

# ELECTRONIC APPENDIX

This is the Electronic Appendix to the article

**A cryptic taxon of Galápagos tortoise in conservation peril**

by

Michael A. Russello, Scott Glaberman, James P. Gibbs, Cruz Marquez,  
Jeffrey R. Powell and Adalgisa Caccone

*Biol. Lett.* ([doi:10.1098/rsbl.2005.0317](https://doi.org/10.1098/rsbl.2005.0317))

Electronic appendices are refereed with the text; however, no attempt is made to impose a uniform editorial style on the electronic appendices.

# ELECTRONIC APPENDIX A

This is the Electronic Appendix to the article

## A cryptic taxon of Galápagos tortoise in conservation peril

By

Russello *et al.*

### Materials and Methods

#### *Population genetic analyses*

Haplotypic diversity ( $h$ ; Nei 1987) was calculated based on mitochondrial DNA (mtDNA) control region (CR) sequences. Levels of genetic divergence between samples were calculated with the fixation index  $\phi_{st}$  (Excoffier et al. 1992) using the Kimura 2-parameter (K2P) genetic distance with a gamma value of 0.5 (empirically-determined for CR sequences only; Beheregaray et al. 2004). Significance of  $\phi_{st}$  for all possible pairwise population comparisons was assessed using 2,000 permutations. Tests for significant geographic structure among populations on Santa Cruz were conducted using analysis of molecular variance. All of the above analyses were executed in ARLEQUIN (Schneider et al. 2000). MtDNA sequence alignments (CR, 12S, 16S, cyt b) for the three Santa Cruz lineages were further employed to identify diagnostic nucleotide sites by means of population aggregation analysis (Davis & Nixon 1992). The presence of characters fixed and different between populations was used as evidence to diagnose distinct units. A

mtDNA CR haplotype tree was reconstructed for the population level sampling using PAUP\*4.0b10 (Swofford 2002) under maximum parsimony and employing a heuristic search with 1,000 random addition replicates and TBR branch swapping.

Genotypic diversity within populations was calculated as the mean number of alleles per locus, observed and expected heterozygosity at nine microsatellite loci using GENEPOP (Raymond & Rousset 1995). Levels of nuclear DNA differentiation among populations were estimated by: 1) pairwise population comparisons of the fixation index using the estimator  $\Theta$  (Weir & Cockerham 1984) as calculated in ARELQUIN (Schneider et al. 2000); and 2) assignment tests of individual tortoises based on a Bayesian method (Rannala & Mountain 1997) implemented in GENECLASS (Cornuet et al. 1999). Assignments were conducted using a simulation procedure with 10,000 randomly generated genotypes. Tortoises with a likelihood < 5% of belonging to their sampled population were not assigned to that locality.

Type I error rates for analyses involving multiple simultaneous comparisons were corrected using the sequential Bonferroni procedure (Rice 1989).

### *Phylogenetic analyses*

Consensus DNA sequences for 26 operational taxonomic units including extinct and all extant *G. nigra* lineages, and three outgroups were aligned in Clustal X (Thompson et al. 1997) employing default settings for gap opening and extension cost, and corrected by eye. Bayesian phylogenetic analysis for the combined dataset (2332 aligned nucleotide characters) was conducted in MrBayes 3.0b4 (Ronquist & Huelsenbeck 2003) assuming a mixed-model of nucleotide substitution. Specifically, data were partitioned into “coding”

(cyt b) and “non-coding” (12S, 16S, CR), and the best-fit model of nucleotide substitution was selected for each partition following a series of hierarchical likelihood ratio tests (hLRTs) as implemented in Modeltest (Posada & Crandall 1998). A codon substitution model (HKY+G) was applied to the coding partition and a standard 4X4 model (HKY+I+G) was assumed for the non-coding partition. Both partitions were allowed to have different overall rates. The analysis ran four simultaneous chains for  $2.0 \times 10^6$  total generations with trees sampled every 100 generations for a total 20,000 trees. The first 2,000 trees were discarded as burn-in samples and the remaining 18,000 trees were used to construct a majority-rule consensus tree and derive posterior probability values. Additional measures of nodal support were calculated using bootstrap re-sampling under a maximum likelihood (ML) optimality criterion assuming a general HKY+I+G model for the complete data set as selected by hLRTs. ML topologies were subjected to 500 bootstrap replicates with the starting tree for TBR branch swapping obtained by neighbor-joining.

In order to further test hypotheses of monophyly, additional ML searches were performed, constraining the analyses to include a: 1) monophyletic *porteri* (La Caseta and Cerro Fatal); and 2) monophyletic Santa Cruz lineages (La Caseta, Cerro Fatal and Cerro Montura). ML phylogenetic analyses were executed as described above while enforcing topological constraints and searching for compatible trees. ML topologies resulting from the constrained and unconstrained searches were compared using the Shimodaira-Hasegawa log-likelihood test (Shimodaira & Hasegawa 1999) as implemented in PAUP\*4.0b10 (Swofford 2002).

Divergence times of the Santa Cruz lineages both within and among island(s) were calculated based on mtDNA CR pairwise K2P genetic distances (selected as described above) and a mutation rate of 3.4% per million years. The rate of evolution was previously estimated using the r8s program (Sanderson 2003) and two calibration points to allow scaling of times to real units: 1) assigned root of the tree 3.3 mya which is the approximate subaerial age of the oldest island in Galápagos (e.g. Española); 2) enforced a maximum age of 700,000 years (approximate age of emergence of Isabela) for a clade including all Isabela populations (Beheregaray et al. 2004).

## References

- Beheregaray, L. B. et al. 2004 Giant tortoises are not so slow: Rapid diversification and biogeographic consensus in the Galápagos. *Proc. Natl Acad. Sci. U.S.A.* **101**, 6514-6519.
- Cornuet, J. M., Piry, S., Luikart, G., Estoup, A. & Solignac, M. 1999 New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.
- Davis, J. I. & Nixon, K. C. 1992 Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* **41**, 421-435.
- Excoffier, L., Smouse, P. E. & Quattro, J. M. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics* **131**, 479-491.
- Nei, M. 1987 *Molecular Evolutionary Genetics* (Columbia University Press, New York).

- Posada, D. & Crandall, K. A. 1998 MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.
- Rannala, B. & Mountain, J. L. 1997 Detecting immigration by using multilocus genotypes. *Proc. Natl Acad. Sci. U.S.A.* **94**, 9197-9201.
- Raymond, M. & Rousset, F. 1995 Genepop (Version-1.2) – population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248-249.
- Rice, W. R. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223-225.
- Ronquist, F. & Huelsenbeck, J. P. 2003 MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.
- Sanderson, M. 2003 r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301-302.
- Schneider S., Kueffer J., Roessli D., Excoffier L. 2000 *Arlequin Ver. 2.000: A software for population genetic data analysis* (Genetics and Biometry Laboratory, University of Geneva, Switzerland).
- Shimodaira, H. & Hasegawa, M. 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114-1116.
- Swofford, D. L. 2002 *PAUP\*: Phylogenetic Analysis Using Parsimony, v.4.0b10* (Illinois Natural History Survey, Champaign).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997 The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Weir, B. S. & Cockerham, C. C. 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.