Reconstructing the phylogeny of new world monkeys (platyrrhini): evidence from multiple non-coding loci

Xiaoping Wang\textsuperscript{a,b}, Burton K. Lim\textsuperscript{c}, Nelson Ting\textsuperscript{d}, Jingyang Hu\textsuperscript{a,e,f}, Yunpeng Liang\textsuperscript{a}, Christian Roos\textsuperscript{g,*}, and Li Yu\textsuperscript{a,*}

\textsuperscript{a}State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, China, \textsuperscript{b}School of Life Sciences, Yunnan University, Kunming, China, \textsuperscript{c}Department of Natural History, Royal Ontario Museum, Toronto, ON, Canada, \textsuperscript{d}Department of Anthropology and Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon, USA, \textsuperscript{e}State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, \textsuperscript{f}Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China, and \textsuperscript{g}Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg, Göttingen, Germany

*Address correspondence to Li Yu and Christian Roos. E-mail: yuli@ynu.edu.cn, croos@dpz.eu.

Handling editor: Fuwen Wei

Received on 14 March 2018; accepted on 12 September 2018

Abstract

Among mammalian phylogenies, those characterized by rapid radiations are particularly problematic. The New World monkeys (NWMs, Platyrhini) comprise 3 families and 7 subfamilies, which radiated within a relatively short time period. Accordingly, their phylogenetic relationships are still largely disputed. In the present study, 56 nuclear non-coding loci, including 33 introns (INs) and 23 intergenic regions (IGs), from 20 NWM individuals representing 18 species were used to investigate phylogenetic relationships among families and subfamilies. Of the 56 loci, 43 have not been used in previous NWM phylogenetics. We applied concatenation and coalescence tree-inference methods, and a recently proposed question-specific approach to address NWM phylogeny. Our results indicate incongruence between concatenation and coalescence methods for the IN and IG datasets. However, a consensus was reached with a single tree topology from all analyses of combined INs and IGs as well as all analyses of question-specific loci using both concatenation and coalescence methods, albeit with varying degrees of statistical support. In detail, our results indicated the sister-group relationships between the families Atelidae and Pitheciidae, and between the subfamilies Aotinae and Callithrichinae among Cebidae. Our study provides insights into the disputed phylogenetic relationships among NWM families and subfamilies from the perspective of multiple non-coding loci and various tree-inference approaches. However, the present phylogenetic framework needs further evaluation by adding more independent sequence data and a deeper taxonomic sampling. Overall, our work has important implications for phylogenetic studies dealing with rapid radiations.

Key words: coalescence, concatenation, non-coding nuclear genes, primates, phylogenetics
The New World monkeys (NWMs, Platyrrhini) are a group of arboreal primates distributed in South and Central America, ranging from southern Mexico to northern Argentina (Kinze 1997; Mittermeier et al. 2013). The evolutionary history of NWMs is characterized by rapid bursts of diversification at the family levels that occurred within a 10 million year window (Kay et al. 2008; Hodgson et al. 2009; Fleagle 2013). Taxonomically, NWMs are divided into the 3 families Pitheciidae with the 2 subfamilies Pitheciinae (uakaris and sakis) and Callitrichinae (titi monkeys), Atelidae with the 2 subfamilies Atelesina (spider and woolly monkeys, and muriquis) and Alouatiinae (howler monkeys), and Cebidae with the 3 subfamilies Cebinae (capuchin and squirrel monkeys), Aotinae (night monkeys) and Callithrichinae (marmosets and tamarins) (Fleagle 2013). Aotinae and Callithrichinae are sometimes elevated to the family level (Mittermeier et al. 2013), but we follow here a NWM classification with 3 families and 7 subfamilies.

Over the last decades, a cascade of molecular phylogenetic studies of NWMs using mitochondrial and nuclear DNA sequences has been conducted (Schneider et al. 1993, 1996; Poux and Douzery et al. 2004; Opazo et al. 2006; Schrago 2007; Fabre et al. 2009; Osterholz et al. 2009; Wildman et al. 2009; Chatterjee et al. 2009; Perelman et al. 2011; Springer et al. 2012; Finstermeier et al. 2013; Jameson-Kiesling et al. 2015; Aristide et al. 2015). Owing to this rapid radiation and recent speciation events, the phylogeny of NWMs is still incompletely resolved as earlier studies revealed contradictory branching patterns among and within the 3 NWM families (Opazo et al. 2006; Fabre et al. 2009; Osterholz et al. 2009; Wildman et al. 2009; Chatterjee et al. 2009; Perelman et al. 2011; Perez et al. 2012; Ting and Sterner 2013; Finstermeier et al. 2013; Jameson-Kiesling et al. 2015; Aristide et al. 2015; Delgado et al. 2016). All 3 alternative sister group relationships among Pitheciidae, Atelidae, and Cebidae have been proposed (Figure 1Aa–Ac) (Schneider et al. 1993, 1996; Poux and Douzery et al. 2004; Opazo et al. 2006; Fabre et al. 2009; Osterholz et al. 2009; Wildman et al. 2009; Chatterjee et al. 2009; Perelman et al. 2011; Springer et al. 2012; Finstermeier et al. 2013; Schrago et al. 2014; Jameson-Kiesling et al. 2015; Aristide et al. 2015). Similar to the branching patterns among families, the phylogenetic relationships among the 3 subfamilies of the Cebidae (Aotinae, Callithrichinae, and Cebinae) are also not well resolved, with earlier studies providing support for each of the 3 possible relationships (Figure 1Ba–Bc) (Poux and Douzery et al. 2004; Opazo et al. 2006; Schrago 2007; Fabre et al. 2009; Wildman et al. 2009; Chatterjee et al. 2009; Springer et al. 2012; Finstermeier et al. 2013; Jameson-Kiesling et al. 2015; Aristide et al. 2015). The poor resolution and discordances among gene trees coupled with short internal branches are consistent with a recent and rapid radiation of NWMs (Kay et al. 2008; Hodgson et al. 2009; Fleagle 2013).

Here, we used 56 nuclear non-coding loci, including 33 introns (INs) and 23 intergenic (IGs) regions, from representatives of all NWM families and subfamilies to test previous phylogenetic hypotheses. The majority of loci (43) have not been used in previous NWM phylogenetic studies, and most are from INs, a class of non-coding DNA less commonly employed (Schneider et al. 1996) compared with the widely used IGs regions (Wildman et al. 2009; Jameson-Kiesling et al. 2015). In addition to traditional concatenation tree reconstruction methods, we applied 2 coalescent-based species-tree estimation methods for the resolution of higher-level relationships in NWMs. The coalescence methods have been thought to suffer less from analytical biases relative to concatenation methods in the case of rapid radiations by accounting for differences between gene and species trees (Swenson and El-Mabrouk 2012; Roch and Warnow 2015). In addition, we utilized a recently developed question-specific approach for reducing incongruence associated with large data sets and tree-inference methods in phylogenomics (Chen et al. 2015).

Materials and Methods

Material

Blood and tissue samples of 20 NWMs from 18 species representing all 3 NWM families and 7 subfamilies were obtained from the zoos in Cologne, Gettorf, Kunning, Landau, Romagne, Rostock, Stockholm, and Toronto (Table 1). Blood samples were taken during routine health checks, whereas muscle samples were obtained from deceased specimens. Blood samples were immediately subjected to DNA extraction after arrival in the laboratory, whereas tissue was stored frozen in 96% ethanol before further processing.

Data sets and laboratory work

Total genomic DNA from blood or tissue was isolated using standard proteinase K or phenol/chloroform extraction (Sambrook et al. 1989). We amplified and sequenced a total of 56 nuclear non-coding loci (Table 2). Forty-three of these loci were taken from a study investigating the phylogenetic relationships among colobine genera (leaf-eating Old World monkeys) (Wang et al. 2012), whereas the other 13 loci derived from a study on NWMs (Wildman et al. 2009). PCR conditions and primer sequences are shown in Supplementary Table S1. The amplified DNA fragments were purified and sequenced in both directions with an ABI PRISM™ 3700 DNA or 3130xL sequencer following the manufacturer’s protocol.

Figure 1. Alternative phylogenetic relationships that have been proposed among (A) NWM families and (B) subfamilies of the Cebidae family.
Table 1. Information about investigated species, their origin, and genbank accession numbers.

<table>
<thead>
<tr>
<th>Family</th>
<th>Subfamily</th>
<th>Species</th>
<th>Common name</th>
<th>Genbank Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cebidae</td>
<td>Callithrichiinae</td>
<td>Callithrix jacchus</td>
<td>Common Marmoset</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cebella pygmaea</td>
<td>Pygmy Marmoset</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leontopithecus rosalia</td>
<td>Golden Lion Tamarin</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saginus bicolor</td>
<td>Pied Tamarin</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Callimico goeldii</td>
<td>Goeldi’s Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td>Aotinae</td>
<td></td>
<td>Aotus azarae</td>
<td>Azara’s Night Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td>Cebinae</td>
<td></td>
<td>Cebus capricus</td>
<td>Colombian White-faced Capuchin</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sapajus apella</td>
<td>Guianan Brown Capuchin</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saimiri sciureus</td>
<td>Guianan Squirrel Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saimiri boliviensis</td>
<td>Black-capped Squirrel Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Atelinae</td>
<td>Ateles paniscus</td>
<td>Red-faced Back Spider Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ateles fusciceps</td>
<td>Black-headed Spider Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lagothrix lagotricha</td>
<td>Humboldt’s Woolly Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Alouattinae</td>
<td>Alouatta caraya</td>
<td>Paraguyan Howler Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pithecia pithecia</td>
<td>White-faced Saki</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chiroptotes albinalis</td>
<td>Red-nosed Bearded Saki</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cacahou sylvestris</td>
<td>Bald Uakari</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td>Callicebinae</td>
<td></td>
<td>Plecturocebus cupreus</td>
<td>Coppery Titi Monkey</td>
<td>KY458990-KY459995</td>
</tr>
<tr>
<td>Hominidae</td>
<td>Macaca mulatta</td>
<td>Rhesus Macaque</td>
<td>rhesMac2</td>
<td>UCSC Genome Browser</td>
</tr>
<tr>
<td></td>
<td>Pongo abelii</td>
<td>Sumatra Orangutan</td>
<td>ponAbel2</td>
<td>UCSC Genome Browser</td>
</tr>
<tr>
<td></td>
<td>Pan troglodytes</td>
<td>Chimpanzee</td>
<td>panTro2</td>
<td>UCSC Genome Browser</td>
</tr>
<tr>
<td></td>
<td>Homo sapiens</td>
<td>Human</td>
<td>hg18</td>
<td>UCSC Genome Browser</td>
</tr>
</tbody>
</table>

In the case of poor performance of direct sequencing resulting from complex DNA structures or tandem repeats, PCR products were cloned into the PMD18-T Vector and transformed into ultracompetent E. coli cells (TaKaRa Biotechnology Co., Ltd. Dalian, China). Five positive clones per ligation reaction were sequenced. All sequences were checked and queried in BLAST to assess homology. For some species, PCR attempts failed to produce sequence data. These sequences were excluded from the corresponding independent gene analyses and treated as missing data in the combined analyses. In total, 1006 newly determined non-coding sequences have been generated in this study (GenBank Accession Numbers KY458990-KY459009; Table 1). To expand our dataset, orthologous sequences generated in this study (GenBank Accession Numbers KY458990-KY459995) were downloaded from GenBank and used as outgroups.

Alignments and sequence characterization

Sequences were aligned using Muscle 3.8.31 (Edgar 2004) under default settings. All 56 genes were analyzed separately and concatenated. The concatenated alignment was divided into 3 datasets: 1) 33 INs combined, 2) 23 IGs regions combined, and 3) INs and IGs combined. All alignments were visually corrected, and poorly aligned positions and indels were removed with Gblocks 0.91 b (Castrasana 2000) using default settings. Statistical attributes of the nucleotide sequence datasets were estimated with MEGA 7 (Kumar et al. 2016) and DAMBE 7.0.5 (Xia 2017) was used to check for substitution saturation.

Phylogenetic analyses based on consensus and coalescent methods

Phylogenetic trees for individual and concatenated loci were reconstructed with maximum-likelihood (ML) and Bayesian methods in RAxML 8.0.12 (Stamatakis 2006; Silvestro and Michalak 2010) and Mr Bayes 3.2.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), respectively. The best-fit models of sequence evolution for each locus were selected under the Akaike Information Criterion (AIC) (Akaike 1974; Posada and Buckley 2004) with jModeltest 1.1.0 (Posada 2008, 2009). The chosen models and their parameters were applied to infer both ML and Bayesian trees. For tree reconstructions based on concatenated datasets, each locus was treated as a separate partition with its own substitution model. ML trees were calculated with the heuristic algorithm, 10 random-addition sequence replicates and TBR branch swapping. Tree reliability was assessed using a bootstrap (BS) analysis with 100 replicates (Felsenstein 1985). For Bayesian analyses, we used 3-heated chains and a single cold chain in all Metropolis-coupled Markov chain Monte Carlo (MCMC) runs. We performed 3 independent runs for each dataset, each for 2 million generations with parameter sampling every 100 generations. The average standard deviation of split frequencies was close to 0.001 when the runs were finished. The first 25% of the trees were discarded as burn-in. A 50% majority-rule consensus of post burn-in trees was constructed to summarize the posterior probability (PP) for each split.

In addition to the traditional concatenation methods, we applied 2 coalescent-based species-tree estimation methods, that is, Accurate Species TRee Algorithm (ASTRAL-II) (Mirarab et al. 2014) and Species Tree estimation using Average Ranks of coalescence (STAR) (Liu et al. 2009). The ASTRAL analyses used the unrooted gene trees as the input file and the Maximum Quartet Support Species Tree (MQSST) was searched. The STAR analyses, conducted in STRAW (Shaw et al. 2013), were performed with multilocus BSs (Sco 2008) to estimate statistical support. For both analyses, individual gene trees for each of the non-coding sequences were estimated using RAxML 8.0.12 under the GTR+G model with 1000 BS replicates.
Phylogenetic analyses based on question-specific approach

Increasing the number of investigated loci does not always allow for better resolution of phylogenetic relationships, particularly when single locus analyses reveal contrasting results. In such cases, building and investigating question-specific datasets may be a more powerful approach to resolving questionable branching patterns (Chen et al. 2015). Chen et al. (2015) proposed 2 question-specific strategies to resolve such problematic nodes. In the “hypothesis-control approach” loci whose gene trees do not support any of the hypotheses for given question are removed, whereas in the “node-control strategy” only loci are selected whose gene trees recover a control node. The second approach is more relaxed than the first (Chen et al. 2015), and hence we selected the more strict “hypothesis-control approach” to address the branching patterns among NWM families and cebid subfamilies. Accordingly, loci whose phylogenetic trees do not support any of the 3 hypotheses for the relationships among platyrhine families (Figure 1Aa–Ac) and cebid subfamilies (Figure 1Ba–Bc) were removed. The resulting 2 question-specific datasets were used to conduct phylogenetic analyses as described above.

Divergence time estimation

Divergence times for the NWM radiation were estimated using a relaxed lognormal molecular clock in BEAST 2.4.7 (Drummond and Rambaut 2007). We assumed a GTR+I+G model of sequence evolution with 4 rate categories. Uniform priors were employed for GTR substitution parameters (0, 100), the gamma shape parameter (0, 100) and the proportion of invariant sites parameter (0, 1). The uncorrelated lognormal relaxed molecular clock model was used to estimate substitution rates for all nodes in the tree, with uniform priors on the mean (0, 100) and standard deviation (0, 10) of this clock model. We employed the Yule (pure-birth process) of speciation as the tree prior and a UPGMA tree to construct a starting tree.

We applied 6 calibration points that were used in earlier studies (Springer et al. 2012; Benton et al. 2015; Byrne et al. 2016) and derived from the fossil record. For all 6 nodes, a uniform distribution prior was selected. We used 1) the origin of Anthropoidea: minimum = 33.9 million years ago (Ma), maximum = 66.0 Ma (Benton et al. 2015), 2) Homo–Pan split: minimum = 5.11 Ma (Springer et al. 2012), maximum = 10.0 Ma (Benton et al. 2015), 3) the origin of Hominidae: minimum = 11.6 Ma, maximum = 28.5 Ma (Springer et al. 2012), 4) the origin of Catarrhini: minimum = 24.44 Ma, maximum = 34 Ma (Benton et al. 2015), 5) the origin of Pitheciidae: minimum = 15.7 Ma, maximum = 26.0 Ma (Byrne et al. 2016) and 6) the origin of Cebinae: minimum = 12.5 Ma, maximum = 26.0 Ma (Byrne et al. 2016). Posterior distributions of parameters were approximated by sampling from 2 independent MCMC analyses. Each analysis ran for 100 million generations with parameters logged every 1000 generations. Convergence was assessed in Tracer 1.5 (Rambaut and Drummond 2009) after excluding the first 25% as burn-in. A consensus chronogram with node height distribution was generated with TreeAnnotator 2.4.7 and visualized with FigTree 1.4.3 (Rambaut 2012).

For studies with animals

All applicable institutional and/or national guidelines for the care and use of animals were followed. No animals were sacrificed for this study. Blood samples were taken during routine health checks by experienced veterinarians and not specifically for this study. Tissue samples were obtained from deceased animals. All research adhered to the legal requirements of the countries in which research was conducted. The study was carried out in compliance with the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and the principles of the American Society of Primatologists for the ethical treatment of nonhuman primates.

Results

Characteristics of the nuclear non-coding data

General information about the 56 nuclear non-coding regions from 18 NWM species is summarized in Table 2. Alignments for individual loci varied from 422 bp (chr2-8) to 1251 bp (2p22.3). After the removal of poorly aligned positions and indels alignment lengths varied from 234 bp (chr3-5) to 773 bp (ENC15). Alignments differed also in the number of parsimony-informative sites, ranging from 0 (chr4-2) to 96 (chr18-3). An A-T bias (average ratio = 56.4%) was apparent in most loci, as typically observed in non-coding regions (Yu and Zhang 2005; Matthee et al. 2007; Wang et al. 2012). The optimal model of sequence evolution varied by locus, suggesting different evolutionary histories of individual loci. Pairwise K2P distances among all NWM species ranged from 1.2% (chr18-3) to 11.7% (chr18-3), with an average of 5.6%.

The concatenation of all non-coding regions (alignment 3) recovered a total of 25, 204 sites, of which 2741 (10.9%) were parsimony-informative. The average K2P distance is 4.8%. The 33 concatenated INs (alignment 1) resulted in 14, 320 sites, comprising 1584 (10.9%) parsimony-informative sites and showing an average K2P distance of 4.2%. In comparison, the concatenated 23 IGs (alignment 2) resulted in 10, 684 sites, of which 1157 (10.8%) were parsimony-informative and the average K2P distance was 5.2%. Therefore, a slightly slower evolutionary rate for INs relative to IGs was observed, consistent with the fact that there are selective constraints on INs but not on IGs due to the presence of pre-mRNA secondary structures (Ometto et al. 2006). We found that there was no evidence of substitution saturation across our INs and IGs dataset based on the Iss statistic in DAMBE, as evidenced by the significantly lower values of Iss (index of substitution saturation) than Iss.c (critical value for symmetrical tree topology) (IN: 0.2584 < 0.8558, P = 0, P < 0.001; IG: 0.2898 < 0.8473, P = 0, P < 0.001).

Phylogenetic relationships based on the consensus and coalescent methods

To test different hypotheses about the branching pattern among NWM families and Cebidae subfamilies, different datasets (INs, IGs, and INs and IGs combined) were investigated with concatenation (ML and BA) and coalescent-based tree-inference methods (ASTRAL and STAR) (Figures 2 and 3). On family level, a consistent tree topology was obtained from all 3 datasets and the 4 applied methods, although the sister grouping of Atelidae and Pitheciidae was weakly supported. For relationships among subfamilies within the Cebidae, weakly supported and inconsistent results were obtained for IN and IG datasets using the concatenation and coalescence tree-inference methods. Even the 2 coalescent-based tree reconstruction methods resulted in inconsistent tree topologies for the same dataset (ASTRAL: Figure 2a, b; STAR: Figure 2a, c). However, statistical support for the nodes of interest, that is, the relationships among families and cebid subfamilies, were generally low, except for the tree topology derived from the IG dataset where strong support was found for a sister grouping of Atelidae and...
matchA and matchB, loci that match branching patterns presented in figure 4, respectively.
Pitheciidae (hypothesis in Figure 1Ac), and Aotinae and Callithricinae within Cebidae (hypothesis in Figure 1Ba), at least in ML and Bayesian reconstructions (Figure 2c: ML BS: 78% and 95%, Bayesian PP: 0.95 and 0.98, respectively).

Phylogenetic analyses inferred from combining all non-coding regions using both concatenation (ML and BA) and coalescence (ASTRAL and STAR) methods resulted in an identical tree topology (Figure 3). All analyses suggested a division of NWMs into the 3 monophyletic families Cebidae, Atelidae and Pitheciidae, and consistently supported the sister grouping of Atelidae and Pitheciidae to the exclusion of Cebidae, corresponding to hypothesis Ac (ML BS: 89%, Bayesian PP: 1.00, STAR BS: 68%, ASTRAL BS: 71%). Within Cebidae, the subfamilies Aotinae and Callithricinae form a clade to the exclusion of Cebinae in all analyses, corresponding to hypothesis Ba (ML BS: 70%, Bayesian PP: 0.97, STAR BS: 51%, ASTRAL BS: <50%).

Phylogenetic relationships based on question-specific approach

The individual analyses of 56 non-coding regions produced a variety of tree topologies with low levels of nodal support, probably owing to the limited phylogenetic information harbored in a single region (Supplementary Figure S1). For interfamilial and inter-subfamilial relationships of NWMs, the analyses of 22 loci (Table 2 and Figure 4) do not support any of the 6 hypotheses presented in Figure 1. In accordance with the question-specific strategy (Chen et al. 2015), these “non-matching” loci were excluded, with the aim of improving the signal strength of the data for the questions of interest. The resulting data sets comprise 30 loci for the interfamilial relationships (13,796 bp in total; referred to as matchA dataset hereafter) and 24 for the inter-subfamilial relationships (11,108 bp in total; referred to as matchB dataset hereafter) (Table 1). MatchA and matchB contain 1,711 bp (12.4%) and 1,386 bp (12.5%)
Among mammalian phylogenies, those characterized by rapid species radiations have long been a challenging problem. Our study based on a set of nuclear non-coding loci, including INs and IGs regions, using both concatenation and coalescence tree-inference methods as well as a question-specific approach, provides insights into the phylogenetic relationships among NWM families and subfamilies.

In our study, we obtained consistent branching patterns among NWM families from IGs and INs and different tree-inference methods, but different relationships were recovered for the 3 Cebidae subfamilies (Figure 2). However, a consensus tree supporting hypotheses Ac and Ba was consistently recovered from all the analyses of the combined IG and IN datasets (Figure 3), albeit with varying degrees of statistical support. Among the 3 NWM families, a closer affinity between Pitheciidae and Atelidae than either is to Cebidae was obtained, that is, hypothesis Ac. This result is in agreement with studies using nuclear protein-coding loci and combined nuclear and mitochondrial loci (Schneider et al. 1993, 1996; Harada et al. 1995; Opazo et al. 2006; Schrago 2007), but disagrees with some other studies relying on nuclear protein-coding genes (Poux and Douzery 2004; Perelman et al. 2011) and genomic segments (Schrago et al. 2014), mitochondrial genome data (Finstermeier et al. 2013), non-coding genes (Wildman et al. 2009; Jameson-Kiesling et al. 2015), transposable elements analyses (Osterholz et al. 2009) and the combined datasets of different classes of genes (Fabre et al. 2009; Chatterjee et al. 2009; Springer et al. 2012; Aristide et al. 2015), in which 2 alternative hypotheses Aa or Ab were suggested instead. In this study, the proposals of hypotheses Aa and Ab were not found in any of the datasets (Figures 2 and 3). In contrast, the results from the concatenation and coalescence analyses supported hypothesis Ac. Notably, the nodal supports for hypothesis Ac for the combined IG dataset are relatively high in ML and Bayesian analyses (BS: 78%, PP: 0.95).

Depending on the taxa examined and analytical methods used, previous studies have supported each of the 3 hypotheses of the relationships among the 3 subfamilies of Cebidae (Figure 1B). In the present study, a sister grouping of Aotinae and Callithrichinae to the exclusion of Cebinae, that is, hypothesis Ba, was recovered from the concatenation analyses and STAR analyses based on the IG dataset (Figure 2c) as well as those from all the analyses of the combined INs and IGs (Figure 3) with high nodal supports in the concatenation analyses. This result is in agreement with previous nuclear and mitochondrial analyses (Perelman et al. 2011; Perez et al. 2012; Springer et al. 2012; Finstermeier et al. 2013; Jameson-Kiesling et al. 2015; Aristide et al. 2015). The alternative hypothesis Bc, that is, the sister grouping of Callithrichinae and Cebinae, was found in concatenation and STAR analyses of INs and ASTRAL analyses of IGs (Figure 2a) but with low nodal support in all cases. The hypothesis Bb for the grouping of Aotinae and Cebinae is only recovered here in ASTRAL analyses of INs (Figure 2b).

Previous phylogenetic studies of NWMs have been based mainly on concatenation tree-inference methods (Schneider et al. 1993, 1996; Poux and Douzery et al. 2004; Opazo et al. 2006; Schrago 2007; Fabre et al. 2009; Osterholz et al. 2009; Wildman et al. 2009; Chatterjee et al. 2009; Perelman et al. 2011; Springer et al. 2012; Finstermeier et al. 2013; Jameson-Kiesling et al. 2015; Aristide et al. 2015). A coalescence-based method (*BEAST) has been applied by Perez et al. (2012) using data from published studies (Opazo et al. 2006; Wildman et al. 2009; Perelman et al. 2011), whereas Schrago et al. (2014) analyzed 92Mbp of genomic segments of a limited
number of samples using STAR and MPEST, which resulted in an unresolved tree topology. In our study, analyses of the independent nuclear non-coding datasets from previous studies using both traditional concatenation (ML and Bayesian) and 2 recently-developed summary coalescence methods (ASTRAL and STAR) provide an opportunity to examine their application in addressing the phylogenetic resolution among NWMs. Intriguingly, our analyses suggest a consensus tree on interfamilial relationships and an incongruence concerning inter-subfamilial relationships between concatenation and coalescence phylogenetic results in the case of IN and IG analyses, respectively (Figure 2). Phylogenetic inconsistent between different studies may be caused by different markers, incomplete lineage sorting and different tree-building methods. Phylogenetic incongruence between concatenation and coalescence trees has been reported in mammalian orders (McCormack et al. 2012; Song et al. 2012; Tsagkogeorga et al. 2013; Kumar et al. 2013; Giarla and Esselstyn 2015), snakes (Pyron et al. 2014; Ruane et al. 2014), birds (Haddrath and Baker 2012; Fuchs et al. 2013; Jarvis et al. 2014) and plants (Springer and Gatesy 2012; Zhao et al. 2013; Zhong et al. 2013, 2014; Xi et al. 2013, 2014; Wickett et al. 2014; Tang et al. 2015; Simmons and Gatesy 2015; Nater et al. 2015). It has been thought that the probability of the occurrence of such conflicting signals would increase when splitting times between taxa are short (McCormack et al. 2012; Song et al. 2012; Tsagkogeorga et al. 2013; Kumar et al. 2013; Xi et al. 2013, 2014; Pyron et al. 2014; Tang et al. 2015; Nater et al. 2015), as typical for platyrhines. It is noted that when all non-coding regions (INs and IGs) are combined, a congruence supporting hypotheses Ac for interfamilial and Ba for inter-subfamilial relationships from both tree-inference methods was retrieved (Figure 3). Perez et al. (2012); Perez and Rosenberger (2014) mentioned that the discrepancy between coalescence and concatenation methods in resolving the rapid radiation events among NWMs is not unexpected most likely as a result of incomplete lineage sorting. However, it should be also mentioned that the support values from the coalescence analyses in the present study are low (Figures 2 and 3), in contrast to the generally high support from the concatenation analyses. In fact, the coalescent-based analyses which use summary methods that estimate the species tree by combining individual gene trees have been thought to suffer from insufficient phylogenetic signal in the case of short gene regions for which the alignments will increase gene tree estimation error (Pollock et al., 2002; Roch and Warnow 2015). For our dataset, 56 individual loci were taken from previous studies (Wildman et al. 2009; Wang et al. 2012) and not all sequence length of these individual loci are longer than 1 kb. So we assume that the coalescent-based analyses of our dataset are more likely to suffer from insufficient phylogenetic signal in the case of very short gene regions, which may lead to the low nodal supports observed in coalescent-based analyses. An increasing number of informative loci used may increase the power of coalescence methods for our dataset to further phylogenetic resolution. We suggested that the dataset used maybe a more important factor for the phylogenetic studies dealing with the family and subfamily-level relationships of NWMs given that the combined dataset and the question-specific loci dataset retrieve the consistent relationships regardless of the tree-building methods.

Chen et al. (2015) developed a question-specific approach which operates by selecting those gene sequences that yield support for one of several predefined hypotheses, with the aim of concentrating the phylogenetic signal for a specific question and not allowing it to be swamped by individual gene history. By alleviating the incongruences associated with data size and the tree inference method, the authors demonstrated the enhanced performance of their method for resolving problematic relationships within jawed vertebrates. Interestingly, using 2 question-specific datasets in our study recovered a single tree in favor of the combined non-coding regions trees, regardless of the tree-inference methods used (Figure 3), providing further support for hypotheses Ac and Ba. Hence, the application of this approach showed its resolving power at the family and subfamilial level among platyrhines. The comparison of different datasets examined here found that the 2 question-specific datasets matchA and matchB demonstrated slightly higher variable sites (25.2% and 25.6%, respectively), and parsimony-informative sites (12.4% and 12.5%, respectively), than the INs (22.8% and 10.9%), IGs (23.5% and 10.8%) and all combined dataset (23.1% and 10.9%). Thus, it seems that question-specific methods may collect more phylogenetic signals to reconstruct the evolutionary history of NWMs.

In conclusion, our study provides support for some previously suggested relationships among families (a sister-group relationship between Pitheciidae and Atelidae) and subfamilies (a sister group relationship between Aotinae and Callitrichinae) within NWMs from the perspective of multiple non-coding loci and various tree-inference methods (STAR and ASTRAL) as well as a question-specific approach. Nonetheless, to clarify the NWMs phylogenetic framework still needs future evaluation by the addition of independent sequence data and a deeper taxonomic sampling.

Acknowledgments
We are grateful to the staff of the zoos in Cologne, Gettorf, Kunming, Landau, Magdeburg, Romagne, Rostock, Stockholm, and Toronto for providing valuable primate samples. This work was supported by the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (Large Research Infrastructure Funding). We would also like to thank Nicole Otto and Christiane Schwarz for help in the laboratory.

Funding
Financial support for this project was obtained from the Yunnan Provincial Education Fund project in Scientific Research to XPW (No. 2015Y003), the National Youth Talent Support Program to LY, and the German Primate Center to CR.

Author contributions
L.Y. and C.R. designed and managed the project. X.P.W., B.K.L., and C.R. helped with sample collection. X.P.W., C.R., and Y.P.L. performed the experiments. X.P.W. and J.Y.H. performed the data analyses. X.P.W., L.Y., C.R., B.K.L., and N.T. wrote the paper. All authors read and approved the final manuscript.

Supplementary material
Supplementary material can be found at https://academic.oup.com/cz.

References


Harada ML, Schneider H, Schneider MPC, Sampaio I, Czelusnak J et al., 1995. DNA analysis on the phylogenetic systematics of New World monkeys: support for the sister-grouping of Cebus and Saimiri from two unlinked nuclear genes. Mol Phylogenet Evol 4:331–349.


