Editorial Manager(tm) for Journal of Fish Biology
Manuscript Draft

Manuscript Number: MSM 10-260R2

Title: EXPLORING THE EFFECT OF MICROSATELLITE SIZE HOMOPLASY ON RECONSTRUCTION OF PHYLOGENETIC RELATIONSHIPS OF PICOTE SPLITFIN Zoogoneticus quitzeoensis.

Short Title: Homoplasies in Zoogoneticus microsatellites

Article Type: Brief Communication

Keywords: Microsatellites; size homoplasies; phylogenetic reconstructions

Corresponding Author: LUIS BOTO, PhD

Corresponding Author's Institution:

First Author: LUIS BOTO, PhD

Order of Authors: LUIS BOTO, PhD;Omar Domínguez-Domínguez, PhD;Ignacio Doadrio

Abstract: This study explores the effects of microsatellite size homoplasies in the reconstruction of phylogenetic relationships and estimates of population parameters as Fixation index (FST) using as a case study a truncated microsatellite from the picote splitfin Zoogoneticus quitzeoensis. The results suggest that the use of imperfect microsatellites may have only a minor effect in phylogenetic and population studies.

Ethics questionnaire for JFB

This questionnaire relates to the Editorial published in JFB **68**, 1-2, which you have been asked to read. Please note that submitted manuscripts will only be considered if the experimental methods employed are ethically justified. PLEASE SUBMIT THE COMPLETED QUESTIONNAIRE WITH YOUR MANUSCRIPT ONLINE THROUGH EDITORIAL MANAGER.

Corresponding author's name: _____LUIS BOTO_____

Question 1: If the fishes have been collected as part of faunal surveys, have the fishes, where feasible, been killed rapidly or returned to the wild after being held in aquaria?

Yes X

No ☐

Question 2: What method was used if they were killed?

Fishes were returned to aquarium

Question 3: If you have undertaken experimental work, has the care and use of experimental animals complied with local and or national animal welfare laws, guidelines and policies?

YesX

No ☐

If 'Yes', state these and provide suitable evidence (*e.g.* for the U.K. a Home Office PPL number is sufficient) that protocols have undergone an ethical review process by an institutional animal care and use (or similar) committee, a local ethics committee, or by appropriately qualified scientific and lay colleagues.

Permissions for fishing in Mexico

Please read the exceptions below (Questions 4 to 7). If any of these exceptions apply to your study, complete the appropriate section. Otherwise leave blank.

If 'No', because these laws do not exist in your country, please state this.

Please read the exceptions below (Questions 4 to 7). If any of these exceptions apply to your study, complete the appropriate section. Otherwise leave blank.

Question 4: Did you use experimental conditions that severely distressed the animals?

Yes ☐

No X

If 'Yes', state the conditions and how they can be justified.

Question 5: Did you use humane endpoints that minimized adverse effects?

Yes **X**
No ☐

Question 6: Have you performed surgical procedures?

Yes ☐
No **X**

If 'Yes', have you suitably described these in your manuscript?

Question 7: If the procedures caused more than slight pain or distress, did you use appropriate sedation, analgesia and anaesthesia, with appropriate post-operative care?

Yes **X**
No ☐

If 'Yes', outline these.

MS 222 anaesthesia

If 'No', did any of your procedures involve sentient, un-anaesthetized animals paralysed by chemical agents such as muscle relaxants?

Yes ☐
No ☐

If 'Yes', provide details. Normally these procedures will be considered unacceptable by *JFB*.

If 'No', did any of the procedures, particularly those that involve lethal endpoints, cause adverse effects or lasting harm to a sentient animal?

Yes ☐
No ☐

If 'Yes', provide details. Normally these procedures will be considered unacceptable by *JFB* unless any harm caused can be justified against the benefit gained.

Exploring the effect of microsatellite size homoplasy on reconstruction of phylogenetic relationships of picote splitfin *Zoogoneticus quitzeoensis*

L. BOTO*†, O. DOMÍNGUEZ-DOMÍNGUEZ‡ AND I. DOADRIO*

***Dpto. Biodiversidad y Biología Evolutiva. Museo Nacional Ciencias Naturales. CSIC. C/ José Gutiérrez Abascal 2, 28006, Madrid Spain and ‡ Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, 58088, Morelia, Michoacán, México*

Running title: Homoplasies in *Zoogoneticus quitzeoensis* microsatellites

†Author to whom correspondence should be addressed: Tel.: +34 914111328; email: mcnb119@mncn.csic.es

This study explores the effects of microsatellite size homoplasies in the reconstruction of phylogenetic relationships and estimates of population parameters as Fixation index (F_{ST}) using as a case study a truncated microsatellite from the picote splitfin *Zoogoneticus quitzeoensis*. The results suggest that the use of imperfect microsatellites may have only a minor effect in phylogenetic and population studies.

Key words: microsatellites, size homoplasies, phylogenetic reconstructions

A recent paper (Domínguez-Domínguez *et al.,* 2008) showed that the endangered picote splitfin *Zoogoneticus quitzeoensis* (Bean, 1898) populations can be split into two lineages that originated 3.3 My ago, as demonstrated by both mtDNA and microsatellites (microsatellite dataset from Domínguez-Domínguez *et al.,* 2007). They also reported discrepancies in mitochondrial and microsatellite data affecting the relationships of more recently diverged populations within lineage I. While mtDNA separated Moloya and Magdalena populations from others within this lineage (La Luz, Orandino, and Platanera), microsatellites relate them to the Orandino population. The discrepancy between these two genetic markers was interpreted as being caused by differences in genetic drift effects and stochastic processes affecting the genetic structure of the populations.

Despite the suggestion that size homoplasy in microsatellites (presence of alleles with equal length but different evolutionary history) may not be a significant problem in many population genetic analyses (Estoup *et al.,* 2002), the topic remains under active discussion (Ellegren, 2004).

Taking into account that the ZT1.6 microsatellite used in Domínguez-Domínguez *et al.*(2007) presents a truncated repeat structure (TG)nGG(TG)n in *Z. quitzeoensis* (Boto & Doadrio, 2003), as well as the fact that these imperfect or truncated microsatellites allow homoplasies to be detected more easily than do perfect microsatellites (Estoup *et al.,* 2002), this study explores whether discrepancies between mtDNA and microsatellite inference of phylogenetic relationships could be explained by homoplasies at this locus.

Twenty sequences were analysed from the different sized alleles present in homozygotes of the *Z. quitzeoensis* populations studied by Domínguez-Domínguez *et al.* (2007). Alleles were selected to obtain a random representation of the

populations and allele sizes and to maximise the Moloya and Magdalena population sample. Table I shows the allelic frequencies for the ZT1.6 locus as well as the sample size from the original dataset. Amplified PCR products from homozygote individuals in the Domínguez-Domínguez *et al.* (2007) dataset have been sequenced directly using the specific forward primer (Boto & Doadrio, 2003). All of the sequences matched the expected length, which was determined by the number of repeats and base pairs in the flanking region.

The results (Table II) showed that ZT1.6 alleles of the same size present a repeat structure in the Moloya/Magdalena populations different from the other populations within lineage I in Domínguez-Domínguez *et al.* (2008). This confirms that these are size-homoplasious alleles. Moreover, whereas alleles from other populations revealed a incremental variation pattern, alleles from Moloya/Magadalena, which only differed by a single repetition, showed a different repeat structure. In all sequenced alleles, the repeat flanking sequence was identical to the one present in the original clone of *Z. tequila* (Accession number AY102709).

Considering that the sequenced alleles are the most frequent alleles in the Moloya/Magdalena populations, with only two other alleles present at low frequencies, these results also suggest that analysis of phylogenetic relationships among lineage I populations could be affected by the use of this microsatellite. To explore this idea a re-analysis of the original microsatellite dataset was performed with and without the ZT1.6.microsatellite. Neighbour-joining trees using D_A distance (modified Cavalli-Sforza distance) and 5000 bootstrap replications were created using POPTREE2 software (Takezaki *et al.,* 2010). The results showed no differences in lineage I phylogenetic relationships. The same Orandino/Moloya-

Magdalena topology described by Domínguez-Domínguez *et al.* (2008) was obtained with or without the inclusion of the ZT1.6 microsatellite (Fig. 1).

The fixation indices (F_{ST}) for all population pairs were recalculated using GENEPOP (Raymond & Rousset, 1995) (version 4.0.10) both including and excluding the ZT1.6 microsatellite. The results (Table III) show that the microsatellite presence or absence has a small effect on F_{ST} values (range 0.00-0.08) in pairwise comparisons among populations of lineage I (Moloya, Magdalena, Platanera, Orandino, and La Luz). This suggests that the differences observed by Domínguez-Domínguez *et al.* (2008) when comparing mitochondrial and microsatellite markers with respect to phylogenetic relationships among lineage I populations of *Z. quitzeoensis* cannot be explained by the allele size homoplasies affecting the ZT1.6 microsatellite. Nevertheless, the possibility that the other loci could also be homoplasic cannot be ruled out.

Estoup *et al.* (2002) have proposed that molecularly accessible size homoplasy (MASH) makes up only a portion of the size homoplasy present in microsatellites. When considering the time elapsed since the divergence of *Z. quitzeoensis* lineages and homoplasy dependence on factors such as mutation rate, effective population size, and between-population divergence time, there is a high probability that the other perfect microsatellites may be homoplasic, although this homoplasy may not be molecularly accessible. The splitting in the same lineages I and II using mitochondrial DNA sequences or microsatellites (Domínguez-Domínguez *et al.,* 2008) disagrees with this interpretation, unless the similar topology obtained with these markers is simply coincidental leading to spurious results.

The authors think that the detected ZT1.6 homoplasies in the more recently diverged lineage I populations could reflect, as suggested by Domínguez-

Domínguez *et al.* (2008), a very close relatedness and a very different recent evolutionary history for those populations inhabiting water bodies that have undergone severe reductions in size as is the case with Moloya and Magdalena. In this sense, the results, which show the presence of private alleles for the ZT1.6 locus in Moloya/Magdalena, reinforce the status of these populations as an Operative Conservation Unit, as proposed by Domínguez-Domínguez *et al.* (2007) In addition, results of the current study add to the sparse bibliography with respect to the effect of homoplasies in reconstructing phylogenetic relationships.

Several studies have detected size homoplasy in microsatellites in different organisms (Primmer & Ellegren, 1998; Viard *et al.,* 1998; Angers *et al.,* 2000; Van Oppen *et al.,* 2000; Culver *et al.,* 2001; Adams *et al.,* 2004; Yokoyama *et al.,* 2004; Lia *et al.,* 2007; Barkley *et al.,* 2009; Machado *et al.,* 2010). Only a few of these studies deal directly with the effect that these homoplasies have on the reconstruction of phylogenetic relationships or on estimation of population parameters (Viard *et al.,* 1998; Angers *et al.,* 2000; Adams *et al.,* 2004; Barkley *et al.,* 2009; Machado *et al.,* 2010) and have generally reached contradictory conclusions.

Adams *et al.* (2004) found a small effect of the use of microsatellites with size homoplasy on the F_{ST} analogue R_{ST} in the tropical tree *Corythophora alta.* Angers *et al.* (2000) describe a small effect of homoplasies on the structure of the fresh water snail *Bulinus truncatus* populations. Conversely, Viard *et al.* (1998) found microsatellite size homoplasy affected the structure of *Bulinus truncatus* and *Bombus terrestris* populations. These authors also found that the use of homoplasious microsatellites may affect the reconstruction of phylogenetic relationships in *Apis mellifera*. Two recent papers (Barkley *et al.,* 2009; Machado *et*

133 *al.,* 2010) suggest however that homoplasic microsatellites have a moderate effect
134 on the reconstruction of phylogenetic relationships or population genetic studies.
135 Barkley *et al.* (2009) suggest that the use of this type of microsatellite has a slight
136 effect on the reconstruction of phylogenetic relationships within the genus *Citrus* and
137 Machado *et al.* (2010), analysing *Drosophila antonietae* populations, claim that
138 neither size homoplasy nor null alleles represent significant problems for population
139 genetic analyses, since they are compensated for by the high degree of
140 microsatellite polymorphisms.

141 The use of different methodological approaches makes comparisons difficult.
142 The present study revealed size homoplasy in a truncated microsatellite locus of *Z.*
143 *quitzeoensis*, the presence or absence of which does not seem to affect the topology
144 of phylogenetic relationships among populations of one of the two lineages
145 described for this species, and suggests a small effect of the use of microsatellites
146 exhibiting size homoplasies in phylogenetic reconstruction and F_{ST} values.

147 These results are in agreement with the observations of Adams *et al.* (2004),
148 Barkley *et al*. (2009), Machado *et al.* (2010), and with the Estoup *et al.* (2002)
149 proposal that homoplasic microsatellites are not a significant problem in many types
150 of population analyses, especially if mutation models that assume homoplasy are
151 used to calculate genetic distances, as suggested by Barkley *et al.* (2009).

152 Although that the other perfect microsatellites used in the study by Domínguez-
153 Domínguez *et al.* (2008) could also be homoplasic and the recovered phylogenetic
154 relationships spurious (even though the two main and ancient lineages are
155 recovered by both microsatellite and mitochondrial markers) the results support the
156 possibility that size homoplasy could be the product of a different evolutionary history

affecting populations such as those from Moloya and Magdalena, which inhabit water bodies that have undergone severe depletion in recent times.

This work was partially funded by project CGL2010-15237/BOS from Spanish MEC. The authors thank E. Bulmer for help with the English text and A. G. Sáez for advice during the preparation of the manuscript. English text has been further edited by Lucidus Consultancy. In addition, authors thank the anonymous referees for suggestions for improvements to the paper.

References

Adams, R. I., Brown, K. M. & Hamilton, M. B. (2004). The impact of microsatellite electromorph size homoplasy on multilocus population structure estimates in a tropical tree (*Corythophora alta*) and an anadromous fish (*Morone saxatilis*). *Molecular Ecology* **13**, 2579–2588. doi 10.1111/j.1365-294X.2004.02256.x

Angers, B., Estoup, A. & Jarne, P. (2000). Microsatellite size homoplasy, SSCP, and population structure: A case study in the freshwater snail *Bulinus truncatus*. *Molecular Biology and Evolution* **17**, 1926-1932.

Barkley, N. A., Krueger, R. R., Federici, C. T. & Roose, M. L. (2009). What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. Plant Systematics and Evolution **282**, 71-86. doi 10.1007/s00606-009-0208-2

Boto, L. & Doadrio, I. (2003). Polymorphic microsatellites in two different species of the genus *Zoogoneticus* Meek, 1902 (Goodeidae, Actynopterygii). Molecular Ecology Notes **3**, 70-72. doi:10.1046/j.1471-8286.2003.00353.x

Culver, M., Menotti-Raymond, M. A. & O'Brien, S. J. (2001). Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*). *Molecular Biology and Evolution **18***, 1151-1156.

Domínguez-Domínguez, O., Boto, L., Alda, F., Pérez-Ponce de Leon, G. & Doadrio, I. (2007). Human impacts on drainages of the Mesa Central, México, and its genetic effects on an endangered fish, *Zoogoneticus quitzeoensis*. *Conservation Biology* **21**,168-180. doi:10.1111/j.1523-1739.2006.00608.x

Domínguez-Domínguez, O., Alda, F., Pérez-Ponce de Leon, G., García-Garitagoitia, J. L. & Doadrio, I. (2008). Evolutionary history of the endangered fish *Zoogoneticus quitzeoensis* (Bean, 1898) (Cyprinodontiformes, Goodeidae) using a sequential approach to phylogeography based on mitochondrial and nuclear DNA data. *BMC Evolutionary Biology* **8**:161

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Review in Genetics* **5**, 435-445

Estoup, A., Jarne, P. & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604. doi: 10.1046/j.1365-294X.2002.01576.x

Lia, V. V., Bracco, M., Gottlieb, A. M., Poggio, L. & Confalonieri, V. A. (2007).Complex mutational pattern and size homoplasy at maize microsatellite loci. *Theoretical and Applied Genetics* **115**, 981-991. doi 10.1007/s00122-007-0625-y

206 Machado, L. P. B., Mateus, R. P., Sene, F. M. & Manfrin, M. H. (2010). Microsatellite
207 allele sequencing in population analyses of the South American cactophilic
208 species *Drosophila antonietae* (Diptera: Drosophilidae). *Biological Journal of*
209 *the Linnean Society* **100**, 573-584.doi 10.1111/j.1095-8312.2010.01441.x
210 Primmer, C. R. & Ellegren, H. (1998). Patterns of molecular evolution in avian
211 microsatellites. *Molecular Biology and Evolution* **15**, 997-1008.
212 Raymond M. & Russet, F. (1995). GENEPOP (Version 1.2): Population genetics
213 software for exact test and ecumenicism. *J. Heredity* **86**, 248-249.
214 Takezaki, N., Nei, M. & Tamura, K. (2010). POPTREE2: Software for constructing
215 population trees from allele frequency data and computing other population
216 statistics with Windows interface. *Molecular Biology and Evolution* **27**, 747-
217 752.doi:10.1093/molbev/msp312
218 http://www.kms.ac.jp/~genomelb/takezaki/poptree2/index.htm
219 Van Oppen, M. J. H., Rico, C., Turner, G. F. & Hewitt, G. M. (2000). Extensive
220 homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a
221 complex microsatellite locus in lake Malawi cichlids. *Molecular Biology and*
222 *Evolution* **17**, 489-498.
223 Viard, F., Franck, P., Dubois, M-P., Estoup, A. & Jarne, P. (1998). Variation of
224 microsatellite size homoplasy across electromorphs, loci, and populations in
225 three invertebrate species. *Journal of Molecular Evolution* **47**, 42-51.
226 Yokoyama, J., Fukuda, T., Yokoyama, A. & Nakajima, M. (2004). Extensive size
227 homoplasy at a microsatellite locus in the Japanese bumblebee, *Bombus*
228 *diversus*. *Entomological Science* **7**, 189-197.
229
230

231

232

Table

TABLE I. Allele frequencies for ZT1.6 microsatellite locus in *Zoogoneticus quitzeoensis* extracted from Domínguez-Domínguez *et al*. (2007) dataset

Population	Allele frequencies [a]	Number of samples
San Francisco del Rincón	**230**=0.03; **234**=0.03; **236**=0.05; **240**=0.16; **242**=0.34; **250**=0.39	19
Belisario	**232**=0.03; **236**=0.53; **240**=0.34; **244**=0.11	19
Platanera	**236**=0.21; **238**=0.08; **242**=0.71	12
Orandino	**232**=0.10; **236**=0.25; **238**=0.20; **240**=0.05; **242**=0.20; **246**=0.15; **252**=0.05	10
San Cristobal	**230**=0.04; **232**=0.08; **234**=0.04; **236**=0.21; **240**=0.54; **244**=0.08	12
Mintzita	**232**=0.06; **236**=0.12; **240**=0.71; **244**=0.12	17
La Luz	**236**=0.93; **240**=0.07	19
Moloya	**238**=0.45; **240**=0.45; **250**=0.10	10
Magdalena	**238**=0.56; **240**=0.43; **246**=0.01	7

a Allele lengths(bp) are in bold

Table

Table II. Repeat structure for zt1.6 alleles in *Zoogoneticus quitzeoensis*. Moloya and Magdalena populations are highlighted in bold

Allele length	*n*	Structure	Population
236	1	$(TG)_7GG(TG)_9$	Orandino
236	1	$(TG)_7GG(TG)_9$	La Luz
236	1	$(TG)_7GG(TG)_9$	Belisario
238	1	$(TG)_7GG(TG)_{10}$	Orandino
238	**3**	**$(TG)_2GG(TG)_{15}$**	**Moloya**
238	**2**	**$(TG)_2GG(TG)_{15}$**	**Magdalena**
240	2	$(TG)_7GG(TG)_{11}$	San Cristobal
240	1	$(TG)_7GG(TG)_{11}$	Mintzita
240	**2**	**$(TG)_9GG(TG)_9$**	**Moloya**
240	**1**	**$(TG)_9GG(TG)_9$**	**Magdalena**
242	1	$(TG)_7GG(TG)_{12}$	San Francisco
242	1	$(TG)_7GG(TG)_{12}$	Platanera
244	1	$(TG)_7GG(TG)_{13}$	Belisario
246	1	$(TG)_7GG(TG)_{14}$	Orandino
250	1	$(TG)_7GG(TG)_{16}$	San Francisco

n, number of sequenced alleles

Table

TABLE III. Estimated pairwise comparisons of F_{ST} for *Zoogoneticus quitzeoensis* populations, including (above the diagonal) or excluding (below the diagonal) ZT1.6 microsatellite

	MAG	MOL	PLA	LUZ	ORA	SFR	BEL	SCR	MIN
MAG		0.02	0,34	0.36	0.13	0.24	0.32	0.288	0.28
MOL	0.04		0.32	0.36	0.11	0.24	0.30	0.27	0.26
PLA	0.30	0.28		0.41	0.22	0.29	0.37	0.37	0.38
LUZ	0.28	0.27	0.35		0.27	0.40	0.34	0.40	0.41
ORA	0.13	0.10	0.24	0.24		0.22	0.24	0.24	0.25
SFR	0.21	0.23	0.32	0.33	0.24		0.18	0.15	0.18
BEL	0.33	0.31	0.36	0.36	0.26	0.11		0.04	0.06
SCR	0.32	0.31	0.36	0.37	0.27	0.09	0.02		0.04
MIN	0.29	0.28	0.34	0.36	0.24	0.10	0.01	0.06	

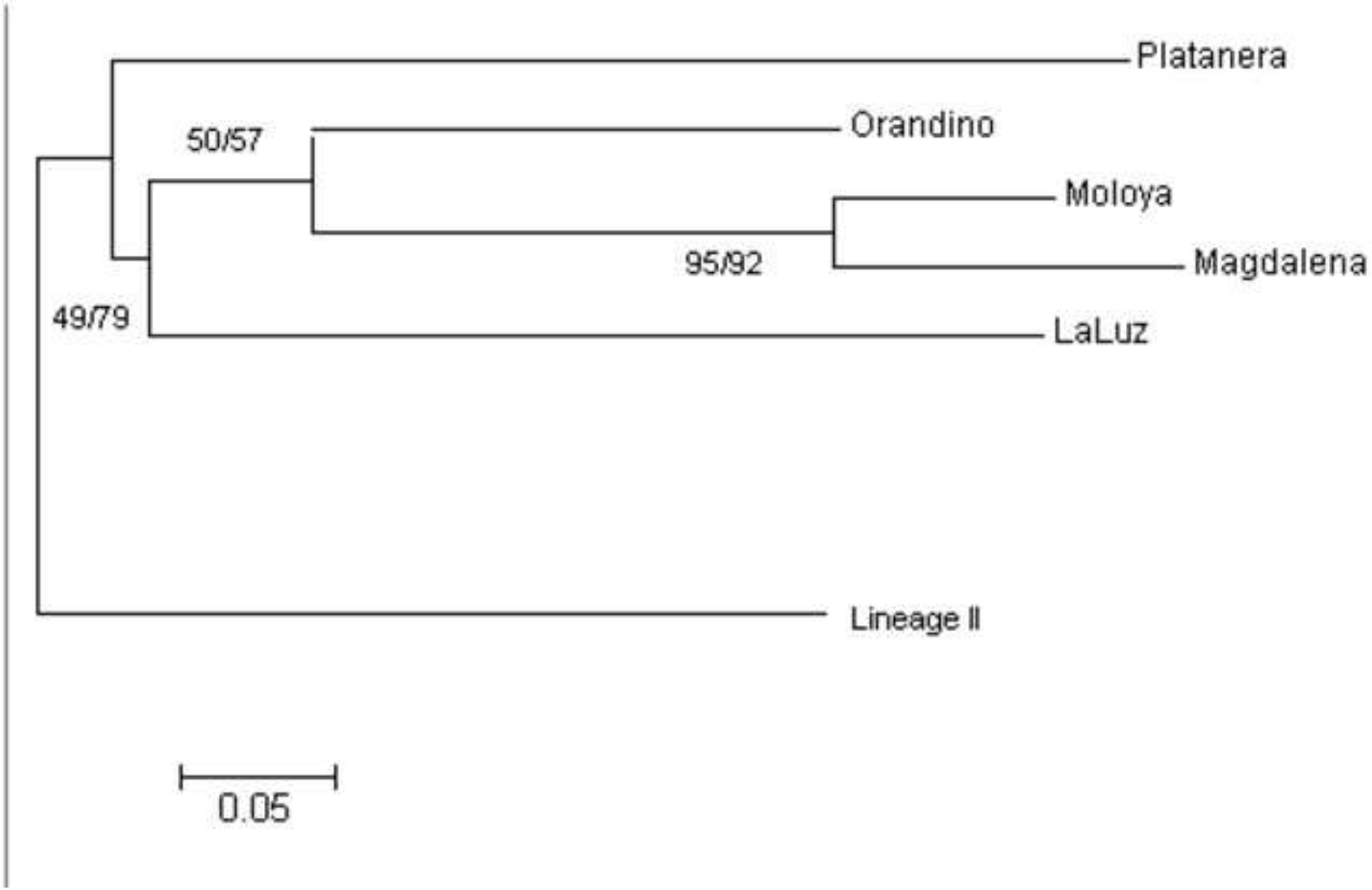
MAG, Magdalena; MOL, Moloya; PLA, Platanera; LUZ, La Luz; ORA, Orandino; SFR, San Francisco del Rincón; BEL, Belisario; SCR, San Cristobal; MIN, Mintzita.

Figure Captions

Caption Fig. 1.

Neighbour-joining tree showing the same topology in lineage I (Domínguez-Domínguez *et al*., 2008) populations, both including and excluding the ZT1.6 microsatellite from the analysis. Branch numbers show bootstrapping support with and without the presence of the ZT1.6 microsatellite.

Figure
Click here to download high resolution image



***Copyright form**

This piece of the submission is being sent via mail.