Data in Brief 5 (2015) 1078–1091

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data supporting a molecular phylogeny of the hyper-diverse genus *Brueelia*

Sarah E. Bush [a,*], Jason D. Weckstein [b,1], Daniel R. Gustafsson [a], Julie Allen [c], Emily DiBlasi [a], Scott M. Shreve [c,2], Rachel Boldt [c], Heather R. Skeen [b,3], Kevin P. Johnson [c]

[a] *Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112, USA*
[b] *Field Museum of Natural History, Science and Education, Integrative Research Center, 1400 S. Lake Shore Drive, Chicago, IL 60605, USA*
[c] *Illinois Natural History Survey, University of Illinois, 1816 South Oak Street, Champaign, IL 61820, USA*

ARTICLE INFO

Article history:
Received 1 October 2015
Received in revised form
14 October 2015
Accepted 20 October 2015
Available online 2 November 2015

Keywords:
Brueelia
Lice
Songbirds
Host-specificity
Phylogenetic reconstruction
Macroevolution

ABSTRACT

Data is presented in support of a phylogenetic reconstruction of one of the largest, and most poorly understood, groups of lice: the *Brueelia*-complex (Bush et al., 2015 [1]). Presented data include the voucher information and molecular data (GenBank accession numbers) of 333 ingroup taxa within the *Brueelia*-complex and 30 outgroup taxa selected from across the order Phthiraptera. Also included are phylogenetic reconstructions based on Bayesian inference analyses of combined COI and EF-1α sequences for *Brueelia*-complex species and outgroup taxa.

© 2015 Elsevier Inc.. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

DOI of original article: http://dx.doi.org/10.1016/j.ympev.2015.09.015

* Corresponding author.
E-mail address: bush@biology.utah.edu (S.E. Bush).
[1] Current address: Department of Ornithology and Department of Biodiversity, Earth, and Environmental Sciences, Academy of Natural Sciences of Drexel University, 1900 Benjamin Franklin Parkway, Philadelphia, PA 19103, USA.
[2] Current address: Department of Biology, Centre College, Danville, KY 40422, USA.
[3] Current address: Biology Department, Loyola University Chicago, Chicago IL 60660, USA.

http://dx.doi.org/10.1016/j.dib.2015.10.022
2352-3409/© 2015 Elsevier Inc.. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications table

Subject area	Biology, genetics and genomics
More specific subject area	Phylogenetics
Type of data	Specimen matrix, phylogenetic reconstruction
How data was acquired	Phylogenetic reconstruction using Bayesian inference methods
Data format	Raw, analyzed
Experimental factors	n/a
Experimental features	n/a
Data source location	worldwide
Data accessibility	Within this article, sequences available in GenBank

Value of the data

- Evolutionary history of feather lice in the Brueelia-complex was reconstructed.
- Data support re-recognition of historic genera, and erection of several new genera.
- Associations of lice with geography and host-family are correlated with phylogeny.
- Host association and geographic origin of each sequenced specimen are provided.

1. Data, materials and methods

The data presented herein supports a phylogenetic reconstruction of the *Brueelia*-complex; these data complement the companion article by Bush et al. [1].

1.1. Sampling

We sampled a total of 333 louse specimens belonging to the *Brueelia*-complex (Supplemental Table 1). These lice were sampled from 250 bird species belonging to 66 bird families and five orders (Passeriformes, Coraciiformes, Cuculiformes, Piciformes, and Trogoniformes). Sampled lice include 38 known species and 211 lice that represent either new species of lice or new host associations. These samples were collected from 23 countries and all continents except Antarctica. An additional 30 outgroup taxa for rooting the phylogeny were selected to represent nested sister taxonomic relationships within the family Philopteridae [2,3]. These 30 louse outgroup species were from 27 host species, in 17 host families, collected from 9 countries.

Lice were collected either from euthanized bird specimens using ethyl acetate fumigation or from live birds dusted with pyrethrum powder [4,5]. Care was taken to make sure that individual hosts were kept separate at all times and to clean all working surfaces between fumigation. Lice were collected by the authors and colleagues during field-work conducted over several decades and were stored in vials of 95% ethanol, usually in ultracold (-80 °C) freezers.

1.2. DNA extraction, amplification and alignment

DNA was extracted from lice using either the Qiagen DNeasy micro-kit (Valencia, California, USA) following the manufacturer's protocol as described by Valim and Weckstein [6], or the Qiagen DNeasy tissue kit (Valencia, California, USA) following the manufacture's protocol as described by Johnson et al. [7]. After DNA was extracted from individual lice, the exoskeletons were retained and mounted on microscope slides [8]. These voucher slides were used to identify each specimen to genus using the keys in Price et al. [9]. Specific-level identifications were based on original descriptions, specific keys if possible, and comparison with identified slide mounted material. Voucher slides are deposited in

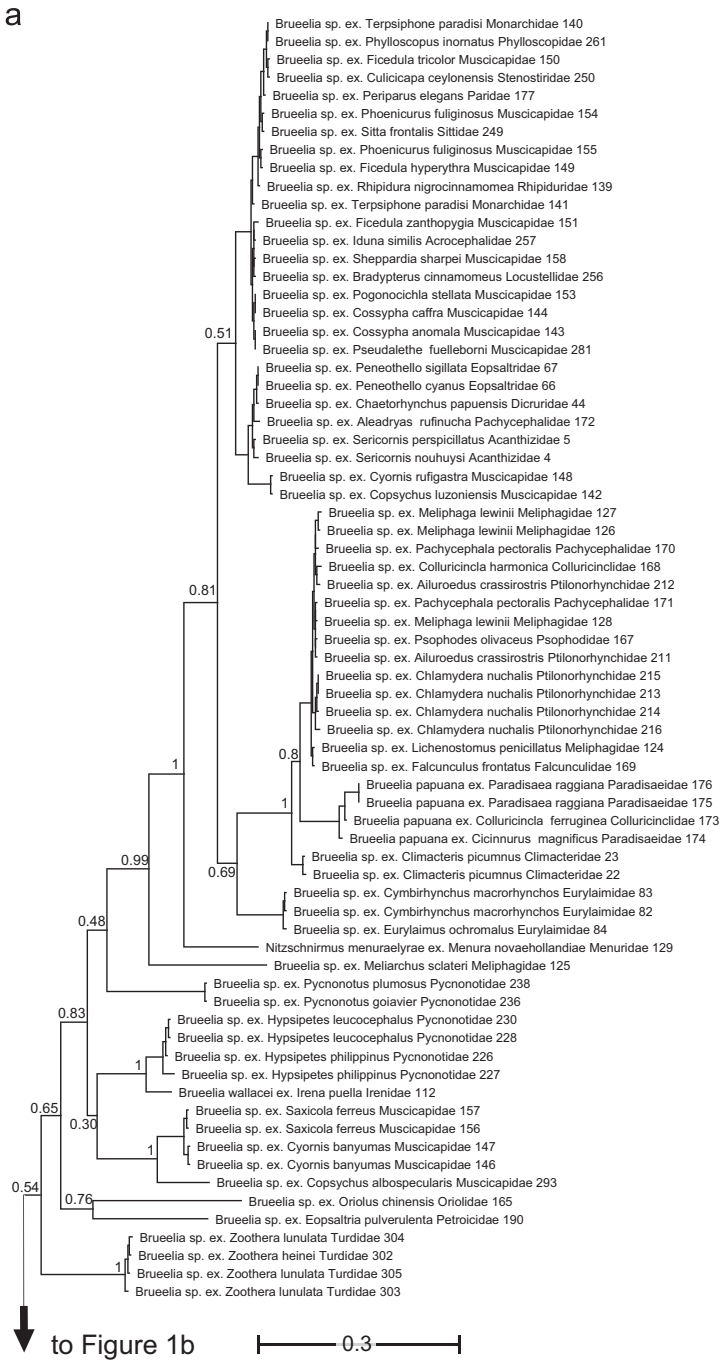

Fig. 1. Consensus tree from Bayesian analysis of combined COI and EF-1α sequences for *Brueelia*-complex species and outgroup taxa. Branches proportional to substitutions per site for the consensus tree (scale indicated). Numbers associated with nodes are posterior probabilities for the clade from a 10 million generation MCMC analysis, sampled every 1000 generations and excluding the first 1 million generations as burn-in (values < 0.5, and values associated with short terminal branches not shown here; all support values > 0.5 are shown on Fig. 2). Numbers after taxonomic names refer to Supplemental Table 1. Louse taxonomy follows the classification of Price et al. [9] and subsequent publications. Host taxonomy follows Clements et al. [21] and Dickinson et al. [22]: host genus, species, and family are all indicated. Tree partitioned into six portions (a-f).

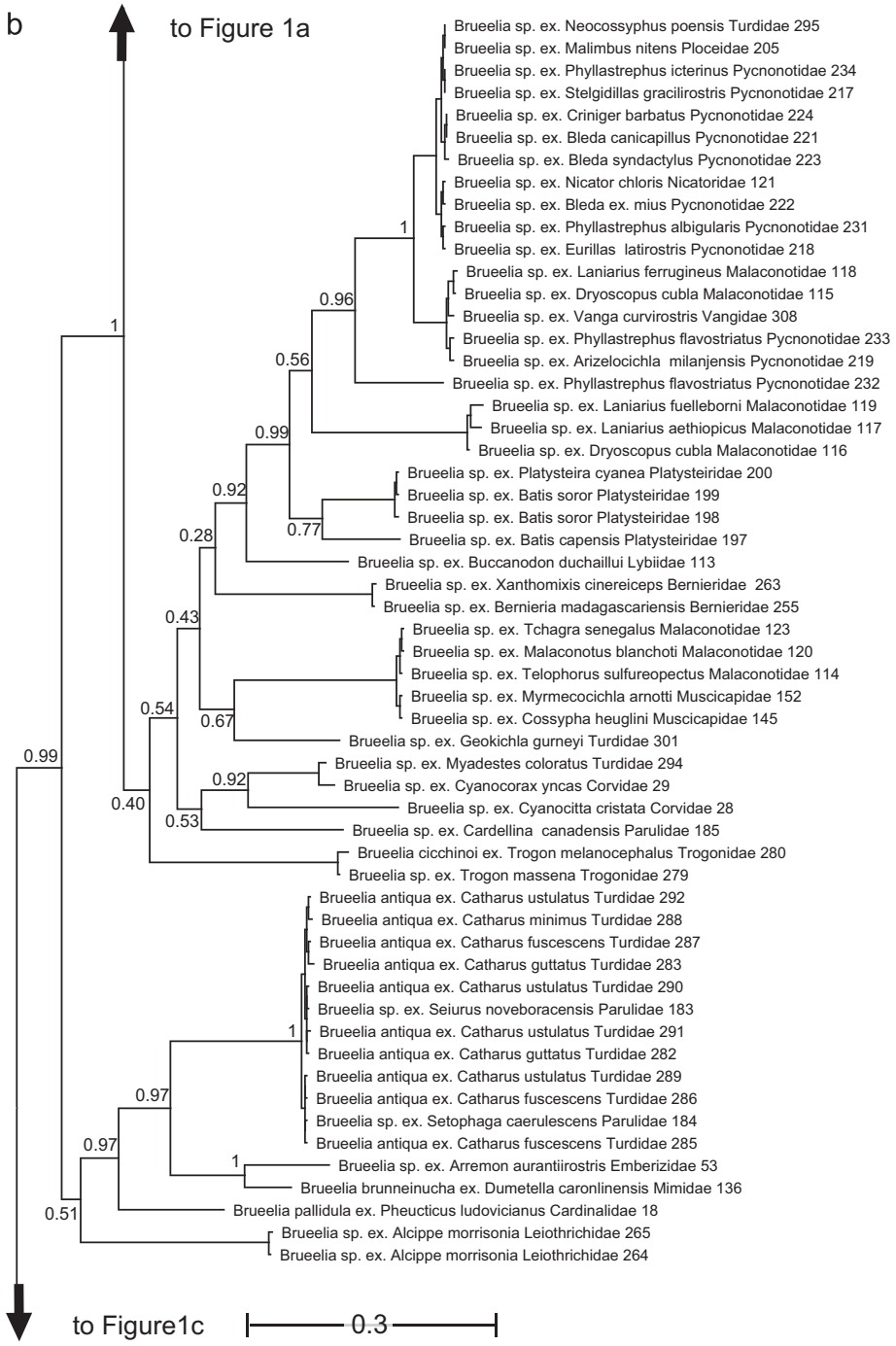

Fig. 1. (*continued*)

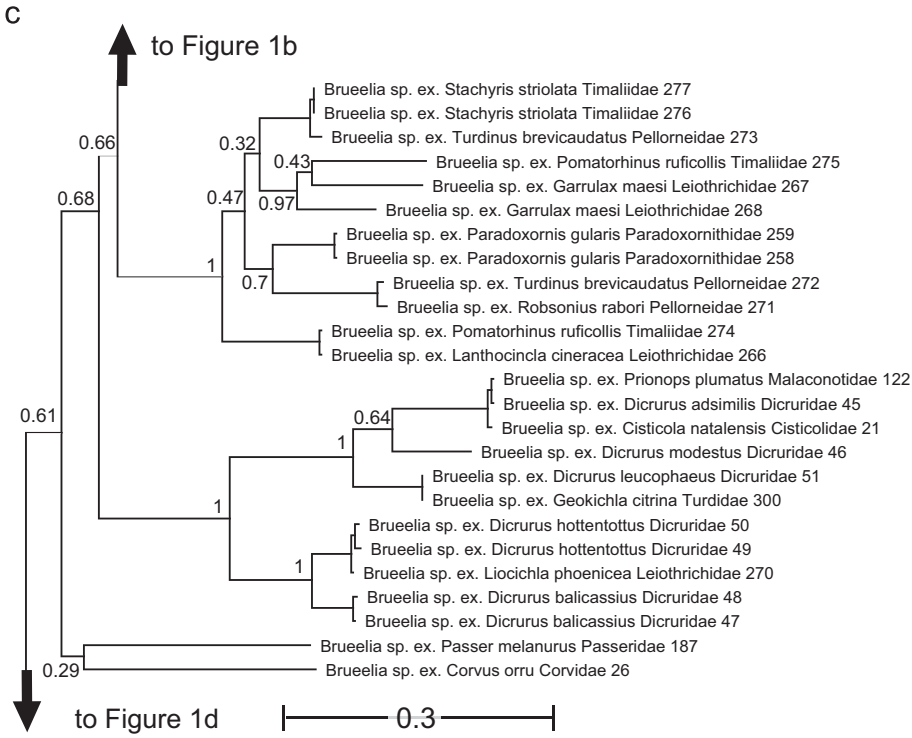

Fig. 1. (*continued*)

the Illinois Natural History Survey Insect Collection (INHS), Price Institute for Parasite Research at the University of Utah (PIPeR), and Field Museum of Natural History (FMNH) (Supplemental Table 1).

Portions of one mitochondrial (COI) and one nuclear gene (EF-1α) were selected because these genes have successfully resolved phylogenies of closely related groups of parasitic lice and more distantly related "bark lice" [3,10–13]. We used PCR to amplify and sequence portions of the mitochondrial cytochrome oxidase I (COI; 379 bp) and the nuclear gene elongation factor 1a (EF1 α; 347 bp) using published amplification and sequencing protocols [12,13]. Purified PCR products were cycle sequenced using ABI Big Dye (Applied Biosystems, Foster City, California) and run on an ABI Prism 3730 DNA sequencer (Applied Biosystems). Raw sequence data were trimmed, edited, and reconciled using Sequencher 5.0.1 (Genecodes CO., Ann Arbor, Michigan) or Geneious (version 7.0.3, Biomatters LTD). Both genes are protein coding and therefore we were able to easily align them by eye according to codons. These aligned gene sequences were then concatenated for phylogenetic analysis.

1.3. Phylogenetic analyses

The final sequence alignment was analyzed using PartitionFinder (v1.1.1; [14,15]), an open source python script that selects the best-fit partitioning schemes and models of molecular evolution for phylogenetic analysis. We tested whether the two genes (COI, EF1 α) should be analyzed together under the same model and parameters or as two separate partitions. We tested only these two partitions because separating each of these genes by codon would only provide 100 bps for each partition, a very small amount of sequence for estimating parameters and would likely result in over-parameterization. The PartitionFinder analysis found that a single partition and GTR+I+G model of molecular evolution best fit the data, using both AICc and BIC criterion. Using these parameters,

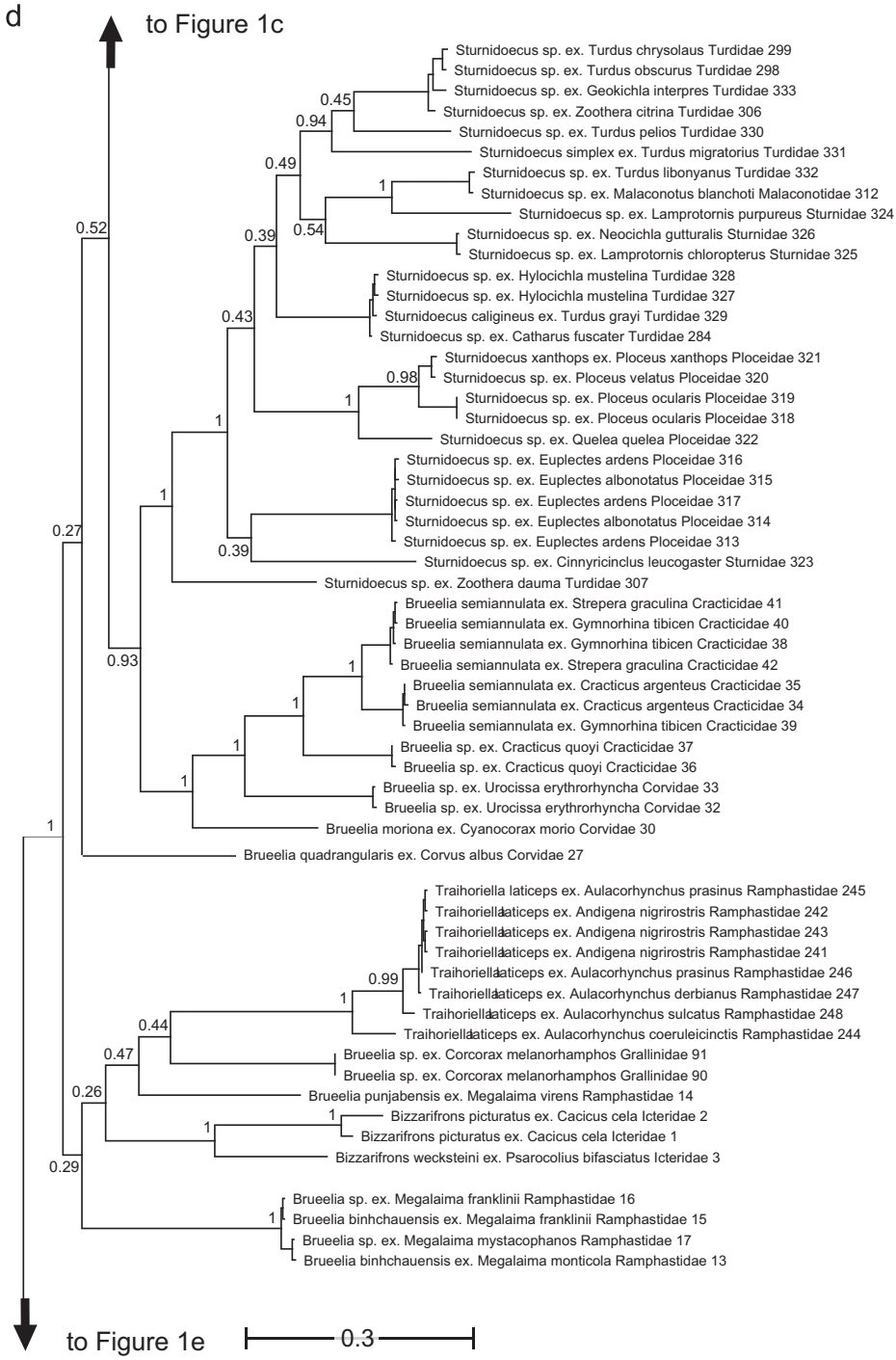

Fig. 1. (*continued*)

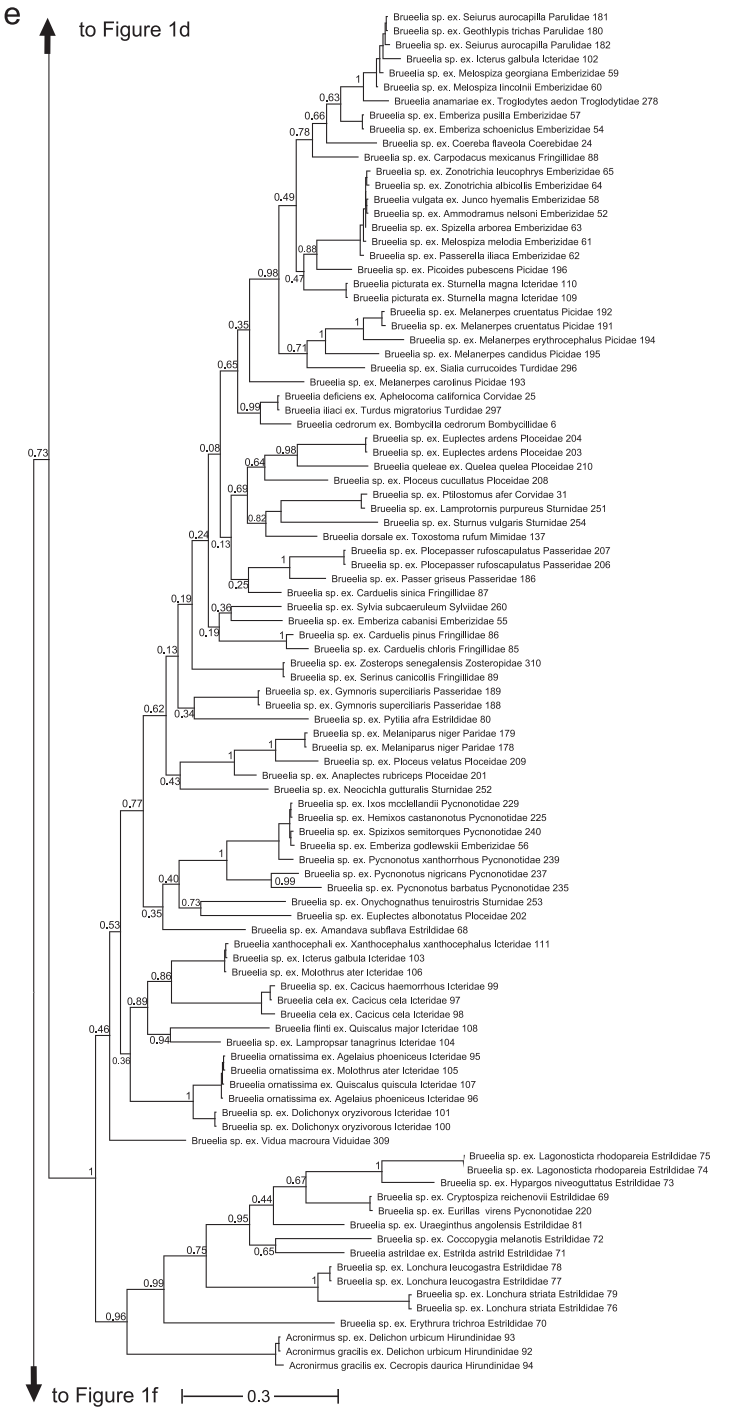

Fig. 1. (*continued*)

Fig. 1. (*continued*)

which were estimated as part of the analysis, and a flat Dirichlet prior for state frequencies, we ran a Bayesian analysis in MrBayes 3.2.2 [16–18] for 10,000,000 generations. Each Bayesian analysis included two parallel runs, each with four Markov chains, to ensure that our analyses were not stuck at local optima [19]. Markov chains were sampled every 500 generations, yielding 20,000 parameter point-estimates. We used these 20,000 point-estimates minus the burn-in generations (500 point-estimates, 250,000 generations) to create a 50% majority-rule consensus tree and calculated Bayesian posterior probabilities to assess nodal support. We rooted the Bayesian tree using a nested set of sister taxa within the family Philopteridae [2,12,13,20].

A consensus tree from the Bayesian analysis of combined COI and EF-1α sequences for *Brueelia*-complex is shown in Fig. 1. A cladogram of the consensus tree from the Bayesian analysis is shown in Fig. 2.

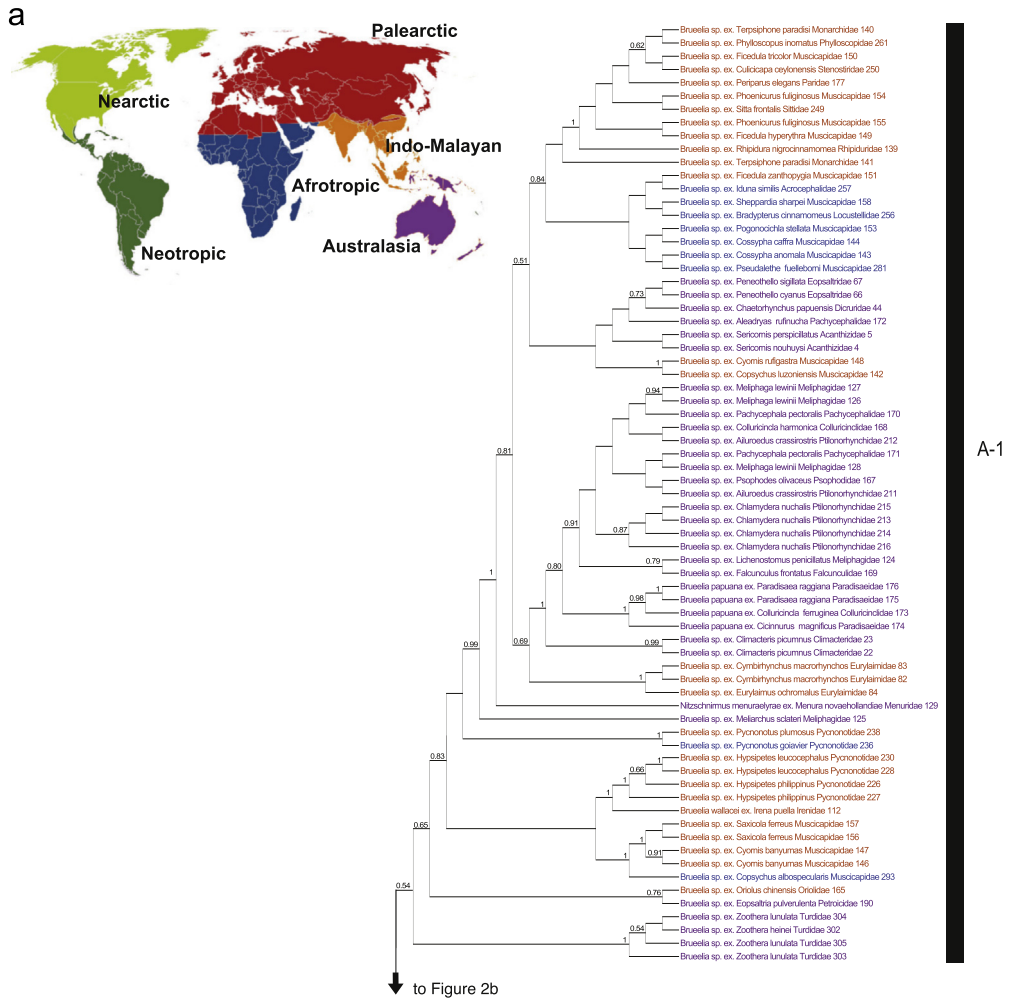

Fig. 2. Cladogram of the consensus tree from a Bayesian analysis of combined COI and EF-1α sequences for *Brueelia*-complex species and outgroup taxa. Numbers associated with nodes are posterior probabilities calculated from 10 million MCMC generations sampled every 1000 and excluding the first 1 million generations as burn-in (values < 0.5 not shown). Taxa colored to indicate geographic origin as indicated in map in 2a. Conventions as in Fig. 1. Tree partitioned into five portions (a–e).

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.dib.2015.10.022. Specifically, Supplemental Table 1, which is a list of studied specimens, their voucher numbers, host associations, geographic origin, and GenBank accession numbers.

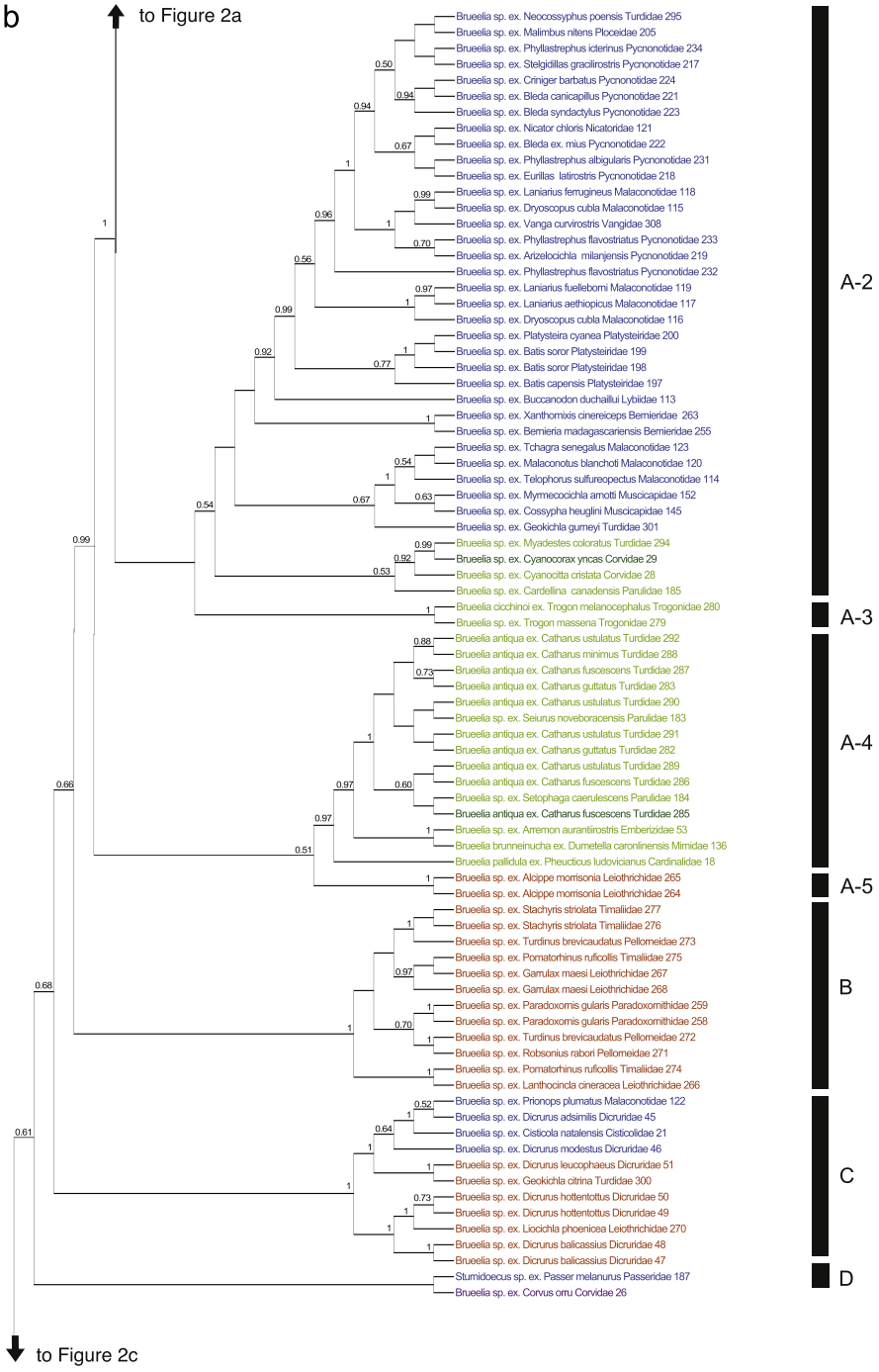

Fig. 2. (*continued*)

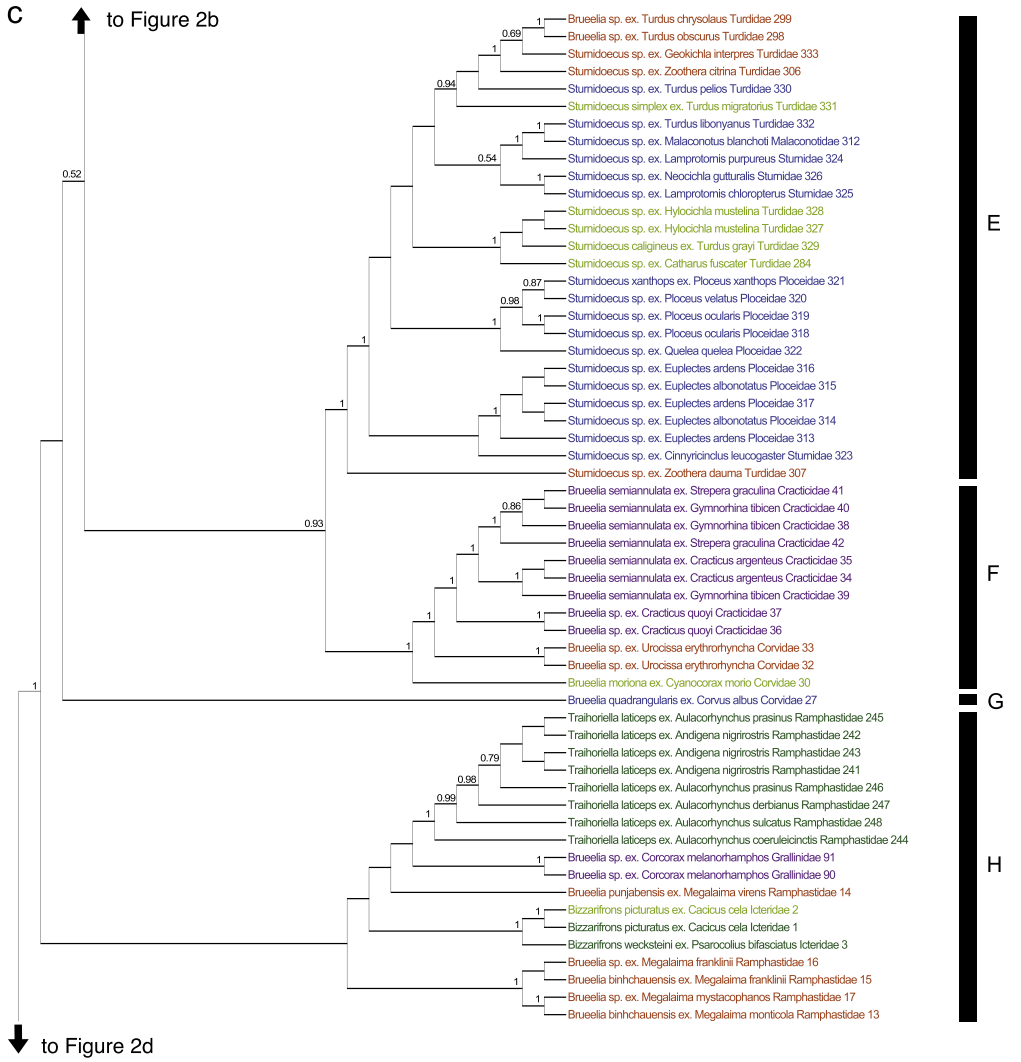

Fig. 2. (*continued*)

References

[1] S.E. Bush, J.D. Weckstein, D.R. Gustafsson, J. Allen, E. DiBlasi, S.M. Shreve, R. Boldt, H.R. Skeen, K.P. Johnson, Unlocking the

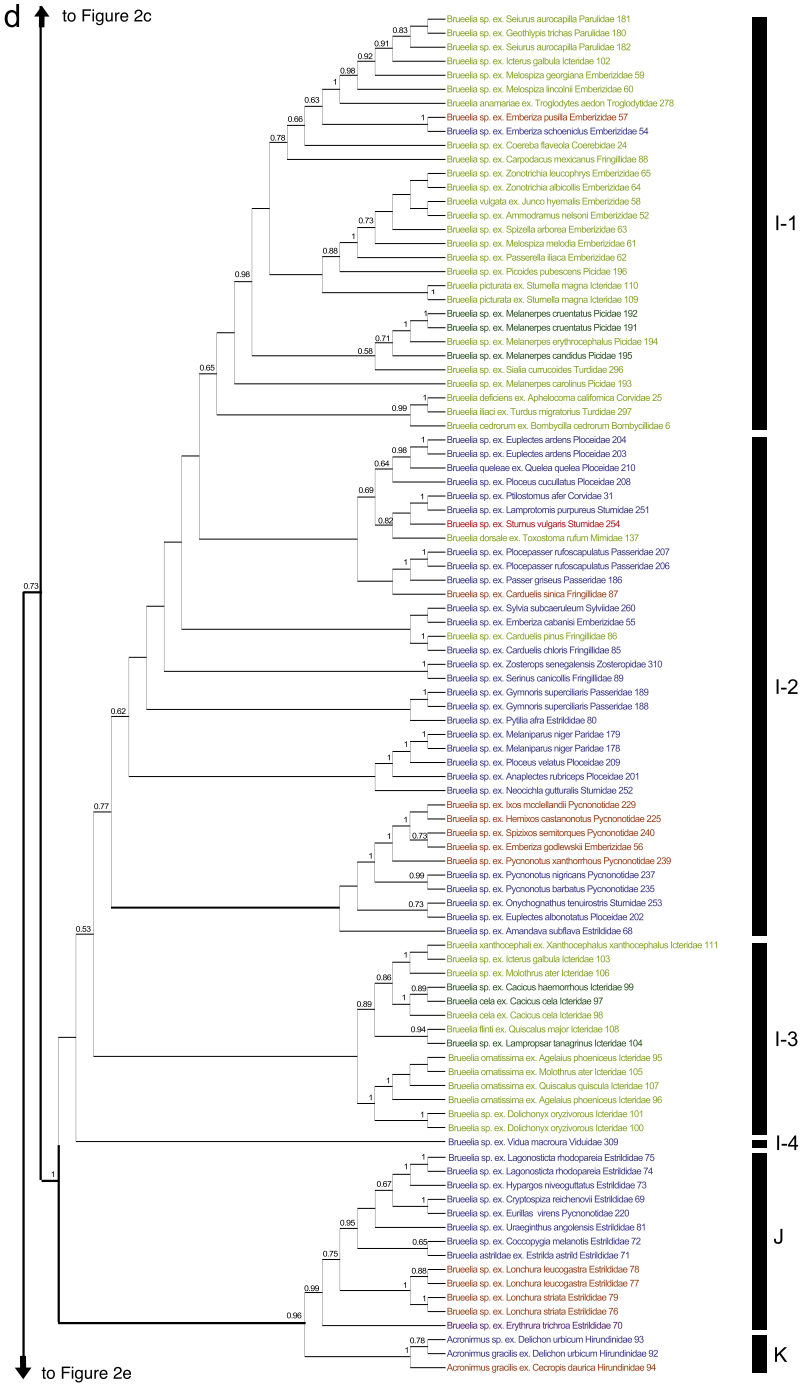

Fig. 2. (*continued*)

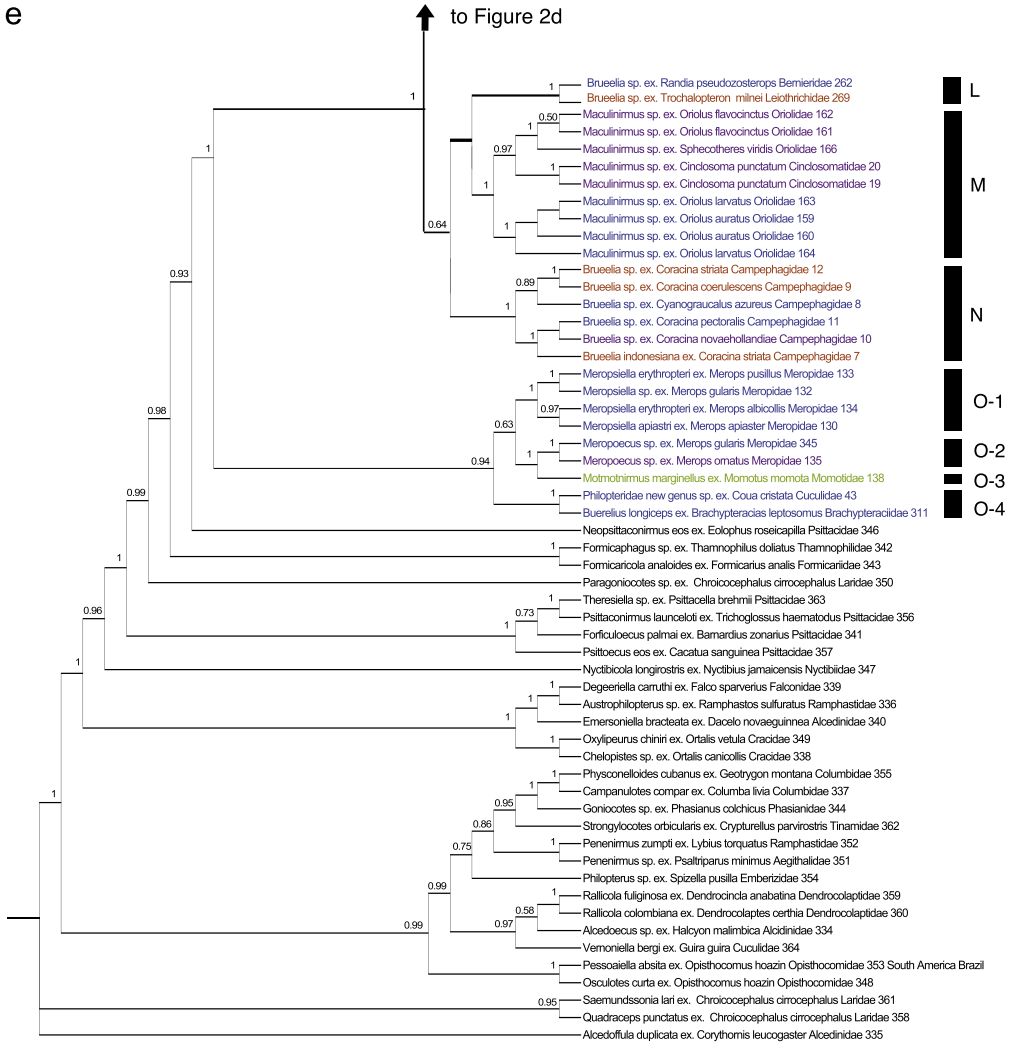

Fig. 2. (*continued*)

black box of feather louse diversity: a molecular phylogeny of the hyper-diverse genus *Brueelia*, Mol. Phylogenet. Evol. (2015), in press.

[2] R.H. Cruickshank, K.P. Johnson, V.S. Smith, R.J. Adams, D.H. Clayton, R.D.M. Page, Phylogenetic analysis of partial sequences of elongation factor 1*a* identifies major groups of lice (Insecta: Phthiraptera), Mol. Phylogenet. Evol. 19 (2001) 202–215.

[3] V.S. Smith, T. Ford, K.P. Johnson, P.C.D. Johnson, K. Yoshizawa, J.E. Light, Multiple lineages of lice pass through the K–Pg boundary, Biol. Lett. 7 (2011) 782–785.

[4] D.H. Clayton, D.M. Drown, Critical evaluation of five methods for quantifying chewing lice (Insecta: Phthiraptera), J. Parasitol. 87 (2001) 1291–1300.

[5] C. Bueter, J.D. Weckstein, K.P. Johnson, J.M. Bates, C.E. Gordon, Comparative phylogenetic histories of two louse genera found on *Catharus* thrushes and other birds, J. Parasitol. 95 (2009) 295–307.

[6] M.P. Valim, J.D. Weckstein, Two new species of *Brueelia* Kéler, 1936 (Ischnocera, Philopteridae) parasitic on Neotropical trogons (Aves, Trogoniformes), ZooKeys 128 (2011) 1–13.

[7] K.P. Johnson, R.J. Adams, D.H. Clayton, Molecular systematics of Goniodidae (Insecta: Phthiraptera), J. Parasitol. 87 (2001) 862–869.

[8] R.L. Palma, Slide-mounting of lice: a detailed description of the Canada balsam technique, N. Z. Entomol. 6 (1978) 432–436.
[9] R.D. Price, R.A. Hellenthal, R.L. Palma, K.P. Johnson, D.H. Clayton, The Chewing Lice: World Checklist and Biological Overview, Illinois Natural History Survey, 2003.
[10] K.P. Johnson, J.D. Weckstein, C.C. Witt, R.C. Faucett, R.G. Moyle, The perils of using host relationships in parasite taxonomy: phylogeny of the *Degeeriella* complex, Mol. Phylogenet. Evol. 23 (2002) 150–157.
[11] K.P. Johnson, R.H. Cruickshank, R.J. Adams, V.S. Smith, R.D.M. Page, D.H. Clayton, Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera), Mol. Phylogenet. Evol. 26 (2003) 231–242.
[12] K.P. Johnson, K. Yoshizawa, V.S. Smith, Multiple origins of parasitism in lice, Proc. R. Soc. Lond. B 271 (2004) 1771–1776. http://dx.doi.org/10.1098/rspb.2004.2798.
[13] V.S. Smith, R.D.M. Page, K.P. Johnson, Data incongruence and the problem of avian louse phylogeny, Zool. Scr. 30 (2004) 239–259. http://dx.doi.org/10.1111/j.0300-3256. 2004.00149.x.
[14] R. Lanfear, B. Calcott, S.Y.W. Ho, S. Guindon, PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses, Mol. Biol. Evol. 29 (2012) 1695–1701.
[15] R. Lanfear, B. Calcott, D. Kainer, C. Mayer, A. Stamatakis, Selecting optimal partitioning schemes for phylogenomic datasets, BMC Evolut. Biol. 14 (2014) 82.
[16] J.P. Huelsenbeck, F. Ronquist, MrBayes: Bayesian inference of phylogeny, Bioinformatics 17 (2001) 754–755.
[17] F. Ronquist, J.P. Huelsenbeck, MRBAYES 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 (2003) 1572–1574.
[18] F. Ronquist, M. Teslenko, P.V.D. Mark, D.L. Ayres, A. Darling, S.H. Hohana, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, MrBayes 3.2: efficient Bayesian phylogenetic inference and model selection across a large model space, Syst. Biol. 61 (2012) 1–4.
[19] J.P. Huelsenbeck, J.P. Bollback, Empirical and hierarchical Bayesian estimation of ancestral states, Sys. Biol. 50 (2001) 351–366.
[20] K.P. Johnson, S.M. Shreve, V.S. Smith, Repeated adaptive divergence of microhabitat specialization in avian feather lice, BMC Biol. 10 (2012) 52. http://dx.doi.org/10.1186/1741-7007-10-52.
[21] J.F. Clements, T.S. Schulenberg, M.J. Iliff, D. Roberson, T.A. Fredericks, B.L. Sullivan, C.L. Wood, The eBird/Clements checklist of birds of the world: Version 6.8. Downloaded from 〈http://www.birds.cornell.edu/clementschecklist/download/〉, 2014.