



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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



Data Article

Genome statistics and phylogenetic reconstructions for Southern Hemisphere whelks (Gastropoda: Buccinulidae)

Felix Vaux^{a,*}, Simon F.K. Hills^a, Bruce A. Marshall^b,
Steve A. Trewick^a, Mary Morgan-Richards^a

^a Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand

^b Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand

ARTICLE INFO

Article history:

Received 29 June 2017

Received in revised form

1 November 2017

Accepted 3 November 2017

Available online 7 November 2017

ABSTRACT

This data article provides genome statistics, phylogenetic networks and trees for a phylogenetic study of Southern Hemisphere Buccinulidae marine snails [1]. We present alternative phylogenetic reconstructions using mitochondrial genomic and 45S nuclear ribosomal cassette DNA sequence data, as well as trees based on short-length DNA sequence data. We also investigate the proportion of variable sites per sequence length for a set of mitochondrial and nuclear ribosomal genes, in order to examine the phylogenetic information provided by different DNA markers. Sequence alignment files used for phylogenetic reconstructions in the main text and this article are provided here.

© 2017 The Authors. 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: <https://doi.org/10.1016/j.ympbev.2017.06.018>

* Corresponding author.

E-mail address: felixvaux.evolution@gmail.com (F. Vaux).

¹ Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand.

<https://doi.org/10.1016/j.dib.2017.11.021>

2352-3409/© 2017 The Authors. 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
More specific subject area	Phylogenetics; Genetics; Evolutionary Biology
Type of data	Table, text file, graph, figure
How data was acquired	High-throughput and Sanger DNA sequencing
Data format	Text file format for DNA sequence alignments and phylogenetic trees is.nex (nexus) and.tree respectively.
Experimental factors	Total DNA was extracted from specimens using CTAB buffer. DNA was paired-end sequenced using the high-throughput Illumina HiSeq. 2500 platform. Short-length DNA sequences were amplified via PCR and Sanger sequenced.
Experimental features	mtDNA genome and 45 S nuclear ribosomal DNA sequences were assembled using reference sequences. Sequences were aligned with gaps and poorly-aligned positions removed. Phylogenetic trees were constructed using Bayesian (BEAST 1.8.3) and Maximum-Likelihood methods (RAxML 8.2.8). The unrooted phylogenetic network of some alignments was investigated using SplitsTree 4.
Data source location	Most specimens originate from New Zealand waters, some were collected from the coasts of Australia, Japan, USA (California), and the UK.
Data accessibility	Interactive.nwk (Newick) tree files are provided here and with the main article [1].

Value of the data

- Summary statistics for whole mitochondrial DNA sequences and 45S nuclear ribosomal genes are presented because such information for gastropods is currently rare, and base bias is known to influence phylogenetic inferences.
- Phylogenetic reconstructions (from short-length DNA sequence data) presented here include multiple buccinid and buccinulid taxa not included in the main-text trees, and may be useful to future evolutionary studies of Neogastropoda.
- DNA sequence variation and phylogenetic trees are provided because Southern Hemisphere taxa are currently under-sampled.
- The proportion of variable DNA sites for a selection of mitochondrial and nuclear genes from buccinulid whelks are compared. This information can improve genetic marker selection for future molluscan studies.

1. Data

The data presented here originates from a phylogenetic study of Southern Hemisphere whelks [1], which refers to a group of marine snails that can be classified within the taxonomic families Buccinulidae or Buccinidae [2–8]. The classification of these gastropod snails depends upon a biogeographic hypothesis and an assumption of reciprocal monophyly between the majority of lineages in the Northern and Southern Hemispheres [3,6–8]. Results from our study indicated that Buccinulidae and Southern Hemisphere whelks are not monophyletic [1].

32 putative buccinulid and buccinid marine snails, as well as three fasciolariid snails used as a phylogenetic outgroup, were high-throughput sequenced on the Illumina 2500 platform. Sequence data was assembled to provide mitochondrial (mtDNA) genomic and 45S nuclear ribosomal DNA (rDNA) sequence data for most taxa, although some individuals failed to successfully sequence for the entire mtDNA or rDNA. This data was complemented with short-length sequence data from the mitochondrial 16S RNA and *cox1* genes and nuclear ribosomal 28S RNA gene. This short-length

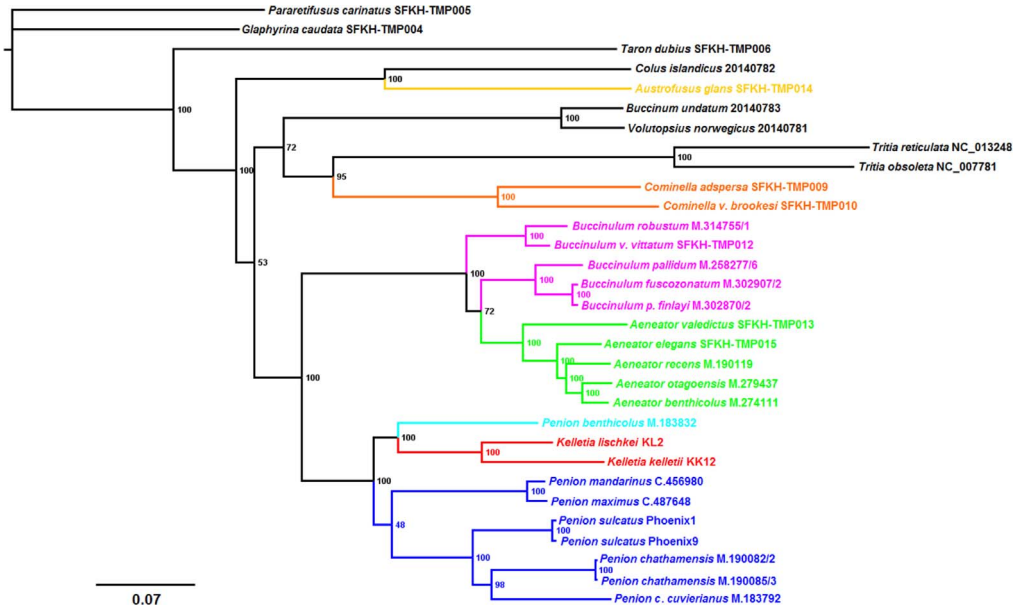


Fig. 1. Maximum-likelihood mtDNA phylogeny of buccinid and buccinulid whelks. A maximum-likelihood derived phylogeny generated using RAxML 8.2.8 [9], based on an alignment of 31 concatenated mitochondrial genome sequences (11,128 bp incorporating protein-encoding, tRNA and rRNA genes). No partitions were used. No outgroup or monophyly was enforced for this tree. Genera putatively belonging to Buccinulidae are shown in different colours.

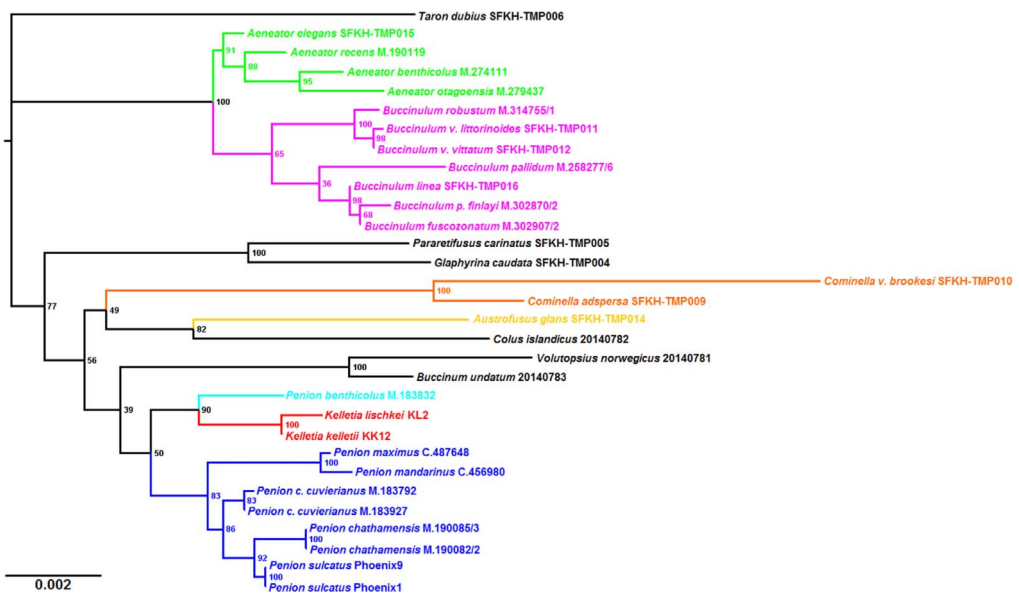


Fig. 2. Maximum-likelihood 45S rDNA phylogeny of buccinid and buccinulid whelks. A maximum-likelihood derived phylogeny generated using RAxML 8.2.8 [9], based on a 4667 bp alignment of 31 concatenated nuclear rDNA gene sequences (18S, 5.8S, 28S rRNA). No partitions were used. No outgroup or monophyly was enforced for this tree. Genera putatively belonging to Buccinulidae are shown in different colours.

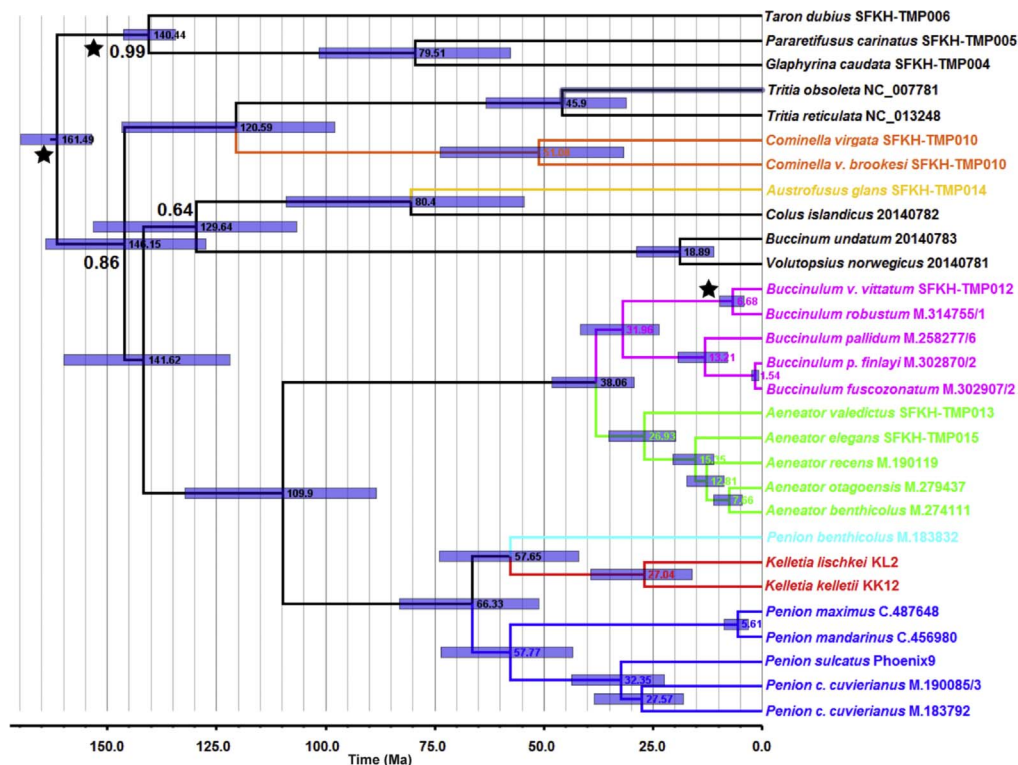


Fig. 3. Bayesian calibrated mtDNA phylogeny of buccinid and buccinulid whelks. A Bayesian phylogeny based on an alignment of 25 concatenated mitochondrial genome sequences (incorporating protein-encoding, tRNA and rRNA genes), which has been fossil calibrated to estimate divergence dates among the whelk lineages. Two sequence partitions were used: 1) protein-encoding and tRNA genes (10,635 bp), and 2) tRNA genes (1065 bp) using the GTR + I + G and HKY + I + G substitution models respectively [10,11]. Black stars indicate splits that fossil calibrated. Tree root height was calibrated using the earliest known buccinoid fossils [12], and fossil calibrations were also used for the earliest Fascioliariidae (un-enforced outgroup) [13,14], and the earliest known occurrence of the tip branch *Buccinulum v. vittatum* [15]. BEAST 1.8.3 [16] using and MCMC length of 100 million, 1000 sample frequency and a 10% burn-in was used to generate this phylogeny. Node labels are estimated median divergence dates with the 95% highest posterior density (HPD) range shown as a blue bar. Posterior support values are also shown at nodes, but only if support was < 1.0. Putative buccinulid genera are shown in different colours.

sequence data was acquired via PCR amplification and Sanger sequencing using universal primers. Sequence alignments used for analyses presented in the main text are attached to this paper.

Using these sequence alignments, we present maximum-likelihood and Bayesian phylogenetic reconstructions for the sampled buccinulid whelks. These phylogenetic trees are alternative reconstructions that can be compared to trees presented in the main text. Splits networks are also estimated using the mtDNA genomic and nuclear ribosomal RNA (18S, 5.4S, 28S) sequence data. The proportion of variable sites per sequence length for a set of mitochondrial and nuclear ribosomal genes is investigated as well, which provides insight towards marker information for recent and distant evolutionary change in neogastropods (Figs. 1 and 9).

2. Experimental design, materials and methods

The DNA extraction, purification, sequencing method and routine for sequence assembly is provided in the main text [1]. The main text also explains how the figures presented here were

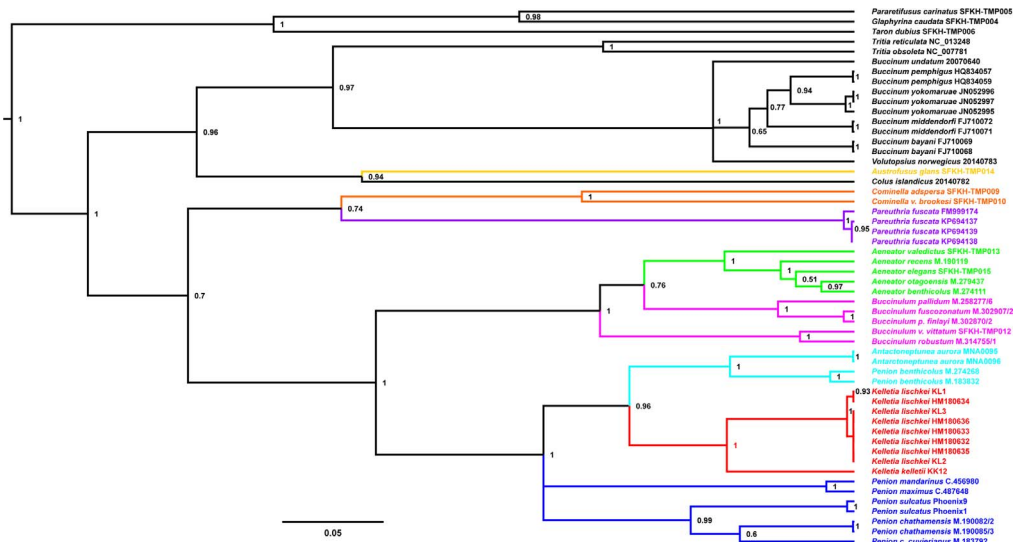


Fig. 4. Bayesian *cox1* phylogeny of buccinid and buccinulid whelks. A Bayesian phylogeny based on a 439 bp alignment of mitochondrial *cox1* gene sequences obtained from 54 individual marine snails. The GTR + I + G substitution model was used [11]. The phylogeny was produced using a Bayesian method (100 million MCMC, 10% burn-in, 1000 sample frequency, node labels are posterior support values), via BEAST 1.8.3 [16]. For this tree no outgroup was specified explicitly but reciprocal monophyly was enforced for the Fasciolaridae and Buccinidae/Buccinulidae/Nassariidae. Genera putatively belonging to Buccinulidae are shown in different colours.

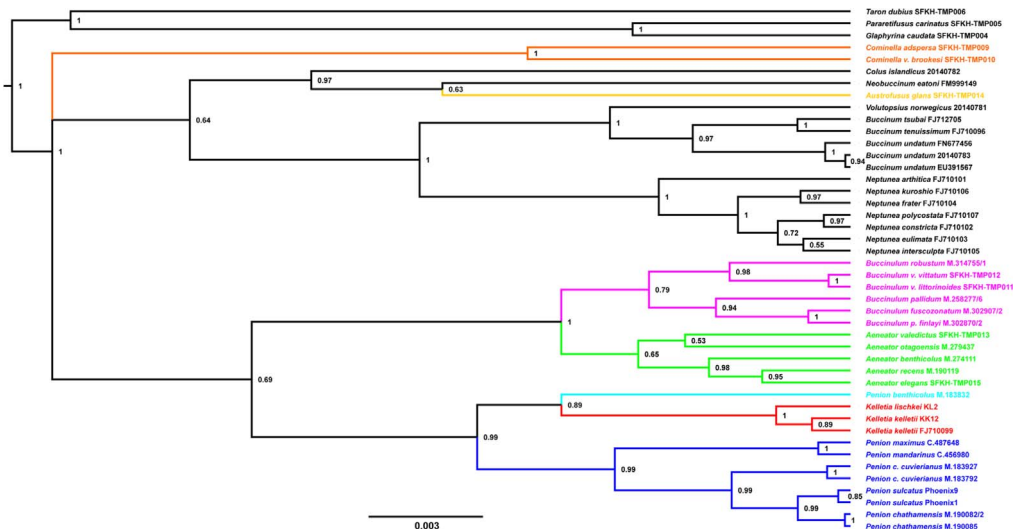


Fig. 5. Bayesian 28S RNA phylogeny of buccinid and buccinulid whelks. A Bayesian phylogeny based on an alignment of nuclear ribosomal 28 S RNA gene sequences obtained from 44 individual marine snails (1486 bp). The GTR + I + G substitution model was used [11]. The phylogeny was produced using a Bayesian method (100 million MCMC, 10% burn-in, 1000 sample frequency, node labels are posterior support values), via BEAST 1.8.3 [16]. For this tree no outgroup was specified explicitly but reciprocal monophyly was enforced for the Fasciolaridae and Buccinidae/Buccinulidae/Nassariidae. Genera putatively belonging to Buccinulidae are shown in different colours.

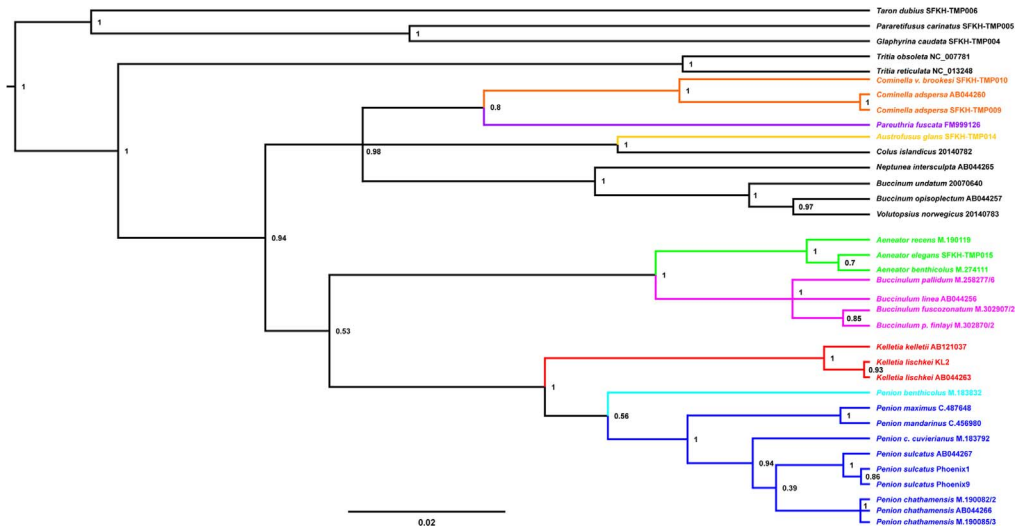


Fig. 6. Bayesian 16S RNA phylogeny of buccinid and buccinulid whelks. A Bayesian phylogeny based on an alignment of mitochondrial 16S RNA gene sequences obtained from 35 individual marine snails (868 bp). The GTR + I + G substitution model was used [11]. The phylogeny was produced using a Bayesian method (100 million MCMC, 10% burn-in, 1000 sample frequency, node labels are posterior support values), via BEAST 1.8.3 [16]. For this tree no outgroup was specified explicitly but reciprocal monophyly was enforced for the Fascioliariidae and Buccinidae/Buccinulidae/Nassariidae. Genera putatively belonging to Buccinulidae are shown in different colours.

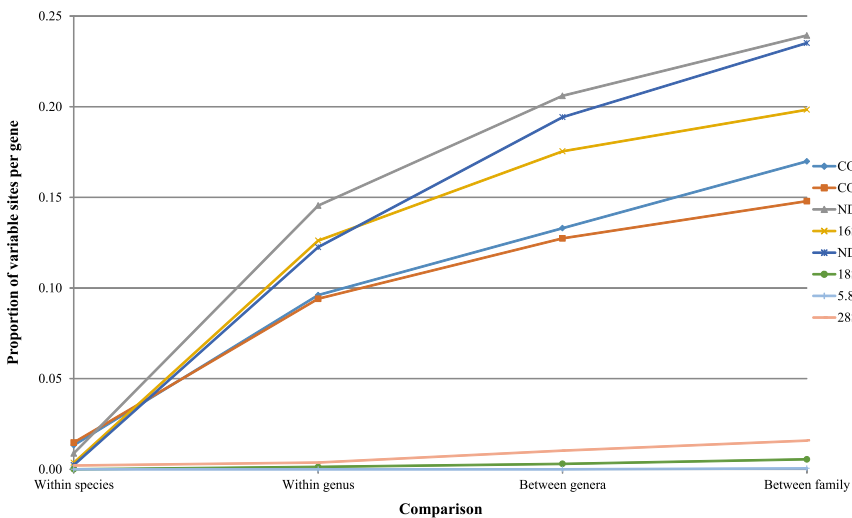


Fig. 7. Proportion of variable sites at increasingly deep levels of divergence. The proportion of variable sites per sequence length (bp) for a selection of mtDNA and nuclear rDNA genes reflects different rates of DNA substitution. Values were calculated using Geneious 9.1.3 [17]. The trends plotted effectively represent change in the phylogenetic information provided by each gene for different levels of investigation. Average numbers of variable sites were used for groups in genus and family-level comparisons. For example, we used the average number of differences for all sampled whelk (Buccinidae/Buccinulidae) taxa from all sampled Fascioliariidae taxa. Sampling from *Aeneator*, *Buccinulum* and *Penion* was used to estimate generic-level differences as these groups contained more than two specimens. Likewise, only *P. sulcatus*, *P. chathamensis*, and *P. c. curvierianus* were used for within-species estimates as these taxa were sampled twice. Since read coverage varies for some genes, not all individuals were included for estimates made for each gene.

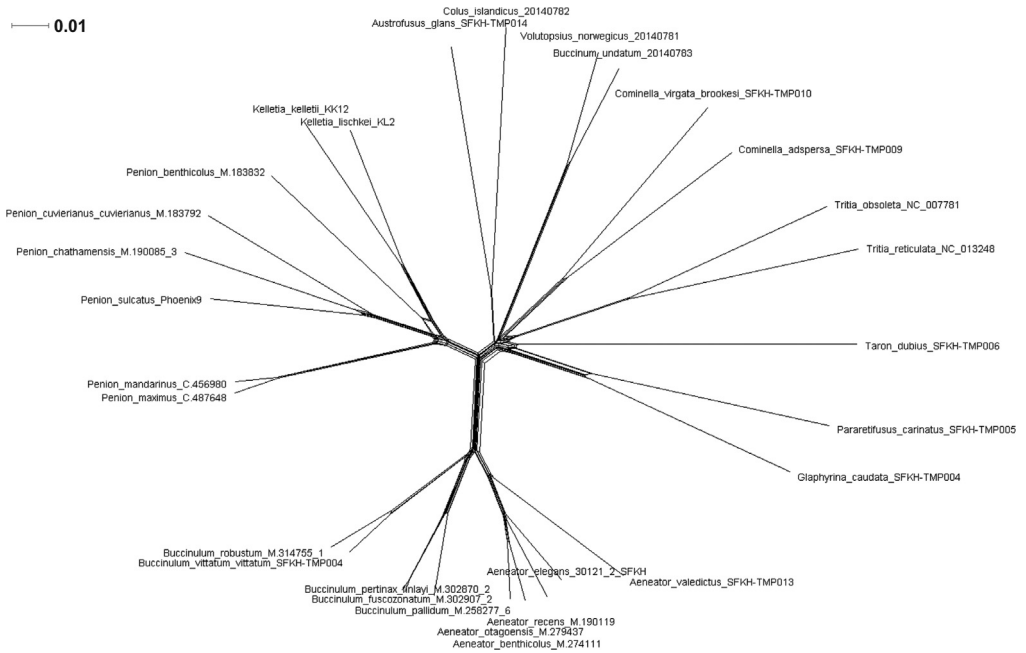


Fig. 8. Splits network illustrating alternative phylogenetic signal in mtDNA sequence data for marine snails. The splits network of based on an alignment of 31 concatenated mitochondrial genome sequences (incorporating protein-encoding, tRNA and rRNA genes; 11,128 bp). Splits were generated using the Neighbor-Net algorithm in SplitsTree 4 [18]. The splits network presents a generalisation of all of possible topological solutions for the phylogenetic signal contained in the underlying sequence data, but it does not quantify the likelihood of alternative phylogenetic relationships. Edge length is proportional to split weight, and box structures within the network indicate signal for alternative topologies in the underlying sequence data.

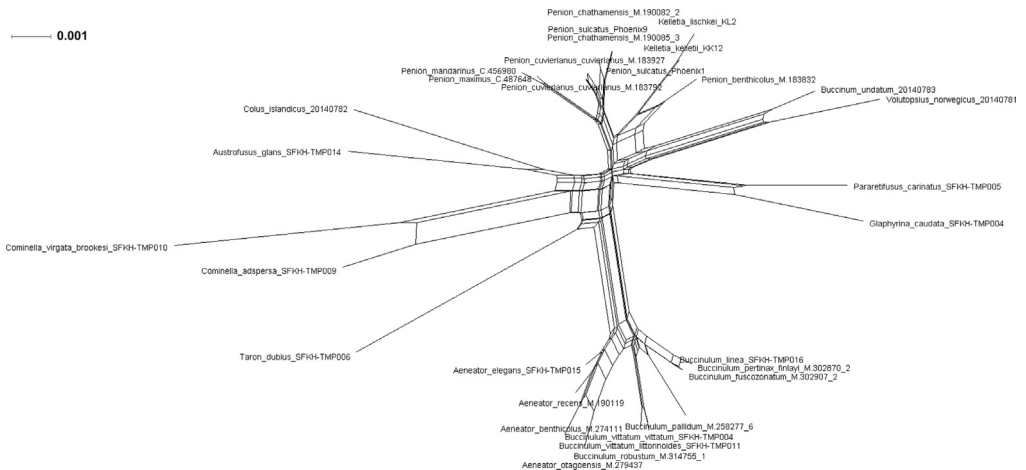


Fig. 9. Splits network for illustrating alternative phylogenetic signal in 45S rDNA sequence data for marine snails. The splits network of based on a 4667 bp alignment of 31 concatenated nuclear rDNA gene sequences (18S, 5.8S, 28S rRNA genes). Splits were generated using the Neighbor-Net algorithm in SplitsTree 4 [18]. The splits network presents a generalisation of all of possible topological solutions for the phylogenetic signal contained in the underlying sequence data, but it does not quantify the likelihood of alternative phylogenetic relationships. Edge length is proportional to split weight, and box structures within the network indicate signal for alternative topologies in the underlying sequence data.

Table 1

A summary of statistics for the length and nucleotide composition for the concatenated DNA sequences for the nuclear ribosomal RNA genes 18S, 5.8S and 28S (the internal transcribed spacer regions are not included). All listed specimens were newly sequenced for this study.

Species	Museum ID	Concatenated nuclear rDNA 18S, 5.8S, 28S					
		Length (bp)	% A	% C	% G	% T	GC bias
<i>Pararetifusus carinatus</i>	SFKH-TMP005	5337	23	24.5	30.0	22.2	54.5
<i>Glaphyrina caudata</i>	SFKH-TMP004	5339	23	24.5	30.0	22.2	54.5
<i>Taron dubius</i>	SFKH-TMP006	5339	23	24.7	30.1	22.0	54.8
<i>Austrofuscus glans</i>	SFKH-TMP014	5338	24	24.4	30.0	22.2	54.4
<i>Colus islandicus</i>	20140782	5334	24	24.5	30.0	22.2	54.5
<i>Volutopsius norwegicus</i>	20140781	5338	24	24.4	30.0	22.3	54.4
<i>Buccinum undatum</i>	20140783	5339	24	24.3	30.0	22.3	54.3
<i>Cominella adspersa</i>	SFKH-TMP009	5339	23	24.6	30.0	22.0	54.6
<i>Cominella v. brookesi</i>	SFKH-TMP010	5339	21	24.9	30.3	21.7	55.2
<i>Buccinulum fuscozonatum</i>	M.302907/2	5340	22	24.8	30.1	22.0	54.9
<i>Buccinulum linea</i>	SFKH-TMP016	5340	22	24.8	30.1	22.0	54.9
<i>Buccinulum v. littorinoides</i>	SFKH-TMP011	5340	22	24.7	30.1	22.0	54.8
<i>Buccinulum pallidum</i>	M.258277/6	5340	22	24.7	30.2	21.9	54.9
<i>Buccinulum p. finlayi</i>	M.302870/2	5340	22	24.7	30.1	22.0	54.8
<i>Buccinulum robustum</i>	M.314755/1	5340	22	24.8	30.1	21.9	54.9
<i>Buccinulum v. vittatum</i>	SFKH-TMP004	5340	22	24.7	30.1	22.0	54.8
<i>Aeneator benthicolus</i>	M.274111	5340	22	24.6	30.1	22.1	54.7
<i>Aeneator elegans</i>	SFKH-TMP015	5340	22	24.7	30.1	22.0	54.8
<i>Aeneator otagoensis</i>	M.279437	5340	22	24.7	30.2	21.9	54.9
<i>Aeneator recens</i>	M.190119	5340	22	24.6	30.1	22.1	54.7
<i>Penion benthicolus</i>	M.183832	5337	23	24.4	30.0	22.3	54.4
<i>Kelletia kelletii</i>	KK12	5337	24	24.4	29.9	22.3	54.3
<i>Kelletia lischkei</i>	KL2	5337	24	24.3	29.9	22.4	54.2
<i>Penion mandarinus</i>	C.456980	5339	24	24.4	29.9	22.2	54.3
<i>Penion maximus</i>	C.487648	5339	25	24.4	29.9	22.2	54.3
<i>Penion sulcatus</i>	Phoenix9	5339	23	24.4	30.0	22.3	54.4
<i>Penion sulcatus</i>	Phoenix1	5339	23	24.4	30.0	22.3	54.4
<i>Penion chathamensis</i>	M.190085/3	5339	23	24.4	29.9	22.3	54.3
<i>Penion chathamensis</i>	M.190082/2	5339	23	24.4	29.9	22.3	54.3
<i>Penion c. cuvierianus</i>	M.183792	5339	23	24.3	30.0	22.3	54.3
<i>Penion c. cuvierianus</i>	M.183927	5339	23	24.3	30.0	22.4	54.3

Table 2

A summary of the statistics for the length and nucleotide composition for the mitochondrial genomes newly sequenced as part of this study. Specimens marked with one asterisk (*) exhibit drops in read coverage for some small regions, for example *K. kelletii* has 54 bp missing from *cox1*. Specimens marked with two asterisks (**) have genomes with large gaps in genome coverage for some regions, such as *B. v. vittatum* that has 266, 151 and 64 bp missing from the ATP6, *cox1* and ND2 genes respectively.

Species	Museum ID	mtDNA genome					
		Length (bp)	% A	% C	% G	% T	GC bias
<i>Pararetifusus carinatus</i>	SFKH-TMP005	15204	31.5	14	15.0	40.1	28.4
<i>Glaphyrina caudata</i>	SFKH-TMP004	15235	31.5	13	14.6	40.7	27.9
<i>Taron dubius</i>	SFKH-TMP006	15189	29.3	15.7	17.0	38.0	32.7
<i>Austrofuscus glans</i>	SFKH-TMP014	15195	31.1	14.5	15.3	39.1	29.8
<i>Colus islandicus</i>	20140782	15158	30.5	14.9	15.8	38.7	30.7
<i>Volutopsius norwegicus</i>	20140781	15232	29.3	15.7	16.5	38.4	32.2
<i>Buccinum undatum</i>	20140783	15231	29.5	15.6	16.3	38.7	31.9
<i>Cominella adspersa</i>	SFKH-TMP009	15251	30.4	15.7	16.0	38.0	31.7
<i>Cominella v. brookesi</i>	SFKH-TMP010	15263	29.6	15.9	16.7	37.8	32.6
<i>Buccinulum fuscozonatum</i>	M.302907/2	15246	30.2	14.8	15.8	39.1	30.6

Table 2 (continued)

Species	Museum ID	mtDNA genome					GC bias
		Length (bp)	% A	% C	% G	% T	
<i>Buccinulum pallidum</i>	M.258277/6	15247	30.9	14.1	15.2	39.7	29.3
<i>Buccinulum p. finlayi</i>	M.302870/2	15247	30.1	14.8	15.9	39.1	30.7
<i>Buccinulum robustum</i>	M.314755/1	15244	29.6	15.2	16.1	39.0	31.3**
<i>Buccinulum v. vittatum</i>	SFKH-TMP012	15244	29.6	15.1	16.4	38.9	31.5**
<i>Aeneator benthicolus</i>	M.274111	15254	30.4	14.7	15.7	39.2	30.4
<i>Aeneator elegans</i>	SFKH-TMP015	15254	30.3	14.6	15.8	39.3	30.4
<i>Aeneator otagoensis</i>	M.279437	15249	30.3	14.7	15.5	39.5	30.2*
<i>Aeneator recens</i>	M.190119	15264	30.0	14.9	16.0	39.1	30.9
<i>Aeneator valedictus</i>	SFKH-TMP013	15258	29.3	15.8	16.7	38.2	32.5
<i>Penion benthicolus</i>	M.183832	15229	29.5	16.2	17.0	37.3	32
<i>Kelletia kelleitii</i>	KK12	15104	29.3	16.0	17.1	37.6	31*
<i>Kelletia lischkei</i>	KL2	15225	29.6	16.1	16.8	37.5	32.9
<i>Penion mandarinus</i>	C.456980	15250	30.4	15.1	16.2	38.3	31.3
<i>Penion maximus</i>	C.487648	15249	30.6	15.1	16.0	38.2	31.1
<i>Penion sulcatus</i>	Phoenix9	15227	29.2	16.0	17.2	37.5	32
<i>Penion sulcatus</i>	Phoenix1	15227	29.2	16.1	17.2	37.4	33
<i>Penion chathamensis</i>	M.190085/3	15227	28.6	16.8	18.0	36.7	34.8
<i>Penion chathamensis</i>	M.190082/2	15228	28.5	16.8	18.0	36.7	34.8
<i>Penion c. cuvierianus</i>	M.183792	15235	28.6	16.9	17.8	36.7	34.7
<i>Penion c. cuvierianus</i>	M.183927	15241	28.3	17.1	18.0	36.6	35.1**

generated, including the software and settings used. Legends for tables and figures presented below specify which sequence alignments were used (again referenced in the main text) (Tables 1 and 2).

Acknowledgements

This work was supported by the Royal Society of New Zealand Te Aparangi Marsden Fund grant (12-MAU-008), a Ministry of Business, Innovation and Employment Te Tipu Putaiao Postdoctoral Fellowship (CONT-22922-TTP-MAU), and a funding contribution from the Department Of Conservation Taxonomic and Threat Status Information fund (RIF 4718).

This work includes samples collected as part of two Antarctic survey projects: TAN0402: A biodiversity survey financed by the former New Zealand Ministry of Fisheries.

TAN0802: Funded by the New Zealand Government under the New Zealand International Polar Year Census of Antarctic Marine Life Project (Phase 1: So001IPY; Phase 2; IPY2007-01).

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2017.11.021>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2017.11.021>.

References

- [1] F. Vaux, S.F.K. Hills, B.A. Marshall, S.A. Trewick, M. Morgan-Richards, A phylogeny of Southern Hemisphere whelks (Gastropoda: Buccinulidae) and concordance with the fossil record, *Mol. Phylogenet. Evol.* 114 (2017) 367–381.
- [2] J. Thiele, Die antarktischen Socken und Muscheln, *Dtsch. Südopolar-Exped. 1901–1903* 13 (1912) 183–285.
- [3] H.J. Finlay, The recent Mollusca of the Chatham Islands, *Trans. NZ Inst.* 59 (1928) 232–286.
- [4] A.W.B. Powell, The recent and Tertiary species of the genus *Buccinulum* in New Zealand, with a review of related genera, *Trans. NZ Inst.* 60 (1929) 57–101.
- [5] W. Wenz, *Gastropoda. Teil 1: allgemeiner Teil und Prosobranchia*, in: G.H. Schindewolf (Ed.), *Handbuch der Paläozoologie*, Borntraeger, Berlin, 1941.
- [6] A.W.B. Powell, Antarctic and subantarctic Mollusca: Pelecypoda and Gastropoda, *Discov. Rep.* 26 (1951) 47–196.
- [7] M.G. Harasewych, Y.I. Kantor, A revision of the Antarctica genus *Chlanidota* (Gastropoda: Neogastropoda: Buccinulidae), *Proc. Biol. Soc. Wash.* 112 (1999) 253–302.
- [8] S. Hayashi, The molecular phylogeny of the Buccinidae (Caenogastropoda: Neogastropoda) as inferred from the complete mitochondrial 16S rRNA gene sequences of selected representatives, *Molluscan Res.* 25 (2005) 85–98.
- [9] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics* 30 (2014) (1312–131).
- [10] M. Hasegawa, K. Kishino, T. Yano, Dating the human-ape splitting by a molecular clock of mitochondrial DNA, *J. Mol. Evol.* 22 (1985) 160–174.
- [11] S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences, *Lect. Math. Life Sci.* 17 (1986) 57–86.
- [12] A. Kaim, A.L. Beisel, Mesozoic gastropods from Siberia and Timan (Russia). Part 2: Neogastropoda and Heterobranchia, *Pol. Polar Res.* 26 (2005) 41–64.
- [13] E.C. Allison, Middle Cretaceous Gastropoda from Punta China, Baja California, Mexico, *J. Paleol.* 29 (1955) 400–432.
- [14] S. Tracey, J.A. Todd, D.H. Erwin, Mollusca: Gastropoda, in: M.J. Benton (Ed.), *The Fossil Record*, Chapman & Hall, London, 1993, pp. 131–167.
- [15] A.G. Beu, P.A. Maxwell, Cenozoic Mollusca of New Zealand, *New Zealand Geol. Surv. Bull.* (1990) 58.
- [16] A.J. Drummond, M.A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7, *Mol. Biol. Evol.* 29 (2012) 1969–1973.
- [17] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics* 28 (2012) 1647–1649.
- [18] D.H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies, *Mol. Biol. Evol.* 23 (2006) 254–267.