A multi-locus phylogeny suggests an ancient hybridization event between Campephilus and melanerpine woodpeckers (Aves: Picidae)

Jérôme Fuchs a,b,c,d,* , Jean-Marc Pons a,b , Liang Liu e , Per G.P. Ericson f , Arnaud Couloux g , Eric Pasquet a,b

a UMR7205 «Origine, Structure et Evolution de la Biodiversité, Département Systématique et Evolution, Muséum National d'Histoire Naturelle, 55 Rue Buffon, 75005 Paris, France
b UMS MNHN/CNRS 2700 Outils et Méthodes de la Systématique Intégrative (OMSI), Muséum National d'Histoire Naturelle, 57 rue Cuvier, F-75231 Paris Cedex 05, France
c DST/NRF Centre of Excellence at the Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, Cape Town, South Africa
d Department of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Science Building, University of California, Berkeley, CA 94720-3160, USA
e Department of Statistics and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
f Department of Vertebrate Zoology and Molecular Systematics Laboratory, Swedish Museum of Natural History, P.O. Box 50007, SE-104 05 Stockholm, Sweden
g Genoscope, Centre National de Séquençage, 2, rue Gaston Crémieux, CPS706, 91057 Evry Cedex, France

ARTICLE INFO

Article history:
Received 12 October 2012
Revised 26 January 2013
Accepted 13 February 2013
Available online 26 February 2013

Keywords:
Campephilus
Gene tree
Hybridization
Introgression
Phylogeny
Species tree

ABSTRACT

The ever increasing number of analysed loci in phylogenetics has not only allowed resolution of some parts of the Tree of Life but has also highlighted parts of the tree where incongruent signals among loci were detected. Previous molecular studies suggested conflicting relationships for the New World genus Campephilus, being either associated to the Megapicini or Dendropocini. Yet, the limited number of analysed loci and the use of the concatenation approach to reconstruct the phylogeny prevented the disen- tanglement of lineage sorting and introgression as causal explanation of this topological conflict. We sequenced four mitochondrial, nine autosomal and three Z-linked loci and used a method that incorporates population level processes into the phylogenetic framework to understand which process (lineage sorting of genetic polymorphism or hybridization/introgression) best explains this conflict. Our analyses revealed that the autosomal FGB intron-7 and to a lesser extent the Z-linked loci have a different phylo- genetic history from the mitochondrial loci and some other nuclear loci we analysed. We suggest that this conflicting pattern is the result of introgression consecutive to a hybridization event at the time when members of the Campephilus and melanerpine (Melanerpes and Sphyrapicus) lineages colonized the New World. The case of Campephilus highlights that the mitochondrial genome does not always carry the 'wrong' phylogenetic signal after a past hybridization event. Indeed, we here emphasise that the sig- nature of such event can also be detected in the nuclear genome. With the ongoing increase in the num- ber of loci analysed in phylogenetic studies, it is very likely that further cases will be discovered. Our current results indicate that (1) the genus Campephilus is related to the Asian genera Blythipicus, Chrysoco- laptes and Reinwardtipicus, in accordance with morphological data and (2) that the nuclear genome of Campephilus is likely the mixture of two unrelated lineages. Yet, further work with a denser sampling of loci is necessary to evaluate the extent of the Sphyrapicus/Melanerpes lineage nuclear genome that introgressed into the Campephilus genome.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The greatly increased number of loci analysed in phylogenetic and phylogeographic studies has not only allowed resolution of certain parts of the Tree of Life (e.g., Hackett et al., 2008) but has also highlighted parts of the tree where incongruent signals among loci or data sets exist (e.g., Jeffroy et al., 2006). Analytical issues aside, at least four non-exclusive biological processes may explain the occurrence of incongruent nodes in phylogenetic trees: horizontal gene transfer, gene duplication, hybridization or introgression and lineage sorting (Funk and Omland, 2002). Horizontal gene transfer is a well-known phenomenon in prokaryotes (Boucher et al., 2003; but see also Moran and Jarvik, 2010 for examples involving eukaryotes). The inclusion of paralogous sequences in phylogenetic studies may lead to erroneous estimations of species trees; the multiple copies that originated from a duplication event affect the haplotype/allele inference and encompass different mutational constraints (e.g., Bensasson et al., 2001). Hybridization followed by introgression could also have effects on phylogenetic reconstruction although not all parts of the genome have the same
sensitivity to hybridization. The maternally inherited mitochondrial genome may be more prone than the bi-parentally inherited nuclear genome to be in conflict with the species tree, especially if hybridization is unidirectional or selection acts on it (Takahata and Slatkin, 1984; Ballard and Whitlock, 2004). A fourth main process, the stochastic sorting of gene lineages at speciation events is particularly problematic when multiple cladogenetic events occur in a short period of time (Rokas and Carroll, 2006). In the latter case, it is very likely that gene trees will not reflect the species tree (e.g., Felsenstein, 1979; Pamilo and Nei, 1988) and performing the analysis using a concatenation approach may not resolve this problem (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007). The woodpeckers (Piciformes: Piciniae) comprise a strongly supported monophyletic group of 180 species (25 genera) that is distributed in all major biogeographic regions except Australasia and Antarctica. A solid framework of phylogeny and biogeography of woodpeckers has been created over the last few years and four primary lineages are now usually recognised (Hemicircus, Dendrocopini, Malarpicini, Megapicini; Webb and Moore, 2005; Benz et al., 2006; Fuchs et al., 2007, 2008). Yet, Fuchs et al. (2007) also reported a conflict between the mitochondrial and nuclear genome concerning the relationships of the New World genus Campephilus. Indeed, mitochondrial loci (Webb and Moore, 2005; Fuchs et al., 2007) strongly favoured Campephilus being nested within the Megapicini, an otherwise Indo-Malayan endemic clade comprising three genera (Blythipicus, Chrysocolaptes and Reinwardtipicus). In contrast, the two nuclear loci analysed by Fuchs et al. (2007) recovered discordant results: Campephilus was closely related to the Dendrocopini, a lineage comprising the New World sapsuckers (Sphyrapicus) and melanerpine woodpeckers (Melanerpes), as well as the members of the widely distributed pied woodpecker assemblage (Dendrocopos, Dendrocopinae, Picoides, Veniliornis). Myoglobin intron-2 (MB) suggested that Campephilus is the sister-group to the whole Dendrocopini clade, whereas beta-fibrinogen intron-7 (FGB) nested it within the Dendrocopini, as sister-group to the Sphyrapicus/Melanerpes clade. Because three out of the four primary lineages of woodpeckers (Dendrocopini, Malarpicini, Megapicini) appeared in a relatively short period of time (Fuchs et al., 2007), this conflict could be explained by both: (1) lineage sorting of polymorphic loci at the time of the earliest splits within the Piciniae, or by (2) an ancient hybridization involving members of the new Dendrocopini and Megapicini with further introgression of some parts of the genome in one of the lineages.

Here we aim to understand whether the conflicting signal encountered concerning the phylogenetic relationships of Campephilus could be the result of lineage sorting of genetic polymorphism at speciation events or the consequence of a past hybridization event. For this purpose, we sequenced additional mitochondrial and nuclear loci for a representative set of woodpecker genera and analysed the data using a method that incorporates population level processes into the phylogenetic framework.

2. Materials and methods

2.1. Sampling of taxa and loci

We sampled representatives of all primary Picinae lineages (Supplementary Table 1, Webb and Moore, 2005; Benz et al., 2006; Fuchs et al., 2007). We sampled three species of Campephilus representing two of the three primary lineages within this genus (Fleischer et al., 2006); the third lineage (principalis-impertullis) is composed of two species that are thought to be extinct. Representatives of the Picumninae and Jyninae were included as proximate outgroups (e.g., Fuchs et al., 2007). Trees were rooted with a representative of the Indicatoridae, the sister group of the Picidae (e.g., Ericson et al., 2006; Hackett et al., 2008).

We sampled loci that are located on different chromosomes on the chicken genome in order to have unlinked loci to reconstruct the phylogeny (Supplementary Table 2). Comparison of the gene mapping on the ongoing assembly of the zebra finch genome, a species closer to woodpeckers than the chicken (Hackett et al., 2008), indicates that all the loci we sequenced for this project are located on the same chromosomes in the chicken and in the zebra finch. Extraction, PCR-amplifications, and cycle sequencing made use of standard protocols and are identical to those reported in Fuchs et al. (2007, 2008).

2.2. Sequence alignment and analyses

Alignment was performed by eye for all loci using Se-Al v2.0a11 (Rambaut, 2007). In cases where length variance heterozygosity was detected within an individual, we only used the allele without the autapomorphic indel for the concatenated analyses. The alignments are available from the first author upon request. Since FGB had a very different phylogenetic signal from all other loci we analysed, we re-aligned FGB7 by using its reverse-complement sequence (Landan and Graur, 2007). We retrieved an alignment that was very similar to the initial one (differences involved how some poly-T and poly-A regions were aligned, but the topologies resulting from the analyses of the two alignments were identical, tree not shown). This result suggests that misalignment cannot explain the conflicting signal of FGB7.

For each locus, the most appropriate substitution models were determined using the Bayesian Information Criterion (Posada and Buckley, 2004) as implemented in Topal v2.5 (Milne et al., 2009). Prior to the phylogenetic analyses, we tested for recombination in the nuclear loci using the GARD (Kosakovsky Pond et al., 2006) algorithm implemented in the HyPhy package on the Datamonkey webserver (http://www.datamonkey.org; Delport et al., 2010).

Phylogenetic analyses including individual gene trees and a concatenated approach were conducted using maximum likelihood and Bayesian inference, as implemented in RAxML v7.0.4 (Stamatakis et al., 2006; Stamatakis et al., 2008), MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2003; Ronquist and Huelsenbeck, 2003), and Mr-EST (Liu et al., 2010). For the MrBayes analyses, four Metropolis-coupled MCMC chains (one cold and three heated) were run for five to thirty million iterations with trees sampled every one hundred iterations.

We used the Bayes Factor (BF) to determine if partitioning the mitochondrial data set by gene and codon position had a better fit to the data than without any partition scheme; a value greater than 4.6 for ln BF was considered as strong evidence against the simpler model (M0) (Jeffreys, 1961; Nylander et al., 2004; Brown and Lemmon, 2007).

To ensure that we reached the target posterior distribution, we checked that the potential scale reduction factor (PSRF) approached 1.0 for all parameters and that the average standard deviation of split frequencies converged towards zero. We also used Tracer v1.5 (Rambaut and Drummond, 2007) to ascertain that our sampling of the posterior distribution had reached a sufficient effective sample size (ESS).

For the species tree analyses, we used a new pseudo-likelihood method, MP-EST, that incorporates population level processes into the phylogenetic framework by using the coalescent to estimate a species tree (topology and branch lengths in coalescent units) based on a set of gene trees. This new method incorporates uncertainty of gene trees using a two-step bootstrap technique (Liu et al., 2010). The bootstrap gene trees generated from RAxML for the 16 loci were used as the input data to build MP-EST trees. The MP-EST trees were then summarized by a majority-rule consensus tree
constructed by the subroutine **Consense** in **Phylip** (Felsenstein, 2005). In addition, the MP-EST analysis was conducted for two reduced data sets by excluding (1) the locus FGB7, and (2) the locus FGB8 and the three Z-linked loci.

### 2.2.2. Topology test

We used **BEAST** v1.6 (Drummond et al., 2002, 2006; Drummond and Rambaut, 2007) to estimate for each locus the coalescent times for the genera *Campephilus* and *Melanerpes/Sphyrapicus*. We assigned the best fitting model to each of the thirteen loci (mitochondrial genes were considered a single locus). As a calibration point we used the *Hemicircus*remaining Picinae split, modelled as a normal distribution with the mean and standard deviation set to 13.4 mya and 1.7, respectively (Fuchs et al., 2007). We assumed a normal distribution with the mean and standard deviation set to 2.2.

### 3. Results

The length of the alignments, best fit models, model parameters and likelihood scores for each locus are indicated in Table 1.

### 3.1. Mitochondrial loci

The mitochondrial fragments analysed correspond to the positions 4007–5047 (ND2), 5466–6160 (CO1), 8015–8698 (ATP6) and 9551–9902 (ND3) of the *Dryocopus pileatus* mitochondrial genome (GenBank accession number NC008546, Gibb et al. 2007). We could not obtain the CO1 sequence for *Hemicircus canente*, ND3 for *Picoides minor*, and ATP6 for *Indoracinus maculatus*. Since we could not obtain ATP6 for the outgroup, we used the ATP6 sequence of another *Indoracinus* species (*I. minor*). All species but *Melanerpes carolinus* and *M. flavifrons* possess the pyrimidine insertion (T for *Indoracinus maculatus*, *M. formicivorus*, and *Reinwardtia validus*, C for all other species analyzed) at position 174 in ND3 reported for several clades of birds (Mindell et al. 1998); this extra-nucleotide was removed prior to phylogenetic analyses. The concatenated mitochondrial sequences retained for analyses were 2271 bp long. The 50% majority rule consensus tree from the partitioned Bayesian analyses of the mitochondrial data set (Fig. 1, Supplementary Fig. 1; 10 partitions, harmonic mean −ln = 30796.77) recovered the four primary lineages highlighted in previous analyses of mitochondrial sequences: *Hemicircus*, Dendropicini, Malarpicini, Megapicini (including *Campephilus*); relationships among these four lineages did not receive significant posterior probabilities (Fig. 1). Constraining *Campephilus* to be related to the Dendropicini had a significantly worse harmonic mean than without constraint (ln = 66.638).

### 3.2. Nuclear loci: Individual

Sequence data were gathered from nine autosomal loci (total length: 6329 bp). We could not obtain the *mos* sequence for *D. pileatus*, *S. varius* and *M. formicivorus*. The IRBP sequence of *I. maculatus* contained a 28 bp insertion that disrupted the reading frame. We were not sure if this sequence represents a paralog or a homologous sequence that became non-functional in the Indicatoridae. Consequently, we did not include this sequence in the analyses and rooted the tree with the sequence of *Jynx torquilla* in the IRBP.

---

**Table 1**

Sequence length of loci, best fit model as estimated using **TOPALI v2.5** (Milne et al., 2009) and the BIC criterion, **MRBAYES** and likelihood scores. **Bayesian Likelihood** values refer to the harmonic mean. **NA** means Non-Applicable. **Asterisks** indicate significant **Bayes Factor** values when comparing partitioned and non-partitioned models. **OTUs**: Operational Taxonomic Units.

<table>
<thead>
<tr>
<th>MitDNA locus</th>
<th>ND2</th>
<th>CO1</th>
<th>ATP6</th>
<th>ND3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td></td>
<td>1041</td>
<td>695</td>
<td>684</td>
</tr>
<tr>
<td>OTUs</td>
<td>41</td>
<td></td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>Model</td>
<td>GTR+I</td>
<td></td>
<td>GTR+I</td>
<td>GTR+I</td>
</tr>
<tr>
<td>Model 1st pos.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2nd pos.</td>
<td></td>
<td>Trn+I</td>
<td></td>
<td>Trn+I</td>
</tr>
<tr>
<td>Model 3rd pos.</td>
<td></td>
<td>GTR+I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>−1402.53</td>
<td></td>
<td>−7110.29</td>
<td>−8948.57</td>
</tr>
<tr>
<td>BI partitioned</td>
<td>−13488.50</td>
<td>−7116.20</td>
<td>−8591.16</td>
<td>−4056.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Autosomal locus</th>
<th>MB</th>
<th>GAPDH</th>
<th>Mos</th>
<th>TGFb2</th>
<th>PER</th>
<th>FGB5</th>
<th>FGB7</th>
<th>PEPCk</th>
<th>IRBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td></td>
<td>719</td>
<td>463</td>
<td>605</td>
<td>611</td>
<td>820</td>
<td>679</td>
<td>952</td>
<td>706</td>
</tr>
<tr>
<td>OTUs</td>
<td>41</td>
<td>47</td>
<td>38</td>
<td>41</td>
<td>48</td>
<td>45</td>
<td>42</td>
<td>46</td>
<td>40</td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td>KSI+F</td>
<td>GTR+I</td>
<td>Trn+I</td>
<td>GTRef+I</td>
<td>GTR+I</td>
<td>GTR+I</td>
<td>KSI+I</td>
<td>KSI+I</td>
</tr>
<tr>
<td>Model 1st pos.</td>
<td>NA</td>
<td>NA</td>
<td>HKY+I</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Trn</td>
</tr>
<tr>
<td>Model 2nd pos.</td>
<td>NA</td>
<td>NA</td>
<td>KSI</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>KSI+I</td>
</tr>
<tr>
<td>Model 3rd pos.</td>
<td>NA</td>
<td>NA</td>
<td>KSI+I</td>
<td>NA</td>
<td>NA</td>
<td>KSI+I</td>
<td>NA</td>
<td>KSI+I</td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>−3065.13</td>
<td>−2139.41</td>
<td>−2270.38</td>
<td>−2554</td>
<td>−3732.6</td>
<td>−2798.67</td>
<td>−4322.15</td>
<td>−3468.99</td>
<td>−2468.71</td>
</tr>
<tr>
<td>BI partitioned</td>
<td>NA</td>
<td>NA</td>
<td>−2202.47</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>−2216.55</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Z-linked locus</th>
<th>MUSK</th>
<th>BRM</th>
<th>ACO1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>593</td>
<td></td>
<td>391</td>
</tr>
<tr>
<td>OTUs</td>
<td>42</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>Model</td>
<td>KSI+I</td>
<td></td>
<td>GTRef+I</td>
</tr>
<tr>
<td>BI</td>
<td>−2993.47</td>
<td>−1752.09</td>
<td>−5376.79</td>
</tr>
<tr>
<td>BI partitioned</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
analyses. We did not find any evidence of recombination in the nuclear loci using the GARD algorithm.

Analyses performed on the individual loci were generally congruent and provided topologies that are more in agreement with
the mitochondrial tree (Supp. Figs. 2–10), although the relationships of *Campephilus* were generally poorly supported. The one noticeable exception involves FGB7, where a clear conflicting signal appeared, as *Campephilus* was closely related to *Sphyrapicus* and *Melanerpes* and not to the Megapicini, as in the mitochondrial tree. Constraining *Campephilus* to be related to the Megapicini had a significantly worse mean harmonic value than without the constraint in the FGB7 gene tree (ln BF = 22.74), marginally in the MB (ln BF = 4.89) gene tree but not in the GAPDH (ln BF = 3.419), mos (ln BF = 0.786), FGB5 (ln BF = 3.092), IRBP (ln BF = 4.235), PER (ln BF = 3.743) and TGFB2 (ln BF = 0.08) gene trees. Constraining *Campephilus* to be related to the Dendropicini had a significantly worse harmonic mean than without constraint for GAPDH (ln BF = 8.192), MB (ln BF = 4.75), PER (ln BF = 22.411) but not in FGB5 (ln BF = 2.465), IRBP (ln BF = 2.023) and mos (ln BF = −0.207).

### 3.3. Nuclear loci: Autosomal concatenated

The tree resulting from the analyses of the concatenated autosomal data set was well resolved with primary clades (Dendropicini, Malarpicini, Megapicini) being support by bootstrap values 85% and posterior probabilities of 0.99–1.0 (Fig. 2). Analysing all loci, the genus *Campephilus* was the sister group of the Dendropini (BP = 86, PP = 1.0). Excluding FGB7 had a strong effect on the relationships of *Campephilus* as it became sister (BP = 83, PP = 1.0) to the Dendropini/Malarpicini clade (BP = 54, PP = 0.59).

#### 3.4. Z-linked loci

Sequence data were gathered from three Z-linked loci (total length: 2088 bp). Only a few nodes were resolved in the Bayesian analyses of the BRM locus (harmonic mean −ln = 1752.09, Suppl. Fig. 11). *Campephilus* was related to *Sphyrapicus/Melanerpes* (BP = 77, PP = 0.81), whereas *Blythipicus* and *Chrysocolaptes/Remwardtipicus* were part of the general polytomy. More supported nodes were found in the MUSK and ACO1 gene trees. In MUSK (Supplementary Fig. 12), *Campephilus* was the second lineage to branch off in the Picinae (BP < 50, PP = 0.92); the *Melanerpes/Sphyrapicus* clade was sister to the Dendropini (BP = 91, PP = 1.0) and the Megapicini was sister to the whole Melanerpini/Dendropicini/Malarpicini clade (BP = 78, PP = 1.0). In ACO1 (Supplementary Fig. 13), *Campephilus* clustered with the Megapicini and Malarpicini in a polytomy (BP = 70, PP = 0.99) whereas the *Melanerpes/Sphyrapicus* clade grouped with the Dendropini (BP = 99, 0.01).
The topology resulting from the concatenated analyses of the three Z-linked loci is very similar to the MUSK topology but the support values at nodes were generally increased (Fig. 3). Campephilus was the second lineage to branch off in the Picinae.
(BP = 90, PP = 1.0); the Melanerpes/Sphyrapicus clade was sister to the Dendrocolapini (BP = 100, PP = 1.0) and the Megapicini was sister to the whole Melanerpesini/Dendrocolapini/Malarpicini clade (BP = 53, PP = 0.99). Constraining Campephilus to be related to the Megapicini tree had not a significantly worse mean harmonic in BRM (ln BF = 2.7881) and MUSK (ln BF = 2.465). Constraining Campephilus to be related to the Dendrocolapini had a significantly worse mean harmonic value than without constraint for ACO1 (ln BF = 38.402) and MUSK (ln BF = 23.332). In the concatenated analyses of the Z-linked loci, constraining Campephilus to be related to Megapicini or the Dendrocolapini had a worst mean harmonic value than the best topology (In BF = 5.455 and ln BF = 64.108, respectively) although only marginally in the former.

3.5. All data: Species tree and concatenated approaches

The MP-EST tree constructed from 16 loci (Fig. 4a) is very similar to the concatenation tree (Supplementary Fig. 14), with the exception that clade formed by the Dendrocolapini and Malarpicini is unresolved in the MP-EST tree, while it is highly supported in the concatenation tree (PP = 1.0, Supplementary Fig. 14). The MP-EST trees for the reduced data sets with either FGB7 or FGB7 and the Z-linked loci excluded also show unresolved relationships among the primary lineages (Campephilus, Dendrocolapini, Malarpicini, Melanerpesini, and Megapicini) with bootstrap support values being less than 0.4 (Fig. 4b and c) but there is now a tendency for Campephilus to cluster as the sister group of the Megapicini. The analyses performed on the concatenated data sets excluding FGB7 (Supplementary Fig. 15) or FGB7 and Z-linked loci excluded (Supplementary Fig. 16) yielded trees that were similar to their concatenated equivalents, excepted that the sister-group relationships between Campephilus and the Megapicini only appeared when FGB7 and the Z-linked loci were excluded.

Using Phybase (Liu and Yu, 2010), we tested through a simulation (10,000 gene trees) the null hypothesis that the observed clade (Campephilus related to Melanerpes/Sphyrapicus) at locus FGB7 is due to lineage sorting. The gene trees were simulated under the coalescent model from the MP-EST trees constructed by excluding FGB7 and the Z-link loci (12 loci, Fig. 4c). If the clade (Campephilus, Melanerpes/Sphyrapicus) can be explained under coalescent, we expect to see many simulated gene trees supporting the group (Campephilus, Melanerpes/Sphyrapicus). Otherwise, we reject the null hypothesis and conclude that the pattern is caused by hybridization. The result shows that the proportion of the simulated gene trees supporting the group (Campephilus, Melanerpes/Sphyrapicus) is 0.026, which allow us to reject the hypothesis that the conflict concerning the relationships of Campephilus could be explained with lineage sorting.

3.6. Coalescent time analyses

Results from the dating analyses performed with BEAST indicate that the coalescent times between the genera Campephilus and Melanerpes/Sphyrapicus vary substantially among genes (from 6.5 to 12.7 mya) although the 95% HPD are overlapping (Table 2). The FGB7 locus gave the second most recent coalescent time for the divergence between Campephilus and Melanerpes/Sphyrapicus (Table 2, Fig. 5). The two loci (FGB and BRM) that gave the most recent coalescent time for Campephilus and Melanerpes/Sphyrapicus are the two loci that support their direct sister-group relationships.

4. Discussion

Our analyses based on a method that incorporates lineage sorting as an explicit model, provided a well resolved phylogeny of the woodpeckers and highlighted that at least one autosomal marker (FGB7), and to a lesser extent the Z-loci, carry divergent signal with respect to the other autosomal and mitochondrial loci we analysed. Below we discuss in detail which processes can best explain this incongruence.

4.1. FGB7 and the Z-linked loci as markers with a different history

Our analyses revealed that FGB7 and the Z-linked loci have a different phylogenetic signal than all other autosomal loci, including
Campephilus informative concerning the relationships of the neighbouring FGB5 we analysed (although FGB5 was not very

Fig. 5. Times to most recent common ancestor for the genera Campephilus Melanerpes/Sphyrapicus. Estimates were obtained using BEAST v1.6. The best fitting model was assigned to each of the eight loci. The Hemricites/remaining Picinae split was used as a calibration point (normal distribution, mean: 13.4 mya and standard deviation 1.7 mya). We assumed a Yule Speciation Process for the tree prior and locus specific molecular clock models (see Table 2).

Table 2
Coalescent times (Time to Most Recent Ancestor) for the genera Campephilus and Sphyrapicus/Melanerpes obtained using BEAST v1.6. The best molecular clock model, as inferred from Bayes Factor, is indicated between parentheses. Estimates and their 95% HPD are given in million years before present.

<table>
<thead>
<tr>
<th>Locus</th>
<th>TMRCA (95% HPD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC1 (strict clock, ln B = 3.591)</td>
<td>11.2 (7.9–14.5)</td>
</tr>
<tr>
<td>BRM (lognormal, ln B = 13.153)</td>
<td>7.3 (3.1–12.9)</td>
</tr>
<tr>
<td>FGB5 (strict clock, ln B = 1.746)</td>
<td>11.8 (7.9–15.7)</td>
</tr>
<tr>
<td>FGB7 (strict clock, ln B = 3.223)</td>
<td>8.5 (5.6–11.7)</td>
</tr>
<tr>
<td>GAPDH (strict clock, ln B = 0.474)</td>
<td>8.8 (5.7–12.2)</td>
</tr>
<tr>
<td>IRBP (strict clock, ln B = 0.065)</td>
<td>10.7 (6.4–15.0)</td>
</tr>
<tr>
<td>Mitochondrial (4 partitions) (lognormal, ln B = 158.66)</td>
<td>12.0 (8.4–15.7)</td>
</tr>
<tr>
<td>mos (lognormal, ln B = 9.654)</td>
<td>11.3 (6.6–15.7)</td>
</tr>
<tr>
<td>MUSK (lognormal, ln B = 4.15)</td>
<td>12.3 (8.5–15.7)</td>
</tr>
<tr>
<td>Myoglobin (strict clock, ln B = 1.527)</td>
<td>10.5 (6.9–14.1)</td>
</tr>
<tr>
<td>PECK (strict clock ln BF = 2.735)</td>
<td>10.2 (7.0–13.5)</td>
</tr>
<tr>
<td>PER (strict clock, ln B = 1.268)</td>
<td>12.3 (9.0–16.0)</td>
</tr>
<tr>
<td>TGF52 (lognormal, ln B = 4.936)</td>
<td>9.0 (5.4–12.8)</td>
</tr>
</tbody>
</table>

the neighbouring FGB5 we analysed (although FGB5 was not very informative concerning the relationships of Campephilus with respect to the Melanerpin and Megapicini). The FGB introns, located on macrochromosome 4 of the chicken and zebra finch genome, have been used extensively over the past decade to resolve various parts of the avian tree of life (e.g. Fain and Houde, 2004), especially in woodpeckers (e.g. Benz et al., 2006; Fuchs et al., 2006, 2007, 2008; Prychitko and Moore, 1997). Recently, FGB7 has been used to propose two divisions within the Neovaves; Coronaves and Metaves (Fain and Houde, 2004). Further studies using additional nuclear loci (Ericson et al., 2006; Hackett et al., 2008) or partial or complete mitochondridial genomes (Brown et al., 2008; Morgan-Richards et al., 2008) could not recover these clades and also highlight some conflicting signal between FGB7 and other markers. The most obvious explanation would be paralogy in some clades. Yet, Morgan-Richards et al. (2008) cloned pcr-products from one ‘Metaves’ and one ‘Coronaves’ and found no evidence of gene duplication in any of the species. Furthermore, the FGB7 locus is a single copy in the chicken and Zebra Finch genomes as well as in vertebrate genomes as a whole, suggesting that paralogy is unlikely. We did not found evidence of recombination in FGB7 using the GARD algorithm. Consequently, only hybridization or lineage sorting of polymorphic loci could explain the incongruences observed among loci.

4.2. Conflicting signal for the relationships of Campephilus: hybridization or lineage sorting?

Introgressive hybridization and stochastic sorting of ancestral polymorphisms can result in similar topological incongruence between a gene tree and a species tree. The coalescent time between species for gene sequences subject to introgressive hybridization should be shorter than that for gene sequences not subject to introgressive hybridization (Holder et al., 2001; Zhang and Sota, 2007). Yet, when the introgression event occurred close to the speciation time, the coalescent time for gene sequences subject to hybridization should be very close or in the range of the variation we should expect from stochastic sorting of ancestral polymorphism. In the latter case, evidence of hybridization would be difficult to put forward unless the two involved lineages are not directly related. This is exactly what we observe for the genera Campephilus and Melanerpes/Sphyrapicus. Indeed, we showed that at least one autosomal locus (FGB7) carries a conflicting signal with respect to all the other loci we analysed. This pattern could be explained by an early hybridization event between Campephilus and the ancestor of the Melanerpes/Sphyrapicus clade with some parts of the nuclear genome of Melanerpes/Sphyrapicus being retained in the Campephilus lineage (Fig. 6). The coalescent simulations we performed allowed us to reject that the pattern observed in the FGB7 gene tree could be due to lineage sorting (p = 0.026). Furthermore, the hybridization hypothesis is also favoured over incomplete lineage sorting because it involves two distantly related lineages, based on our multi-loci estimate of the species tree. The two lineages that are involved are currently well differentiated behaviourally and morphologically. Melanerpes and Sphyrapicus are generalist woodpeckers feeding on sap and ants, whereas Campephilus is highly specialized in extracting larvae from wood trunks (Short, 1982). It is unknown how these two lineages were differentiated morphologically and behaviourally at the time when the hybridization event occurred as no fossils from that period have been described in North America. Based on our divergence time analyses on the mtDNA data set, the Melanerpes/Sphyrapicus lineage colonized the New World slightly before Campephilus (10.5 mya versus 8.2 mya). The hybridization event could have occurred during a very brief time window, which would correspond to the split between (1) Campephilus and the Indo-Malayan Megapicini (8.3 myrs, 95% HPD: 5.5–11.0) and the first diversification event in Campephilus (5.5 myrs, 95% HPD: 3.5–7.7) and (2) between the split between Melanerpes/Sphyrapicus and the Dendropicini (10.5 myrs, 95% HPD: 7.1–13.7) and the first diversification event for the Melanerpi (8.2 myrs, 95% HPD: 5.5–10.9) (Fig. 6). This would leave a very narrow time window of about 0.1 mya for the hybridization to have happened. This hybridization event would have occurred only 3–4 myrs after Melanerpi/Dendropicini and Campephilus/Megapicini diverged from each other (Fuchs et al., 2007). Among the nine autosomal loci we sequenced, only FGB7 gave a very strong conflicting signal, which would indicate that only the Chromosome 4, or part of it, was introgressed in the Campephilus genome. From the data available so far, all remaining part of the autosomal genome represents more likely the true evolutionary history of Campephilus. The reason for this is unknown, and until more data are gathered from further loci located on this chromosome, it is not possible to know if the whole chromosome 4 carries the same phylogenetic information or if recombination admixed the evolutionary history of this chromosome (in our data
Fig. 6. Schematic representation of how the data would support the hybridization/introgression hypothesis. The topology is identical to Fig. 1 (mtDNA, 10 partitions) and the divergence times were obtained using the mtDNA data set (four partitions and an uncorrelated lognormal molecular clock for each partition). The grey window on the phylogeny indicates the time frame where the hybridization would have happened. Two primary scenarios that differ by their initial assumptions are depicted: one scenario that assumes unidirectional hybridization due to selective mate choice (female Campephilus for male Melanerpes/Sphyrapicus) (1), and one scenario that assumes hybridization in both ways (2 and 3). The potential genomic composition of the first two generations after the hybridization event(s) (F1 and backcrosses with pure lineages) are shown. A common assumption to all models is that Haldane's rule (Hybrid sterility and inviability of the heterogametic sex) applies. One possibility that FGB7 from Melanerpes/Sphyrapicus could have introgressed in Campephilus is through recombination (and/or) selection. Note that the whole chromosome 4 could have introgressed during the meiosis or just a small part of the chromosome that includes FGB7 due to recombination. Scenarios 1, 2 and 3a are compatible with the data we observe. C and MS refer to Campephilus and Melanerpes/Sphyrapicus, respectively. For illustration purposes, we illustrate the segregation of the genome with four different autosomal chromosomes. Genomic composition of the results from the backcrosses of F1 with pure is averaged across possibilities (no recombination). ‘No signal in any loci for the Melanerpini being associated to the Megapicini.'
set, the tightly linked FGB5 is uninformative concerning the relationships of Campephilus). Fibrinogen is a protein involved in blood clotting (Doolittle, 1984); therefore there are no obvious indications that fibrinogen itself was the target of selection in two closely related lineages of woodpeckers, although selection could have acted on a closely linked locus. Currat et al. (2008) showed, using spatially explicit simulations, that: (1) massive introgression of neutral genes takes place during the colonisation of an occupied territory by an invading species if interbreeding is not severely prevented between the invading and the local species and (2) that introgression occurs almost exclusively from the local to the invading species. All these conclusions fit perfectly into the hybridization hypothesis (Fig. 6) as (1) introgression only occurred in the lineage that colonized the New World later (Campephilus) and (2) the time that elapsed since the divergence between the Melanerpes/Sphyrapicus and Campephilus and their putative hybridization (3–4 myrs) was largely in the time window reported for hybridization events in birds (lineages having diverged from each other up to 17 myrs; Price and Bouvier, 2002, see also Mallet et al., 2007 for Heliconius butterflies). Organelles, such as mitochondria, are often found to be more sensitive to introgression due to their inheritance mode and to potential selective advantage (Ballard and Whitlock, 2004; Melo-Ferreira et al., 2005; Ropiquet and Hassinan, 2006; Bossu and Near, 2009; Spinks and Shaffer, 2009). Unlike all these previous studies, our results suggest that some parts of the nuclear genome, but not the mitochondria, can become introgressed. Therefore, hybridization in the Campephilus case was either asymmetrical, as a result of sexual selection, involving only females from the Campephilus lineage and males from the Melanerpes/Sphyrapicus lineage, or hybrid individuals from the alternative combination were either not viable or were unable to back-cross as a result of cytoplasmic incompatibility (Borge et al., 2005a, 2005b).

In accordance with the Haldane's rule it is also possible that only heterogametic female hybrids were strongly counter-selected (Coyne et al., 1991). In addition, recent simulation studies have predicted that markers associated with the most dispersing sex (female in the case of birds) are less likely to be introgressed (Petit and Excoffier, 2009). In birds, there are numerous case studies showing that nuclear introgression rates are higher than mitochondrial introgression rates (Petit and Excoffier, 2009). The Z-linked loci that we sequenced suggested a different set of relationships for Campephilus, as the sister group of all Picidae but Hemicircus. This would imply that the Z chromosome, or some part of it, could have been introgressed from Melanerpes/Sphyrapicus in the Campephilus lineage as well. Yet, rates of molecular evolution, extent of historical gene flow and selection may be very variable across the Z-chromosome (Putman et al., 2007; Borge et al., 2005a), potentially due to its role in reproductive isolation (Saether et al., 2007). Therefore, more loci from this chromosome are needed before any definitive conclusion can be drawn about the phylogenetic signal carried by the Z chromosome.

4.3. Update concerning the phylogenetic relationships among the woodpeckers

Our analyses provide a new perspective on the phylogenetic relationships among woodpeckers, with all relationships among genera being supported by posterior probabilities of one in the simultaneous analysis of all available loci but FGB7 (see above). Our study suggests that the genus Campephilus is in fact related to the Megapicini, as was suggested by morphological data (Swierczewski and Raikow, 1981). Previous studies, each based on a different combination of three loci and using traditional concatenation methods, could not resolve with confidence the relationships among the three primary clades, the Dendropicini, Malarpicini and Megapicini (Webb and Moore, 2005; Benz et al., 2006; Fuchs et al., 2007). Using an expanded data set (16 loci vs 3 in Fuchs et al., 2007) and new tree building algorithms did not help to resolve the relationships among the primary Picinae lineages (Malarpicini, Dendropicini/Melanerini and Megapicini), suggesting that the primary lineages diversified in a short period of time.

Acknowledgments

We are very grateful to M. Flannery (CAS), J. Bates, S. Hackett, and D. Willard (FMNH), R. Brumfield, D. Dittmann and F. Sheldon (LSUMZ), A. Johnsen and J. Lifjeld (NHMO), M. Braun and J. Dean (USNM), S. Birks (UWBM), J. Fjeldså and J.B. Kristensen (ZMUC) and C. Spottiswoode (U. Cambridge) for kindly providing us tissues and R.C.K. Bowie, A.C. Carnaval, R. Pereira and S.M. Rovito for giving constructive comments. Laboratory help was provided by A. Tillier, C. Bonillo and J. Lambourdière (MNHN), M. Irestedt and P. Eldenäs (NRM). We also thank D. Zuccon for sharing unpublished IRBP primers. This work was supported by the Plan Plurifomation ‘Etat et structure phylogénétique de la biodiversité actuelle et fossile’, the ‘Consortium National de Recherche en Génomique’, and the ‘Service de Syntomatique Moléculaire’ of the Muséum National d’Histoire Naturelle (IFR 101). It is part of the Agreement No. 2005/67 between the Genoscope and the Muséum National d’Histoire Naturelle on the project ‘Macrophylogeny of life’ directed by Guillaume Lecontre. Laboratory work at NRM was supported by the Swedish Research Council (Grant No. 621-2004-2913 to PE). We acknowledge the support of a Synthesys grant made available by the European Community–Research Infrastructure Action under the EFS Structuring the European Research Area Programme (SE-TAF 746 to JF). LL was support by a NSF grant (DMS-1222745). This research was supported by a postdoctoral fellowship to JF from the DST/NRF Centre of Excellence at the Percy FitzPatrick Institute.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.02.014.

References


