



Conservation phylogenetics and computational species delimitation of Neotropical primates

Beatriz Mello, Júlio F. Vilela, Carlos G. Schrago*

Department of Genetics, Federal University of Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:

Primate conservation
South America
PTP
GMYC
Molecular dating
RelTime

ABSTRACT

Accurate definition of species boundaries plays a crucial role in the management and conservation of biological diversity. Although the science of genetics has been largely used to unveil cryptic biodiversity, phylogenetic theory has not been fully explored to address this issue. Recently, single-locus approaches to delimit evolutionary significant units (ESUs) have been widely used in several groups of organisms, presenting a valuable tool to test the uniqueness of biological clades. In this context, this study aimed to test the performance of coalescent-based delimitation methods using Neotropical primates, which is a lineage that includes several endangered species. This was achieved by evaluating the congruence between inferred ESUs and the nominal species recognized by taxonomists based on single-locus data. The results suggested that the molecular marker that best replicates standard taxonomy is the *COI* gene. GMYC presented the best overall performance at preventing lumping profiles, and the methodological combination that led to the lowest lumping frequency was the ST-GMYC using the timescale inferred on RelTime. The highest matching frequencies were produced by the ST-GMYC based on a BEAST timescale inferred with a Yule prior for the branching process. Overall, CSD approaches produced higher frequencies of matches and splits than lumps. The implementation automated the delimitation of evolutionary units in a circularity-free system, which is valuable for decision-making policies, enabling the refinement of an entirely taxonomy-based procedure. The results also highlighted the importance of phylogeny-based analyses for the discipline of conservation. Importantly, our conclusions were based on a Neotropical primates dataset, with several endangered species that inhabit degraded areas, which has made this clade a focus of conservation policies in recent decades.

1. Introduction

Species delimitation has always played a central role in theoretical biology. However, it also stands out as a major endeavor of applied biological sciences, because accurate definition of species boundaries is crucial for management and conservation of biological diversity (Isaac et al. 2004; Peterson and Navarro-Sigüenza 1999; Rojas-Soto et al. 2010; Sites and Crandall 1997; Sites and Marshall 2003). In recent decades, the science of genetics has featured preeminently in establishing new methods capable of unveiling the cryptic biodiversity and setting standards for species delimitation (Hoban et al. 2013; Pérez-Portela et al. 2013). On the other hand, phylogenetics, the primary biological discipline of evolutionary patterns, has had a secondary role in conservation and environmental policy-making (Davies et al. 2008; Vázquez and Gittleman 1998). This situation is contradictory given the wide range of applications of phylogenetic theory in evolutionary biology (Baum and Shaw 1995; Eldredge and Cracraft 1980; Felsenstein

1988; O'Meara et al. 2006; Paradis 2005; Takahata 1989).

Recently, however, authors have increasingly incorporated phylogeny-aware analyses in biological conservation issues (Bálint et al. 2011; Bickford et al. 2007; Moritz 1995; Pettengill and Neel 2008; Thuiller et al. 2015). One of these new methodological advances is the development of computational species delimitation (CSD) algorithms that can delimit or validate evolutionary significant units (ESUs) (*sensu* Ryder 1986) based on the phylogenetic species concept (Eldredge and Cracraft 1980). CSD has been a valuable tool for detecting the limits between species from prokaryotes to eukaryotes (Carstens and Dewey 2010; Fujisawa et al. 2016; Meier-Kolthoff et al. 2013; Yang and Rannala 2010). Once biodiversity conservation stands as one of the main global challenges for the 21st century (Gibson et al. 2011; Jetz et al. 2007; Sala et al. 2000), these methodologies present a practical approach to infer ESUs based on molecular sequence data in a phylogenetic framework. CSD methods are diverse, and recent developments incorporating the coalescent theory provide a powerful test of

* Corresponding author at: Universidade Federal do Rio de Janeiro, Instituto de Biologia, Departamento de Genética, CCS, A2-092, Rua Prof. Rodolpho Paulo Rocco, SN, Cidade Universitária, Rio de Janeiro, RJ CEP: 21941-617, Brazil.

E-mail address: carlos.schrago@gmail.com (C.G. Schrago).

<https://doi.org/10.1016/j.biocon.2017.11.017>

Received 10 May 2017; Received in revised form 17 October 2017; Accepted 16 November 2017

0006-3207/ © 2017 Elsevier Ltd. All rights reserved.

alternative hypotheses towards independence of lineages (Fujita et al. 2012).

The growing number of molecular sequences generated by massively parallel sequencing platforms has allowed the acquisition of individual-level data for multiple loci, thus enabling the use of coalescent-based CSD methods (Ence and Carstens 2011; Yang and Rannala 2010). However, obtaining molecular markers per individual to conduct CSD may be a hard task, as it involves extensive fieldwork or a large collaborative framework to acquire samples. Moreover, handling data from multiple loci can be computationally demanding for large-scale CSD analyses (Fujita et al. 2012; Zhang et al. 2011).

As an alternative to these drawbacks, single-locus CSD methodologies are useful to establish ESUs by assembling publicly available data for several terminals in a phylogenetic tree. Currently, two major single-locus CSD approaches have been widely used, namely, the general mixed Yule coalescent (GMYC) (Fujisawa and Barraclough 2013) and the Poisson tree process (PTP) (Zhang et al. 2013). Both are based on coalescent theory and aim to identify the transition from a population-level coalescent process to a macroevolutionary speciation process in the phylogenetic tree. The main difference between these single-locus strategies is that GMYC finds the process breakpoint in ultrametric trees, modeling speciation events by time units, whereas PTP models speciation in units of substitution per sites, eliminating the need for ultrametricity. Apart from traditional taxonomy, such methods may be viewed as auxiliary approaches to test the richness of ESUs, mainly in highly endangered biological clades, where delimitation of an evolutionarily unattached lineage is fundamental for conservation management (Agapow et al. 2004; Mace 2004; Moritz 1994).

In this context, the Neotropical region, which harbors one of the highest levels of terrestrial biodiversity on earth, holds several threatened species due to habitat loss and consequences of global warming (Cincotta et al. 2000; Gibson et al. 2011; Jantz et al. 2015; Malcolm et al. 2006; Myers et al. 2000). The degradation of biomes in South and Central America has intensified in recent decades, and it impacts the preservation of terrestrial biodiversity (Carrete et al. 2009; Santibáñez and Santibáñez 2007; Strassburg et al. 2012). This condition is aggravated for the conservation of large mammals, such as Neotropical primates (NP, parvorder Platyrrhini) (Boyle 2008; Costa et al. 2005; Lopes and Ferrari 2000; Rylands et al. 1997), which have been the focus of intense conservation policies for decades (Kierulff et al. 2012).

The greatest species richness of NP is located in Brazil (Wilson and Reeder 2005), and ca. 30% of platyrrhine species (44 species) are currently categorized as endangered (EN) or critically endangered (CR), with five genera comprising more than 50% of their species diversity classified as threatened (*Ateles*, *Brachyteles*, *Chiropotes*, *Lagothrix*, and *Leontopithecus*) (The IUCN Red List of Threatened Species 2015). Moreover, the number of NP nominal species has been increasing at a higher rate than any other primate taxa (Isaac et al. 2004). NP are thus an excellent case study to be used as a proxy for evaluation of single-locus CSD methods, because the richness of the inferred ESUs can be readily compared to the units defined by standard taxonomy. In this sense, such comparison might corroborate the uniqueness of red list species or spotlight lineages that were previously unrecognized as endangered, highlighting the significance of phylogeny-aware methods of biodiversity conservation. Furthermore, although several CSD methods were proposed, no evaluation of their performance has ever been conducted for large mammals.

Currently, most CSD methods rely on dated phylogenies, as is the case for the GMYC single-locus approach. However, molecular dating of phylogenetic trees is not trivial, because a reliable fossil record is needed. For NP, the fossil record is scarce, and fossil findings are chronologically and geographically biased (Fleagle 2013). Moreover, molecular dating has been shown to be tricky in datasets combining both intraspecific and interspecific diversity because of rate dependency effects (Ho et al. 2005). A Bayesian framework involves an additional difficulty in dealing with topological prior for both

population-level gene genealogies (the coalescent) and species-level macroevolutionary events (e.g., the birth–death process) (Ritchie et al. 2016a, 2016b). Therefore, as ultrametric trees are required for methodologies like GMYC, it is relevant to assess the performance of distinct molecular dating methods for CSD.

In this work, we aimed to evaluate the performance of coalescent-based species delimitation methodologies using single-locus datasets and to test the congruence between inferred ESUs and the nominal species recognized by taxonomists. We used NP as a case study to explore the consequences of applying this approach to a highly endangered clade. Therefore, besides defining species boundaries per se, we argue that, with the current availability of tons of molecular sequence data, computational delimitation of evolutionary lineages might be helpful for governmental decisions and biodiversity conservation per se. We tested the impact of the single-locus CSD methodology and dating method, as well as the extent of taxonomic sampling and the chosen molecular markers at inferring ESUs and recovering matches with traditional taxonomy.

2. Material and methods

2.1. Dataset composition and study design

As our primary goal was to test the impact of single-locus species delimitation methods on the number and distribution of Platyrrhini evolutionary units, we chose cytochrome b (*Cyt-b*) and cytochrome c oxidase subunit 1 (*COI*) sequences to address this question. These genes are widely sampled in mammals as a result of years of cumulative research efforts and projects such as the "Barcode of Life" (<http://www.barcodeoflife.org/>) (Hebert et al. 2003; Luo et al. 2011). Therefore, all platyrrhine sequences available in GenBank (Benson et al. 2005) for these two mitochondrial molecular markers were downloaded. In total, 491 sequences of *Cyt-b* were retrieved, representing 64 species and 16 genera. For *COI*, 89 sequences were retrieved, representing 33 species and 16 genera. The species diversity included accounts for approximately 50% and 26% of NP described species (Wilson and Reeder 2005).

These two mitochondrial molecular markers were then used to investigate the consequences of using shallower or deeper taxonomic sampling in CSD. To do so, we conducted analyses on two distinct types of datasets. The first datasets consisted of the complete sampling of sequences for NP for both *Cyt-b* and *COI*; the second datasets were limited to the genus-level sampling, also for both *Cyt-b* and *COI*. This was done to investigate whether the size of the phylogeny, tree depth, and tree shape would affect the results. If CSD methods are robust, we expect to find no significant difference between the results generated based on distinct datasets. Thus, for the sake of clarity, two terminologies were employed to distinguish between the analyses performed on those datasets: i) parvorder-level and ii) genus-level.

For both *Cyt-b* and *COI* parvorder-level dataset analyses, we used as outgroups sequences from *Pan troglodytes*, *Macaca mulatta*, *Homo sapiens*, *Gorilla gorilla*, and *Pongo abelii*. For genus-level datasets, we restricted the analyses to genera with a significant amount of nucleotide sequences. Therefore, sequences of both *Cyt-b* and *COI* from *Aotus*, *Ateles*, *Callicebus*, *Callithrix*, *Lagothrix*, *Saguinus*, *Saimiri*, and *Sapajus* were used to conduct single-locus CSD. For *Alouatta*, *Chiropotes*, *Cebus*, and *Cacajao*, only *Cyt-b* sequences were employed in genus-level analyses. One representative of every genus of the same family was chosen to compose the outgroups. Importantly, when identical sequences (haplotypes) were found, they were pruned to a single copy for the analyses, as it may negatively impact the delimitation of ESUs because of zero-length terminal branches (Fujisawa and Barraclough 2013; Monaghan et al. 2009).

All *Cyt-b* and *COI* alignments were carried out via amino acid sequences in MAFFT (Katoh and Standley 2013), with default options and visually inspected. Phylogenetic inference was implemented in both

Download English Version:

<https://daneshyari.com/en/article/8847623>

Download Persian Version:

<https://daneshyari.com/article/8847623>

[Daneshyari.com](https://daneshyari.com)