

A Case Study Using Mitochondrial Genomes of the Order Diprotodontia (Australasian Marsupials) Suggests that Single Ortholog is Not Sufficient for Phylogeny

Victor CC Wang^{1,2}, Nur Jannah Kamarudin^{1,2}, Xue Ting Tan^{1,2}, Avettra Ramesh^{1,2}, Shermaine SM Chew^{1,2}, Madhurya V Murthy^{1,2}, Nikita V Yablochkin^{1,2}, Karthiga Mathivanan^{1,2} and Maurice HT Ling^{1,2,3*}

¹Department of Applied Sciences, Northumbria University, United Kingdom

²School of Life Sciences, Management Development Institute of Singapore, Singapore

³HOHY PTE LTD, Singapore

*Corresponding Author: Maurice HT Ling, School of Life Sciences, Management Development Institute of Singapore, Singapore.

Received: July 24, 2020; Published: August 31, 2020

Abstract

All organisms exist today descended from a common ancestor and phylogenetic tree is a common means to analyze such evolutionary histories. Currently, orthologs are routinely used to construct phylogenetic trees. However, the number of orthologs required to determine the evolutionary history of a set of organisms is not clear. In this case study, we compare the generated phylogenetic trees from one ortholog against that of the complete set of orthologs using 13 mitochondrial genes of the 24 species from the Order Diprotodontia. Using the phylogenetic tree generated from the complete set of orthologs as benchmark, our results suggest that using single ortholog may result in distinctly different phylogenies as compared to benchmark and the average number of branch points from multiple single orthologs is significantly different (paired t-statistic = 8.01, p-value = 3.27e-14) from benchmark. This suggests that phylogenetic analysis from single ortholog or multiple single orthologs is not likely to reflect actual evolutionary history and the complete set of orthologs is required.

Keywords: Mitochondrial Genomes; Diprotodontia; Phylogeny

Introduction

Evolution is the change in heritable traits and characteristics of biological species over successive generations and is widely observable in laboratory and natural populations [1]. The four major processes of evolution [2] are gene flow [3], genetic drift [4], mutations [5] and natural selection [6]. Inherently, evolution suggests that all known life forms on earth share a common ancestry - a last universal common ancestor (LUCA) [7], which was proposed by Charles Darwin in 1859 the *On the Origin of Species* [8] where he states that "*Therefore I should infer from analogy that probably all the organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed*". This is supported by studies [9,10] constructing a universal phylogenetic tree using ribosomal RNA sequence data, showing LUCA at the root of eubacteria, archaebacteria and eucarya. The similarities among all known modern-day organisms indicate that they have diverged through evolution from the LUCA [11] and phylogenetic analysis is commonly used to determine the evolutionary histories and relationships among modern-day species [12,13].

Phylogenetic trees can be constructed based on phenotypic traits before the availability of genetic and molecular data [14] or molecular sequences [15] since its availability in the second half of 20th century [16]. Numerous phylogenetic trees based on phenotypic traits were still being produced until the late twentieth century [17]. Phylogenetic trees based on molecular data are considered to be more reliable as compared to morphological phylogenetic trees due to convergent evolution [15], which is the independent evolution of similar

traits in distinct lineages giving rise to analogous structures or functions [18]. Examples of analogous structure include wings of insects, birds and bats, and the echolocating system of marine mammals and bats [19]. Phylogenetic analysis has many applications; such as, identification of potential drugs and vaccine candidates against diseases [20,21], guiding conservation policies [22] and understanding the evolutionary history of organisms [23,24].

Phylogenetic analysis using molecular sequences require the use of orthologs [25], which is a term coined by Walter Fitch [26] to refer to genes or peptides descended from a common ancestor. However, the number of orthologs required is not clear - is one ortholog or the entire set of orthologs required to determine the evolutionary history of a set of organisms? Therefore, using 13 mitochondrial genes of the 24 species from the Order Diprotodontia, this study compares the generated phylogenetic trees from one ortholog against that of the complete set of orthologs. Our results suggest that using single ortholog may result in distinctly different phylogenies as compared to using the complete set of orthologs.

Materials and Methods

Mitochondrial genome sequences

Twenty-four (Supplementary table S1) complete mitochondrial sequences of Australasian marsupials (Order Diprotodontia) were identified from GenBank; namely, (i) *Dactylopsila trivirgata* (DAT, Accession NC_008134.1), (ii) *Distoechurus pennatus* (DPE, Accession NC_008145.1), (iii) *Lagorchestes conspicillatus* (LCO, Accession KY996508.1), (iv) *Lagostrophus fasciatus* (LFA, Accession NC_008447.1), (v) *Lagorchestes hirsutus* (LHI, Accession NC_008136.1), (vi) *Macropus bernardus* (MBE, Accession KY996505.1), (vii) *Macropus fuliginosus* (MFU, Accession NC_039717.1), (viii) *Macropus giganteus* (MGI, Accession KY996502.1), (ix) *Macropus robustus* (MRO, Accession NC_001794.1), (x) *Macropus rufus* (MRU, Accession KY996501.1), (xi) *Notamacropus agilis* (NAG, Accession KY996507.1), (xii) *Notamacropus irma* (NIR, Accession KY996503.1), (xiii) *Notamacropus parma* (NPA, Accession KY996504.1), (xiv) *Notamacropus rufogriseus* (NRU, Accession KY996499.1), (xv) *Petaurus breviceps* (PBR, Accession NC_008135.1), (xvi) *Phalanger vestitus* (PVE, Accession NC_008137.1), (xvii) *Phascolarctos cinereus* (PCI, Accession NC_008133.1), (xviii) *Potorous tridactylus* (PTR, Accession NC_006524.1), (xix) *Pseudocheirus peregrinus* (PPE, Accession NC_006519.1), (xx) *Petrogale xanthopus* (PXA, Accession KY996509.1), (xxi) *Tarsipes rostratus* (TRO, Accession NC_006518.1), (xxii) *Trichosurus vulpecula* (TVU, Accession NC_003039.1), (xxiii) *Vombatus ursinus* (VUR, Accession NC_003322.1) and (xxiv) *Wallabia bicolor* (WBI, Accession KY996500.1). For each organism, complete mitochondrial genomic sequences and mitochondrial coding sequences were obtained.

Phylogenetic analysis

Coding sequences (DNA sequences) across the organisms were compared to identify for missing coding sequences in one or more organisms. Common coding sequences were concatenated to generate an overall phylogeny using Clustal Omega (27) and Simply Phylogeny (27) from EMBI-EBI, using default parameters. This acts as consensus evolutionary path. This is followed by generating phylogenies for each of the coding sequences using Clustal Omega (27) and Simply Phylogeny (27) from EMBI-EBI, using default parameters. These phylogenies by coding sequences were compared against previously estimated consensus evolutionary path for potentially different evolutionary paths.

Phylogenetic branch scoring

Each phylogenetic tree was scored by branching points to represent the divergence between the twenty-four species. The scoring of branch points provides insights and understanding of the evolutionary relationships and histories of the twenty-four species. The scoring method is illustrated in Figure 1. Using organisms A and E as an example, the position of organism A and organism E can both be used as starting points to score branch points. However, the results vary depending on the different positions of organisms. With organism A as

starting point, the score of branch points from organism A to organisms E is 3. With organism E as a starting point, the score is 1. Hence, the minimum score of 1 will be taken.

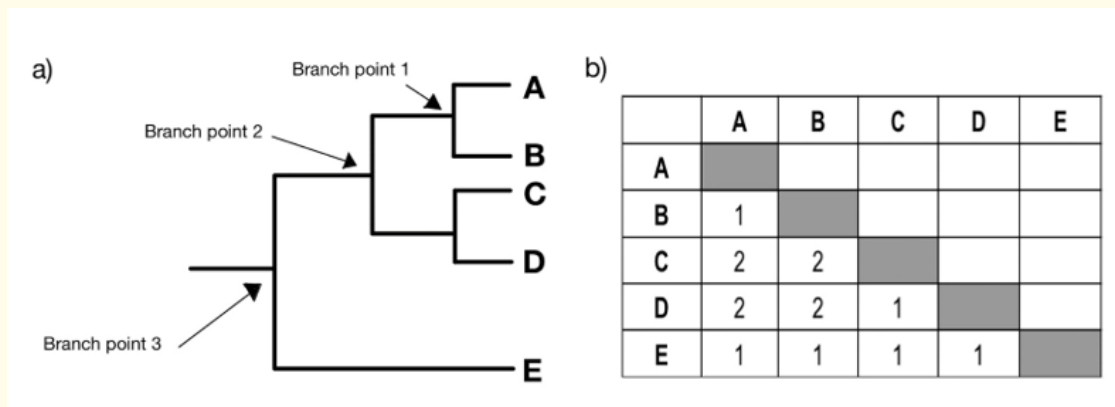


Figure 1: Example of Phylogeny Scoring. (a) The scoring of phylogenetic tree branch points between organism A and organism E. Branch point 1 represents the common ancestor of organisms A and B while branch point 2 represents the common ancestor of organisms A, B, C and D. Branch point 3 represents the common ancestor of organisms; A, B, C, D and E. (b) Table of ancestry showing the scoring of phylogenetic tree branch points of organisms A, B, C, D and E. According to the table, it can be deduced that organisms A and B, C and D, A and E, B and E, C and E, as well as organisms D and E, were the most closely related based on the branching score of 1. A branching score of 2 suggested that organisms A and C, A and D, B and C, as well as organisms B and D, were the least closely related.

Results and Discussion

Similar organization in mitochondrial genomes

An analysis of the 24 mitochondrial genomes shows that all 24 mitochondria genomes consist of 13 protein coding genes, which is consistent to the findings of Janke., *et al.* [28], in the same order for all organisms (Figure 2). The order of protein coding genes is (i) NADH dehydrogenase subunit 1 (ND1; 955 to 956 bp), (ii) NADH dehydrogenase subunit 2 (ND2; 1041 to 1049 bp), (iii) cytochrome c oxidase subunit I (COX1; 1541 to 1553 bp), (iv) COX2 cytochrome c oxidase subunit II (COX2; 681 to 692 bp), (v) ATP synthase F0 subunit 8 (ATP8; 206 to 215 bp), (vi) ATP synthase F0 subunit 6 (ATP6; 678 to 680 bp), (vii) cytochrome c oxidase subunit III (COX3; 783 bp), (viii) NADH dehydrogenase subunit 3 (ND3; 338 to 351 bp), (ix) NADH dehydrogenase subunit 4L (ND4L; 296 to 296 bp), (x) NADH dehydrogenase subunit 4 (ND4; 1374 to 1377 bp), (xi) NADH dehydrogenase subunit 5 (ND5; 1808 to 1817 bp), (xii) NADH dehydrogenase subunit 6 (ND6; 494 to 506 bp) and (xiii) cytochrome b (CYTB; 1139 to 1145 bp).

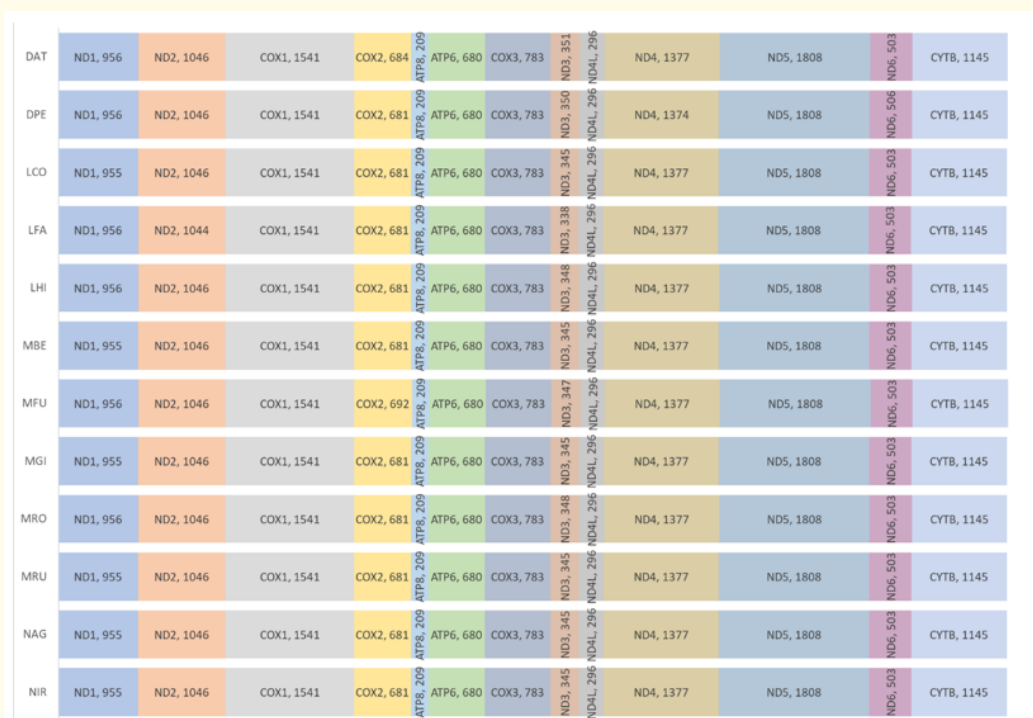


Figure 2A: Sequence order of the first 12 mitochondrial genomes.

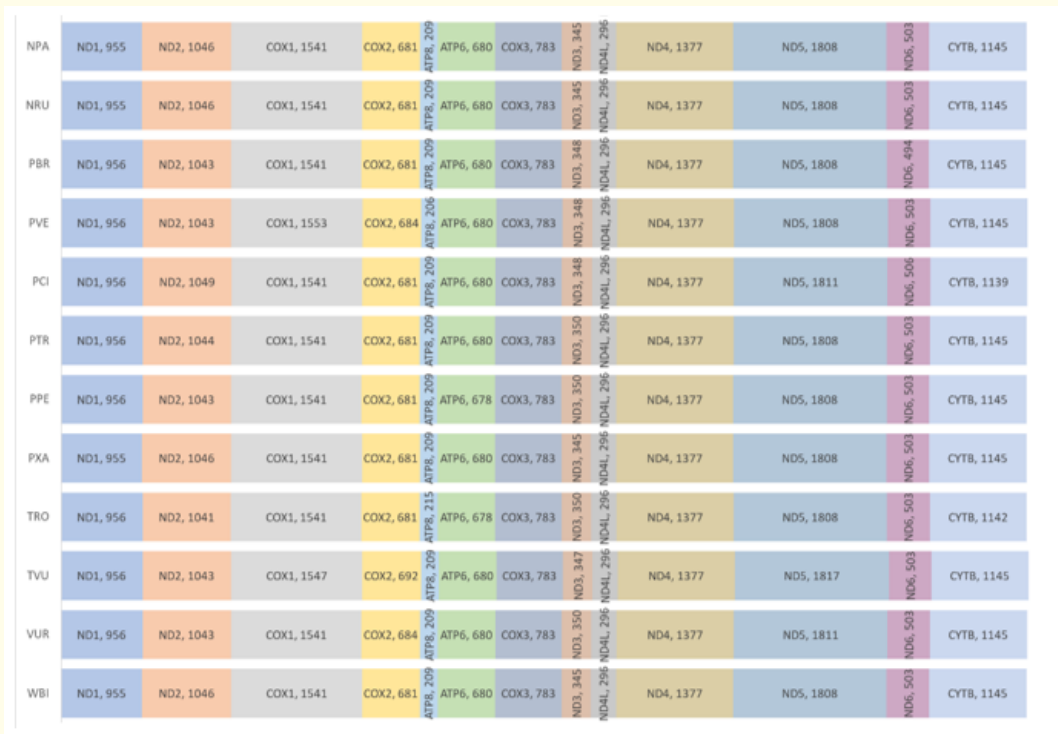


Figure 2B: Sequence order of the second 12 mitochondrial genomes.

Phylogeny depends on ortholog used

Each of the 13 orthologous protein coding genes across 24 organisms will result in an ortholog-specific phylogenetic tree and the complete set of genes will result in one phylogenetic tree, resulting in 14 phylogenetic trees. The phylogenetic tree generated from whole genome has been used to represent evolutionary history of organisms [29-32]. Hence, we used the phylogenetic tree generated from the entire set of orthologous protein coding genes in mitochondria as the benchmark.

Our results show that phylogenetic trees from single orthologs can vary from each other (Figure 3b to 3d, Supplementary Figures S1 to S13). For example, TRO is most closely related to PPE when NADH dehydrogenase subunit 4L (ND4L) was used (Figure 3b) but most closely related to PBR when NADH dehydrogenase subunit 5 (ND5) was used (Figure 3c) for phylogenetic tree building. When NADH dehydrogenase subunit 2 was used (Figure 3d), TRO is most closely related to a sub-branch of 6 organisms (DAT, DPE, PBR, PCI, PPE, TVU, and VUR). More importantly, our results show that phylogenetic tree from single orthologs (Figure 3b to 3d) differs to that of complete set of orthologs (Figure 3a) as TRO is likely most closely related to DPE.

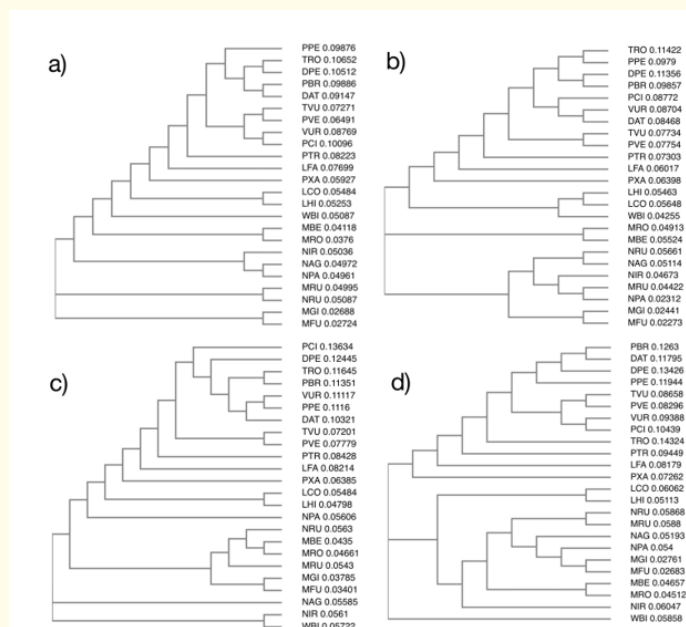


Figure 3: Phylogenetic trees from 3 orthologs and complete genome. (a) Phylogenetic tree according to the complete genome. (b) Phylogenetic tree according to NADH dehydrogenase subunit 4L. (c) Phylogenetic tree according to NADH dehydrogenase subunit 5. (d) Phylogenetic tree according to NADH dehydrogenase subunit 2.

Branch scoring of 24 organisms resulted in 276 pairwise comparisons of 14 phylogenetic trees (13 single ortholog-based trees and 1 complete ortholog-based tree); showing the minimum, average, and maximum branch scores from the 13 single ortholog-based trees, and comparing that to the branch scores from complete ortholog-based tree (Figure 4). Using paired t-test on the comparisons (n = 276), our results show that the minimum (t = 17.56, p-value = 7.89e-47), and maximum (t-statistic = 22.01, p-value = 1.32e-62) branch scores from the 13 single ortholog-based trees are significantly different to the branch scores from complete ortholog-based tree. This suggests that both minimum and maximum differs substantially from the benchmark; hence, not a good representation of evolutionary history.

Given that minimum and maximum scores are affected by outliers, it may be plausible to use average branch scores from several single ortholog-based trees. Our results show that the average branch score is also significantly different ($t = 8.01$, $p\text{-value} = 3.27e-14$) from that of complete ortholog-based tree. This implies caution for the average of trees approach [33,34] and suggests that the complete set of orthologs approach may be more reliable.

Conclusion

While it is intuitively known that the support for evolutionary history of organisms is proportional to the number of orthologs used [31], it is not clear how many ortholog(s) is/are required to give a reliable phylogeny that is representative of the species phylogeny; that is, the actual evolutionary history of the species in study? Our results suggest that the complete set of orthologs is required as both single ortholog-based phylogenetic tree and the average phylogenetic tree from multiple single ortholog-based phylogenetic tree differ significantly from the phylogenetic tree from the complete set of orthologs.

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Ashraf MA and Sarfraz M. "Biology and Evolution of Life Science". *Saudi Journal of Biological Sciences* 23.1 (2016): S1-S5.
2. Star B and Spencer HG. "Effects of Genetic Drift and Gene Flow on the Selective Maintenance of Genetic Variation". *Genetics* 194.1 (2013): 235-244.
3. Ellstrand NC and Rieseberg LH. "When Gene Flow Really Matters: Gene Flow in Applied Evolutionary Biology". *Evolutionary Applications* 9.7 (2016): 833-836.
4. Kennedy DA and Dwyer G. "Effects of Multiple Sources of Genetic Drift on Pathogen Variation within Hosts". *PLOS Biology* 16.6 (2018): e2004444.
5. Hershberg R. "Mutation The Engine of Evolution: Studying Mutation and Its Role in the Evolution of Bacteria". *Cold Spring Harbor Perspectives in Biology* 7.9 (2015): a018077.
6. Dykhuizen D. "Thoughts Toward a Theory of Natural Selection: The Importance of Microbial Experimental Evolution". *Cold Spring Harbor Perspectives in Biology* 8.3 (2016): a018044.
7. Weiss MC., et al. "The Last Universal Common Ancestor between Ancient Earth Chemistry and the Onset of Genetics". *PLOS Genetics* 14.8 (2018): e1007518.
8. Darwin C. "On the Origin of Species". Harvard University Press (1964).
9. Woese CR., et al. "Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya". *Proceedings of the National Academy of Sciences* 87.12 (1990): 4576-4579.
10. Woese C. "The Universal Ancestor". *Proceedings of the National Academy of Sciences of the United States of America* 95.12 (1998): 6854-6859.
11. Koonin EV and Wolf YI. "The Common Ancestry of Life". *Biology Direct* 5.1 (2010): 64.

12. Alcaraz LD., *et al.* "Understanding the Evolutionary Relationships and Major Traits of Bacillus through Comparative Genomics". *BMC Genomics* 11 (2010): 332.
13. Caetano Wyler S and Naciri Y. "Evolutionary Histories Determine DNA Barcoding Success in Vascular Plants: Seven Case Studies Using Intraspecific Broad Sampling of Closely Related Species". *BMC Evolutionary Biology* 16.1 (2016): 103.
14. Lee MSY and Palci A. "Morphological Phylogenetics in the Genomic Age". *Current Biology* 25.19 (2015): R922-R929.
15. Zou Z and Zhang J. "Morphological and Molecular Convergences in Mammalian Phylogenetics". *Nature Communications* 7.1 (2016): 12758.
16. Bradley BJ. "Reconstructing Phylogenies and Phenotypes: A Molecular View of Human Evolution". *Journal of Anatomy* 212.4 (2008): 337-353.
17. Wanninger A. "Morphology is Dead - Long Live Morphology! Integrating MorphoEvoDevo into Molecular EvoDevo and Phylogenomics". *Frontiers in Ecology and Evolution* (2015).
18. Sackton TB and Clark N. "Convergent Evolution in the Genomics Era: New Insights and Directions". *Philosophical Transactions of the Royal Society B* 374.1777 (2019): 20190102.
19. Hao Y., *et al.* "Genomic Insights into the Adaptive Convergent Evolution". *Current Genomics* 20.2 (2019): 81-89.
20. Llanes A., *et al.* "Betacoronavirus Genomes: How Genomic Information has been Used to Deal with Past Outbreaks and the COVID-19 Pandemic". *International Journal of Molecular Sciences* 21.12 (2020).
21. Ciccozzi M., *et al.* "The Phylogenetic Approach for Viral Infectious Disease Evolution and Epidemiology: An Updating Review". *Journal of Medical Virology* 91.10 (2019): 1707-1724.
22. Pabijan M., *et al.* "Evolutionary Principles Guiding Amphibian Conservation". *Evolutionary Applications* 13.5 (2020): 857-878.
23. Boisard J and Florent I. "Why the -Omic Future of Apicomplexa Should Include Gregarines". *Biology of the Cell* 112.6 (2020): 173-185.
24. Giribet G and Edgecombe GD. "The Phylogeny and Evolutionary History of Arthropods". *Current Biology* 29.12 (2019): R592-R602.
25. Altenhoff AM., *et al.* "Inferring Orthology and Paralogy". In: Anisimova M, editor. *Evolutionary Genomics: Statistical and Computational Methods*. New York, NY: Springer New York (2019): 149-175.
26. Fitch WM. "Distinguishing Homologous from Analogous Proteins". *Systematic Zoology* 19.2 (1970): 99-113.
27. Madeira F., *et al.* "The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019". *Nucleic Acids Research* 47.1 (2019): W636-W641.
28. Janke A., *et al.* "The Marsupial Mitochondrial Genome and the Evolution of Placental Mammals". *Genetics* 137.1 (1994): 243-256.
29. Rannala B and Yang Z. "Phylogenetic Inference using Whole Genomes". *Annual Review of Genomics and Human Genetics* 9 (2008): 217-231.
30. Richter M and Rosselló-Móra R. "Shifting the Genomic Gold Standard for the Prokaryotic Species Definition". *Proceedings of the National Academy of Sciences of the United States of America* 106.45 (2009): 19126-191231.
31. Sims GE., *et al.* "Whole-Genome Phylogeny of Mammals: Evolutionary Information in Genic and Nongenic Regions". *Proceedings of the National Academy of Sciences of the United States of America* 106.40 (2009): 17077-17082.

32. Yu X and Reva ON. "SWPhylo - A Novel Tool for Phylogenomic Inferences by Comparison of Oligonucleotide Patterns and Integration of Genome-Based and Gene-Based Phylogenetic Trees". *Evolutionary Bioinformatics Online* 14 (2018): 1176934318759299.
33. Stone EA and Sidow A. "Constructing a Meaningful Evolutionary Average at the Phylogenetic Center of Mass". *BMC Bioinformatics* 8.1 (2007): 222.
34. Brown DG and Owen M. "Mean and Variance of Phylogenetic Trees". *Systematic Biology* 69.1 (2020): 139-154.

Supplementary Materials

No.	Organism Name	Symbol	Common Name	Accession Number
1	<i>Dactylopsila trivirgata</i>	DAT	Striped Possum	NC_008134.1
2	<i>Distoechurus pennatus</i>	DPE	Feather-tailed Possum	NC_008145.1
3	<i>Lagorchestes conspicillatus</i>	LCO	Spectacled Hare Wallaby	KY996508.1
4	<i>Lagostrophus fasciatus</i>	LFA	Banded Hare Wallaby	NC_008447.1
5	<i>Lagorchestes hirsutus</i>	LHI	Rufous Hare Wallaby	NC_008136.1
6	<i>Macropus bernardus</i>	MBE	Black Wallaroo	KY996505.1
7	<i>Macropus fuliginosus</i>	MFU	Western Grey Kangaroo	NC_039717.1
8	<i>Macropus giganteus</i>	MGI	Eastern Grey Kangaroo	KY996502.1
9	<i>Macropus robustus</i>	MRO	Common Wallaroo	NC_001794.1
10	<i>Macropus rufus</i>	MRU	Red Kangaroo	KY996501.1
11	<i>Notamacropus agilis</i>	NAG	Agile Wallaby	KY996507.1
12	<i>Notamacropus irma</i>	NIR	Western Brush Wallaby	KY996503.1
13	<i>Notamacropus parma</i>	NPA	Parma Wallaby	KY996504.1
14	<i>Notamacropus rufogriseus</i>	NRU	Red-necked Wallaby	KY996499.1
15	<i>Petaurus breviceps</i>	PBR	Sugar Glider	NC_008135.1
16	<i>Phalanger vestitus</i>	PVE	Stein's Cuscus	NC_008137.1
17	<i>Phascolarctos cinereus</i>	PCI	Koala	NC_008133.1
18	<i>Potorous tridactylus</i>	PTR	Long-nosed Potoroo	NC_006524.1
19	<i>Pseudocheirus peregrinus</i>	PPE	Common Ringtail Possum	NC_006519.1
20	<i>Petrogale xanthopus</i>	PXA	Yellow-footed Rock Wallaby	KY996509.1
21	<i>Tarsipes rostratus</i>	TRO	Honey Possum	NC_006518.1
22	<i>Trichosurus vulpecula</i>	TVU	Common Brushtail Possum	NC_003039.1
23	<i>Vombatus ursinus</i>	VUR	Common Wombat	NC_003322.1
24	<i>Wallabia bicolor</i>	WBI	Swamp Wallaby	KY996500.1

Table S1: Listing of organisms.

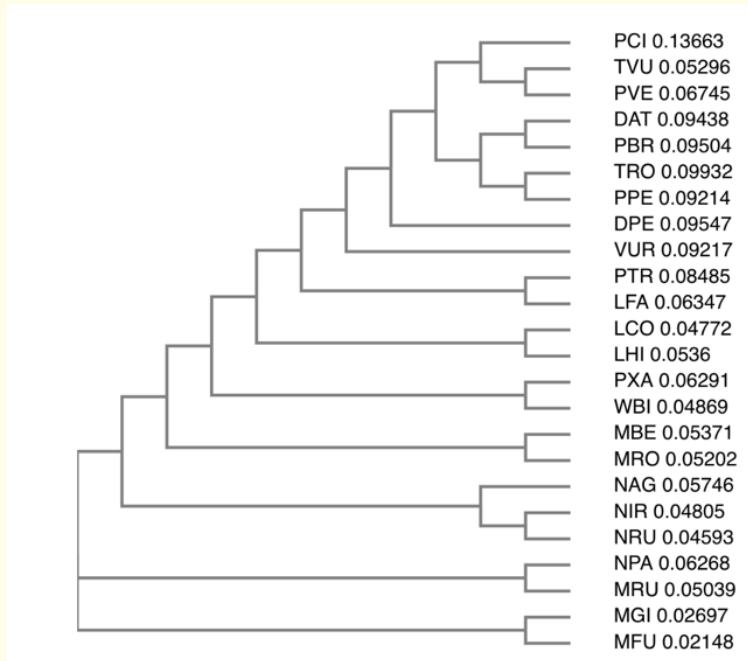


Figure S1: Phylogeny According to ATP synthase F0 subunit 6.

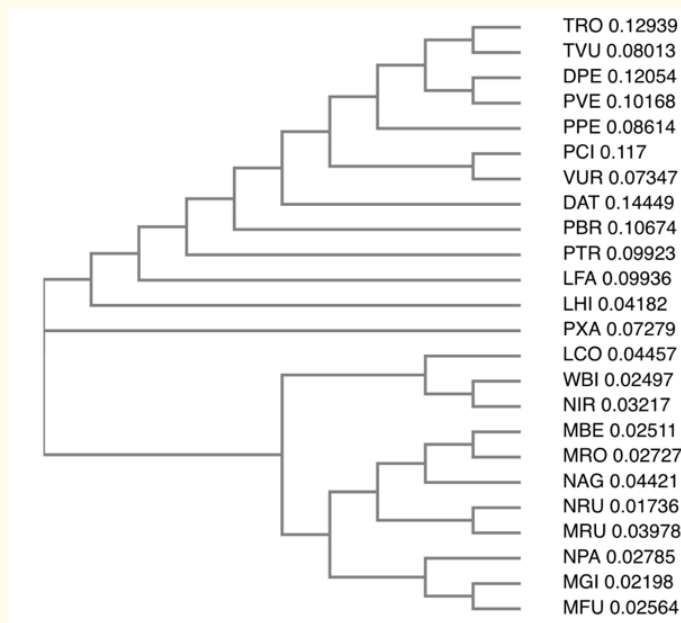


Figure S2: Phylogeny according to ATP synthase F0 subunit 8.

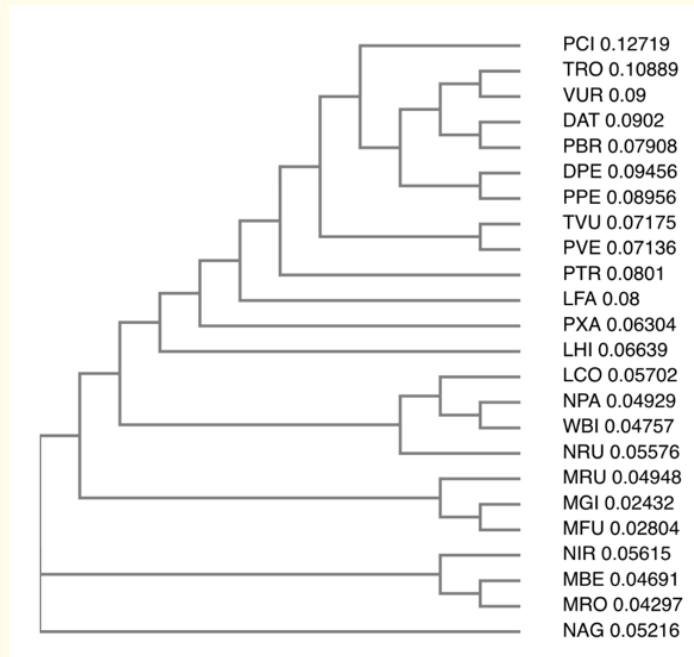


Figure S3: Phylogeny according to cytochrome b.

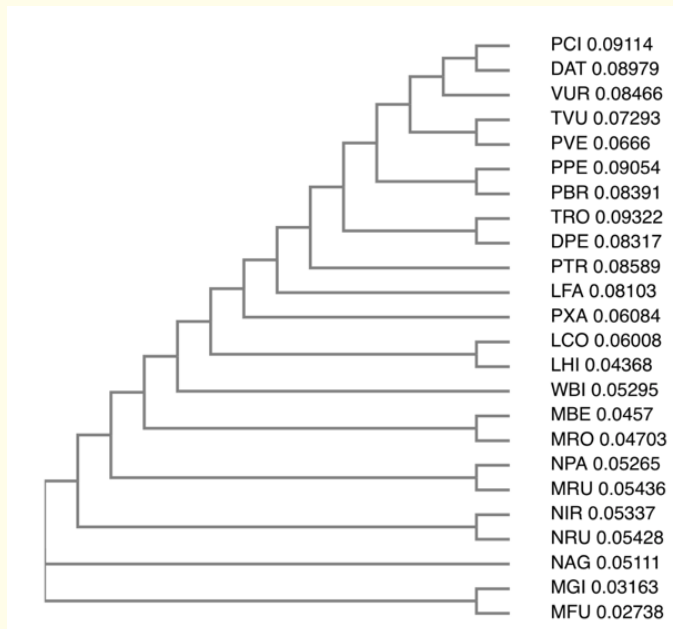


Figure S4: Phylogeny according to cytochrome c oxidase subunit 1.

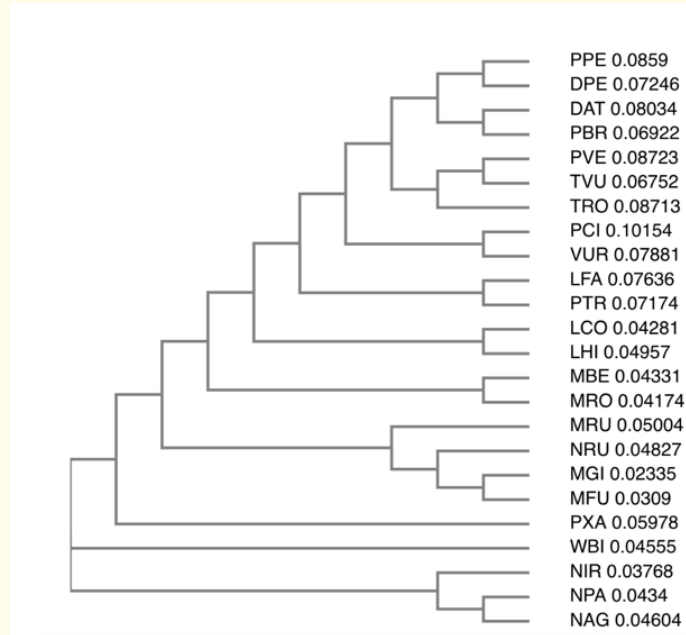


Figure S5: Phylogeny according to cytochrome c oxidase subunit 2.

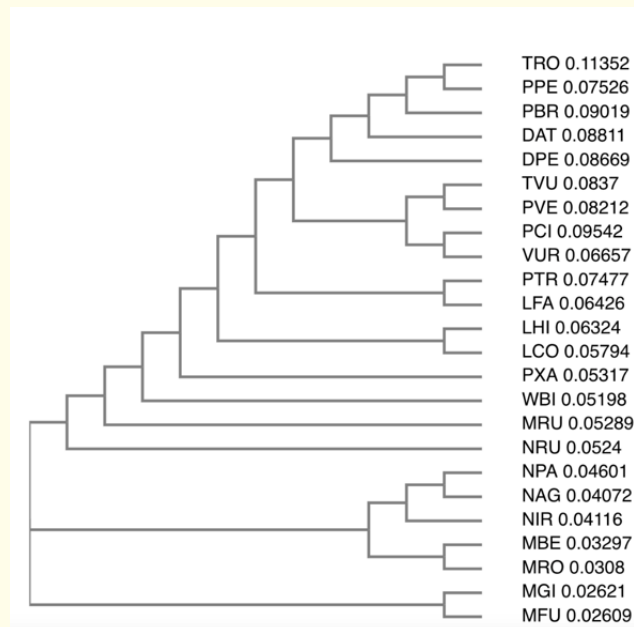


Figure S6: Phylogeny according to cytochrome c oxidase subunit 3.

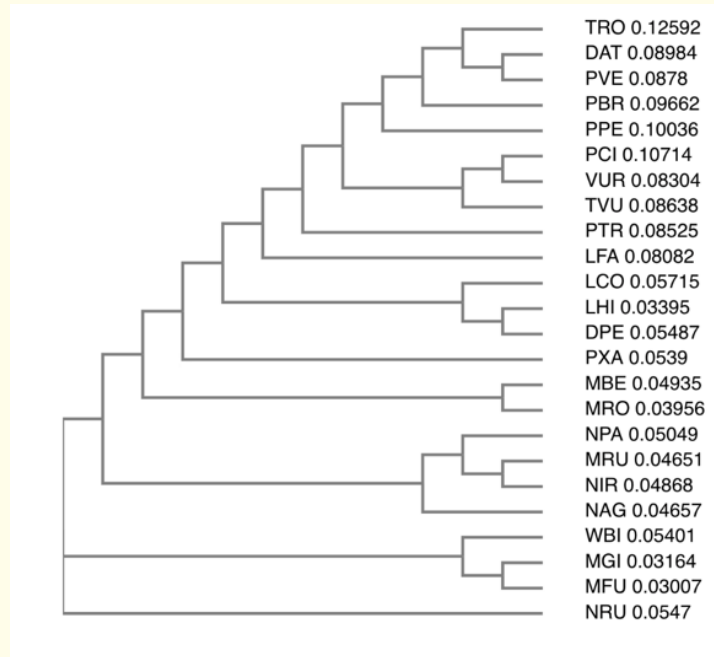


Figure S7: Phylogeny according to NADH dehydrogenase subunit 1.

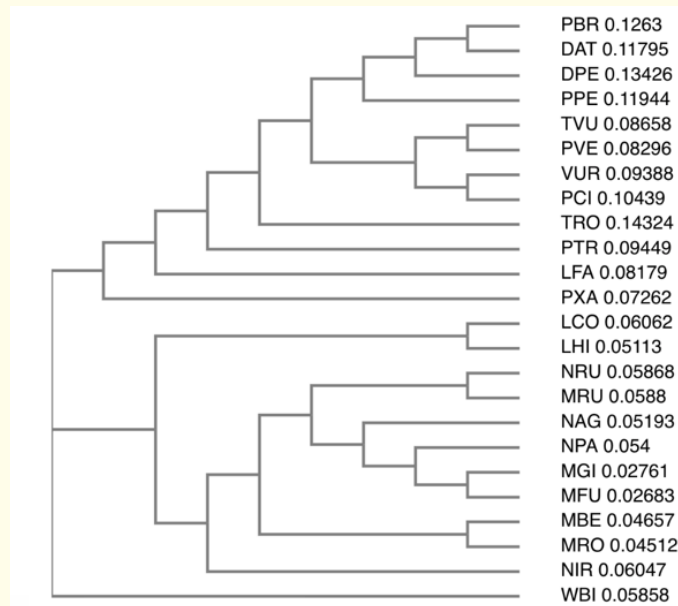


Figure S8: Phylogeny according to NADH dehydrogenase subunit 2.

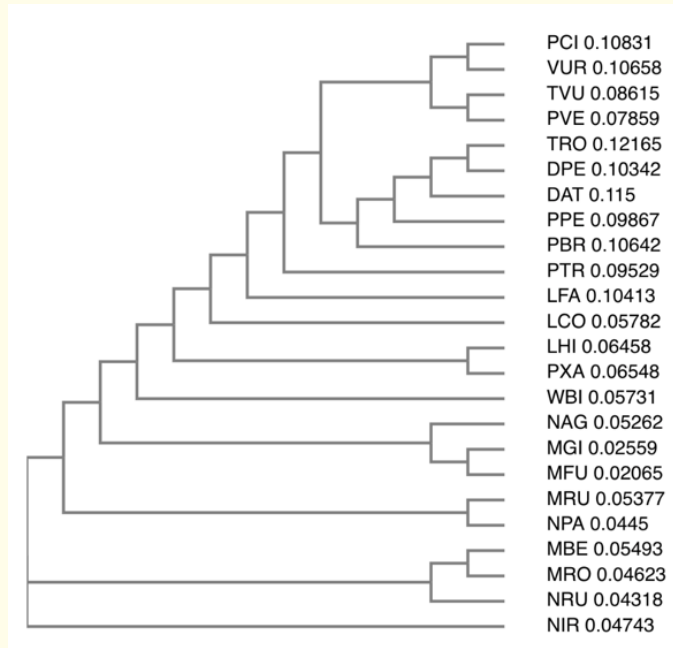


Figure S9: Phylogeny according to NADH dehydrogenase subunit 3.

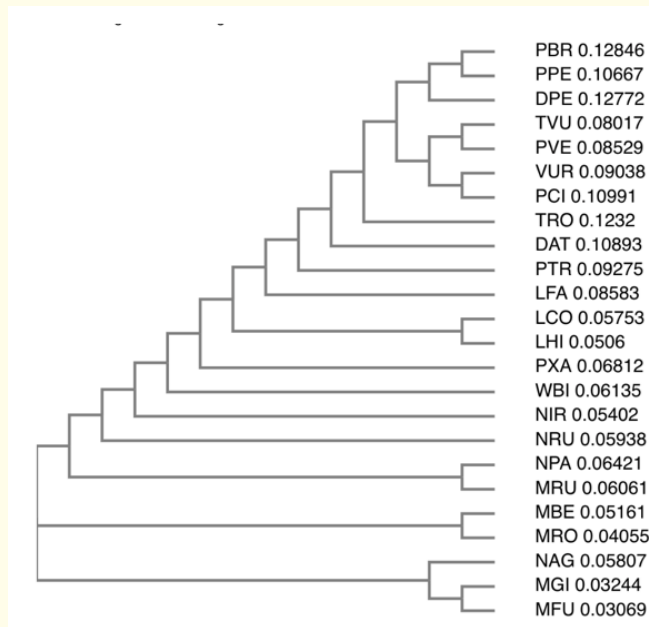


Figure S10: Phylogeny according to NADH dehydrogenase subunit 4.

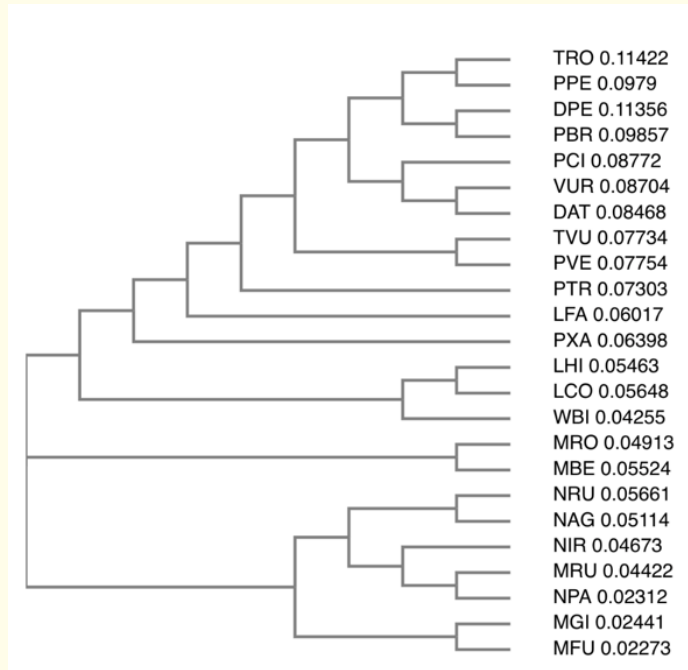


Figure S11: Phylogeny according to NADH dehydrogenase subunit 4L.

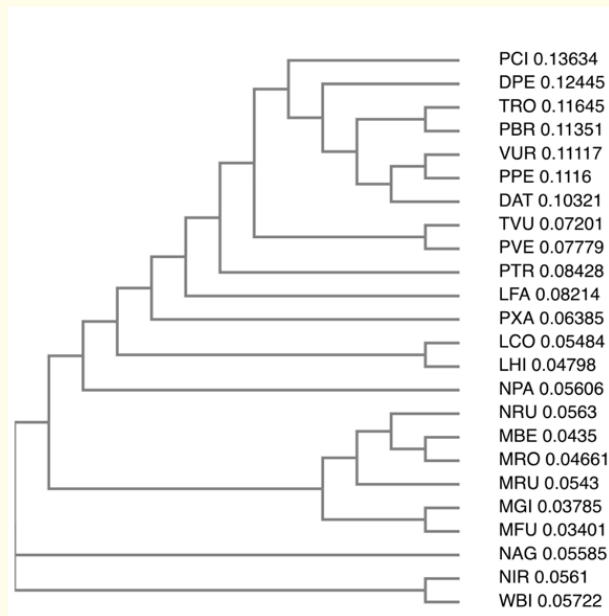


Figure S12: Phylogeny according to NADH dehydrogenase subunit 5.

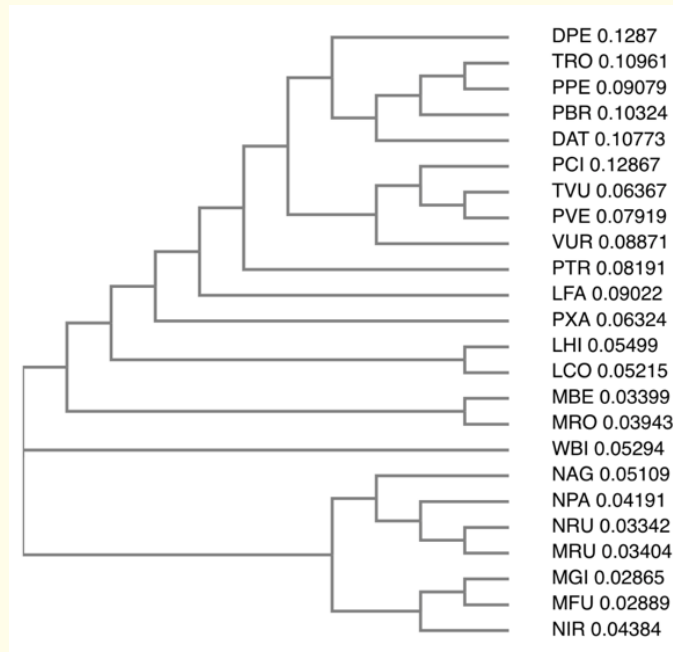


Figure S13: Phylogeny according to NADH dehydrogenase subunit 6.

Volume 5 Issue 9 September 2020

©All rights reserved by Maurice HT Ling, et al.