

Testing heterogeneous base composition as potential cause for conflicting phylogenetic signal between mitochondrial and nuclear DNA in the land snail genus *Theba* Risso 1826 (Gastropoda: Stylommatophora: Helicoidea)

Andre Böckers¹ · Carola Greve² · Rainer Hutterer² · Bernhard Misof² · Martin Haase¹

Received: 20 December 2015 / Accepted: 17 May 2016 / Published online: 27 May 2016
© Gesellschaft für Biologische Systematik 2016

Abstract Several pitfalls can mislead phylogenetic analyses based on molecular data, including heterogeneous base composition. Previous work has revealed conflicting topologies in analyses of the land snail genus *Theba* Risso 1826 based on mitochondrial cytochrome oxidase subunit I (COI) and nuclear AFLP data, respectively. However, the third codon positions of COI had heterogeneous base composition, prompting the present investigation asking specifically if this was the cause for the mito-nuclear discordance. For a potentially better resolution of the mitochondrial data, we also sequenced a fragment of 16S rRNA, the loop sections of which proved to have inhomogeneous base frequencies as well. In partitioned phylogenetic analyses, we compared topologies generated from the original data to those based on alignments in which the heterogeneous partitions were RY-coded and to a LogDet transformed distance analysis. In addition, we tested whether conventional Bayesian analyses would reconstruct the original topology from inhomogeneous data simulated based on this original topology. All our analyses, regardless of whether we accounted for heterogeneous base frequencies or not, revealed very similar topologies, confirming previous findings. Thus, the phylogenetic signal of mtDNA in the land snail

genus *Theba* appeared to be robust despite considerable inhomogeneity of base composition. Therefore, the discordance of mitochondrial and nuclear topologies is probably real and most likely a consequence of incomplete lineage sorting.

Keywords Base composition · Conflicting phylogenetic signal · Heterogeneity · LogDet · Non-stationary evolution · RY-coding

Introduction

Phylogenetic analyses based on molecular data can be misled by a variety of pitfalls such as model misspecification (Posada and Buckley 2004), long branch attraction (Felsenstein 1978), or heterogeneous base composition (Lockhart et al. 1994; Mooers and Holmes 2000; Jermiin et al. 2004) to name a few. Heterogeneous base composition may suggest relatedness of lineages which share similar nucleotide frequencies by chance and not by common descent. Compositional heterogeneity has been reported on different levels of phylogenetic divergence and may not only affect nucleotides but also amino acids (Foster and Hickey 1999; Singh et al. 2009; Nesnidal et al. 2010). There is controversy over the severity of the effects of divergent nucleotide or amino acid frequencies on the accuracy of phylogenetic reconstruction (Rosenberg and Kumar 2003; Jermiin et al. 2004). Nevertheless, as non-stationary data violate the assumptions of standard reconstruction methods, a number of approaches have been developed to account for this issue including LogDet distances and specific substitution models for maximum likelihood analyses (Galtier and Gouy 1995; Boussau and Gouy 2006; Dutheil and Boussau 2008). In addition, RY-

Electronic supplementary material The online version of this article (doi:10.1007/s13127-016-0288-0) contains supplementary material, which is available to authorized users.

✉ Martin Haase
martin.haase@uni-greifswald.de

- ¹ Zoologisches Institut und Museum, Ernst-Moritz-Armdt—Universität Greifswald, Soldmannstraße 23, 17489 Greifswald, Germany
- ² Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

coding has been suggested as a remedy for heterogeneous nucleotide frequencies (Phillips and Penny 2003).

In our ongoing phylogenetic and biogeographic analyses of the land snail genus *Theba*, which naturally occurs in NW Africa, the Canary and Selvagem Islands, as well as on the Iberian Peninsula (Gittenberger and Ripken 1987; Greve et al. 2010; Däumer et al. 2012), we have encountered a number of problems and contradictory results. According to our initial analysis based on fragments of the mitochondrial cytochrome oxidase subunit I (COI) and the internal transcribed spacer 1 of the nuclear ribosomal RNA complex (ITS1), the genus evolved on the Canary Islands and back-colonized the continents. The phylogenetic signal was dominated by COI, however, the third codon positions were inhomogeneous which had to be corrected by RY-coding (Greve et al. 2010), possibly at the cost of information (Sauer and Hausdorf 2010). Subsequently, we analyzed amplified fragment length polymorphisms (AFLPs) and considerably more specimens, which turned the topology upside-down suggesting the origin of the genus in NW Africa and dispersal to the Canary and Selvagem Islands as well as the Iberian Peninsula (Haase et al. 2014). In the same paper, we conducted an analysis based solely on COI and the same set of specimens. Again, homogeneity of the third codon positions had to be established by RY-coding and the resulting topology was similar to the AFLP topology, however, with a different continental clade as a sister group to the remaining clades. In contrast to the AFLP tree, the basal nodes were extremely poorly supported.

In general, mito-nuclear discordance is commonly encountered in phylogenetic analyses and mostly attributed to incomplete lineage sorting, introgression, or unresolved taxonomy (e.g., Avise 1994; Funk and Omland 2003). Alternatively, factors including selection or sex-related asymmetries such as female-biased dispersal are considered (Toews and Brelsford 2012). However, systematic biases in sequence evolution are rarely questioned in this context.

In the present paper, we asked whether the topological ambiguities were due to (1) lack of resolution of COI and/or (2) the heterogeneity of base composition. In order to potentially increase mitochondrial information and resolution, we sequenced a fragment of 16S rRNA. In many phylogenetic analyses on comparable taxonomic levels, 16S rRNA has proved to evolve more conservatively than COI did and thus to provide more information on deeper levels (e.g., Fiorentino et al. 2010; Zielske et al. 2011; Johnson et al. 2012; Palsson et al. 2014). To control for the effects of inhomogeneous base frequencies, we conducted LogDet-distance, maximum parsimony (MP), and maximum likelihood (ML) analyses as well as Bayesian inference (BI), the latter three based on both the original data as well as on RY-coded data. The Bayesian approaches also included analyses allowing for heterogeneous evolutionary rates among lineages (Drummond et al. 2006). With the exception of LogDet, we conducted our analyses

based on optimally partitioned data in order to retrieve the maximum information (Phillips and Penny 2003; Lanfear et al. 2012). ML analyses implementing models that take compositional heterogeneity into account were not feasible because of the size of the data set and/or their restriction to unpartitioned alignments (Galtier and Gouy 1998; Boussau and Gouy 2006; Dutheil and Boussau 2008). In a second approach, we tested whether conventional Bayesian analyses would reconstruct the original topology of *Theba* from inhomogeneous data simulated based on this original topology.

Material and methods

Material and DNA sequencing

Our analyses included 172 of the 182 specimens of *Theba* analyzed by Haase et al. (2014) (Table 1). We used existing COI sequences and newly sequenced a fragment of 16S rRNA (see below) from the stored DNA extracts, which did not work for ten individuals. The outgroup comprised *Cochlicella acuta* (O. F. Müller 1774; Geomitridae), *Cornu aspersum* (O. F. Müller 1774; Helicidae), *Drusia deshayesii* (Moquin-Tandon 1848; Parmacellidae; formerly in *Parmacella*, see Martínez-Ortí & Borredà 2013), *Obelus despreauxii* (D'Orbigny 1839; Geomitridae), and *Trochoidea pyramidata* (Draparnaud 1805; Geomitridae). For the latest suprageneric classification adopted here, see Razkin et al. (2015). The 16S rRNA fragment was amplified using the primers 16Scs1 and 16Sma2 developed by Chiba (1999). Polymerase chain reactions (PCRs) were performed in a total volume of 11 µl containing 1 µl 10× BH4 reaction buffer (BIOLINE GmbH, Luckenwalde, Germany), 4.4 mM of MgCl, 0.3 pM of each primer, 0.2 mM of dNTP, 0.4 µl of BSA (1 %), 0.2 U of DNA-polymerase (BIOLINE), 50 ng DNA, and dd water. The PCR profile comprised an initial denaturation at 95 °C for 3 min, 35 cycles including denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, and elongation at 72 °C for 1 min, and a final extension at 72 °C for 7 min. PCR products were cleaned using Exonuclease I (New England Biolabs GmbH, Frankfurt/Main, Germany) and Shrimp-Alkaline-Phosphatase (Promega, Madison, WI, USA). Cycle sequencing was performed using the Big Dye Terminator Ready Reaction Mix v3.1 (Applied Biosystems, Carlsbad, CA, USA) and the PCR primers. After cleaning with CleanSEQ (Beckman Coulter, Beverly, MA, USA), sequences were read in both directions on an ABI 3130xl Genetic Analyzer.

Sequence editing and alignment

Sequences were edited in DNA Baser Sequence Assembler 4.16 (Heracle BioSoft SRL) and initially aligned together with a structure annotated sequence of *Albinaria turrita*

Table 1 Material sequenced

Label	Species	N	C	Country: Region	Coordinates	COI GenBank	16S rRNA GenBank
AG.1.1.	<i>Cornu aspersum</i>	1	–	France: Cap Ferret	N 44° 38.783' W 01° 15.048'	HM034454	KU521591
CA.1.	<i>Cochlicella acuta</i>	1	–	Spain: Andalusia	N 36° 43.500' W 04° 24.947'	KU521584	KU521588
DD.1.	<i>Drusia deshayesii</i>	1	–	Morocco: Nador	N 35° 11.162' W 03° 09.324'	KU521583	KU521587
OD.1.	<i>Obelus despreauxii</i>	1	–	Spain: GC	N 27° 52.167' W 15° 23.832'	KU521585	KU521589
TP.1.	<i>Trochoidea pyramidata</i>	1	–	Italy: Siena	N 43° 18.980' E 11° 19.559'	KU521586	KU521590
EA.1.	<i>T. p. pisana</i>	2	4	Spain: Andalusia	N 36° 16.712' W 06° 05.732'	HM034570 HM034571	KU521616 KU521617
ESA.09.1.	<i>T. p. arietina</i>	3	4	Spain: Andalusia	N 36° 38.110' W 06° 10.847'	KF582641-KF582643	KU521606 KU521607 KU521618
ESA.09.10.	<i>T. s. helicella</i>	1	2	Spain: Andalusia	N 36° 50.479' W 02° 22.177'	KF582662	KU521736
ESA.09.12.	<i>T. guttenbergeri</i>	4	4	Spain: Andalusia	N 36° 43.095' W 02° 38.163'	KF582634-KF582637	KU521609-KU521612
ESA.09.2.	<i>T. andalusica</i>	4	4	Spain: Andalusia	N 36° 36.621' W 06° 16.876'	KF582617-KF582620	KU521593 KU521594 KU521614 KU521619
ESA.09.3.	<i>T. p. arietina</i>	7	4	Spain: Andalusia	N 36° 36.786' W 06° 17.317'	KF582644-KF582650	KU521613 KU521615 KU521628- KU521631 KU521644
ESA.09.4.	<i>T. andalusica</i>	7	4	Spain: Andalusia	N 36° 00.716' W 05° 36.266'	KF582621-KF582627	KU521604 KU521622 KU521624- KU521626 KU521645 KU521649
ESA.09.5.	<i>T. p. pisana</i>	1	4	Spain: Andalusia	N 36° 03.650' W 05° 43.235'	KF582656	KU521651
ESA.09.6.	<i>T. andalusica</i>	3	4	Spain: Andalusia	N 36° 01.852' W 05° 36.790'	KF582628 KF582630 KF582631	KU521605 KU521623 KU521627
ESA.09.8.	<i>T. p. pisana</i>	1	4	Spain: Andalusia	N 36° 17.714' W 05° 16.351'	KF582663	KU521650
ESA.1.	<i>T. andalusica</i>	1	4	Spain: Andalusia	N 36° 45.674' W 03° 39.860'	KF597306	KU521643
ESM.1.	<i>T. p. pisana</i>	1	4	Spain: Andalusia	N 36° 43.368' W 04° 24.687'	KF582657	KU521646
FU.1.	<i>T. sp. 1b</i>	2	3	Spain: FU	N 28° 24.772' W 13° 55.700'	JN408081 JN408082	KU521686 KU521689
FU.10.	<i>T. sp. 5 "Sand"</i>	2	3	Spain: FU	N 28° 10.541' W 14° 12.442'	JN408104 JN408105	KU521656 KU521657
FU.11.	<i>T. sp. 5 "Sand"</i>	2	3	Spain: FU	N 28° 08.909' W 14° 15.119'	JN408106 JN408107	KU521653 KU521658
FU.12.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 06.127' W 14° 17.080'	JN408108	KU521659
FU.13.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 04.256' W 14° 25.286'	JN408128	KU521663
FU.15.	<i>T. sp. 5 "Sand"</i>	2	3	Spain: FU	N 28° 04.918' W 14° 28.311'	JN408110 JN408111	KU521654 KU521655
FU.17.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 06.498' W 14° 23.277'	JN408130	KU521659
FU.18.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 04.580' W 14° 25.699'	JN408132	KU521664
FU.19.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 03.983' W 14° 19.824'	JN408134	KU521667
FU.2.	<i>T. sp. 1b</i>	2	3	Spain: FU	N 28° 25.250' W 14° 00.684'	JN408084 JN408085	KU521683 KU521687
FU.20.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 04.547' W 14° 20.094'	JN408112	KU521666
FU.21.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 04.962' W 14° 20.279'	JN408114	KU521668
FU.22.	<i>T. cf. clausoinflata "Rock"</i>	1	3	Spain: FU	N 28° 05.278' W 14° 20.482'	JN408117	KU521665
FU.23.	<i>T. cf. clausoinflata "Top"</i>	1	3	Spain: FU	N 28° 05.809' W 14° 20.999'	JN408119	KU521659
FU.25.	<i>T. cf. clausoinflata "Top"</i>	1	3	Spain: FU	N 28° 06.059' W 14° 21.380'	JN408121	KU521662
FU.26.	<i>T. sp. 1a</i>	1	3	Spain: FU	N 28° 44.549' W 13° 49.519'	JN408123	KU521674
FU.27.	<i>T. sp. 1a</i>	1	3	Spain: FU	N 28° 40.596' W 13° 55.260'	JN408124	KU521670
FU.4.	<i>T. sp. 1a</i>	1	3	Spain: FU	N 28° 42.292' W 13° 52.870'	JN408088	KU521672
FU.5.	<i>T. sp. 1a</i>	2	3	Spain: FU	N 28° 36.800' W 13° 55.547'	JN408091 JN408092	KU521673 KU521675
FU.6.	<i>T. sp. 1b</i>	3	3	Spain: FU	N 28° 31.478' W 14° 00.068'	JN408093-JN408095	KU521684 KU521688 KU521690
FU.7.	<i>T. sp. 1b</i>	2	3	Spain: FU	N 28° 19.761' W 14° 03.207'	JN408096 JN408097	KU521682 KU521685
FU.8.	<i>T. sp. 4</i>	3	3	Spain: FU	N 28° 15.306' W 14° 11.162'	JN408099-JN408101	KU521706-KU521708
FU.9.	<i>T. sp. 4</i>	2	3	Spain: FU	N 28° 12.792' W 14° 13.079'	JN408102 JN408103	KU521704 KU521705
GC.10.	<i>T. p. pisana</i>	1	4	Spain: GC	N 27° 51.291' W 15° 25.432'	KF582654	KU521634
GC.12.	<i>T. p. pisana</i>	1	4	Spain: GC	N 28° 05.857' W 15° 28.671'	HM034574	KU521635
GC.13.	<i>T. arinagae</i>	1	3	Spain: GC	N 27° 52.167' W 15° 23.832'	JN408143	KU521669
GC.3.	<i>T. grasseti</i>	2	3	Spain: GC	N 28° 02.424' W 15° 45.864'	HM034542 JN408139	KU521677 KU521678
GC.4.	<i>T. grasseti</i>	3	3	Spain: GC	N 28° 09.552' W 15° 25.992'	HM034543 JN408142 KF582638	KU521679-KU521681
GC.5.	<i>T. p. pisana</i>	1	4	Spain: GC	N 28° 02.058' W 15° 26.067'	KF582653	KU521633
GC.7.	<i>T. p. pisana</i>	1	4	Spain: GC	N 27° 51.291' W 15° 25.432'	KF582652	KU521652
GC.8./GC.11.	<i>T. geminata</i>	2	3	Spain: GC	N 28° 01.992' W 15° 45.432'	JN408136 KF582633	KU521693 KU521694
IS.1.	<i>T. p. pisana</i>	1	4	Israel: Netzer Sereni	N 31° 55.467' E 34° 49.702'	HM034587	KU521640
LZ.10.	<i>T. impugnatata</i>	2	1	Spain: LZ	N 29° 11.690' W 13° 27.628'	HM114304 HM114305	KU521762 KU521763
LZ.12. ^a	<i>T. geminata</i>	2	3	Spain: LZ	N 29° 02.981' W 13° 33.793'	JN408159 JN408160	KU521695 KU521696
LZ.12. ^a	<i>T. sp. 2</i>	1	1	Spain: LZ	N 29° 02.981' W 13° 33.793'	JN408161	KU521750
LZ.13.	<i>T. sp. 2</i>	1	3	Spain: LZ	N 29° 11.794' W 13° 29.541'	JN408162	KU521701
LZ.14.	<i>T. sp. 2</i>	1	1	Spain: LZ	N 29° 05.156' W 13° 34.230'	JN408164	KU521749
LZ.15.	<i>T. sp. 2</i>	1	1	Spain: LZ	N 29° 06.249' W 13° 37.880'	JN408168	KU521748
LZ.16.	<i>T. sp. 2</i>	1	1	Spain: LZ	N 29° 06.717' W 13° 39.378'	JN408170	KU521755
LZ.17.	<i>T. sp. 2</i>	2	1	Spain: LZ	N 28° 56.692' W 13° 42.149'	JN408172 JN408173	KU521751 KU521753
LZ.18.	<i>T. sp. 2</i>	2	1	Spain: LZ	N 28° 55.113' W 13° 46.506'	JN408174 JN408175	KU521752 KU521754
LZ.20.	<i>T. sp. 2</i>	1	1	Spain: LZ	N 29° 03.803' W 13° 40.787'	JN408177	KU521756
LZ.21. ^a	<i>T. geminata</i>	2	3	Spain: LZ	N 28° 59.431' W 13° 37.352'	JN408180 JN408182	KU521691 KU521692
LZ.21. ^a	<i>T. sp. 2</i>	1	1	Spain: LZ	N 28° 59.431' W 13° 37.352'	JN408181	KU521747
LZ.23. ^a	<i>T. sp. 1a</i>	2	3	Spain: LZ	N 28° 52.667' W 13° 49.961'	JN408183 JN408184	KU521671 KU521676

Table 1 (continued)

Label	Species	N	C	Country: Region	Coordinates	COI GenBank	16S rRNA GenBank
LZ.23. ^a	<i>T. sp. 2</i>	1	1	Spain: LZ	N 28° 52.667' W 13° 49.961'	JN408185	KU521746
LZ.3.	<i>T. geminata</i>	1	3	Spain: LZ	N 29° 08.481' W 13° 28.922'	JN408144	KU521702
LZ.4.	<i>T. geminata</i>	1	3	Spain: LZ	N 29° 07.672' W 13° 30.847'	JN408146	KU521698
LZ.5. ^a	<i>T. geminata</i>	1	3	Spain: LZ	N 29° 07.893' W 13° 30.912'	JN408148	KU521699
LZ.5. ^a	<i>T. impugnata</i>	5	1	Spain: LZ	N 29° 07.893' W 13° 30.912'	HM114299-HM114303	KU521757-KU521761
LZ.6.	<i>T. geminata</i>	1	3	Spain: LZ	N 29° 09.082' W 13° 30.022'	JN408152	KU521700
LZ.7.	<i>T. geminata</i>	1	3	Spain: LZ	N 29° 11.099' W 13° 30.064'	JN408154	KU521703
LZ.8.	<i>T. geminata</i>	1	3	Spain: LZ	N 29° 12.799' W 13° 28.869'	JN408156	KU521697
M.1.	<i>T. p. pisana</i>	3	4	Morocco: Bouknadel	N 34° 07.600' W 06° 45.039'	HM034560 HM034562 KF582658	KU521595-KU521597
M.13.	<i>T. p. pisana</i>	1	4	Morocco: Rabat	N 34° 00.414' W 06° 49.271'	HM034557	KU521598
M.16.	<i>T. subdentata</i>	1	2	Morocco: Essaouira	N 31° 27.313' W 09° 40'49.9	HM034501	KU521738
M.19.	<i>T. subdentata</i>	1	2	Morocco: Sidi Mogdoul	N 31° 29.300' W 09° 46.000'	KF582659	KU521739
M.2.	<i>T. p. pisana</i>	1	4	Morocco: Rabat	N 33° 59.593' W 06° 50.987'	HM034556	KU521599
M.20.	<i>T. p. ampullacea</i>	1	4	Morocco: Sidi Mogdoul	N 31° 29.300' W 09° 46.000'	HM034564	KU521592
M.24.	<i>T. subdentata</i>	3	2	Morocco: Tiznit	N 29° 48.071' W 09° 38.078'	HM034526 HM034528 KC526931	KU521728-KU521730
M.25.	<i>T. solimae</i>	2	2	Morocco: Bou Tzalam	N 29° 26.766' W 09° 41.688'	HM034494 KC526932	KU521741 KU521742
M.27.	<i>T. solimae</i>	1	2	Morocco: Oued Draa	N 28° 30.983' W 10° 57.443'	HM034495	KU521740
M.28.	<i>T. sp. 3</i>	1	2	Morocco: Lagune de Khnifiss	N 27° 54.566' W 12° 23.311'	HM034473	KU521725
M.29.	<i>T. chudeaui</i>	1	2	Western Sahara: Laayoune	N 26° 51.394' W 13° 28.168'	HM034485	KU521711
M.30.	<i>T. chudeaui</i>	2	2	Western Sahara: Laayoune	N 27° 06.723' W 13° 23.878'	HM034486 HM034487	KU521709 KU521710
M.31.	<i>T. chudeaui</i>	1	2	Western Sahara: Laayoune	N 26° 41.451' W 13° 33.116'	HM034488	KU521712
M.32.	<i>T. chudeaui</i>	2	2	Western Sahara: Laayoune	N 26° 39.180' W 13° 38.775'	HM034489 HM034490	KU521713 KU521714
M.34.	<i>T. sp. 3</i>	1	2	Western Sahara: Tarfaya	N 27° 18.702' W 13° 04.194'	HM034475	KU521724
M.35./M.37.	<i>T. sp. 3</i>	2	2	Western Sahara: Tarfaya	N 27° 16.263' W 13° 07.498'	HM034481 KC526947	KU521717 KU521718
M.36.	<i>T. sp. 3</i>	1	2	Morocco: Tan Tan	N 27° 55.271' W 12° 18.169'	HM034480	KU521726
M.39.	<i>T. sacchii</i>	5	2	Morocco: Tan Tan	N 28° 08.679' W 11° 15.814'	HM034472 KC526937- KC526940	KU521719- KU521723
M.41.	<i>T. p. ampullacea</i>	3	4	Morocco: Safi	N 31° 56.583' W 09° 25.876'	HM034595 HM034566 KF582640	KU521602 KU521603 KU521608
M.42.	<i>T. p. cantinensis</i>	2	4	Morocco: Cap Meddouza	N 32° 32.411' W 09° 17.016'	HM034567 HM034568	KU521600 KU521601
M.44.	<i>T. subdentata</i>	1	2	Morocco: Tamri	N 30° 42.492' W 09° 50.277'	HM034513	KU521732
M.45.	<i>T. subdentata</i>	1	2	Morocco: Tamanar	N 31° 01.572' W 09° 40.239'	KF582660	KU521737
M.46.	<i>T. subdentata</i>	1	2	Morocco: Oued Souss	N 30° 26.071' W 08° 53.734'	KF582661	KU521733
M.47.	<i>T. subdentata</i>	1	2	Morocco: Tamri	N 30° 42.492' W 09° 50.277'	HM034519	KU521731
M.48.	<i>T. sp. 3</i>	1	2	Morocco: Dchira	N 27° 01.717' W 13° 03.106'	HM034483	KU521715
M.49.	<i>T. p. pisana</i>	1	4	Morocco: Rif	N 35° 08.619' W 02° 25.472'	KF582655	KU521620
M.50.	<i>T. p. pisana</i>	3	4	Morocco: Nador	N 35° 06.247' W 02° 44.757'	HM034601-HM034603	KU521621 KU521632 KU521638
M.51.	<i>T. subdentata</i>	1	2	Morocco: Oued Souss	N 30° 25.104' W 09° 02.716'	HM034521	KU521734
M.52.	<i>T. subdentata</i>	1	2	Morocco: Taroudant	N 30° 26.883' W 08° 53.255'	HM034523	KU521735
M.55.	<i>T. subdentata</i>	1	2	Morocco: Oued Massa	N 30° 03.544' W 09° 39.123'	HM034529	KU521727
M.56.	<i>T. sp. 3</i>	1	2	Western Sahara: Tarfaya	N 27° 16.263' W 13° 07.498'	HM034484	KU521716
MAL.1.	<i>T. p. pisana</i>	1	4	Malta: San Gwann Valley	N 35° 54.620' W 14° 28.778'	KF582651	KU521647
ND.1.	<i>T. p. pisana</i>	1	4	Netherlands: Domburg	N 51° 33.788' W 03° 29.356'	HM034576	KU521639
SA.1.	<i>T. p. pisana</i>	1	4	South Africa: Western Cape	S 34° 10.823' E 22° 08.064'	HM034599	KU521642
SEL.1.	<i>T. macandrewiana</i>	3	1	Portugal: Selvagem Grande	N30°08.567' W15° 51.917'	JN408140 JN408141 KF582639	KU521743-KU521745
T.1.	<i>T. p. pisana</i>	1	4	Tunesia: Djerba	N 33° 50.508' E 11° 00.039'	HM034579	KU521648
ZYP.1.	<i>T. p. pisana</i>	2	4	Cyprus: Limassol	N 34° 42.332' E 33° 00.969'	HM034583 HM034584	KU521636 KU521637
ZYP.2.	<i>T. p. pisana</i>	1	4	Cyprus: Limassol	N 34° 41.589' E 33° 05.060'	HM034585	KU521641

First five taxa belong to the outgroup

FU Fuerte Ventura, GC Gran Canaria, LZ Lanzarote, C Clade, N number of specimens

^a samples containing more than one species

using CLUSTAL W (Thompson et al. 1994). The sequence of *A. turrita* was originally retrieved from the European Ribosomal Database (de Rijk et al. 2000; van de Peer et al. 2000), which is no longer maintained. The secondary structure of *A. turrita* served as seed for a structure-informed alignment made in RNAsalsa 0.8.1 (Stocsits et al. 2009). This was then trimmed to 856 base pairs (bp) in BioEdit 7.2.5 (Hall 1999)

and concatenated with the 630 bp alignment of the COI fragment (Haase et al. 2014). Aliscore 2.0 (Misof and Misof 2009; Kück et al. 2010) did not detect random similarity; therefore, no masking was necessary. We then defined five partitions: stems and loops of 16S rRNA and the three codon positions of COI (see below). These were separately tested for homogeneity of base frequencies excluding constant sites as proxies for

invariant sites (Lockhart et al. 1996) using the χ^2 test implemented in PAUP* 4b10 (Swofford 2003). Loops and third codon positions turned out to have heterogeneous base composition ($\chi^2 = 864.44$, $df = 528$, $P < 0.001$; $\chi^2 = 1707.56$, $df = 528$, $P < 0.001$). Saturation of substitutions was tested for each partition in DAMBE 5.3.105 (Xia 2013) based only on fully resolved sites as recommended by the program. Saturation may have been problematic only for the third codon positions of COI and then only if the underlying tree was considered unsymmetrical. However, DAMBE simulates saturation indices for Xia et al.'s (2003) test only for up to 32 taxa. Therefore, for considerably larger datasets such as ours, interpretation remains somewhat ambiguous in general. As Aliscore did not detect noisy positions, we considered that lack of phylogenetic signal was not an issue in our data.

Phylogenetic analyses of empirical data

We conducted analyses (1) ignoring heterogeneity of base composition, (2) accounting for heterogeneity of base composition, and (3) accounting for heterogeneity of substitution rates. The first group of analyses comprised MP, ML, and BI. MP was conducted in PAUP* 4b10 with 500 replicates, stepwise addition, and random starting trees. We applied TBR branch swapping and restricted each replicate to 1 million rearrangements. Robustness was assessed by 1000 bootstrap replicates. For ML, we used Garli 2.0 (Zwickl 2006) running 500 replicates for both finding the optimal trees and bootstrapping. BI was performed in MrBayes 3.2.2 (Ronquist et al. 2012) over 8 million generations, saving every 100th tree with a burnin of 25 %. To account for heterogeneity of base frequencies, we constructed a BioNeighbor-joining tree based on LogDet distances (Lockhart et al. 1994) in PAUP* 4b10, removing invariant sites in proportion to frequencies estimated from constant sites. The proportion of invariant sites was estimated in jModeltest v2.1.4 (Darriba et al. 2012). In addition, we recoded the heterogeneous partitions (loops and third codon positions) using R for purines and Y for pyrimidines (RY-coding) and repeated MP, ML, and BI analyses. While the RY-coded loops indeed became homogeneous, the third codon positions remained heterogeneous. Finally, we conducted Bayesian tree reconstructions also in BEAST 1.8.0 (Drummond et al. 2012), implementing the log-normal uncorrelated relaxed molecular clock and a birth-death model as tree prior. We jointly summarized four independent analyses with each 20 million generations, every 1000th tree sampled, and a burnin of 10 %. BI was repeated with the RY-coded alignment, as well. Convergence of parameter estimates in both types of Bayesian analyses were controlled by ensuring that effective sample sizes were larger than 200 as indicated in

Tracer 1.6 (Rambaut et al. 2014) and based on the criteria implemented in the respective programs. Prior to the analyses based on substitution models, Partition Finder 1.1.0 (Lanfear et al. 2012), comparing all possible combinations of up to five partitions, confirmed the above partitioning scheme as optimal and selected appropriate models based on the Bayesian information criterion (Table 2).

Phylogenetic analyses of simulated data

In order to test whether inhomogeneous base composition may have influenced the topology of the mitochondrial tree of *Theba*, we simulated 100 alignments with five partitions of the original length based on the original

Table 2 Best fitting substitution models implemented in maximum likelihood (ML) and two types of Bayesian analyses (MrBayes, BEAST)

Analysis	Partition	Substitution model
ML	16S rRNA stems	TIM + I + Γ
	16S rRNA loops	HKY + Γ
	COI 1st codon position	TrN + I + Γ
	COI 2nd codon position	K81 + Γ
	COI 3rd codon position	GTR + Γ
MrBayes	16S rRNA stems	Doublet model, Nst = 2 + Γ
	16S rRNA loops	Nst = 2 + I + Γ
	COI 1st codon position	Nst = 6 + I + Γ
	COI 2nd codon position	Nst = 1 + Γ
MrBayes, RY-coded	COI 3rd codon position	Nst = 6 + Γ
	16S rRNA loops	Binary model, + Γ
	COI 3rd codon position	Binary model, + Γ
BEAST	16S rRNA stems	TN93 + I + Γ
	16S rRNA loops	HKY + Γ
	COI 1st codon position	TN93 + I + Γ
	COI 2nd codon position	GTR + Γ
	COI 3rd codon position	GTR + Γ
ML, backbone simulations	Partition 1	HKY + Γ
	Partition 2	HKY + Γ
	Partition 3	TIM + Γ
	Partition 4	JC
	Partition 5	K81uf + Γ
MrBayes, simulations	Partition 1	Nst = 2 + Γ
	Partition 2	Nst = 2 + Γ
	Partition 3	Nst = 6 + Γ
	Partition 4	Nst = 1
	Partition 5	Nst = 6 + Γ

In contrast to the programs Garli (ML) and BEAST, MrBayes explicitly allows to model binary data such as the RY-coded partitions. Note that these models differ according to the programs used

topology and reconstructed the trees using MrBayes. We did that in Indelible 1.03 (Fletcher and Yang 2009) based on a reduced taxon set comprising five individuals per clade and the outgroup *Cornu aspersum* in order to save computation time. Indelible allows the simulation of sequences under non-stationary conditions. The backbone tree was constructed in an ML framework with Garli after model fitting with jModeltest (Table 1). As the base of the tree was unresolved, we introduced a branch with length of 0.15 separating outgroup from ingroup. The remaining topology was fully resolved. For the partitions corresponding to those with heterogeneous base frequencies in the original data, partitions 2 (loops) and 5 (third codon positions), we fitted separate substitution models to the four main clades and all older branches based on the results of jModeltest for the original data (see configuration file in Appendix 1). The models used for the five partitions in reconstructions with MrBayes are again listed in Table 2. Every 100th tree of a total of 1 million generations was sampled with a burnin of 25 %. Convergence of parameter estimates was monitored as stated above.

Results

In our presentation of the results, we focus on the inter-relationships of the four main clades. Relationships within these clades are not considered. After RY-coding, only the base composition of the loops was no longer heterogeneous in contrast to the third codon positions ($\chi^2 = 134.33$, $df = 176$, $P = 0.999$; $\chi^2 = 293.15$, $df = 176$, $P < 0.001$). Figure 1 shows the LogDet tree with collapsed main clades. A Bayesian

analysis with unmanipulated clades is given in Supplement 1. Clade 1 consisted of snails from the Selvagem Islands and Lanzarote, clade 2 was composed of sequences exclusively from the Canary Islands, clade 3 contained mainly samples from NW Africa, and clade 4 snails from NW Africa as well as Europe. The reconstructions based on original data and RY-coded data, respectively, are summarized in Fig. 2. All tree reconstructions gave very similar results, with an ingroup significantly supported only by the LogDet and BEAST analyses. The four main clades were, however, largely well supported and most methods revealed clade 1 as a robust sister group to the remaining three clades. Only both ML analyses showed a polytomy instead of nodes 2 and 3. Except for the LogDet and BEAST analyses, all approaches reconstructed clades 2 and 4 as sister group; however, only MrBayes recovered this with significant support. In the BEAST analysis, node 3 was a polytomy and only the LogDet and the RY-coded BEAST analyses reconstructed clades 2 and 3 as sister taxa, however, with negligible support. Based on RY-coded sequences, MrBayes also recovered node 3 as polytomy which also included parts of clade 3. In general, the approaches supposed to mitigate the effects of heterogeneous base composition did not influence the gross topology, i.e., the relationships of the main clades. RY-coding largely resulted in weaker resolution.

The Bayesian reconstructions based on 100 simulated data sets were highly concordant. In at least 95 cases, the scaffold topology was recovered, with the exception of the root node and one node within clade 2 (Figs. 3 and 4). This suggests that the phylogenetic signal largely remained unambiguous despite introducing heterogeneity of base composition in two partitions.

Fig. 1 BioNeighbor-joining tree based on LogDet distances with bootstrap support values. Ingroup clades collapsed. For labels of outgroup taxa see Table 1. Expanded clades are shown in Supplement 1. Scale bar = substitutions per site

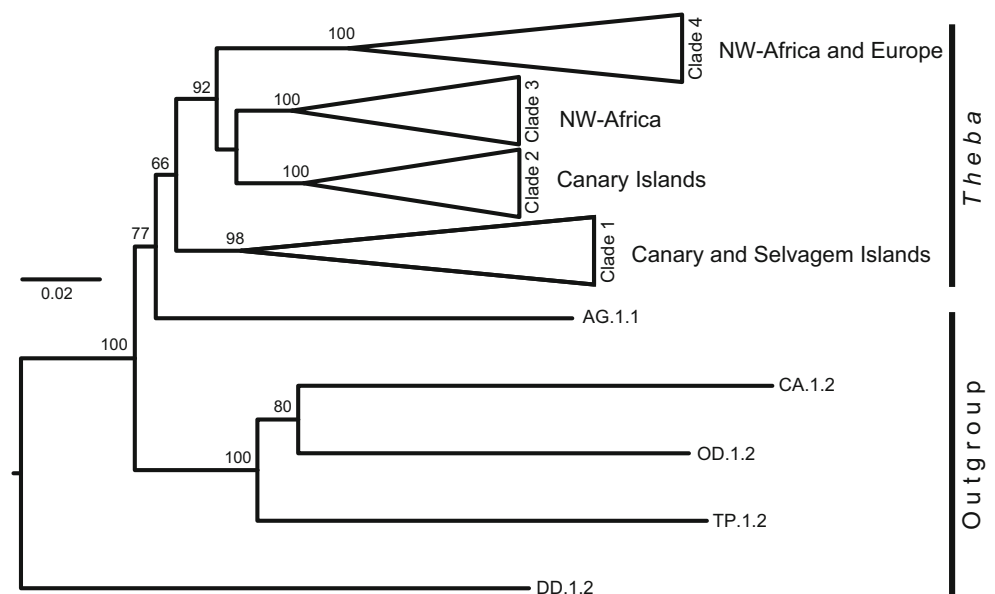
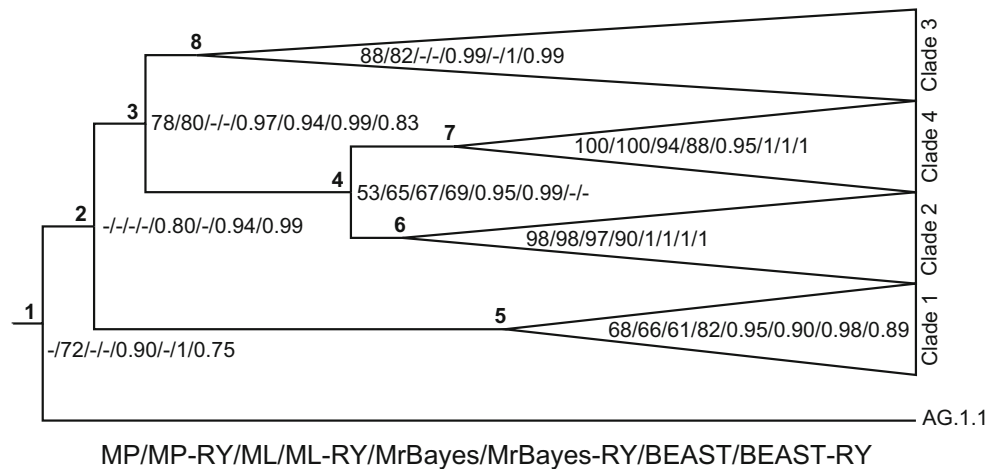


Fig. 2 Fifty percent majority rule consensus tree composed of maximum parsimony (MP), maximum likelihood (ML), as well as Bayesian trees reconstructed in MrBayes and BEAST, respectively. All analyses were conducted with original and RY-coded data. Support values from these analyses are indicated in order below tree (bootstrap values or posterior probabilities). Four outgroup taxa were pruned from tree. Nodes were numbered for reference to the text

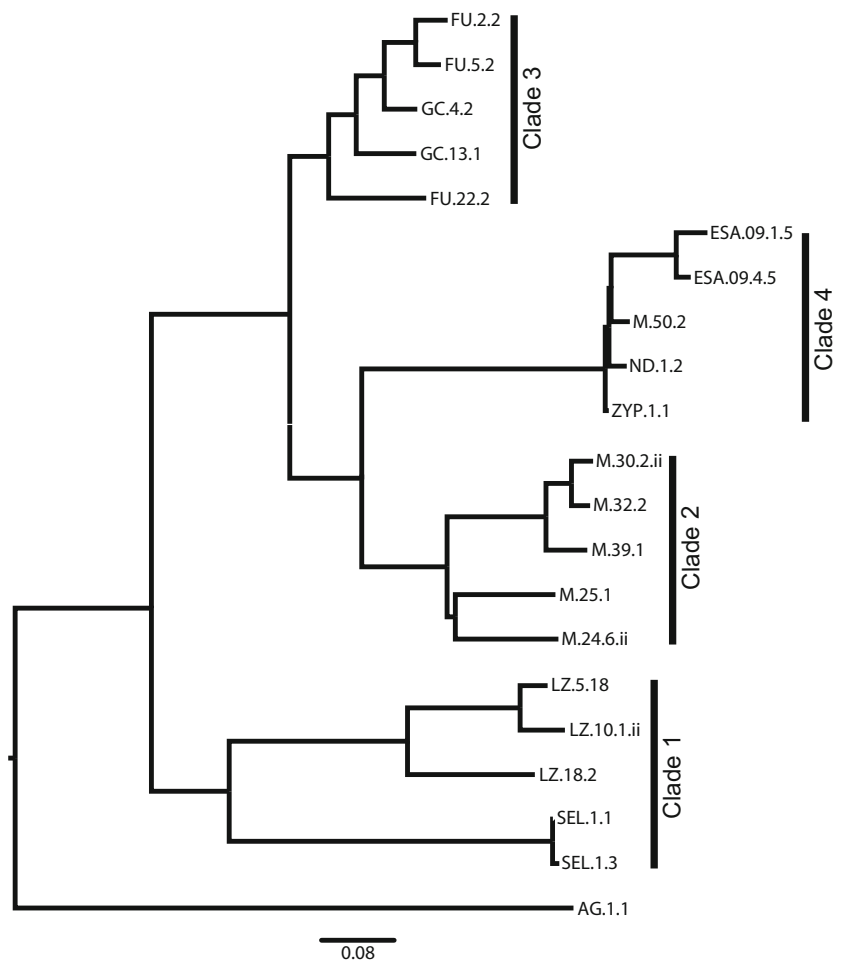


Discussion

Just like the protein coding COI, the newly sequenced 16S rRNA gene exhibited segments with homogeneous as well as heterogeneous base composition. It appears that in *Theba* sections of mitochondrial DNA underlying stronger constraints such as the first and second codon positions or

stems evolve rather conservatively with regard to base frequencies, while selectively more neutral sections such as third codon positions and loops show higher variation in substitution patterns. Thus, by generating more data, we even increased the proportion of sites with inhomogeneous base composition as the loop sections comprised 75 % of the entire 16S rRNA fragment.

Fig. 3 Model tree for sequence simulations generated in a ML framework and based on a reduced taxon set (see Table 1 for taxon labels)



However, the general picture of tree reconstruction was the same for the standard approaches as well as those taking inhomogeneous base frequencies into account. Augmenting the mitochondrial data by 16S rRNA slightly increased the support for the deeper nodes of the *Theba* phylogeny compared to our foregoing analyses (Greve et al. 2010; Haase et al. 2014). However, in accordance with Greve et al. (2010), the topology still suggested an origin of the genus on the Selvagem and Canary Islands with subsequent colonization of the continents in contrast to the AFLP data. In addition, some relationships among the main clades remained ambiguous. The poorly supported topology of the COI tree in Haase et al. (2014), with the Moroccan-Mediterranean clade corresponding to our clade 4 as sister group to the remaining clades, may have been due to over-parameterization in RAxML (Stamatakis 2006) implementing GTR as substitution model. We repeated the analysis of the foregoing paper in a different version of RAxML offering also HKY85 and K80, which have fewer parameters. While the HKY85 topology corresponded well to the one based on GTR, the topology based on the most simple model K80 was indeed very similar to the one reported by Greve et al. (2010) (data not shown). Well into the age of phylogenomics, it is now generally accepted that increasing the number of sites increases the accuracy of phylogenetic reconstructions. This has also been observed in studies investigating the effects of heterogeneous base composition (Rosenberg & Kumar 2003; Jermini et al. 2004; Betancur-R et al. 2013), suggesting that adding a second sequence reduced the ambiguity in the

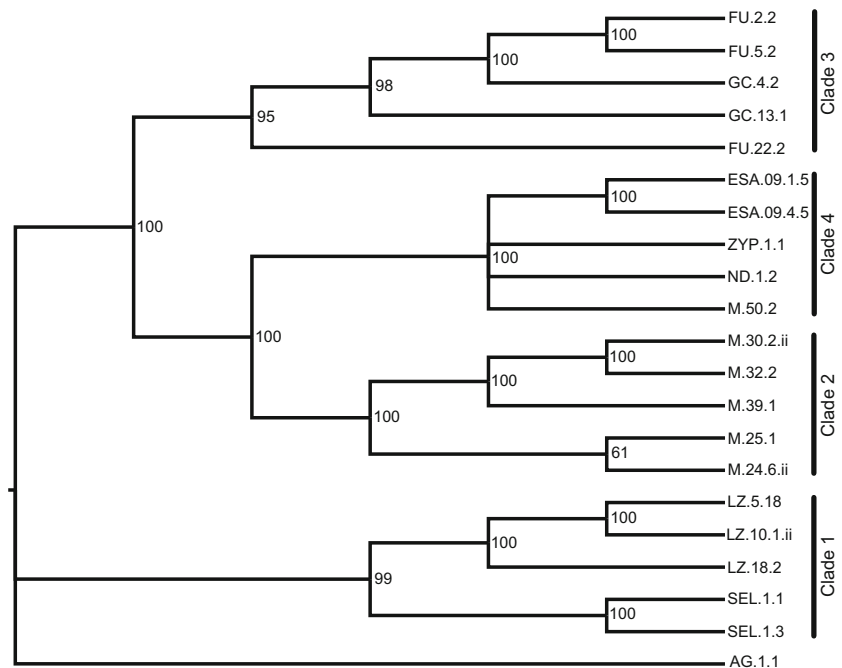
phylogenetic signal of COI and resulted in a more accurate and robust reconstruction.

Comparing trees based on RY-coding with those reconstructed from empirical data, we observed both the desired effect, i.e., improved resolution with respect to support (Phillips and Penny 2003; Ishikawa et al. 2012), as well as nodes that received less support. The latter was probably due to loss of information (Sauer and Hausdorf 2010).

Our simulations confirmed the reconstructions based on real data. The Bayesian analyses assuming stationary evolutionary processes were not misled by the introduction of inhomogeneous base composition and recovered the original tree topology that was used to simulate sequence evolution. The only exceptions were lack of support for a single node within one of the main clades and for the root node. In conclusion, the phylogenetic signal of mtDNA in the land snail genus *Theba* appeared to be robust despite considerable inhomogeneity of base composition. A Bayesian analysis of the original data excluding the inhomogeneous partitions was considerably less resolved (not shown), confirming the information content of the excluded data.

The case of *Theba* is concordant with several other phylogenetic studies which have not been affected by heterogeneous base frequencies (Rosenberg and Kumar 2003). Conditions under which compositional heterogeneity becomes a problem have only rarely been investigated. Simulations suggested that extreme changes in base frequencies are necessary to mislead phylogenetic analyses (Van Den Bussche et al. 1998; Conant & Lewis 2001) or that inhomogeneous base frequencies in combination with other confounding effects such as rate heterogeneity among lineages

Fig. 4 Fifty percent majority rule consensus tree with consensus indices based on 100 summary trees from analyses of simulated data in MrBayes. For labels of taxa see Table 1



may generate problems (Ho & Jermiin 2004). Jermiin et al. (2004) showed that short internal branches may not be recovered if base composition is not homogeneous across taxa. As the internal branches of our mitochondrial *Theba* phylogeny had considerable lengths, this may explain why compositional heterogeneity had no detrimental effects.

In general, the effects of non-stationary evolutionary processes on phylogenetic reconstruction still appear to be poorly understood and are probably highly dependent on the actual data. Finally, the incongruence in the phylogenetic signal of mitochondrial and nuclear data in *Theba* is probably real and most likely a consequence of incomplete lineage sorting (see Toews and Brelsford 2012).

Acknowledgments We thank Christel Meibauer for help in the laboratory. The suggestions of two anonymous reviewers and Serena Dool helped in improving an earlier version of the manuscript. Financial support was received from the German Science Foundation DFG (MI 649/7-1, HU 430/2-2).

Appendix 1

Configuration file for indelible in order to simulate 100 alignments with partly heterogeneous base composition.

```
[TYPE] NUCLEOTIDE 1
[SETTINGS]
[ancestralprint] NEW
[output] NEXUS
[phylipextension] phy
[nexusextension] nex
[fastaextension] fas
[randomseed] 1568746
[prinrates] FALSE
[insertaslowercase] TRUE
[markdeletedinsertions] FALSE
[printcodonsasaminoacids] FALSE
[fileperrep] FALSE
[MODEL] mStems
[submodel] HKY 3.1210
[statefreq] 0.3352 0.1610 0.2997 0.2042
[rates] 0 0.2290 4
[MODEL] mCOI_Pos1
[submodel] TIMef 10.9424 0.0003 0.5327
[rates] 0 0.2310 4
[MODEL] mCOI_Pos2
[submodel] F81
[statefreq] 0.4361 0.2203 0.1280 0.2156
[rates] 0 0 0
[MODEL] mCOI_Pos3_Clade1
[submodel] TIM 1.0000 0.0041 0.2911
[statefreq] 0.5030 0.0642 0.3577 0.0752
[rates] 0 0 0
```

```
[MODEL] mCOI_Pos3_Clade2
[submodel] HKY 14.5611
[statefreq] 0.5677 0.0295 0.3292 0.0736
[rates] 0 0 0
[MODEL] mCOI_Pos3_Clade3
[submodel] TIM 1.0000 0.0003 0.1100
[statefreq] 0.5575 0.0212 0.3594 0.0620
[rates] 0 1.1720 4
[MODEL] mCOI_Pos3_Clade4
[submodel] HKY 9.0993
[statefreq] 0.5153 0.1286 0.2430 0.1132
[rates] 0 1.1020 4
[MODEL] mLoops_Clade1
[submodel] HKY 1.6952
[statefreq] 0.3814 0.0900 0.4100 0.1187
[rates] 0 0.2870 4
[MODEL] mLoops_Clade2
[submodel] HKY 1.3060
[statefreq] 0.4166 0.0745 0.3785 0.1305
[rates] 0 0.2310 4
[MODEL] mLoops_Clade3
[submodel] HKY 3.5360
[statefreq] 0.3965 0.0752 0.3930 0.1353
[rates] 0 0.2040 4
[MODEL] mLoops_Clade4
[submodel] HKY 2.2864
[statefreq] 0.3709 0.1251 0.3609 0.1431
[rates] 0 0.1940 4
[MODEL] mLoops_Clade2u4
[submodel] HKY 1.6551
[statefreq] 0.4090 0.0958 0.3696 0.1256
[rates] 0 0.2800 4
[MODEL] mLoops_Clade2u3u4
[submodel] HKY 1.8387
[statefreq] 0.4021 0.0874 0.3848 0.1257
[rates] 0 0.3160 4
[MODEL] mLoops_Clade1u2u3u4ohneAG
[submodel] HKY 1.6923
[statefreq] 0.4055 0.0805 0.3984 0.1157
[rates] 0 0.3260 4
[MODEL] mCOI_Pos3_Clade2u4
[submodel] HKY 9.7875
[statefreq] 0.5431 0.0649 0.3018 0.0903
[rates] 0 1.4340 4
[MODEL] mCOI_Pos3_Clade2u3u4
[submodel] TIM 1.0000 0.0128 0.0574
[statefreq] 0.5546 0.0489 0.3170 0.0795
[rates] 0 1.0980 4
[MODEL] mCOI_Pos3_Clade1u2u3u4ohneAG
[submodel] TIM 1.0000 0.0093 0.0821
[statefreq] 0.5339 0.0521 0.3306 0.0835
[rates] 0 1.0890 4
[MODEL] mLoops_root_alle
```

```
[submodel] HKY 1.7047
[statefreq] 0.4024 0.0846 0.3995 0.1136
[rates] 0 0.3650 4
[MODEL] mCOI_Pos3_root_alle
[submodel] TIM 1.0000 0.0091 0.0881
[statefreq] 0.5292 0.0538 0.3341 0.0828
[rates] 0 1.0190 4
[TREE] Tree1
(AG.1.1.:0.614593,(((SEL.1.1.:0.00033546,SEL.1.3.:
0.00742396):0.355958,(LZ.18.2.:0.140155,(LZ.5.18.:
0.0302689,LZ.10.1.ii.:0.0493119):0.123939):
0.196293):0.0856161,(((ZYP.1.1.:0.00395935,
(ND.1.2.:0.0195846,(M.50.2.:0.0206259,(ESA.09.1.5.:0.03
4363,ESA.09.4.5.:0.016):0.071676):0.00235451):
0.0041263):0.268011,((M.39.1.:0.0454841,(M.30.2.ii.:
0.0238217,M.32.2.:0.0205558):0.0280523):0.108879,
(M.25.1.:0.110189,M.24.6.ii.:0.113481):0.00909306):
0.0939243):0.0792071,(FU.22.2.:0.107376,(GC.13.1.:
0.0661584,(GC.4.2.:0.0363594,(FU.2.2.:0.0351765,
FU.5.2.:0.0277639):0.0346273):0.0308667):0.030331):
0.0428032):0.152313):0.15);
[BRANCHES] Branch_Loops (AG.1.1. #mLoops_
Clade1,(((SEL.1.1.
#mLoops_Clade1,SEL.1.3. #mLoops_Clade1) #mLoops_
Clade1,(LZ.18.2.
#mLoops_Clade1,(LZ.5.18. #mLoops_Clade1,LZ.10.1.ii.
#mLoops_Clade1)
#mLoops_Clade1) #mLoops_Clade1) #mLoops_Clade1,
(((ZYP.1.1.
#mLoops_Clade4,(ND.1.2. #mLoops_Clade4,(M.50.2.
#mLoops_Clade4,(ESA.09.1.5. #mLoops_Clade4,
ESA.09.4.5. #mLoops_Clade4)
#mLoops_Clade4) #mLoops_Clade4) #mLoops_Clade4)
#mLoops_Clade4,((M.39.1.
#mLoops_Clade2,(M.30.2.ii. #mLoops_Clade2,M.32.2.
#mLoops_Clade2)
#mLoops_Clade2) #mLoops_Clade2,(M.25.1.
#mLoops_Clade2,M.24.6.ii.
#mLoops_Clade2) #mLoops_Clade2) #mLoops_Clade2)
#mLoops_Clade2u4,(FU.22.2. #mLoops_Clade3,
(GC.13.1.
#mLoops_Clade3,(GC.4.2. #mLoops_Clade3,(FU.2.2.
#mLoops_Clade3,FU.5.2.
#mLoops_Clade3) #mLoops_Clade3) #mLoops_Clade3)
#mLoops_Clade3)
#mLoops_Clade3) #mLoops_Clade2u3u4)
#mLoops_Clade1u2u3u4ohneAG)#mLoops_root_alle;
[BRANCHES] Branch_COI_Pos3 (AG.1.1.
#mCOI_Pos3_Clade1,(((SEL.1.1.
#mCOI_Pos3_Clade1,SEL.1.3. #mCOI_Pos3_Clade1)
#mCOI_Pos3_Clade1,(LZ.18.2. #mCOI_Pos3_Clade1,
(LZ.5.18.
```

```
#mCOI_Pos3_Clade1,LZ.10.1.ii. #mCOI_Pos3_Clade1)
#mCOI_Pos3_Clade1)
#mCOI_Pos3_Clade1) #mLoops_Clade1,(((ZYP.1.1.
#mCOI_Pos3_Clade4,(ND.1.2.
#mCOI_Pos3_Clade4,(M.50.2. #mCOI_Pos3_Clade4,
(ESA.09.1.5.
#mCOI_Pos3_Clade4,ESA.09.4.5. #mCOI_Pos3_Clade4)
#mCOI_Pos3_Clade4)
#mCOI_Pos3_Clade4) #mCOI_Pos3_Clade4)
#mCOI_Pos3_Clade4,((M.39.1.
#mCOI_Pos3_Clade2,(M.30.2.ii. #mCOI_Pos3_Clade2,
M.32.2.
#mCOI_Pos3_Clade2) #mCOI_Pos3_Clade2)
#mCOI_Pos3_Clade2,(M.25.1.
#mCOI_Pos3_Clade2,M.24.6.ii. #mCOI_Pos3_Clade2)
#mCOI_Pos3_Clade2)
#mCOI_Pos3_Clade2) #mCOI_Pos3_Clade2u4,(FU.22.2.
#mCOI_Pos3_Clade3,(GC.13.1. #mCOI_Pos3_Clade3,
(GC.4.2.
#mCOI_Pos3_Clade3,(FU.2.2. #mCOI_Pos3_Clade3,
FU.5.2. #mCOI_Pos3_Clade3)
#mCOI_Pos3_Clade3) #mCOI_Pos3_Clade3)
#mCOI_Pos3_Clade3)
#mCOI_Pos3_Clade3) #mCOI_Pos3_Clade2u3u4)
#mCOI_Pos3_Clade1u2u3u4ohneAG)#mCOI_Pos3_roo
t_alle;
[PARTITIONS] pStems [Tree1 mStems 174]
[PARTITIONS] pCOI_Pos1 [Tree1 mCOI_Pos1 210]
[PARTITIONS] pCOI_Pos2 [Tree1 mCOI_Pos2 210]
[PARTITIONS] pLoops [Tree1 Branch_Loops 602]
[PARTITIONS] pCOI_Pos3 [Tree1 Branch_COI_Pos3
210]
[EVOLVE]
pStems 100 Stems
pCOI_Pos1 100 COI_Pos1
pCOI_Pos2 100 COI_Pos2
pLoops 100 Loops
pCOI_Pos3 100 COI_Pos3
```

References

- Avice, J. (1994). *Molecular markers, natural history and evolution*. New York: Chapman and Hall.
- Betancur-R, R., Li, C., Munroe, T. A., Ballesteros, J. A., & Ortí, G. (2013). Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Systematic Biology*, 62, 763–785.
- Boussau, B., & Gouy, M. (2006). Efficient likelihood computations with nonreversible models of evolution. *Systematic Biology*, 55, 756–768.
- Chiba, S. (1999). Accelerated evolution of land snails *Mandarina* in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. *Evolution*, 53, 460–471.

- Conant, G. C., & Lewis, P. O. (2001). Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. *Molecular Biology and Evolution*, *18*, 1024–1033.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*, 772–772.
- Däumer, C., Greve, C., Hutterer, R., Misof, B., & Haase, M. (2012). Phylogeography of an invasive land snail: natural range expansion versus anthropogenic dispersal in *Theba pisana pisana*. *Biological Invasions*, *14*, 1665–1682.
- De Rijk, P., Wuyts, J., van de Peer, A., Winkelmans, T., & De Wachter, R. (2000). The European large subunit ribosomal RNA database. *Nucleic Acids Research*, *28*, 177–178.
- D'Orbigny, A. (1839). Mollusques, échinodermes, foraminifères et polypiers, recueillis aux Îles Canaries par MM. Webb et Berthelot. Mollusques. In: P. B. Webb, & S. Berthelot (Eds.), *Histoire Naturelle des Îles Canaries*, tome II, partie 2, Zoologie, livr. 43: 49–72. Paris: Plasson & Renaud.
- Draparnaud, J. P. R. (1805). *Histoire naturelle des mollusques terrestres et fluviatiles de la France*. Paris: Plasson & Renaud.
- Drummond, A. J., Ho, S. Y. W., Philipps, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*, e88.
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973.
- Dutheil, J., & Boussau, B. (2008). Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evolutionary Biology*, *8*, 255–5.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*, *27*, 401–410.
- Florentino, V., Salomone, N., Manganello, G., & Giusti, F. (2010). Historical biogeography of Tyrrhenian land snails: the *Marmorana-Tyrrheniberus* radiation (Pulmonata, Helicidae). *Molecular Phylogenetics and Evolution*, *55*, 26–37.
- Fletcher, W., & Yang, Z. H. (2009). INDELible: a flexible simulator of biological sequence evolution. *Molecular Biology and Evolution*, *26*, 1879–1888.
- Foster, P. G., & Hickey, D. A. (1999). Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, *48*, 284–290.
- Funk, D. J., & Omland, K. E. (2003). Species level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397–423.
- Galtier, N., & Gouy, M. (1995). Inferring phylogenies from DNA sequences of unequal base compositions. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 11317–11321.
- Galtier, N., & Gouy, M. (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Molecular Biology and Evolution*, *15*, 871–879.
- Gittenberger, E., & Ripken, T. E. J. (1987). The genus *Theba* (Mollusca: Gastropoda: Helicidae), systematics and distribution. *Zoologische Verhandlungen*, *241*, 3–59.
- Greve, C., Hutterer, R., Groh, K., Haase, M., & Misof, B. (2010). Evolutionary diversification of the genus *Theba* (Gastropoda: Helicidae) in space and time: a land snail conquering islands and continents. *Molecular Phylogenetics and Evolution*, *57*, 572–584.
- Haase, M., Greve, C., Hutterer, R., & Misof, B. (2014). Amplified fragment length polymorphisms, the evolution of the land snail genus *Theba* (Stylommatophora: Helicidae), and an objective approach for relating fossils to internal nodes of a phylogenetic tree using geometric morphometrics. *Zoological Journal of the Linnean Society*, *171*, 92–107.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.
- Ho, S. Y. W., & Jermiin, L. S. (2004). Tracing the decay of the historical signal in biological sequence data. *Systematic Biology*, *53*, 623–637.
- Ishikawa, S. A., Inagaki, Y., & Hashimoto, T. (2012). RY-coding and non-homogeneous models can ameliorate the maximum-likelihood inferences from nucleotide sequence data with parallel compositional heterogeneity. *Evolutionary Bioinformatics*, *8*, 357–371.
- Jermiin, L. S., Ho, S. Y. W., Ababneh, F., Robinson, J., & Larkum, A. W. D. (2004). The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Systematic Biology*, *53*, 638–643.
- Johnson, M. S., Hamilton, Z. R., Teale, R., & Kendrick, P. G. (2012). Endemic evolutionary radiation of *Rhagada* land snails (Pulmonata: Camaenidae) in a continental archipelago in northern Western Australia. *Biological Journal of the Linnean Society*, *106*, 316–327.
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B., Wägele, J. W., & Misof, B. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, *7*, 10.
- Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701.
- Lockhart, P. J., Steel, M. A., Hendy, M. D., & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution*, *11*, 605–612.
- Lockhart, P. J., Larkum, A. W. D., Steel, M. A., Waddell, P. J., & Penny, D. (1996). Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 1930–1934.
- Martínez-Ortí, A., & Borredà, V. (2013). *Drusia (Escutiella) alexantoni* n. sp. (Gastropoda, Pulmonata, Parmacellidae), a new terrestrial slug from the Atlantic coast of Morocco. *Animal Biodiversity and Conservation*, *36*, 59–67.
- Misof, B., & Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology*, *58*, 21–34.
- Moers, A. Ø., & Holmes, E. D. (2000). The evolution of base composition and phylogenetic inference. *Trends in Ecology & Evolution*, *15*, 365–369.
- Moquin-Tandon, M. A. (1848). Quelques mots sur l'anatomie des mollusques terrestres et fluviatiles. *Actes de la Société Linnéenne de Bordeaux, Série*, *2*(5), 259–264.
- Müller, O. F. (1774). *Vermium terrestrium et fluviatilium, seu animalium infusoriorum, helminthicorum, et testaceorum, non marinorum, succincta historia. Volumen alterum*. Copenhagen: Heineck & Faber.
- Nesnidal, M. P., Helmkampf, M., Bruchhaus, I., & Hausdorf, B. (2010). Compositional heterogeneity and phylogenetic inference of metazoan relationships. *Molecular Phylogenetics and Evolution*, *27*, 2095–2104.
- Palsson, S., Magnúsdóttir, H., Reynisdóttir, S., Jonsson, Z. O., & Ornlófsdóttir, E. B. (2014). Divergence and molecular variation in common whelk *Buccinum undatum* (Gastropoda: Buccinidae) in Iceland: a trans-Atlantic comparison. *Biological Journal of the Linnean Society*, *111*, 145–159.

- Phillips, M. J., & Penny, D. (2003). The root of the mammalian tree inferred from whole mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *28*, 171–85.
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*, 793–808.
- Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). Tracer v1.6. <http://beast.bio.ed.ac.uk/Tracer>
- Razkin, O., Gómez-Moliner, B. J., Prieto, C. E., Martínez-Ortí, Arrébola, J. R., Muñoz, B., Chueca, L. J., & Madeira, M. J. (2015). Molecular phylogeny of the western Palaearctic Helicoidea (Gastropoda, Stylommatophora). *Molecular Phylogenetics and Evolution*, *83*, 99–117.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542.
- Rosenberg, M. S., & Kumar, S. (2003). Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. *Molecular Biology and Evolution*, *20*, 610–621.
- Sauer, J., & Hausdorf, B. (2010). Reconstructing the evolutionary history of the radiation of the land snail genus *Xerocrassa* on Crete based on mitochondrial sequences and AFLP markers. *BMC Evolutionary Biology*, *10*(299), 1–13.
- Singh, N. D., Arndt, P. F., Clark, A. G., & Aquadro, C. F. (2009). Strong evidence for lineage and sequence specificity of substitution rates and patterns in *Drosophila*. *Molecular Biology and Evolution*, *26*, 1591–605.
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihoodbased phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–2690.
- Stocsits, R. R., Letsch, H., Hertel, J., Misof, B., & Stadler, P. F. (2009). Accurate and efficient reconstruction of deep phylogenies from structured RNAs. *Nucleic Acids Research*, *37*, 6184–6193.
- Swofford, D. L. (2003). *PAUP* 4b10. Phylogenetic analysis using parsimony (*and other methods)*. Sunderland: Sinauer.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*, 4673–4680.
- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*, 3907–3930.
- Van den Bussche, R. A., Baker, R. J., Huelsenbeck, J. P., & Hillis, D. M. (1998). Base compositional bias and phylogenetic analyses: a test of the “flying DNA” hypothesis. *Molecular Phylogenetics and Evolution*, *10*, 408–416.
- Van de Peer, Y., de Rijk, P., Wuyts, J., Winkelmans, T., & de Wachter, R. (2000). The European small subunit ribosomal RNA database. *Nucleic Acids Research*, *28*, 175–176.
- Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, *30*, 1720–1728.
- Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, *26*, 1–7.
- Zielske, S., Glaubrecht, M., & Haase, M. (2011). Origin and radiation of rissooidean gastropods (Caenogastropoda) in ancient lakes of Sulawesi. *Zoologica Scripta*, *40*, 221–237.
- Zwickl, D. J. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Ph.D. dissertation, The University of Texas at Austin.