For each species, we selected the 48 microsatellites that contained the best quality repeats with the highest number of tri- tetra- or penta-nucleotide repeats possible and with no small dinucleotide repeats between the primer and the microsatellite sequence to avoid any noise that may interfere with scoring genotypes.

Forward primers were tagged on the 5' end with the universal CAG sequence (CAGTCGGGCGTCATCA) to allow fluorescent labeling of the PCR product. Additionally, a pig-tail (GTTTCTT) was added to the reverse primers to increase the accuracy of genotyping and ensure the consistency of the amplicon size (Brownstein, Carpten & Smith 1996). The resulting 48 pairs of primers were synthetized by Integrated DNA Technologies (www.idtdna.com).

2.3. Loci and primers testing

For each species, the 48 pairs of primers were tested over a set of genomic DNA extracted from eight individuals of the target species. Amplification reactions were carried out in a 8.8 µL

volume comprising 1 µL of DNA, 4.84 µL of 2x Bioline Taq Mastermix, 4.4 pmol of forward CAG-tagged primer, 22 pmol of reverse pig-tail-tagged primer and 22 pmol of CAG primer labeled with a 6-FAM fluorescent dye. The heating cycle parameters used for amplification were 95ºC for 3 min, 37 cycles of denaturation at 94ºC for 15s, annealing for 15s at 57ºC and elongation at 72ºC for 60s. A final extension at 72ºC for 30s was performed. Post-amplification, the PCR products were diluted with water ~~one in 20~~. We added 2 µL of these diluted PCR products to 10 µL of Hi-Di formamide (ABI) and 0.05 µL of GenScan-500 LIZ (ABI) size standard. Samples were denatured at 95ºC for 3 min and sized on the ABI 3730xl capillary sequencer (Applied Biosystems, Carlsbad, California, USA) using the conditions set down by the manufacturer. Chromatograms were analysed using Geneious v 9.0 (Drummond et al. 2011).

Criteria used to select the best loci among the 48 tested for each species included the amplification success rate, peak intensity, the presence or absent of stutter peaks, the polymorphism of the loci, the number of alleles and heterozygosity. The best loci were individually tested against a further 23 samples of the target species.

2.4. Multiplex optimization of PCR

In order to reduce the cost and time of genotyping for further genetic studies, the newly designed microsatellites were combined into multiplex panels of 2 to 4 loci. The panels were set up based on the microsatellite allele-size range. The primers for all the loci of each panel were combined in a single PCR to allow the amplification of several microsatellite loci at the same time. When allele-size ranges overlapped, alternate dyes were employed to allow the discrimination of each locus on the chromatograms. Each of the four ABI dyes FAM, VIC, NED and PET were tailed with a unique ~~sequence~~: M13F (TTTCCCAGTCACGACGTTG), M13V (GCGGATAACAATTTCACACAGG), M13N (TAAAACGACGGCCAGTGC) and M13P (CACAGGAAACAGCTATGACC). The 5' end of the forward primer for the locus was synthetized with the corresponding M13 tail to allow fluorescent labeling of PCR product using a 3-primer protocol (Schuelke 2000). Several multiplex trials were conducted to find the best combination of loci with the optimal concentration of primers and PCR parameters. Primer pairs that failed to amplify within a multiplex were removed from the panels and further optimization focused on the remaining primer pairs. For each species, the multiplex trials were all evaluated

against 8 samples that were the same for those used in the 23 samples above, allowing the consistency to be checked across the templates.

2.5. Genetic variability of the loci

The loci were assayed across several individuals from a single population for each of the three species. Thirty-one *L. johnii* individuals were collected from Bathurst Island (NT, Australia); thirty-four *P. diacanthus* individuals were collected from Wyndham (WA, Australia) and twenty *Le. laticaudis* individuals were collected from Borroloola (NT, Australia). All individuals were assayed as part of the multiplex panels or PCR pooled. The multiplex PCR were comprised of 2-5 µL of DNA depending on the samples (approximately 20 ng total), 8 µL of 2x Bioline Taq Mastermix and various quantities of primers as described in Table 1. Concentrations of the different primers were adjusted to obtain homogenous PCR products revealed by similar intensity chromatogram peaks for each of the dyes within each panel. Final volumes were adjusted with water to bring the total volume to 12 µL. The heating cycle parameters, sizing of the alleles and chromatogram analyses were conducted using the same method as stated above.

The obtained datasets were statistically evaluated. The potential for null alleles, large allele dropout and stuttering to interfere with scoring accuracy was evaluated for each microsatellite locus in each sample using Microchecker v.2.2.3 (Van Oosterhout et al. 2004). The software Arlequin 3.5.2.2 (Excoffier & Lischer 2010) was used to calculate the number of alleles (A), expected (H_e) and observed (H_o) heterozygosity and conduct exact tests of conformance of genotypic proportions to Hardy-Weinberg equilibrium expectations. Estimation of probability values (P_{HW}) employed a Monte Carlo Markov Chain (MCMC) of 10^5 steps and 5.10^4 dememorization. Genotypic equilibrium between pairs of microsatellites (linkage disequilibrium) was tested in Arlequin 3.5.2.2 with 10,000 permutations.

3. Results and Discussion

The sequences of raw reads from NGS data used in this study are available in the eFish Genomic Database Repository of The University of Queensland (*P. diacanthus* DOI 10.14264/uql.2016.306; *L. johnii* DOI 10.14264/uql.2016.307; *Le. laticaudis* DOI 10.14264/uql.2016.308). The IonTorrent runs for *L. johnii* and *P. diacanthus* yielded 1,374,891

and 1,587,789 single reads whereas the Illumina® MiSeq run for *Le. laticaudis* yielded 2,800,640 paired-end reads. The IonTorrent sequencing technology allowed the production of longer reads (range 8-620 bp) compared to the Illumina MiSeq that produced 300 bp reads fixed by the method. The paired-end reads of the MiSeq run were merged to increase their length to 300-575 bp and allow the detection of at least 300 bp length sequences that may contain a microsatellite locus. This resulted in 1,169,198 reads, which is less than what was used to select microsatellite loci in the first two species. The MiSeq run produced higher quality sequences than the IonTorrent (Phred score: 36 vs 29). Quality profiles along the IonTorrent reads showed that the quality of the sequencing decreased with length meaning that the end of the longer reads (> 325 bp) had a lower quality then at their start. QDD Pipe1 detected between 110,000 and 170,000 sequences containing a microsatellite sequence depending on the species (Table 2). This number was independent of the type of NGS platform used. From those sequences QDD Pipe2 removed the low complexity sequences (no BLAST to itself), putative minisatellites (short sequences of repeated nucleotides) and sequences that had BLAST hit to other sequences to only keep the singletons and unique consensus sequences. QDD Pipe3 designed primers for all QDD Pipe2 output reads. The resulting number of sequences that contained a microsatellite sequence and the corresponding primers were given in the final output of QDD pipeline, and varied between 20,000 and 30,000 depending on the species (Table 2). After applying the filtering criteria described previously, ~~97potentially~~ amplifiable microsatellite reads were found for *P. diacanthus*, 121 for *L. johnii* and 103 for *L. laticaudis*. From those microsatellite reads, we selected the ones with the smallest number of repeats and eliminated those with small repeats between the primer and the microsatellite to reach 48 microsatellite loci per species being ultimately tested in the laboratory.

The testing of 144 primer pairs resulted in the selection of 34 polymorphic loci that could be reliably scored and showed consistent amplification success. We selected a final set of 10 loci for *L. johnii*, 11 loci for *P. diacanthus* and 13 loci for *Le. laticaudis* (Table 3). Multiplex panels of microsatellites were developed for the two species *L. johnii* and *P. diacanthus*, and the optimization of each panel resulted in the efficient assay and unambiguous scoring of microsatellites in the two species. The 13 loci for *Le. laticaudis* did not amplify successfully as part of PCR multiplexes using the M13 labeling system. For this species, the loci were all

amplified in individual PCR with the CAG labeling system as described above. The resulting PCR products were then pooled according to the panels described in Table 1 before the ABI run.

Genotypes from 10 microsatellites were obtained from multiplexed PCR for 31 individuals of *L. johnii*. There was no missing data and the number of alleles for each locus varied between 3 and 21 (Table 3). Microchecker indicated there was no evidence of null-alleles for these loci. There were only 2/45 significant tests for linkage disequilibrium between pairs of loci (*Luj076* x *Luj082* and *Luj068* x *Luj072*) and overall deviations from Hardy-Weinberg equilibrium (HWE) were detected at a single locus *Luj094* (p-value=0.048). Heterozygosity was moderate to high for all loci (mean overall loci 0.706 +/- 0.182) and generally similar to expectations (around 0.7 for marine fish, DeWoody & Avise 2000). Genotypes from 11 microsatellites were obtained from multiplexed PCR for 34 individuals of *P. diacanthus*. There was 1.07 % missing data and the number of alleles for each locus varied between 4 and 15 (Table 3). Microchecker indicated there was no evidence of null-alleles for these loci. There were only 2/55 significant tests for linkage disequilibrium between pairs of loci (*Prd020* x *Prd045* and *Prd020* x *Prd049*) and overall deviations from Hardy-Weinberg equilibrium (HWE) were detected at a single locus *Prd045* (p-value=0.046). Heterozygosity was variable and with an overall mean lower than for *L. johnii* (0.685 +/- 0.228). Genotypes from 13 microsatellites were obtained from pooled post PCR products for 20 individuals of *Le. laticaudis*. There was 2.7 % missing data and the number of alleles for each locus varied between 4 and 13 (Table 3). Microchecker indicated there was no evidence of null-alleles for these loci. There were only 2/78 significant tests for linkage disequilibrium between pairs of loci (*Lel012* x *Lel027*) and overall deviations from Hardy-Weinberg equilibrium (HWE) were detected at a single locus *Lel039* (p-value=0.009). Heterozygosity was high and with an overall mean lower than for the two other species (0.830 +/- 0.109).

4. Conclusions

In conclusion, we applied the direct sequencing of a genomic library approach to develop microsatellite loci and it resulted in a significant reduction in laboratory effort compared to traditional protocols for microsatellite discovery (cloning and Sanger sequencing). Merged paired-end reads from the MiSeq platform demonstrated higher quality of reads than the IonTorrent single reads. From the 1-1.5 million raw reads, we selected a reduced number of loci

266 to test (48) and successfully amplified satisfactory polymorphic loci for 10 to 13 of them
267 depending on the species. However, the NGS data revealed the potential for hundreds to
268 thousands of potentially amplifiable microsatellites to be discovered. The microsatellites
269 characterized in this work will be available to explore the population genetics and stock structure
270 of these highly valuable species.
271
272 **Acknowledgements**
273 The authors thank staff from the Western Australian Department of Fisheries and the Northern
274 Territory Department of Primary Industries as well as Indigenous rangers for sample collection.
275 We extend our thanks to Samuel Williams at the Molecular Fisheries laboratory and Sean Corley
276 at the Animal Genetics Laboratory for their assistance with laboratory work.
277
278 **References**
279
280 Abdelkrim J, Robertson BC, Stanton J-AL, and Gemmell NJ. 2009. Fast, cost-effective
281 development of species-specific microsatellite markers by genomic sequencing. *Biotechniques*
282 46:185.
283 Allen GR, Swainston R, and Ruse J. 1997. *Marine fishes of tropical Australia and south-east*
284 *Asia*. Perth: Western Australian Museum.
285 Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data. *Reference*
286 *Source*.
287 Ayvazian S, Chatfield B, and Keay I. 2004. The age, growth, reproductive biology and stock
288 assessment of grass emperor, *Lethrinus laticaudis* in Shark Bay, Western Australia. Department
289 of Fisheries, Research Division, Westrern Australian Marine Research Laboratories.
290 Bowtell B. 1995. Heed jewfish warnings. *Fish and Boat July*.
291 Bowtell B. 1998. Huge schools of black jew about. *Fish and Boat July*.
292 Broderick D, Ovenden J, Buckworth R, Newman S, Lester R, and Welch D. 2011. Genetic
293 population structure of grey mackerel *Scomberomorus semifasciatus* in northern Australia.
294 *Journal of Fish Biology* 79:633-661.
295 Brownstein MJ, Carpten JD, and Smith JR. 1996. Modulation of non-templated nucleotide
296 addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques*
297 20:1004-1006.
298 Castoe TA, Poole AW, Gu W, Jason de Koning A, Daza JM, Smith EN, and Pollock DD. 2010.
299 Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite
300 loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources*
301 10:341-347.
302 Coleman APM. 2004. The national recreational fishing survey: the Northern Territory.
303 Department of Business Industry and Resource Development. Fishery Report 72.
304 DeWoody J, and Avise J. 2000. Microsatellite variation in marine, freshwater and anadromous
305 fishes compared with other animals. *Journal of Fish Biology* 56:461-473.

306 Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse
307 M, and Markowitz S. 2011. Geneious v5. 4.
308 Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform
309 population genetics analyses under Linux and Windows. *Molecular ecology resources* 10:564-
310 567.
311 Gardner MG, Fitch AJ, Bertozzi T, and Lowe AJ. 2011. Rise of the machines–recommendations
312 for ecologists when using next generation sequencing for microsatellite development. *Molecular*
313 *Ecology Resources* 11:1093-1101.
314 Gregory TR. 2001. Animal genome size database. *Available at* http://www.genomesize.com.
315 Grubert M, Saunders T, Martin JM, Lee H, and Walters CJ. 2013. Stock Assessments of Selected
316 Northern Territory Fishes. Northen Territory Government. Fishery Report 110.
317 Grubert MA, Kuhl PJ, and Penn JW. 2010. Ecological risk assessment. Northern Territory
318 coastal line fishery. Norther Territory Department of Resources. Fishery Report 103.
319 Lenanton R, and Potter I. 1987. Contribution of estuaries to commercial fisheries in temperate
320 Western Australia and the concept of estuarine dependence. *Estuaries* 10:28-35.
321 Magoč T, and Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve
322 genome assemblies. *Bioinformatics* 27:2957-2963.
323 Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, and Martin J-F. 2010. QDD: a
324 user-friendly program to select microsatellite markers and design primers from large sequencing
325 projects. *Bioinformatics* 26:403-404.
326 Peters MB, Ovenden JR, Broderick D, Lance SL, Hagen C, and Glenn TC. 2009. Fifteen
327 microsatellite loci for the jungle perch, *Kuhlia rupestris*. *Molecular Ecology Resources* 9:1467-
328 1469.
329 Phelan MJ. 2001. Sciaenid aggregations in northern Australia: an example of successful
330 outcomes through collaborative research. *Fisheries Centre Research Reports*:100.
331 Rao TA, Mohan RL, Chakraborty S, Murty VS, Nair KS, Vivekanandan E, and Raje S. 1992.
332 Stock assessment of sciaenid resources of India. *Indian Journal of Fisheries* 39:85-103.
333 Rutherford ES, Tilmant JT, Thue EB, and Schmidt TW. 1989. Fishery harvest and population
334 dynamics of spotted seatrout, *Cynoscion nebulosus,* in Florida Bay and adjacent waters. *Bulletin*
335 *of Marine Science* 44:108-125.
336 Saunders T, Newman S, Keag M, and Errity C. 2014a. Golden Snapper *Lutjanus johnii*. Status of
337 key Australian fish stocks reports 2014. Fisheries Research and Development Corporation.
338 Canberra. 523-529.
339 Saunders T, Roelofs A, Newman S, and Errity C. 2014b. Black jewfish *Protonibea diacanthus*.
340 Status of key Australian fish stocks reports 2014. Fisheries Research and Development
341 Corporation. Canberra. 344-350.
342 Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature*
343 *biotechnology* 18:233-234.
344 True CD, Loera AS, and Castro NC. 1997. Technical Notes: Acquisition of Broodstock of
345 Totoaba macdonaldi: Field Handling, Decompression, and Prophylaxis of an Endangered
346 Species. *The Progressive fish-culturist* 59:246-248.
347 Van Oosterhout C, Hutchinson WF, Wills DP, and Shipley P. 2004. MICRO‐CHECKER:
348 software for identifying and correcting genotyping errors in microsatellite data. *Molecular*
349 *Ecology Notes* 4:535-538.

Table 1(on next page)

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 *Le.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

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

primer mixes of each panel are mixed in the proportion given in the table with water added to reach 4 µL. – ~~for not applicable~~.

Table 2(on next page)

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

1

Table 3 (on next page)

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	31	161-176	6	0.355	0.414	0.431
		R: GGCTCTGAACCTGGGAGATT								
	Luj094	**F: TCTCAGAGGGTTTGATGCAG (FAM)**	**(AATC)9**	**KX387437**	**31**	**227-239**	**4**	**0.548**	**0.507**	**0.043**
		R: CTTTGGCGCTTTCTATCAGC								
	Luj076	F: CGGGTCGAGTCTGTTTGTGT (VIC)	(AAG)15	KX387436	31	203-230	9	0.774	0.832	0.768
		R: CTTCAGACGGATTAGCAGCA								
	Luj068	F: CCTAGGGTGTCAGTGTCAGTCA (PET)	(AAAG)20	KX387435	31	174-258	17	0.968	0.933	0.106
		R: TGCCTGTATGTTCTCTTGAGC								
	Luj090	F: ATCCTAATGCATCGTGCTTG (FAM)	(AGC)17	KX387444	31	194-257	21	0.903	0.942	0.178
		R: GGCATGTTCTATTGAGGTTGG								
	Luj051	F: TGCAGAGCAACAGAACAACAC (VIC)	(ACTG)10	KX387440	31	172-188	5	0.548	0.604	0.797
		R: CACCTTGCGTTTGCAGTCT								
	Luj114	F: CCATAACTGCTGTTCTGTATCTGG (VIC)	(AGC)9	KX387442	31	282-314	10	0.774	0.806	0.186
		R: AATACGGCAGATCTCGGGTT								
	Luj091	F: TCATTCCCAGGAGCTCAAAT (PET)	(ACAG)12	KX387438	31	219-267	8	0.645	0.834	0.105
		R: AATCGTCACTTTCGACCCAC								
	Luj072	F: ACTCGAAGAACACAGCCCAC (FAM)	(AGC)9	KX387443	31	195-204	3	0.742	0.662	0.725
		R: CACATTTGAATCCTTGCTGG								
	Luj082	F: AAGTACATCGGAGGGCTGAG (VIC)	(ACGAT)12	KX387439	31	220-260	9	0.806	0.858	0.308
		R: TGTTATCAAAGTTCACCGATACAAA								
Protonibea diacanthus	*Prd044*	F: ACAAAGTTTCCTCCTCTGGC (FAM)	(AAG)13	KX387452	34	181-217	8	0.735	0.757	0.444
		R: CACGTTCCATCTTTATTTATTTGC								
	Prd023	F: TCGTGTGAACACTTTGATGC (VIC)	(ATC)11	KX387448	34	292-316	8	0.794	0.834	0.912
		R: CTCGTCTCTGCTCTTGGTCC								
	Prd042	F: TACCTTTGAGATGCGAGCG (NED)	(AGC)12	KX387451	34	233-245	5	0.824	0.687	0.493

		R: GTCAAAGCCATCAATCCAGC								
	Prd012	F: AGGCTGTTTGAACTGCAGGG (PET)	(AAAG)20	KX387445	32	199-263	15	0.906	0.913	0.749
		R: CATGCTGAGCAATATGTGGG								
	Prd046	F: TCATCCTGAGTTTGTGCTGG (FAM)	(AGC)9	KX387454	34	224-236	5	0.324	0.384	0.068
		R: CATGAGTAAGCAGAGCGTGG								
	Prd018	F: ATGAACGGCATCAGTCAGC (NED)	(ACAG)9	KX387446	34	179-207	8	0.971	0.803	0.763
		R: CGTCTGATAAACAGCACTGCC								
	Prd020	F: CAATGTTCTGCAAGAGCTGC (PET)	(ATC)11	KX387447	33	189-216	8	0.970	0.787	0.176
		R: TCAAATGTCAAAGTCCAGTCC								
	Prd045	F: GTCTATCCATGTTCCAGCCC (PET)	(ATCC)11	KX387453	33	283-299	5	0.515	0.582	0.053
		R: TCATCCCAAAGTGACCAACC								
	Prd049	F: CCTTGTCCTCCTTTCAGGC (FAM)	(ACC)9	KX387455	34	216-234	4	0.118	0.115	1.000
		R: GGGTCATTAAACATGGCAGC								
	Prd036	F: TCACGTGAAGCGTCTACAGC (VIC)	(AAG)12	KX387450	34	227-257	10	0.765	0.694	0.759
		R: AAAGGAGGAAACACAGAGCC								
	Prd024	**F: AGAGTGTCCGAGTCCAGAGG (PET)**	**(AAG)11**	**KX387449**	**34**	**199-217**	**6**	**0.618**	**0.701**	**0.047**
		R: CAGTACCTGGTGATGGGAGC								
Lethrinus laticaudis	*Lel011*	F: CTGTCGGAGGTAAAGTGCG (FAM)	(AGC)9	KX387422	19	238-265	4	0.684	0.639	0.403
		R: CTCATGGTGTTGAGGATGGG								
	Lel040	F: TGGTTGCAGACAACTGCC (VIC)	(AGC)9	KX387431	20	173-209	10	0.800	0.738	0.990
		R: CTTAAGAGCAGTGATCCAGGC								
	Lel033	F: AGTGCGACAAAGAAATGGC (NED)	(AGAT)16	KX387428	19	175-231	13	0.947	0.915	0.851
		R: CATTTGTCAGTTATGAAACTTGGC								
	Lel012	F: GCGAGGGTCTGCTACTATAGGG (PET)	(AAT)9	KX387423	20	248-320	12	0.800	0.850	0.234
		R: TGTAAAGTGTAAACCACGTCCC								
	Lel013	F: CCTGAACCTGGAGAACTCGG (FAM)	(ATC)12	KX387424	19	251-263	5	0.632	0.789	0.189
		R: ACTGAGGGAGGAGATAAAGGG								

Lel041	F:	CTGCTGTTCTGGGTTGCC (VIC)	(AAT)19	KX387432	19	144-190	12	0.842	0.889	0.082
	R:	CAACAAGCTGTTGGTGTCCC								
Lel032	F:	AAATCTGCATTATGAAATTGGC (NED)	(AAAG)16	KX387427	20	173-222	10	0.950	0.829	0.409
	R:	CAGCTCCTTGAGTTTAGTCCC								
Lel028	F:	CAGTAGCTTTAATAGTTAGGCACCC (PET)	(AAAG)13	KX387426	20	198-230	8	0.850	0.868	0.948
	R:	GGCTGTCCAGAGTGAGGC								
Lel047	F:	AAAGAATGGGAAGAATGACCC (FAM)	(AGAT)11	KX387434	20	151-195	10	0.950	0.879	0.923
	R:	AAGCCAAGTGATTAAGAAACCC								
Lel027	F:	CACTAAGGGTCCATGTTGCC (FAM)	(AAT)22	KX387425	17	197-236	13	0.941	0.886	0.983
	R:	TCTGTAATGAATGATCAAACCG								
Lel044	F:	TTCTACTTGACCCTGGTAGGC (VIC)	(ATCC)11	KX387433	20	163-187	7	0.750	0.791	0.715
	R:	AATGTAATGCCATAAGCGGG								
Lel036	F:	TCCAATTTACACCAAACTAGGC (NED)	(AAAG)15	KX387429	20	163-215	13	0.850	0.921	0.497
	R:	CCGGAATGATCTGCAGGC								
Lel039	**F:**	**CTTGTAGAGTGTCAACGAGGG (PET)**	**(AAT)11**	**KX387430**	**20**	**184-204**	**6**	**0.650**	**0.767**	**0.009**
	R:	CATGATGCAATAACCATCCC								

Loci in bold are those that depart significantly from HWE expectation before any correction.