

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando (2018), Journal of Evolutionary Biology.

1. Extended methods

1.1. Field work and laboratory procedures

Specimens were collected by hand or noose, sacrificed by a pericardiac injection of sodium tiopental Abbot®/Pentovet®, dissected slightly to extract a sample of liver/muscle for molecular study, fixed in 10-20% formalin, and later transferred to 70% ethanol. Tissues were stored in a freezer with 96% ethanol. Voucher specimens are deposited in the herpetological collections LJAMM-CNP of the Centro Nacional Patagónico, Puerto Madryn, Argentina (CENPAT-CONICET, <http://www.cenpat.edu.ar/nuevo/colecciones03.html>), and the M. L. Bean Life Science Museum, Brigham Young University (BYU; <http://mlbean.byu.edu/ResearchCollections/Collections/ReptilesandAmphibians.aspx>).

For details of individuals used, see Appendix A. Genomic DNA was extracted using the Qiagen® DNeasy® 96 Tissue Kit following the protocol provided by the manufacturer. For PCR and sequencing protocols we followed Morando et al. (2003) for the mitochondrial genes (cyt-b and 12S), Camargo et al. (2012) for the anonymous nuclear loci (ANL: A1D, A12D, A4B and A9C), and the nuclear protein-coding loci (NPCL): EXPH5, KIF24, MXRA5, (Portik et al. 2012), DNAH3, PRLR, PNN, SNCAIP (Townsend et al. 2008), and CMOS (Wiens et al. 1999). We amplified all nuclear genes using the touchdown PCR cycle described by Noonan and Yoder (2009), with standard reaction conditions (per sample: 2 µl dNTPs (1.25 mM), 2 µl 5x Taq buffer, 1 µl each primer (10 µM), 1 µl MgCl (25 mM), and 0.1 µl Taq DNA polymerase (5 U/µl; Promega Corp., Madison, WI); 14 ml total reaction volume). All sequences were edited using the program Sequencher v4.8. (™Gene Codes Corporation Inc. 2007), and aligned with MAFFT (Katoh et al. 2002) based on 100 tree rebuilding iterations and a maxirate of 100. Protein-coding genes were translated to amino acids to check for codon errors. Missing data in all cases were coded as “?” (15.19% of the total dataset was missing). For each gene we selected the best-fitting model of evolution using JModelTest v2.1.1 (Darriba et al. 2012) using the Akaike Criterion Information (corrected) (AICc). In all nuclear genes, recombination was tested using RDP: Recombination Detection Program v3.44 (Martin and Rybicki 2000; Heath et al. 2006). Sequences are deposited in GenBank (Accession Nos. xxxx, to be completed upon ms acceptance), and alignments in Dryad data base (xxxx).

1.2. Gene trees and species trees

Gene tree inferences. We estimated each separated gene tree using the full matrix of 127 individuals and 14 loci using BEAST v1.6.1 (Heled and Drummond 2010). The two mitochondrial genes were concatenated to obtain a single gene tree, thus we obtained 13 independent gene trees in total. These trees were used for phylogenetic network estimation, as well as for presenting the mitochondrial gene tree in the main manuscript (Figure 2). Later we ran BEAST during 10×10^6 generations of MCMC and sampled at intervals of 1,000 generations with a 10% burnin. Convergence was diagnosed by observation of ESS values greater than 200 for each parameter estimated with Tracer v1.5 (Rambaut and Drummond 2009). The maximum credibility tree was obtained using TreeAnnotator.

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando (2018), Journal of Evolutionary Biology.

Species tree. The hybridization test used here requires a species tree topology as part of its model. Thus, we generated a species tree using only the focal species as part of the hybridization hypothesis. We included here the total of 14 loci and 95 individuals representing *L. rothi*, *L. boulengeri*, *L. telsen* and *L. tehuelche*. We ran the species tree analyses in *BEAST v1.6.1 (Heled and Drummond 2010) for 500×10^6 generations of MCMC, and sampled at intervals of 10,000 generations (burnin 10%). Convergence was diagnosed by observation of ESS values greater than 200 for each parameter estimated. *BEAST also generates gene trees and saved them as outputs during the same analyses, thus we used these gene trees for XLC statistic calculations and hybridization test.

Divergence times estimation in the mitochondrial tree. We used substitution rates estimated for the mitochondrial regions for the genus *Liolaemus*: estimated as 0.006339 substitution per site per million years for 12S, and 0.019355 substitution per site per million years for cyt-b. We used BEAST 1.6.1 and a strict clock *prior* independently for each mitochondrial region, estimated using a gamma distribution. For the 12S the initial value was set = 0.006339; shape = 1; scale = 0.009, which leads to a median equal to 0.006339. For cyt-b the initial value was set = 0.019355; shape = 1; scale = 0.0174, which leads to a median approximated equal to 0.019355. We used the Yule process as the tree *prior* model, and ran a MCMC analysis for 100×10^6 generations, sampling every 10,000 intervals and a 10% burnin. Convergence was diagnosed by observation of ESS > 200 for parameter estimation.

1.3. Testing the hybridization hypotheses

We set five M parameters (M_1, M_2, M_3, M_4, M_5) as is shown in Fig. 1 at the main manuscript (see also the species tree in newick format in Supplementary Material 2).

The simplest model has $M_{1,2,3,4,5} = 0$ (strict coalescent model); and then we tested all parameters range between 1 to 5. Thus, we tested models with different levels of gene flow between three different pairs of taxa: 1- *L. rothi* x *L. tehuelche* (current and past); 2- *L. rothi* x *L. boulengeri* + *L. telsen* (current and past); and 3- *L. rothi* and the common ancestor of (*L. tehuelche* + *L. boulengeri* + *L. telsen*). We calculate the likelihood of the data given each hypothesis, and selected the most likely model using likelihood ratio test. All used R functions are available in github.com/melisaolave/Olave_etal2017-MEE. We employed 13 gene trees (cyt-b and 12S mitochondrial loci were combined to infer a single gene tree) and a total of 93 samples of *L. rothi*, *L. boulengeri*, *L. telsen* and *L. tehuelche*.

Model construction. Models included a species tree topology and branch lengths (inferred as described above in section 1.2 in Supplementary Material) and migration parameters ($M = N_e m$; where N_e is the effective population size and m is the proportion of migrants).

Branch lengths were converted to coalescent units to make them compatible with the *ms* program (*ms* branch length units = $4N_e$ generations), which was implemented by

multiplying the branch lengths by the population parameter θ ($\theta = 4N_e\mu$; N_e is the effective population size and μ the mutation rate) for simulating nuclear gene trees, and for

mitochondrial gene tree by $\bar{\theta}$. This was done, by setting *brlength.correction* in *get.topo2ms* function.

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando (2018), Journal of Evolutionary Biology.

We used Lamarc v2.1.8 (Kuhner 2006) to estimate θ values in a Bayesian inference of 31,000 MCMC steps, with a 10% burnin. Diagnosis of convergence was made by observing ESS values equal or greater than 200. Here, θ was estimated = 0.05399.

Simulations, extra lineage approximation and the likelihood ratio test were performed using the R functions available in github.com/melisaolave/Olave_etal2017-MEE, described by (Olave et al. 2017a). Specifically, we used the function `get.topo2ms()` to load the species tree and obtain input commands to ms program (Hudson 2002) program, and the `ms()` function to simulate 10,000 gene trees per locus to infer the distribution of expected extra lineages given each proposed model. The number of extra lineages were counted using the `sptree.vs.genetree()` function to automatize the process using Phylonet package (Than and Nakhleh 2009). We obtained the likelihood of the data given each model using the `get.Likelihood()` function. The model was then selected taking the parameter estimations and comparing the likelihood of nested models using a likelihood ratio test implemented in the function `Lratiotest()`.

1.4. Morphological comparisons

Morphologies were quantified using 10 morphometric variables standardly used in taxonomic studies of the genus, including: snout-vent length (**SLV**, distance from the tip of the snout and the posterior margin of the precloacal scales), axilla-groin distance (**AGD**, distance from the armpit of the right front leg to the anterior insertion of the hindlimb), hand length (**HaL**, distance from the base of the wrist and base of the nail of the third digit; measured ventrally), radius-ulna length (**RUL**, distance from the hand to the elbow), humerus length (**HL**, distance from the elbow and the shoulder; ventrally measured), foot length (**FoL**, distance from the base of the heel to the base of the nail of the fourth digit; measured ventrally), tibio-fibula length (**TFL**, distance from knee to the internal angle with the foot), femur length (**FL**, distance from the knee to its pelvic insertion), and knee-knee distance (**KKD**, distance between knees bent a right angles to the abdomen, measured ventrally).

We also included the following 15 color pattern variables: vertebral line melanism (**VLM**, shape of the dorsal spots), number of spots at the left vertebral line (**NSLVL**), number of spots at the right vertebral line (**NSRVL**), presence of a white spot below SVL spots (**PWSVL**), paraventral melanism (**PVM**), number of left paraventral spots (**NLPS**), number of right paraventral spots (**NRPS**), presence of a white spot below PVM spots (**PWSPL**), mean line (**ML**, presence of line between MLV and MPV), scapular melanism (**SM**, spots at both sides (left and right) between the neck and the shoulders), gular melanism (**GM**, ventrally observed), ventral melanism (**VM**), ventral color (**VC**, color ventrally observed), forelimb melanism (**FM**), and hindlimb melanism (**HM**).

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando (2018), Journal of Evolutionary Biology.

2. Extended results:

2.1. Supplementary tables

Locus	Substitution model selected
cyt-b	HKY+I+G
12S	GTR+I+G
CMOS	HKY
DNAH3	HKY+I
EXPH5	HKY+G
KIF24	HKY+I
A12D	GTR+G
A1D	HKY+G
A4B	HKY+G
A9C	HKY+I
MXRA5	HKY+I
PNN	HKY+I
PRLR	HKY+I
SNCAIP	HKY+I

Table S1: Substitution models selected by jModeltest.

Summary statistic	Locus	
	Cyt-b	12S
Median	0.025715	0.01626
HPD	0.0030635-0.0061598	0.0011281-0.025398
Mean	0.028901	0.01922
Standard deviation	0.00027147	0.00010991

Table S2: Results of *posterior* estimation of substitution rates for the *boulengeri-rothi* complexes.

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando.

Species/group	Centroids	
	Axis 1	Axis 2
i.mt (i)	-0.81	-1.88
i.mt (ii)	-0.84	-0.69
<i>L. rothi</i>	-1.18	0.61
<i>L. telsen</i>	1.61	-0.67
<i>L. tehuelche</i>	1.61	0.75
<i>L. boulengeri</i>	1.77	-0.27

Table S3: Centroid coordinates estimated by the multivariate discriminant analysis (Fig. S1).

Group	i.mt (i)	i.mt (ii)	<i>L. rothi</i>	<i>L. telsen</i>	<i>L. tehuelche</i>	<i>L. boulengeri</i>	Total	Error (%)
i.mt (i)	12	0	1	0	0	1	14	14.29
i.mt (ii)	2	2	2	0	0	0	6	66.67
<i>L. rothi</i>	2	3	34	2	1	1	43	20.93
<i>L. telsen</i>	0	0	0	5	0	1	6	16.67
<i>L. tehuelche</i>	0	0	0	0	14	3	17	17.65
<i>L. boulengeri</i>	0	0	0	1	1	15	17	11.76
Total	16	5	37	8	16	21	103	20.39

Table S4: Cross-classification table from the discriminant analysis (Fig. S1). The rows represent the groups provided *a priori* and the columns represent the assignments based on the discriminant analysis. The proportion of miss-classifications is shown in the Error (%) column. While most of the individuals were classified in their original groups, all groups showed some proportion of classification error. For example, some of the *L. rothi* samples were classified as part of the *boulengeri* species complex, with one *L. rothi* sample each associated with *L. tehuelche*, *L. boulengeri* and *L. telsen*, while the reciprocal cases were not inferred. Individuals of the i.mt (i) and (ii) groups are miss-classified in both directions, as well as with *L. rothi* group.

2.2. Supplementary figure

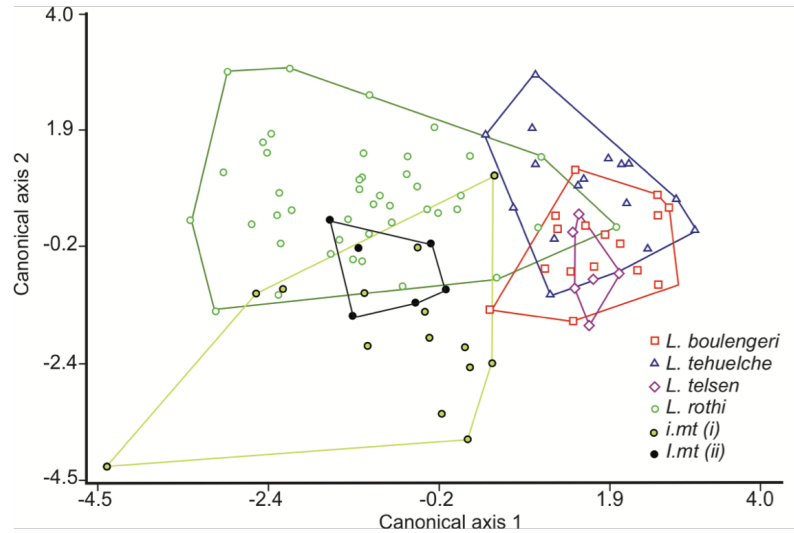


Fig. S1: Discriminate analysis plot estimated for the quantitative variables. Centroids are shown in Table S3 and cross-classification inferences are presenting in Table S4. The i.mt (i) and (ii) represent the hybrid candidates.

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando.

3. References:

- Camargo A, Avila LJ, Morando M, Sites JW Jr (2012) Accuracy and precision of species trees: effects of locus, individual and base pair sampling on inference of species trees in lizards of the *Liolaemus darwini* Group (Squamata , Liolaemidae). *Systematic Biology*, 61, 272-288.
- Darriba D, Taboada GL, Doallo R and Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, 772.
- Heath L, van der Walt V, Varsani A, Martin DP (2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *Journal of Virology*, 80, 11827–11832.
- Heled J, Drummond AJ (2010) Bayesian Inference of species trees from multilocus data research article. *Molecular Biology and Evolution*, 27, 570-580.
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral model. *Bioinformatics*, 337-338.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30, 3059–3066.
- Kuhner MK (2006) Lamarc 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, 22, 768-770.
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, 16, 562-563.
- Morando M, Avila LJ, Sites JW Jr (2003) Sampling strategies for delimiting species: genes, individuals, and populations in the *Liolaemus elongatus–kriegi* complex (Squamata: Liolaemidae) in Andean–Patagonian South America. *Systematic Biology*, 52, 159–185.
- Noonan B, Yoder AD (2009) Anonymous nuclear markers for Malagasy plated lizards (*Zonosaurus*). *Molecular Ecology Resources*, 9, 402–402.
- Olave M, Avila LJ, Sites JW Jr, Morando M (2017a) Detecting hybridization by likelihood calculation of gene tree extra lineages given explicit models. *Methods in Ecology and Evolution* doi: 10.1111/2041-210X.12846.
- Olave M, Avila LJ, Sites JW, Morando M. (2017b). Hidden diversity within the lizard genus *Liolaemus*: Genetic vs morphological divergence in the *L. rothi* complex (Squamata: Liolaeminae). *Molecular phylogenetics and evolution*, 107, 56-63.
- Portik D, Wood PL Jr, Grismer JL, Stanley EL, Jackman TR (2012) Identification of 104 rapidly-evolving nuclear protein-coding markers for amplification across scaled reptiles

Extended Methods and Results of the article “Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*”, by M. Olave, L.J. Avila, J.W. Sites Jr and M. Morando.

using genomic resources. *Conservation Genetic Resources*, 4, 1-10.

Rambaut A, Drummond AJ (2009) Tracer v1.5. Available from

<http://beast.bio.ed.ac.uk/Tracer>.

Than CV, Nakhleh L (2009) Species tree inference by minimizing deep coalescences. *PLoS Computational Biology* 5: e1000501. doi:10.1371/journal.pcbi.1000501

Townsend TM, Alegre RE, Kelley ST, Wiens JJ, Reeder TW (2008) Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Molecular Phylogenetics and Evolution*, 47, 129–142.

Wiens JJ, Reeder TW, Montes de Oca AN (1999) Molecular phylogenetics and evolution in sexual dichromatism among population of the Yarrow’s spiny lizard (*Sceloporus jarrovii*). *Evolution*, 53, 1884–1897.