Development and characterization of 24 polymorphic microsatellites *loci* for the freshwater fish *Ichthyoelephas longirostris* (Pisces: Characiformes) (#9242)

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 / 6 Mar 2016

Files	1 Table file(s) 1 Raw data file(s) Please visit the overview page to **download and review** the files not included in this review pdf.
Declarations	**Involves vertebrate animals.**

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/**

Development and characterization of 24 polymorphic microsatellites *loci* for the freshwater fish *Ichthyoelephas longirostris* (Pisces: Characiformes)

Ricardo M Landínez-García, Edna J Márquez

Ichthyoelephas longirostris (Characiformes: Prochilodontidae) is a Colombian endemic freshwater fish commercially exploited, whose genetics is unknown. This study developed by Illumina sequencing, a set of 24 highly polymorphic microsatellite *loci* for *I. longirostris* (number of alleles per *locus*: 4 - 18, observed heterozygosity: 0.250 - 0.857, expected heterozygosity: 0.401 - 0.935). This is the first report of highly polymorphic molecular markers for analysis of population genetics necessary in the management and conservation of this species.

delete s (microsatellite)

Development and characterization of 24 polymorphic microsatellites *loci* for the freshwater fish *Ichthyoelephas longirostris* (Pisces: Characiformes)

delete Pisces, since fish is already in the title

Ricardo M. Landínez-García[1] and Edna. J. Márquez[1]

[1]Universidad Nacional de Colombia-Sede Medellín-Facultad de Ciencias-Escuela de Biociencias. Antioquia, Colombia.

[1] Author correspondence : Edna. J. Márquez, Calle 59A No 63 – 20 Bloque 19 A Laboratorio 310, Medellín, 050034 Colombia, ejmarque@unal.edu.co; ejmarque@gmail.com.

Abstract

Ichthyoelephas longirostris (Characiformes: Prochilodontidae) is a Colombian endemic freshwater fish commercially exploited, whose genetics is unknown. This study developed by Illumina sequencing, a set of 24 highly polymorphic microsatellite *loci* for *I. longirostris* (number of alleles per *locus*: 4 - 18, observed heterozygosity: 0.250 - 0.857, expected heterozygosity: 0.401 - 0.935). This is the first report of highly polymorphic molecular markers for analysis of population genetics necessary in the management and conservation of this species.

rewrite sentence

change this (Illumina is a name linked to a brand, not proper here)

Subjects: Biodiversity, Conservation Biology, Genetics, Molecular Biology

changed it by large-scale sequencing or next-gen

Keywords; Connectivity, Illumina sequencing, molecular marker, Prochilodontidae

Change this, connectivity is not matter of this paper

Introduction

Ichthyoelephas longirostris (Steindachner 1879), is a Colombian endemic fish of importance in the commercial and subsistence fisheries. However, the population genetics of *I. longirostris* remain unknown, althought this information is necessary to develop management and conservation policies for this species. The polymorphic misrosatellites has been developed in members of Prochilodontidae family belonging to genera *Prochilodus* (Carvalho-Costa, Hatanaka & Galetti Jr., 2006; Barbosa et al., 2008; Rueda et al., 2011) and *Semaprochilodus* (Passos et al., 2010). Nevertheless, this information is complete absent for congeners of *I. longirostris* limiting our ability to resolve fine-scale differentiation and subdivision patterns among populations.

Although the implementation of heterologous *loci* (cross-amplification) is commonly used to study non-model species, this practice present problems such as size homoplasy, negative association with source-target species genetic distance, lowest level of polymorphism, null alleles, broken repeat motifs and amplification of non-orthologous *loci* (Primmer et al., 2005; Barbará et al., 2007; Yue, Balazs & Laszlo, 2010). Consequently, it is highly recommend to develop species-specific molecular markers, which are currently facilitated by next generation sequencing techologies (Castoe et al., 2010; Ekblom & Galindo, 2011). Thus, this study developed *de novo* molecular markers for future populations genetics studies of *I. longirostris*.

Materials & Methods

In this work, superficial genome sequencing was carried out with the Illumina MiSeq v2 instrument using the Nextera library preparation kits. This sequencing process generated paired-end reads of 250 bases that were cleaned using Prinseq-lite v0.20.4 (Schmieder & Edwards, 2011) to eliminate low quality regions at both ends and remove the reads lesser than 50 bases or duplicated. The genome assembly of reads was performed with Abyss v1.3.5 (Simpson et al., 2009) using a kmer 64 and the contigs were analysed with the program PAL_FINDER v.0.02.03 (Castoe et al., 2010) to extract those that contained perfect tri-, tetra- and pentanucleotide microsatellites. The primer-pairs for microsatellite *loci* amplification were designed from their flanking sequences by using the program Primer3 v.2.0 (Rozen & Skaletsky, 2000). Additionally, the potential amplifiable *loci* were submitted to electronic PCRs (Rotmistrovsky, Jang & Schuler, 2004) for verifying *in silico* the correct primer alignment.

A total of 40 microsatellites *loci* were selected for optimization and polymorphism analysis in 15 preserved tissues of *I. longirostris*. For this goal, DNA extraction from samples was performed with the commercial kit GeneJET DNA purification (Thermo Scientific), standard conditions were used for PCRs and the amplicons were electrophoresed and visualized on silver-stained 10% polyacrylamide gel. Polymorphic *loci* were selected based on criteria of amplification in all samples, band resolution, specificity, size (from 100 to 400 pb) and ability to detect heterozygotes in the different samples analysed. Then, a set of 24 polymorphic microsatellite *loci* amplified consistently were selected for further genotyping of 28 preserved tissues of *I.*

66 *longirostris* provided by Integral S.A. (Scientific cooperation agreement between Universidad
67 Nacional de Colombia and Integral S.A., on 19th September 2013).

68

69 PCR reactions containing 1×buffer (Invitrogen), 2-4 ng/µl of template DNA, 2.5% formamide
70 (Sigma) 0.35 pmoles/µl labelled forward primer (either FAM6, VIC, NED or PET, Applied
71 Biosystems), 0.5 pmoles/µl reverse primer (Macrogen), 0.2 mM dNTPs (Thermo Scientific),
72 0.05 U/µl Platinum™ Taq DNA Polymerase (Invitrogen) and 2.3 mM $MgCl_2$. The PCR
73 amplifications were performed on a thermocycler T100 (BioRad) with an initial denaturation
74 step of 95 °C for 3 min, followed by 32 cycles consisting of a denaturation step of 90 °C for 22 s
75 and an annealing step of 59 °C for 16 s. The extension step and a final elongation were absent in
76 this thermal profile. Finally, the PCR products were submitted to electrophoresis on an
77 automated sequencer ABI 3730 XL (Applied Biosystems) using LIZ500 (Applied Biosystems)
78 as internal molecular size. Allelic fragments were denoted according their molecular size and
79 scored using GeneMapper 4.0 (Applied Biosystems).

80

81

82 Tests for Hardy–Weinberg and linkage equilibria and the estimation of the observed (H_O) and
83 expected (H_E) heterozygosities were performed using Arlequin v.3.5.1.3 (Excoffier, Laval &
84 Schneider, 2005). Statistical significance in multiple comparison was adjusted applying the
85 sequential Bonferroni correction (Rice, 1989). The software GenAlEx v.6.501 (Peakall &
86 Smouse, 2006) was used to estimate the average number of alleles per *locus*. Potential
87 genotyping errors were evaluated by using Micro-Checker v.2.2.3 software (Van Oosterhout et

al., 2004). The polymorphism information content (PIC) for each marker was determined using the program PICcalc (Nagy et al., 2012) .

Results and Discussion

All 24 polymorphic microsatellite *loci* selected showed clearly defined peaks and the absence of stutter bands in the electropherogram (Table I). The number of alleles across these *loci* ranged from 4 to 18, with an average number of 8,5 alleles/*locus* and average observed heterozygosity (Ho) de 0.669. This results are concordant with those found in *Prochilodus argenteus* collected in reproductive and non-reproductive periods in the river Sao Francisco in Brasil (Hatanaka, Henrique-Silva & Galetti, 2006; Sanches et al., 2012). However, the level of observed heterozygosity was upper than that observed in *P. argenteus* in both reproductive periods (Ho: 0.100-0.351, Hatanaka *et al.,* 2006; Ho: 0.630-0.659, Sanches *et al.,* 2012).

8.5
It does not need to be concordant, although it could be in similar level
These
Brazil
higher
?
it does not appear higher

Additionally, allelic frequencies of 20 *loci* were concordant with Hardy-Weinberg and linkage equilibria after sequential Bonferroni correction and no evidence of null alleles or scoring errors were detected by Micro-checker. Moreover, the PIC values ranged from 0.375 to 0.871 (average: 0,733) indicating that these markers are highly informative (Botstein et al., 1980). Thus, these microsatellite *loci* are strongly recommended for future studies of diversity and population genetics of *I. longirostris*.

0.733

Acknowledgements

This research was supported by the scientific cooperation agreement between Universidad Nacional de Colombia and Integral S.A., on 19th September 2013 and Universidad Nacional de Colombia, Sede Medellín and Empresas Públicas de Medellín, Grant CT-2013-002443 "Variación genotípica y fenotípica de poblaciones de especies reófilas presentes en el área de influencia del proyecto hidroeléctrico Ituango." The authors thank to the Centro Nacional de Secuenciación Genómica, Universidad de Antioquia (Medellín, Colombia) for assistance in Bioinformatics analysis.

References

Barbará T., Palma-Silva C., Paggi GM., Bered F., Fay MF., Lexer C. 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16:3759–3767.

Barbosa ACDR., Galzerani F., Correa TC., Jr PMG., Hatanaka T. 2008. Description of novel microsatellite *loci* in the neotropical fish *Prochilodus argenteus* and cross-amplification in *P . costatus* and *P . lineatus*. *Genetics and Molecular Biology* 1:357–360.

Botstein D., White LR., Skolnick M., Davis RW. 1980. Construction of a genetic link-age map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314–331.

Carvalho-Costa LF., Hatanaka T., Galetti Jr. PM. 2006. Isolation and characterization of polymorphic microsatellite markers in the migratory freshwater fish *Prochilodus costatus*. *Molecular Ecology Notes* 6:818–819.

Castoe TA., Poole AW., Gu W., Jason de Koning AP., Daza JM., Smith EN., 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.

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

Excoffier L., Laval G., Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1:47–50.

Hatanaka T., Henrique-Silva F., Galetti PM. 2006. Population substructuring in a migratory freshwater fish *Prochilodus argenteus* (Characiformes, Prochilodontidae) from the São

144 Francisco River. *Genetica* 126:153–159.

145 Nagy S., Poczai P., Cernák I., Gorji AM., Hegedűs G., Taller J. 2012. PICcalc: An online
146 program to calculate polymorphic information content for molecular genetic studies.
147 *Biochemical Genetics* 50:670–672.

148 Van Oosterhout C., Hutchinson WF., Wills DPM., Shipley P. 2004. Micro-Checker: Software for
149 identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology*
150 *Notes* 4:535−538.

151 Passos KB., Alencar Leão AS., Oliveira DP., Farias IP., Hrbek T. 2010. Polymorphic
152 microsatellite markers for the overexploited Amazonian fish, *Semaprochilodus insignis*
153 (Jardine and Schomburgk 1841). *Conservation Genetics Resources* 2:231–234.

154 Peakall R., Smouse PE. 2006. Genalex 6: Genetic analysis in Excel, population genetic software
155 for teaching and research. *Molecular Ecology Notes* 6:288−295.

156 Primmer RC., Painter NJ., Koskinen TM., Palo UJ., Merilä J. 2005. Factors affecting avian
157 cross-species microsatellite amplification. *Journal of Avian Biology* 36:348–360.

158 Rice WR. 1989. Analyzing tables of statistical tests. *Evolution* 43:223–225.

159 Rotmistrovsky K., Jang W., Schuler GD. 2004. A web server for performing electronic PCR.
160 *Nucleic Acids Research* 32 :W108–W112.

161 Rozen S., Skaletsky H. 2000. Primer3 on the World Wide Web for general users and for
162 biologist programmers. In: Krawetz S, Misener S eds. *Bioinformatics Methods and*
163 *Protocols: Methods in Molecular Biology*. New Jersey: Humana Press, 365–386.

164 Rueda EC., Sommer J., Scarabotti P., Markariani R., Ortí G. 2011. Isolation and characterization
165 of polymorphic microsatellite *loci* in the migratory freshwater fish *Prochilodus lineatus*
166 (Characiformes: Prochilodontidae). *Conservation Genetics Resources* 3:681–684.

167 Sanches A., Galetti Jr. PM., Galzerani F., Derazo J., Cutilak Bianchi B., Hatanaka T. 2012.
168 Genetic population structure of two migratory freshwater fish species (*Brycon orthotaenia*
169 and *Prochilodus argenteus*) from the Sao Francisco River in Brazil and its significance for
170 conservation. *Latin American Journal of Aquatic Research* 40:177–186.

171 Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets.
172 *Bioinformatics* 27:863–864.

173 Simpson JT., Wong K., Jackman SD., Schein JE., Jones SJ., Birol I. 2009. ABySS: A parallel
174 assembler for short read sequence data genome research. *Genome Research* 19:1117–1123.

175 Yue G-H., Balazs K., Laszlo O. 2010. A new problem with cross-species amplification of
176 microsatellites: Generation of non-homologous products. *Zoological Research* 31:131–40.

Table 1(on next page)

Table 1

Primer sequences and characteristics of 24 polymorphic microsatellite *loci* identified in *Ichthyoelephas longirostris*. H_O and H_E: observed and expected heterozygosity estimated from 28 individuals, respectively; PIC: polymorphic information content; P_{HW}: statistical significance of Hardy–Weinberg equilibrium.

1

Name *locus*	Primer sequence for forward (F) and reverse (R) (5′−3′)	Repeat motif	Number of alelles [size range (bp.)]	H_O	H_E	PIC	P_{HW}
Ilo01	F: TGCATCTGAGCTGATGGAGG R: AGTCTCTCTGCAGGTTGGGG	(AAAAC)n	8 (205-240)	0.857	0.828	0.789	0.326
Ilo03	F: CAGATGCAGCTGAACACGG R: TTGTAAACTGGCAGTGTGTTAAACC	(AAAC)n	5 (170-190)	0.571	0.509	0.411	0.968
Ilo04	F: GAAGCTGGCGAATAGAAGGC R: TGACCTACTGTGAAACTGGGG	(AAAC)n	7 (170-202)	0.526	0.605	0.555	0.125
Ilo05	F:GAAGGAACTGAGGTGCAGGG R:CACATCTCCCTCTGTATCCCC	$(AAAG)_n$	9 (149-189)	0.773	0.791	0.773	0.886
Ilo06	F:TCCGTTGATGTAACAACATTAGCC R:GCTCCCTGTGCTCTTCTGC	$(AAAG)_n$	10 (204-252)	0.741	0.848	0.832	0.420
Ilo08	F: GGTTGGGAGTGCCAGATAGG R: AGTGCAGTGCTCAGTCCAGC	(AAAG)n	8 (190-226)	0.679	0.742	0.702	0.218
Ilo09	F: ATGTTTGTGGCATCACCAGG	(AAATC)n	9 (256-301)	0.821	0.799	0.755	**0.022**

	R: CTGGCAGTGCTACCTCAACC						
Ilo10	F: TACGACAGCTGACTGACCCG	(AAC)n	8 (201-225)	0.714	0.808	0.763	0.693
	R: CCCCTAAGAGACAACCGACC						
Ilo11	F: TGTCGTGTCATGTTGTGTCG	(AACAT)n	5 (243-268)	0.308	0.609	0.548	**0.002**
	R: CCCTGTACATGTCCTTCAGAGC						
Ilo12	F:TTGGACCAGATGTGTTTGCC	$(AACG)_n$	4 (170-186)	0.571	0.689	0.677	0.455
	R:TCCTCAGGCATCCTACTGCC						
Ilo15	F: CATAGTAGTGTCATACAACACCTGTGC	(AATG)n	8 (164-200)	0.714	0.838	0.799	**0.013**
	R: TCATTAACCCGTTTGGTGAGG						
Ilo16	F: AGTGTGCGGGGTTAAACTGC	(AATG)n	8 (172-200)	0.630	0.661	0.602	0.861
	R: CCTGCGGTAGACTGGTAATCC						
Ilo17	F: GCAGATGCTTTGGAGTTCCC	(AATG)n	10 (256-304)	0.857	0.866	0.835	0.051
	R: TGGCATGATTATCAATGGGC						
Ilo18	F: ATAACTCTGCACTTCGGGGC	(AATG)n	5 (261-277)	0.393	0.401	0.375	0.561
	R: ATCTAAACCGCATGTGAGCC						
Ilo20	F:ATTTTCACTCGTCGAAGCCC	$(AGGCT)_n$	8 (168-203)	0.714	0.762	0.749	0.210

	R:TGATGTAAACCACAGGCACG						
Ilo21	F:TCCATAACTTGTTTTGCTGCG	$(AGT)_n$	18 (210-276)	0.75	0.886	0.871	0.232
	R:AATCTATAGTCTGAGAGCAACGGC						
Ilo22	F:AAAACAATGCGCTGAATGC	$(ATAC)_n$	4 (266-286)	0.536	0.647	0.636	0.227
	R:ATGTGTACGTGTATATATGCTGGC						
Ilo23	F:CCAAACTGCTCATTCTGGAGG	$(ATAC)_n$	10 (223-259)	0.857	0.881	0.865	0.680
	R:TGGGACGCTTCTTTAGCTCC						
Ilo24	F:ACTGCACACTTGAGATCTGGG	$(ATCT)_n$	10 (166-214)	0.75	0.86	0.844	0.311
	R:GGTACGTTAGCCAAACAGACTGG						
Ilo26	F:TTAAGAGCTCAGAGCGTGCG	$(ATCT)_n$	11 (109-149)	0.815	0.853	0.837	0.137
	R:TGTTTAGCAACTTATTTATGACCTATGACC						
Ilo29	F:ATCTATCTGACAGACTATCTGTTTATTCC	$(ATCT)_n$	8 (243-271)	0.667	0.861	0.845	**0.071**
	R:GAAGCACTCAGAGACAGACAGG						
Ilo35	F: GGATACCCTAAATTTCCTTTGGG	(TCCG)n	11 (268-352)	0.250	0.935	0.871	**0.000**
	R: GCATCACAGCGTCAAGAACC						
Ilo37	F: CACACAAACACTCATCTTAAAAGTCTCC	(TCTG)n	12 (99-151)	0.821	0.885	0.856	0.241

	R: GACCTGCGGAAAGAGAATGG						
Ilo40	F: CAGAGTTTTGGCCGTGAGG	(TTC)n	8 (137-161)	0.750	0.833	0.795	0.218
	R: CAGGGAGGAGTAGTGTCGGG						

2