Received: 11 April 2020 | Revised: 4 October 2020 | Accepted: 21 November 2020

DOI: 10.1111/aje.12840

ARTICLE

African Journal of Ecology WILEY

Codon-based analysis of selection pressure and genetic structure in the *Psammobates tentorius* (Bell, 1828) species complex, and phylogeny inferred from both codons and amino acid sequences

Zhongning Zhao[1] | Neil Heideman[1] | Margaretha D. Hofmeyr[2]

[1]Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa

[2]Chelonian Biodiversity and Conservation, Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, South Africa

Correspondence
Zhongning Zhao, Department of Zoology and Entomology, University of the Free State, Bloemfontein, South Africa.
Email: orochi19851020@yahoo.com

Funding information
Universiteit van die Vrystaat, Grant/Award Number: A1999/158110; National Research Foundation, Grant/Award Number: IFR150216114248; University of the Western Cape, Grant/Award Number: SNS Grants

Abstract

This study used codon analysis (dN/dS and Tv/Ti) to investigate selection pressure and genetic structure in the highly polymorphic ***Psammobates tentorius*** species complex, and amino acid sequences to construct a phylogeny tree for it. Our results revealed a strong selection signal at node 'C2 + C3', possibly driven by aridity intensification resulting from the development of the Benguela Current. A similar signal was noticed at C3, possibly due to the same driving force. These findings suggest that environmental selection pressure favoured those groups and that further cladogenic events were possible. Selection pressure was also found to be high at C1, C4 and C7, which may indicate that they are also favoured by the current selection pressure. The codon-based phylogeny did not retrieve any potentially undescribed species, but nonetheless provided support for the validity of the seven distinct clades retrieved with the DNA sequence data. The amino acid sequence-based phylogeny generally supported the seven lineages as valid putative species. Investigation at the genomic scale could, however, help to solve the issue. In general, we found the codon, dN, dS, Tv, Ti and amino acid sequence-based phylogenetic inferences useful in species delimitation and recommend their use in species delimitation studies.

KEYWORDS

amino acid sequences, nonsynonymous and synonymous substitution, southern Africa, transition, transversion

Résumé

Cette étude s'est appuyée sur une analyse des codons (dN/ dS et Tv/ Ti) afin d'étudier la pression sélective et la structure génétique au sein du complexe d'espèces *Psammobates tentorius* hautement polymorphe, ainsi que sur les séquences d'acides aminés pour élaborer un arbre phylogénique à ces fins. Nos résultats ont révélé un important signal de sélection au nœud « C2 + C3 », probablement induit par l'intensification de l'aridité résultant du développement du courant de Benguela. Un signal similaire a été remarqué au C3, probablement en raison de la même force motrice. Ces résultats suggèrent que la pression sélective environnementale a favorisé

ces groupes et que d'autres événements cladogénétiques pouvaient également être impliqués. La pression sélective s'est également avérée élevée au C1, C4 et C7, ce qui peut indiquer qu'ils sont également favorisés par la pression sélective actuelle. La phylogénie basée sur les codons n'a pas mis en lumière d'espèces potentiellement non décrites, mais a néanmoins fourni un aide concernant la validité des sept clades distincts, déterminés avec les données de séquence d'ADN. La phylogénie basée sur la séquence d'acides aminés a permis de définir les sept lignées en tant qu'espèces présumées valides. Une étude menée à l'échelle génomique pourrait cependant permettre de résoudre le problème. D'une manière générale, nous avons constaté que les inférences phylogénétiques basées sur les codons, dN, dS, Tv, Ti et les séquences d'acides aminés étaient utiles en matière de délimitation des espèces, et nous recommandons leur utilisation dans le cadre des études réalisées dans ce domaine.

1 | INTRODUCTION

1.1 | Background

South Africa is the only country in the world with four biodiversity hotspots and is home to over 30% of the world's tortoise species (Branch, 2008; Cunningham, 2002; Hofmeyr et al., 2014). This includes the tent tortoise [*Psammobates tentorius* (Bell, 1828) species complex] with its highly variable and confusing carapace and plastron patterns, which has precluded the development of a stable taxonomy for the group for decades (Hewitt, 1933 and 1934; Loveridge & Williams, 1957; Boycott & Bourquin, 2000; Branch, 2008; Hofmeyr et al., 2014). Recently Zhao, Heideman, Grobler, et al. (2020) discovered seven distinct lineages in this species complex using multiple DNA sequence-based phylogenetic analyses (for the distribution of the seven clades and their corresponding morphology, see Figure 1a,b) and species delimitation approaches. The results implied that some of the lineages may be formally described as species. Furthermore, among the seven gene loci used in the study conducted by Zhao, Heideman, Grobler, et al. (2020), the cytochrome b (Cyt-b) and NADH dehydrogenase (ND4) genes were found to be the most informative loci for phylogenetic inference in this species complex. Another recent study conducted by Zhao et al. (2020) revealed that the genetic structure of the seven clades was also supported by multiple microsatellite DNA markers.

1.2 | Selection analyses

Evolutionary theory proposes selection pressure as the main driving force shaping genetic variation among organisms (Poon et al., 2009). In all protein-coding genes, among all mutation events, nonsynonymous substitutions can directly lead to functional protein changes after gene translation (Poon et al., 2009), if the substitution occurs in the exon region without being spliced out during the post-transcription period (Patthy, 1994). Thus, a nonsynonymous substitution (dN) is highly likely to impact the fitness of an organism more than a silent mutation (a synonymous substitution - dS) (Poon et al., 2009; Xia, 2013).

An environmental change may result in one amino acid being replaced by another favourable one under certain selection pressures, and eventually, the favourable amino acid can become the fixed form as the new protein becomes part of a new trait, which survives (Poon et al., 2009; Xia, 2013). This process usually takes place over a long period of time; however, sometimes such replacements can take place much faster in response to a sudden environmental selection pressure change (Xia, 2013). Two kinds of selection pressures are used to measure selection direction and strength, namely negative and positive selection pressure. The relative ratio of dN/dS is an important indicator for measuring selection pressure (Poon et al., 2009). Under negative selection pressure, $dN/dS < 1$, that is certain selection pressures do not favour a mutation, whereas under positive selection pressure, $dN/dS > 1$, that is certain selection pressures favour the mutation (Poon et al., 2009). Among the three-codon positions, each evolves at a different rate and encodes different functional units (Poon et al., 2009; Xia, 2013).

In the vertebrate mitochondrial genome, second codon positions are the functional sites which evolve the slowest of the three codons, and substitution changes at these positions will eventually lead to nonsynonymous substitutions (Xia, 1998). By contrast, third codon positions evolve the fastest, and nucleotide substitutions are often synonymous (thus less functional). First codon position evolution nearly parallels amino acid replacement at both transition and transversion levels, is less functional than evolution at second codon positions and more comparable with evolution at third codon positions (Xia, 1998). Therefore, the substitution rate at first, second and third codon positions typically follows the order 'third' > 'first' > 'second' (Poon et al., 2009; Xia, 1998).

Transitions (Ti's) and Transversions (Tv's) together with the Tv/Ti ratio are also believed to be useful indicators for evaluating

selection strength and direction. The proportion of Ti's is usually much higher than the proportion of Tv's (Poon et al., 2009; Xia, 1998). In this study, we were therefore also interested in testing whether the Ti- and Tv-based partitions were useful in delineating putative species.

1.3 | Amino acid-based phylogenetic inferences

During transcription and translation, only the part of the protein-coding nucleotide sequence sites falling within the exon region will eventually be translated into amino acid sequences and folded into functional proteins after splicing of the post-transcription stage (Clancy, 2008; Darnell, 1978). In addition, studying phylogenies based on amino acid sequences can also uncover valuable signals for deeply understanding evolution (Margoliash & Fitch, 1971). Furthermore, amino acid-based studies can potentially limit the interference of silent mutations in the reconstruction of phylogenetic trees. This tends to be more in line with the evolution of functional macro-molecules (amino acids in this case), than nucleotide polymorphisms based phylogenetic reconstruction alone (Margoliash & Fitch, 1971).

In this study, the two most informative protein-coding mtDNA genes in the *P. tentorius* species complex, namely Cyt-b and ND4 (Zhao, Heideman, Grobler, et al., 2020), were used for performing codon-based selection analyses, as well as phylogenetic reconstruction based on the amino acid data sets (Cyt-b and ND4, separately as well as combined). Since the codon-based selection analyses and amino acid sequence-based phylogenetic reconstruction correlate with natural selection, they may provide meaningful insights into the correlation between selection pressure due to environmental change and cladogenesis. Moreover, this study also aimed to investigate whether the different codons and amino acid sequences generated phylogenetic topologies congruent with those retrieved by Zhao, Heideman, Grobler, et al. (2020) using traditional DNA phylogenetic analyses. In addition, to test whether

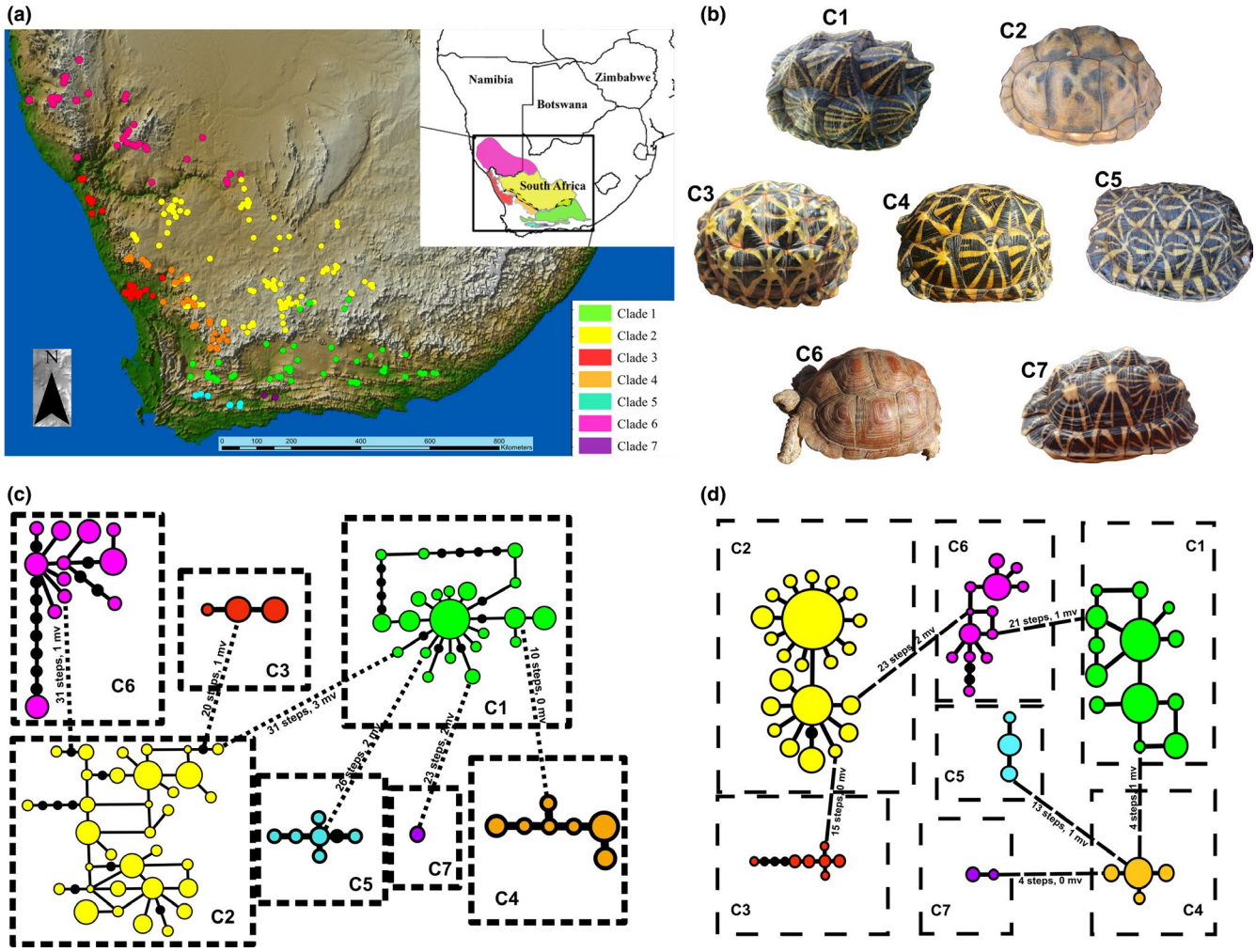

FIGURE 1 (a) map showing the distribution range of the seven clades in relation to geographic topology. The samples used in this study are indicated in colours that correspond to their respective mtDNA lineages. (b) The morphological variation in carapace among the seven lineages. (c) The haplotype network for Cyt-b. (d) The haplotype network for ND4. Note: The same colour scheme was used in networks as in the map [Colour figure can be viewed at wileyonlinelibrary.com]

the codon, dN, dS, Tv, Ti and amino acid-based phylogeny were useful in delineating Operational Taxonomic Units (OTUs) in these slow evolving reptiles.

2 | MATERIALS AND METHODS

2.1 | Selection analyses

The two most informative protein-coding genes, Cyt-b and ND4, involving 423 individuals of the *P. tentorius* species complex from the study of Zhao, Heideman, Grobler, et al. (2020) were downloaded from NCBI Genbank. All specimens along with locality information and accession numbers for the two genes are given as Table S1, and accession numbers of outgroups as Appendix S1: Table S2. Sampling covered almost the entire distribution range of the species complex according to the literature (Hewitt, 1933 and 1934; Loveridge & Williams, 1957; Boycott & Bourquin, 2000; Branch, 2008; Hofmeyr et al., 2014). The program MUSCLE v.3.2 (Edgar, 2004) was used to align the sequences. We first used DnaSP v.5 (Librado & Rozas, 2009) to retain only the haplotype sequences of Cyt-b, ND4 and Cyt-b + ND4 combined alignments. This was for determining the haplotype sequence of each data set (Cyt-b, ND4 and Cyt-b + ND4, respectively), and removing all the excess identical sequences. In order to investigate the evolution of the species complex at each codon position to deeply understand its molecular ecology in terms of selection, DAMBE v.6.1.9 (Xia, 2013) was used to first determine the reading frames and to locate the correct order of the three-codon positions for Cyt-b and ND4, respectively. Both genes did not show any stop codons. Thereafter, each gene partition was subdivided into three-codon partitions. The program PAUP* 4.0 (Swofford, 1998) was used to generate a NJ tree for each partition and to check the tree topology.

The codon partitions of Cyt-b and ND4, and the phylogeny of each codon partition were analysed using BEAST V 2.4.8 (Bouckaert et al., 2014) in a Bayesian inference (BI) analysis. Each partition parameter and substitution model were optimised using the optimal substitution model test results of Zhao, Heideman, Grobler, et al. (2020). The Yule clock model was selected, and the Markov Chain Monte Carlo (MCMC) run for 10 million generations, sampling every 2000 chains and discarding the first 10% as burn-in. We then used Tracer 1.6 (Rambaut et al., 2014) to check whether the effective sample size (ESS) of the majority of parameters reached the threshold value of 200. Additionally, each codon partition was also analysed using a Maximum Likelihood (ML) method with RAxML v.8 (Stamatakis, 2014) implemented in the program RDP4 (Martin et al., 2015) for comparative purposes, and each run with 1,000 nonparametric bootstrap replications. To investigate the substitution patterns for each protein-coding gene, transitions (Ti) and transversions (Tv) were separated into partitions by MEGA6 (Tamura et al., 2013) for the two genes. Each partition was analysed by NJ with bootstrap replications of 1,000.

HYPHY v.2.2.4 (Pond & Muse, 2005) was used to perform selection analyses on Cyt-b and ND4 independently. The dN and dS at each node and branch of the local substitution model were calculated for the entire tree topology using HYPHY v.2.2.4. The dN and dS were separated into different partitions for each gene and the NJ method used to generate a tree for each partition using HYPHY v.2.2.4. Bootstrap support values were calculated using RDP4. A likelihood ratio test (LRT) was conducted to test whether selection among tree topologies was in favour of using the local model (variable rate among sites). A preliminary LRT significantly confirmed the suitability of the local model compared to the global model (uniform rate among sites). Thus, nonsynonymous and synonymous substitution rates (dN and dS, respectively) at each branch of the NJ phylogenetic tree were calculated using the local substitution model 'MG94X HKY85_3x4' on Cyt-b and ND4.

Substitution rates are not constant among sites, so heterotachy was proposed as more realistic at different substitution sites of genes by Lopez et al. (2002). In order to test whether each branch evolved differently the DAMBE software was used to perform a relative rate test, although this approach to molecular clock assumptions is controversial (Scherer, 1989). Selection strength and direction in different branches of phylogenetic trees can also be variable (Poon et al., 2009). Thus, batch file 'TestBranchDNDS.bf' implemented in HYPHY v.2.2.4 was used to test nodes for significant dN/dS ratios when comparison with other parts of the tree.

Finally, in order to compare selection patterns between Cyt-b and ND4 in terms of dN/dS distribution, dN/dS proportions, selection strength of dN/dS and proportion of codons under selection, a 'dNdSDistributionComparison.bf' batch file was used, implemented in HYPHY v.2.2.4 to compare selection pattern differences between the protein-coding genes.

2.2 | Amino acid-based phylogenetic inferences

Protein-coding genes were translated into amino acid sequences using DAMBE. MrBayes v.3.2 (Ronquist et al., 2012) software was used to perform amino acid sequence-based BI phylogenetic analyses with Cyt-b, ND4 and the Cyt-b + ND4 combined data set, respectively. Each gene was translated using the mammalian mitochondrial database, as no reptile-specific amino acid analysis model is available; reptiles being phylogenetically comparatively close relatives of mammals (Gauthier et al., 1988; Hedges & Poling, 1999; McNab, 1978). Each gene alignment was run with a sampling chain length of 12 million generations, and sampled every 5,000 generations, with the first 25% discarded as burn-in. The methods and thresholds used to monitor diagnostic convergence and sample mixing were the average standard deviation of split frequencies (ASDSF) and ESS values (ASDSF < 0.02 and all parameters with ESS values > 200). Additionally, MEGA 6 was used to perform ML analysis on the amino acid sequences (Cyt-b, ND4 and Cyt-b + ND4), each with 1,000 bootstrap replications, using 4 discrete Gamma categories with a full amino acid substitution model, mtREV + I + G. The search method

used was the heuristic NNI method. The initial tree was automatically enforced by maximum parsimony for all the protein-coding genes.

2.3 | Species delimitation analyses

To evaluate the connection between haplotypes and the mutation stages between haplotype groups, we employed TCS v.2.1 (Clement et al., 2000) to reconstruct haplotype groups and PopART (Leigh & Bryant, 2015) to draw median-joining (MJ) networks (Bandelt et al., 1999) for Cyt-b and ND4, independently.

In order to test whether the codons, dN, dS, Tv, Ti and amino acid sequences are useful to clarify the species boundaries among the seven lineages, we used two different species delimitation methods. First, we used the coalescent phylogeny-based Bayesian implementation of the Poisson tree processes (bPTP) species delimitation method (Zhang et al., 2013) to delineate the seven lineages via bPTP Web Server (available at https://species.h-its.org/ptp/). This method was suitable for single locus-based molecular data (Zhang et al., 2013). The bPTP method also relies on the number of substitutions between haplotypes, and it assumes that more molecular variability is expected between species than within species (Zhang et al., 2013). We therefore used our haplotype-based phylogenetic tree to conduct the analysis. For the codon-based data sets (including the 1st, 2nd, 3rd and 1st + 2nd codons), we used the BEAST output tree as input tree. For the dN, dS, Tv and Ti data sets, we used the NJ tree as input tree, and for the amino acid sequences, we used the MrBayes output tree as input tree to perform the analysis. In all cases, the analysis was performed with 500,000 MCMC generations with 100 thinning and 0.1 burn-in.

Additionally, we also performed a MCMC sampling, minimum branch identification (mPTP) species delimitation analysis (Kapli et al., 2017) with the mPTP Web Server, using the same input trees for the different data sets as in the above bPTP method. This approach was also developed based on the PTP model but with multiple rates in conserved and faster ML algorithms.

3 | RESULTS

3.1 | Selection and codon analyses

The Tv partition-based tree of the Cyt-b gene (Appendix S1: Figure S1) failed to retrieve phylogenetic relationships among the clades in general, despite the monophyly of all clades being strongly supported (BP > 70). The Ti tree was almost identical to the full tree (Ti + Tv). Notwithstanding this, the Tv tree showed a higher transversion substitution rate and Tv/Ti ratio at the node where C2 and C3 bifurcated.

The Tv partition of the ND4-based NJ tree successfully retrieved six clades, C1-C6, each with strong support (BP > 70) (Appendix S1: Figure S2), but not for C7. The clusters 'C1 + C4 + C5 + C6' and 'C2 + C3' were strongly supported by the phylogeny based on the full data set (Ti + Tv: BP > 70). Outgroup members were generally well supported in comparison with the full tree, which may imply that Tv of ND4 was sufficiently informative to separate taxa at species level. In the Ti partition, nodes '(C1 + C4) + C7) + C5' and

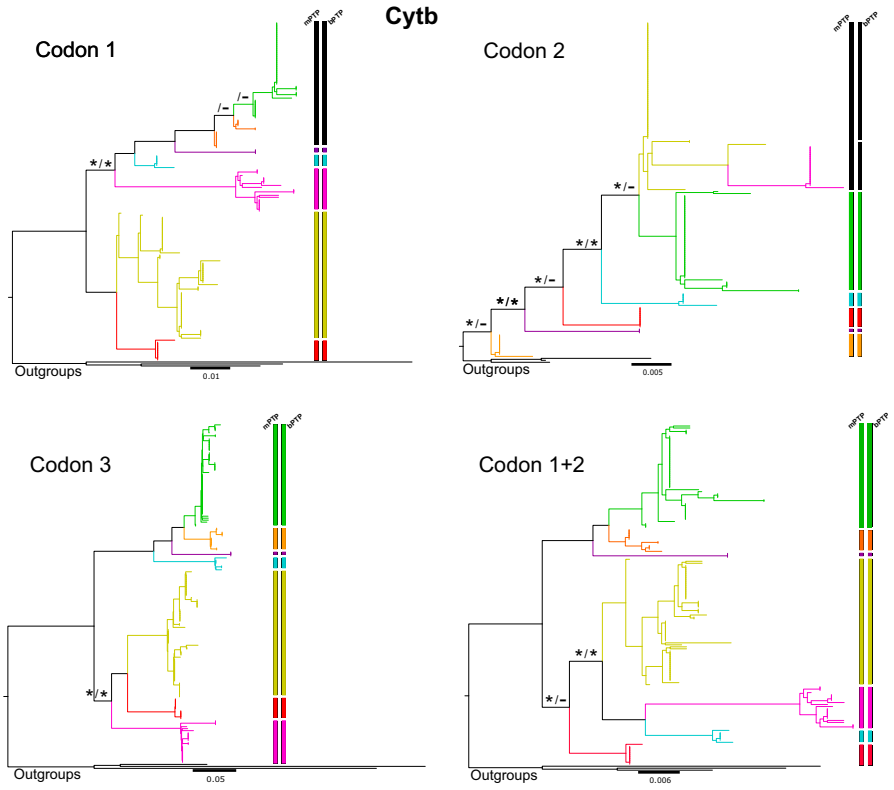

FIGURE 2 MI and BI trees based on different codon partitions for the Cyt-b gene. The colour scheme: 'green' represents C1, 'orange' C4, 'yellow' C2, 'purple' C7, 'aqua' C5, 'red' C3 and 'pink' C6. The same colour scheme was be used in other figures throughout this study. Values on the left side refer to bootstrap support (BP) for the ML analysis, and on the right side the BI posterior probability (PP). The ' * ' denotes weak support: PP < 0.95 or BP < 70, ' - ' denotes an unsupported node, while its absence denotes strong branch support (BP > 70 or PP > 0.95). The barplot at the right side of each tree represents species delimitation results [Colour figure can be viewed at wileyonlinelibrary.com]

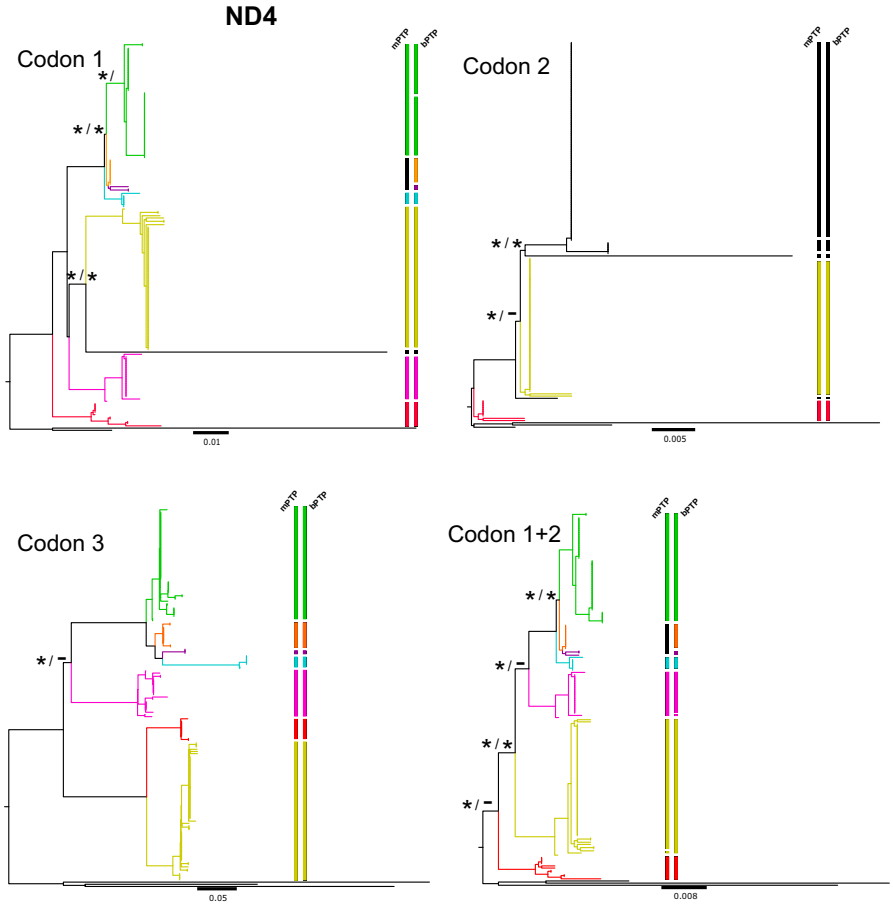

FIGURE 3 MI and BI trees based on different codon partitions for the ND4 gene. Values on the left side refer to bootstrap support (BP) for the ML analysis, and on the right side the BI posterior probability (PP). The ' * ' denotes weak support: PP < 0.95 or BP < 70, ' - ' denotes an unsupported node, while its absence denotes strong branch support (BP > 70 or PP > 0.95). The barplot at the right side of each tree represents species delimitation results [Colour figure can be viewed at wileyonlinelibrary.com]

'(C2 + C3) + C6)' were generally supported which was incongruent with the tree based on the full data set.

For Cyt-b, the codon 1 tree showed strong conflicts (Figure 2) in comparison with the topology of the Cyt-b gene in the traditional phylogenetic analysis. Clade 5 to C7 were strongly supported by both ML and BI analyses (BP > 70 and PP > 0.95), while '(C1 + C4) + C7) + C5' and C6 grouped as a cluster, but with only weak support (PP < 0.95, BP < 70). The monophyly of C1 and C4 was not supported. In terms of evolutionary rate, C6 had the fastest rate of all.

For the codon 2 tree of the Cyt-b gene (Figure 2), five of the expected seven clades were retrieved, and only the monophyly of C2 violated, since C2 and C6 seemed to cluster as one group. The relationships among the clades were generally incongruent with that of the full tree of Cyt-b, since most of the nodes were poorly supported. In terms of the evolutionary rate at the codon 2 position, C5, C6 and C7 were shown to evolve at similar rates but faster compared with the rest of the clades, while two samples of C1 from Matjiesfontein also evolved at a uniquely high rate in comparison with the other clades. Remarkably, in the full tree, the Serrated tent tortoise [*Psammobates oculifer*, (Kuhl, 1,820)] and C6 evolved much faster than the rest. However, at the second codon position, C5, C6 and C7 still showed relatively fast evolutionary rates, while *P. oculifer* seemed to evolve relatively slowly compared with many of the other taxa. Furthermore, although C1 showed a relatively slow evolutionary rate in the full tree, its evolutionary rate was relatively high for the second codon position, which implies higher nonsynonymous mutations.

When comparing the trees of codon 1 + 2 and codon 3 for Cyt-b, their tree topologies differed, with C5 grouping with C6, and cluster (C2 + C3) not being supported. Codon 3 generally had a stronger support value at each node, except for the placement of C6, which grouped with (C2 + C3) with weak support (BP < 70) (Figure 2). All seven clades were generally supported for codons 1, 3 and codon 1 + 2 positions. The support value for each node of the full tree was high and the relationship between nodes strong; only C6 presented a conflict.

For the ND4 gene (Figure 3), the codon 1 partition retrieved seven clades, C1-C7. Clade 2 to C7 each had strong support (PP > 0.95; BP > 70), with only the ML analysis showing weak support for C1 (BP < 70). The node '(C1 + C4) + C7) + C5' was strongly supported (PP > 0.95; BP > 70), while '(C2 + C3) + C6' was not supported (PP > 0.95; BP > 70). The grouping of the leopard tortoise [*Stigmochelys pardalis*, (Bell, 1828)] with C2 was unexpected but could be an artefact due to low substitution information at codon 1.

Codon 2 strongly supported only the monophyly of C2 and C3. The monophyly of *P. tentorius* and *Psammobates* was violated, as *S. pardalis* grouped with 'C1 + C4 + C5 + C6' (although the clade support value was low, PP < 0.95, BP < 70). Outgroup member *S.*

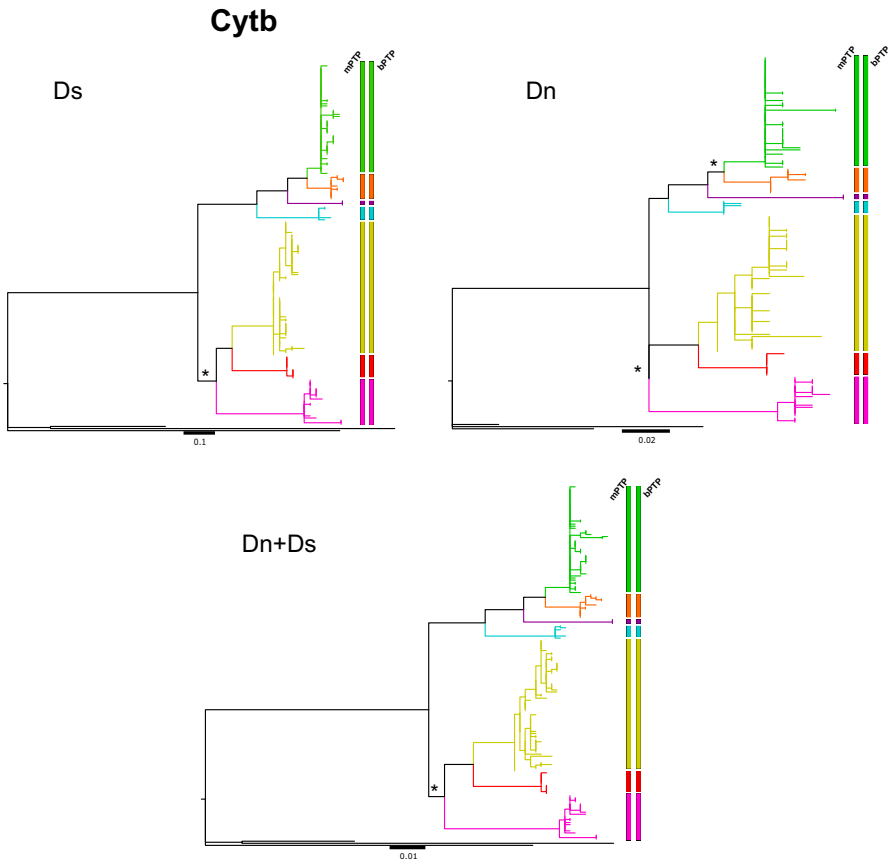

FIGURE 4 NJ trees for the Cyt-b gene for the full partition (dN + dS), dN partition and dS partition, respectively. Bootstrap support is indicated above each branch. The ' * ' denotes weak support: PP < 0.95 or BP < 70, ' - ' denotes an unsupported node, while its absence denotes strong branch support (BP > 70 or PP > 0.95). The barplot at the right side of each tree represents species delimitation results [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 The synonymous rates (dS), nonsynonymous rates (dN) and selection strength (dN/dS) across all nodes; neighbour-joining tree used as backbone tree

Locus	Cyt-b			ND4		
Object	dS	dN	dN/dS	dS	dN	dN/dS
Node						
Pg	0.3654	0.0193	0.053	0.5366	0.0421	0.078
Po	1.0943	0.1044	0.095	1.3339	0.1209	0.091
Pt	0.6044	0.0819	0.136	0.4632	0.0092	0.020
(C2 + C3)+C6	0.0589	0	0.000	NS	NS	NS
(C1 + C4)+C7 + C5	0.1877	0.0081	0.043	0.2228	0.0156	0.070
(C1 + C4)+C7)	0.0978	0.0164	0.168	NS	NS	NS
C1 + C4	0.0621	0.0068	0.110	NS	NS	NS
C2 + C3	0.0503	0.0206	0.410	0.2832	0	0.000
C1	0.0439	0.0171	0.390	0.0216	0.0078	0.361
C2	0.1314	0.0079	0.060	0.1553	0.0086	0.055
C3	0.1737	0.0284	0.164	0.1768	0.0147	0.083
C4	0.0761	0.0194	0.255	0.1342	0	0.000
C5	0.1967	0.0231	0.117	0.2841	0.0078	0.027
C6	0.2793	0.0536	0.192	0.2908	0	0.000
C7	0.1742	0.0565	0.324	0.1009	0	0.000

Abbreviations:: NS, unsupported; Pg, *Psammobates geometricus*; Po, *Psammobates oculifer*; Pt, *Psammobates tentorius*.

pardalis showed higher evolutionary rates at the second codon position. Evolutionary rates within the *P. tentorius* species complex all appeared to be at the same level. Codon 3 strongly supported the seven clades; however, both codon 1 and codon 1 + 2 showed a topological conflict regarding the placement of C3 when compared with the topology of the full tree. Codon 3 showed a topology similar

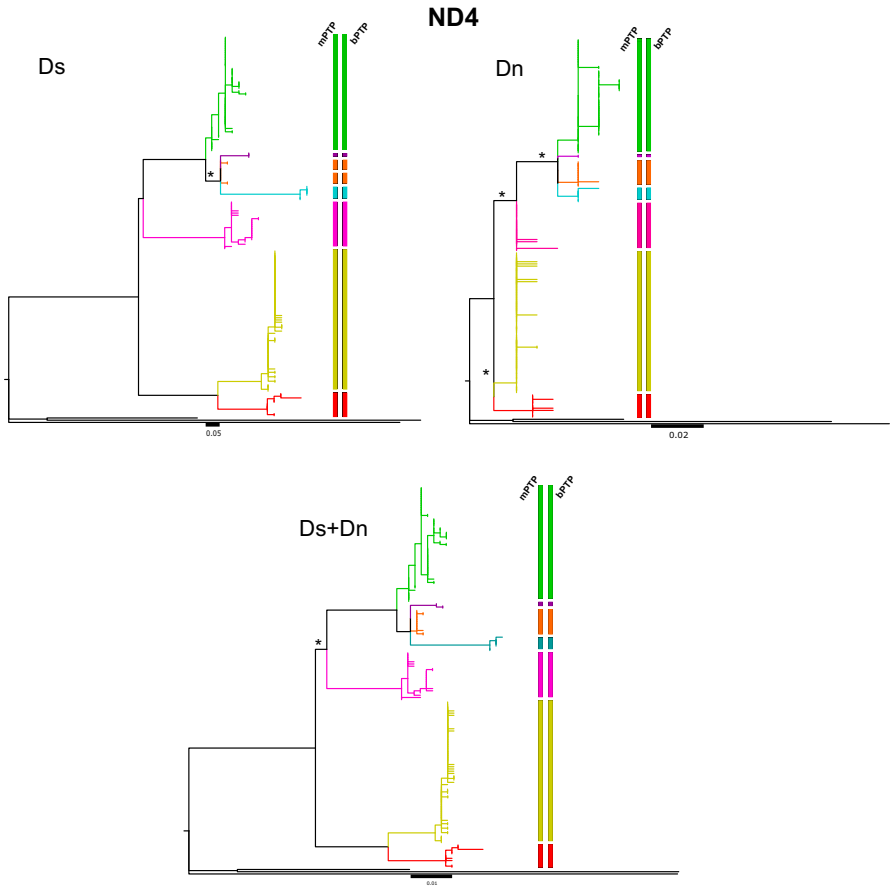

FIGURE 5 NJ trees for the ND4 gene for the full partition (dN + dS), dN partition and dS partition, respectively. Bootstrap support is indicated above each branch. The ' * ' denotes weak support: PP < 0.95 or BP < 70, ' - ' denotes an unsupported node, while its absence denotes strong branch support (BP > 70 or PP > 0.95). The barplot at the right side of each tree represents species delimitation results [Colour figure can be viewed at wileyonlinelibrary.com]

to the full ND4 gene tree with generally strongly supported clusters 'C6 + (C5 + (C1 + C4))' and 'C2 + C3'.

3.2 | dN/dS-based selection analysis

For the Cyt-b gene (Figure 4), no node in the tree topology was detected with a strongly favourable selection signal (Table 1). Nevertheless, the favourable selection strength (dN/dS ratio) at C1, C4, C6, C7 and node (C2 + C3) was found to be significantly greater than that at the rest of the nodes (Table 1), particularly that of (C2 + C3). The nonsynonymous substitution tree still supported all seven clades, but again with the placement of C6 as uncertain.

In the case of the ND4 gene (Figure 5), no positive selection signal (node with dN/dS > 1) was found throughout its tree topology, with the highest selection strength at C1 (Table 1). Nevertheless, the additional codon-based phylogenetic information is still valuable (in terms of determining the clades). The nonsynonymous tree supported the monophyly of all seven clades, but the relationships between them were poorly resolved. Only (C1 + C4) + C7) + C5 was supported as a lineage as well as its sister group (C2 + C3), while the placement of C6 remained unclear (Table 1).

When comparing selection patterns between the Cyt-b and ND4 genes, the dN/dS-based distribution comparison tests revealed: (a) that there was a significant difference in their dN/dS distribution (LR = 24.184, df = 10, p = 0.007); (b) that the selective regimes (dN/dS and proportions) did not differ significantly between them (LR = 2.699, df = 2, p = 0.259); (c) that the selection strengths (dN/dS) were not different across the two genes (LR = 0.378, df = 1, p = 0.828); and (d) that the proportions of codons under selection did not differ significantly between them (LR = 1.958, df = 1, p = 0.162). These results may imply that the two genes were under similar selection pressures.

3.3 | Amino acid phylogenetic analysis

The amino acid phylogenetic tree for Cyt-b (Figure 6) showed strong support for clades C3, C4, C5, C6 and C7 (all with PP > 0.95, BP > 70), while C2 was not supported and C1 was only weakly supported (PP < 0.95, BP < 70). The relationship between C1 and (C4 + C5) was not supported. When comparing the tree topology with the Cyt-b gene tree, topological conflicts occurred at C2, C3 and C4. The amino acid phylogram did not support C2 and C4 as monophyletic groups. Nevertheless, C2 + C6 and C3 still clustered as sister groups with strong support (PP > 0.95, BP > 70). The grouping of C5 with C1, C2, C3, C7 and C6 was unexpected.

The amino acid tree for the ND4 gene (Figure 6) exhibited strong support for clades C3, C4, C5 and C6 (all with PP > 0.95, BP > 70), while C1 and C2 were only weakly supported (PP < 0.95, BP < 70). The relationship between C1, C7 and C4 + C5 was also strongly

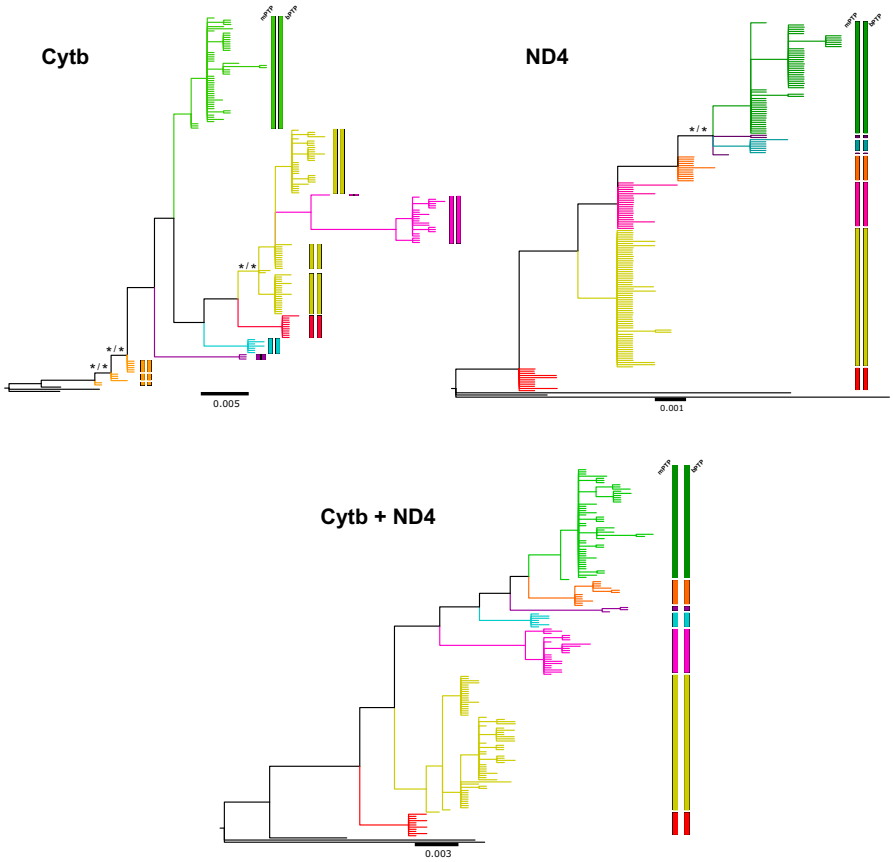

FIGURE 6 Amino acid sequence phylograms for Cyt-b, ND4 and Cyt-b + ND4. The same colour scheme used previously is used to highlight the seven clades. The tree topologies used are the consensus phylograms generated from BI (MrBayes) analysis. The ' * ' denotes weak support: PP < 0.95 or BP < 70, ' - ' denotes an unsupported node, while its absence denotes strong branch support (BP > 70 or PP > 0.95). The barplot at the right side of each tree represents species delimitation results [Colour figure can be viewed at wileyonlinelibrary.com]

supported (PP > 0.95, BP > 70). When comparing the tree topology with the ND4 gene tree, topological conflicts occur at C2 and C3, while the amino acid phylogram did not support C2 as a monophyletic group. Nevertheless, C2 and C3 still clustered as a sister group with strong support (PP > 0.95, BP > 70).

The ND4 and Cyt-b combined amino acid tree (Figure 6) retrieved all seven clades as the DNA sequence data set advocated, each node with strong support (PP > 0.95, BP > 70). The relationship among clades, however, was not congruent when compared with the DNA-based phylogeny. Clade 3 was positioned as a basal clade, rather than clustering with C2 as sister group.

The Transversion/Transition ratio (Tv/Ti) was found to be significantly higher in 'C2 + C3' and C6 for Cytb, while the Tv/Ti ratio was higher in group '(C1 + C4 + C7) + C5', 'C2 + C3' and C3 for the ND4 gene (see Appendix S1: Table S3).

3.4 | Species delimitation analyses

Our haplotype network analysis revealed seven distinctive haplotype groups for both the Cyt-b and ND4 genes that match the seven mtDNA lineages (Figure 1c,d). The mutation steps between C1 and C4 were smaller than that between the other clades.

In general, for the codon-based data sets (Cyt-b and ND4 genes), our species delimitation analyses delineated the seven mtDNA lineages as valid putative species in the case of the 3rd and 1 + 2 codons (Figures 2 and 3), though a few place exhibited minor subdivisions. However, in the case of the 1st codon, the boundary between C1 and C4 was blurred (Figures 2 and 3). The species boundary was especially poorly defined at the conservative 2nd codon, in the Cyt-b gene. Only the species boundaries between C1, C5, C3, C7 and C4 were clearly demarcated, while in the ND4 gene, only C2 and C3 could be differentiated. For the dN and dS partitions, both Cyt-b and ND4 advocated the seven lineages as valid putative taxa, except minor subdivision in C4 for the ND4 gene (Figures 4 and 5). In the amino acid sequences (Figure 6), ND4 and the combined data sets supported the seven lineages as unique putative species, while the Cyt-b gene tended to over-split C2. Lastly, Ti- and Tv-based species delimitation results (for both genes) generally supported the seven lineage as valid putative species.

4 | DISCUSSION

4.1 | Selection analyses

The selection analysis results showed that the dN/dS between Cyt-b and ND4 did not differ significantly with respect to selection regimes (dN/dS and its proportions), selection strengths (global dN/dS) and portions of codons under selection but differed in selection distribution. Thus, it may be concluded that the selection pressure variation on these genes can be disregarded. It may therefore reasonably be deduced that the selection pressures differ on the different phylogenetic groups comes from their respective environments. The amino

acid sequences of Cyt-b and ND4 were therefore combined for the phylogenetic inferences.

The transversion partitions of both Cyt-b and ND4 (Appendix S1: Figures S1–S2) supported the proposition that C2 and C3 were derived from the same common ancestor, and also provided additional support for their sister relationship. Despite the very conservative transversion rate, the transversion tree for both Cyt-b and ND4 generally retrieved seven clades, which matched the clustering scheme revealed in the mtDNA tree. This could be considered as evidence that the seven clades are indeed distinct from each other. In general, the 3rd codon and 1st + 2nd codon partitions of both Cyt-b (Figure 2) and ND4 (Figure 3) retained sufficient levels of polymorphism to support the monophyly of the seven clades. The second codon position of both genes was much less informative, particularly in the ND4 gene, where it did not even support the monophyly of *Psammobates*. The 2nd codon position in Cyt-b did not support the monophyly of C2 + C3, but rather that C2 was mixed with C6 as a single cluster, despite the monophyly of C6 being well supported.

The codon analyses failed to discover more taxa than the traditional DNA phylogenetic analyses. It nevertheless provided additional support for the stability of the seven clades. The unexpected systematic positions of C5 and C7 in the codon-based analyses were perhaps merely artefacts due to insufficient informative sites. Notwithstanding this, it could also be considered as evidence that C5 and C7 are different from C1. As all the substitutions at the 2nd codon position will eventually lead to protein changes (if the substitution takes place in the exon region) (Xia, 1998), the transversion-based phylogeny derived from the Cyt-b gene could indicate that C1, C5, C6 and C7 were under similar selection pressure. However, such an interpretation is difficult to justify given their completely different environments.

The transition and transversion ratio-based analysis supported the distinctiveness of C3 and C6, and also revealed their high degree of genetic isolation, which implies that they may deserve to be treated as separate OTUs to the other clades.

According to the ND4 codon analysis at the 2nd codon (Figure 3), C1, C4, C5 and C6 grouped together with *S. pardalis*. Clade 2 and C3 remained as monophyletic clades, which may imply that in the case of the ND4 gene, C6 (north of Orange River) evolved under similar selection pressures as C1, C4 and C5 (despite the monophyly of these groups not being supported). The monophyly of C2 and C3 at the 2nd codon partition may be regarded as additional support for elevating both of them to species level.

In terms of the dN- and dS-based selection analyses, both genes (Cyt-b and ND4) generally supported the relationships of 'C2 + C3' and 'C1 + C4 + C5', contrary to their observed placement in relation to C6. Again, these results provide additional evidence supporting the proposed 'seven species assumption' in the *P. tentorius* species complex.

For the Cyt-b gene, both dN and dS partitions supported the monophyly of the seven clades, albeit weak in some cases. The dN/dS-based analyses suggest significantly higher positive selection pressure on C3, C6, C7 and the congeneric outgroup *P. oculifer*. The relatively high positive selection strength on C3 could be explained as adaptation to the formation of the Succulent Karoo biome under the influence of the cold Benguela current along the west coast region (Partridge, 1997), the increasing severity of summer droughts (Partridge, 1997) and the global cooling trend after the mid Miocene (Zachos et al., 2001). In addition to these, the formation of the Kamiesberg as a geographic barrier between the West Coast Succulent Karoo (west of the Kamiesberg) and inland Succulent Karoo and Nama Karoo (east of the Kamiesberg) probably also facilitated the isolation between the two populations.

The rapid evolution of *P. oculifer* was also reported by Cunningham (2002), which could have been caused by the geographic isolation from the rest populations. The Orange River maybe function as a vicariance barrier, as *P. oculifer* only occurs north of the Orange River (Branch, 2008; Hofmeyr et al., 2014, 2017), and is generally found in Savannah biomes with relatively high humidity (in terms of rainfall, Branch, 2008; Mucina & Rutherford, 2006). The faster rate of evolution in *P. oculifer* may also be correlated to its unique physiological properties in terms of respiration, thermoregulation or muscle function. Variation in humidity and vegetation type may thus function as additional factors which drive and accelerate cladogenesis in *P. oculifer* (da Silva et al., 2012). These assumptions obviously require verification through further studies. Apart from these, the expansion of the Kalahari Desert (Partridge & Maud, 1987) could also have influenced the diversification of *P. oculifer*. As the phylogenetic results also suggested, there is a sister relationship between *P. oculifer* and *Psammobates geometricus*. The *P. geometricus* occurs exclusively in the Fynbos biome (Branch, 1998 and 2008; Hofmeyr et al., 2014; Hofmeyr et al., 2017), both habitats having higher moisture levels in comparison with the drier Karoo habitat of *P. tentorius*. This may imply that humidity and precipitation also facilitate rapid cladogenesis in *P. oculifer*. However, these assumptions are merely preliminary hypotheses based on molecular evidence and will require testing using experiments or simulations. The apparently higher dN/dS ratio found in C6 may be adaptive responses to the Kalahari desert expansion and increasing aridification in recent time (Partridge & Maud, 1987). Coincidentally, the ND4 gene showed similar results, except that the selection pressure on C6 was not statistically significant. The high level of within-group divergence and high cladogenic radiation potential has been addressed by previous study (Zhao, Heideman, Grobler, et al., 2020), which seems to correspond to the findings of this study.

The selection analysis revealed relatively high (Table 1) selection strength on node (C2 + C3), which may suggest that cladogenesis between C2 and C3 was favoured by current environmental selection. The most likely explanation for this cladogenic event would be the increasing western aridity and the development of winter rainfall due to intensification of the Benguela Current (Hoffmann et al., 2015; Rommerskirchen et al., 2011). These conditions favoured the development of Succulent Karoo vegetation along western South Africa from about 10–8 Ma, although the Succulent Karoo biome became fully established only in the Pliocene (Neumann & Bamford, 2015; Verboom et al., 2009). These findings match results obtained from

calibration dating and biogeographical analyses (Zhao, Heideman, Bester, et al., 2020).

4.2 | Amino acid-based phylogenetic analysis

The results of the amino acid sequence-based phylogeny obtained from the Cyt-b gene (Figure 6) supported the monophyly of all the clades, except C4, which showed a polytomy. Notwithstanding this, significant genetic distance between members of C1 and C4 is still distinguishable. It was expected that the tree topology of the second codon position would be closer to that of the amino acid tree. However, the results did not show clear congruence between the second codon position tree and amino acid tree, for both genes. A possible interpretation of these findings would be that part of the nonsynonymous substitutions came from mutations of the first codon position, though all the substitutions that took place at the second codon position will eventually be translated into changes in amino acid sequences. Interestingly, the amino acid-based phylogeny showed that C5 was sister to 'C2 + C3', rather than to 'C1 + C4'. This may be an indication of the unique evolution of C5 in comparison with 'C1 + C4', which may be considered as additional support for C5 being regarded as a distinct species rather than part of C1.

The amino acid sequences of the ND4 gene (Figure 6) gave results generally compatible with that of the DNA sequences, except that C3 was presented as basal node in the phylogeny, and C4 was not the sister clade of C1 as expected. This again emphasises the distinctiveness of C1, and C2, C3 and C4.

The Cyt-b + ND4 combined amino acid data set confirmed all the clades as being distinct, despite C3 being positioned as basal clade, rather than the sister clade to C2, Again, this phenomenon could be an artefact due to fewer informative sites in comparison with the DNA-based phylogeny. Further genome scale DNA and amino acid (derived from genomic sequence data)-based phylogenetic analyses would be needed to hopefully finally clarify the relationships among the clades.

4.3 | Species delimitation

Our species delimitation analyses generally revealed that the codon-based partition was a useful indicator of species delimitation in the *P. tentorius* species complex, particularly in the 3rd and 1st + 2nd codons. The 2nd codon alone was too conservative for species delimitation due to its conservative nature. Nonetheless, the 2nd codon still provided additional support for demarcating some of the lineages. We found that partitions dN, dS, Ti and Ts were also informative markers for delineating our seven lineages, with only a few cases exhibiting minor subdivisions in certain lineages. Regarding the amino acid sequences, our species delimitation results imply that the functional unit orientated amino acid sequences (after eliminating the influence of silent mutations) were also informative delineators in demarcating the seven lineages into putative species.

In summary, we found that the codon, dN, dS, Ti and Ts partitions, and amino acid sequences were informative and useful in species delimitation in the *P. tentorius* species complex. However, to avoid making premature conclusion, we suggest further studies focusing at a larger scale with both empirical and simulation data which are necessary to verify whether these codon-based partitions and amino acid sequences are accurate markers in species delimitation over a broader species range.

5 | CONCLUSIONS

No further subdivisions were detected from the codon and amino acid-based phylogeny, compared to the phylogenetic analysis results obtained using the traditional molecular methods by Zhao, Heideman, Grobler, et al. (2020). Nonetheless, the codon and amino acid-based phylogeny generally provided additional support for the seven distinct evolutionary lineages, each of which should be regarded as a putative species and valid OTU. The selection analysis found that C3 had the highest fitness under the current selection pressures (based on the dN/dS ratios). This implies that the increasing west coast aridity and the development of winter rainfall due to the intensification of the Benguela Current possibly facilitated the divergence between C2 and C3, and may facilitate further radiation within C3. These assumptions should, however, only be regarded as preliminary, since experimental verification is needed for greater certainty. Finally, the high congruence of the dN/dS ratio-based patterns between the two mtDNA markers (Cyt-b and ND4) may suggest that selection pressure on these two genes is similar.

Finally, this study revealed that the partitions based on codons, dN, dS, Ti, Tv and amino acid sequences are informative and useful for delineating OTUs, and deserve to be considered as a general approach for species delimitation in all organisms with protein-coding genes.

ACKNOWLEDGEMENTS

We thank the National Research Foundation (Grant - IFR150216114248) and the Universities of the Free State (UFS Research Grant: A1999/158110) and the Western Cape (SNS Grants) for providing the funding to carry out the study. William Espenshade from Villanova University (USA) and Beryl Wilson from McGregor Museum (South Africa) were thanked for providing study materials.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

All the NCBI GenBank accession numbers of genetic sequences used in analyses were given in Table S1 and Appendix S1: Table S2.

ORCID

Zhongning Zhao https://orcid.org/0000-0002-6400-4743
Neil Heideman https://orcid.org/0000-0003-2047-1291

REFERENCES

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Boycott, R. C., & Bourquin, O. (2000). *The southern African tortoise book. A Guide to Southern African Tortoises, Terrapins and Turtles*. Revised, expanded edition. Interpak Press.

Branch, B. (1998). Field guide to snakes and other reptiles of southern Africa. Ralph Curtis Books.

Branch, B. (2008). *Tortoise, Terrapins and Turtles of Africa*. Struik Publisher.

Clancy, S. (2008). RNA splicing: Introns, exons and spliceosome. *Nature Education*, *1*(1), 31.

Clement, M., Posada, D. C. K. A., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, *9*(10), 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Cunningham, J. (2002). *A molecular perspective on the family Testudinidae Batsch, 1788*, Ph.D. Dissertation, University of Cape Town.

Darnell, J. E. (1978). Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science*, *202*(4374), 1257–1260. https://doi.org/10.1126/science.364651

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Gauthier, J., Kluge, A. G., & Rowe, T. (1988). Amniote phylogeny and the importance of fossils. *Cladistics*, *4*(2), 105–209. https://doi.org/10.1111/j.1096-0031.1988.tb00514.x

Hedges, S. B., & Poling, L. L. (1999). A molecular phylogeny of reptiles. *Science*, *283*(5404), 998–1001. https://doi.org/10.1126/science.283.5404.998

Hewitt, J. (1933). On the Cape species and subspecies of the genus *Chersinella* Gray. Part I. *Annals of the Natal Museum*, *7*, 255–297.

Hewitt, J. (1934). On the Cape species and subspecies of the genus Chersinella. Part II. *Annals of the Natal Museum*, *7*, 303–352.

Hoffmann, V., Verboom, G. A., & Cotterill, F. P. (2015). Dated plant phylogenies resolve neogene climate and landscape evolution in the Cape Floristic Region. *PLoS One*, *10*(9), e0137847. https://doi.org/10.1371/journal.pone.0137847

Hofmeyr, M. D., Boycott, R. C., & Baard, E. H. W. (2014). *Psammobates tentorius* (Bell, 1828). In M. F. Bates, W. R. Branch, A. M. Bauer, M. Burger, J. Marais, G. J. Alexander, & M. S. De Villiers (Eds.), *Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland* (pp. 70–85). South African National Biodiversity Institute.

Hofmeyr, M. D., Vamberger, M., Branch, W., Schleicher, A., & Daniels, S. R. (2017). Tortoise (Reptilia, Testudinidae) radiations in Southern Africa from the Eocene to the present. *Zoologica Scripta*, *46*, 389–400. https://doi.org/10.1111/zsc.12223

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, *33*(11), 1630–1638. https://doi.org/10.1093/bioinformatics/btx025

Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lopez, P., Casane, D., & Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, *19*(1), 1–7. https://doi.org/10.1093/oxfordjournals.molbev.a003973

Loveridge, A., & Williams, E. E. (1957). Revision of the African tortoises and turtles of the suborder Cryptodira. *Bulletin of the Museum of Comparative Zoology*, *115*, 161–557.

Margoliash, E., & Fitch, W. M. (1971). Amino Acid Sequences and Phylogeny. *Taxon*, *20*, 51–53. https://doi.org/10.2307/1218533

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*, 1. https://doi.org/10.1093/ve/vev003

McNab, B. K. (1978). The evolution of endothermy in the phylogeny of mammals. *The American Naturalist*, *112*(983), 1–21. https://doi.org/10.1086/283249

Mucina, L., & Rutherford, M. C. (2006). *The Vegetation of South Africa, Lesotho and Swaziland*. South African National Biodiversity Institute.

Neumann, F. H., & Bamford, M. K. (2015). Shaping of modern southern African biomes: Neogene vegetation and climate changes. *Transactions of the Royal Society of South Africa*, *70*(3), 195–212. https://doi.org/10.1080/0035919X.2015.1072859

Partridge, T. C. (1997). Evolution of landscapes. In R. M. Cowling, D. M. Richardson, & S. M. Pierce (Eds.), *Vegetation in southern Africa* (pp. 5–20). Cambridge University Press.

Partridge, T. C., & Maud, R. R. (1987). Geomorphic evolution of southern Africa since the Mesozoic. *South African Journal of Geology*, *90*(2), 179–208.

Patthy, L. (1994). Introns and exons. *Current Opinion in Structural Biology*, *4*(3), 383–392. https://doi.org/10.1016/S0959-440X(94)90108-2

Pond, S. L. K., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. In R. Nielsen (Ed.), *Statistical methods in molecular evolution* (pp. 125–181). Springer.

Poon, A. F., Frost, S. D., & Pond, S. L. K. (2009). Detecting signatures of selection from DNA sequences using Datamonkey. In D. Posada (Ed.), *Bioinformatics for DNA Sequence Analysis* (pp. 163–183). Humana Press. https://doi.org/10.1007/978-1-59745-251-9

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). Tracer 1.6. Retrieved from http://beast.bio.ed.ac.uk/tracer

Rommerskirchen, F., Condon, T., Mollenhauer, G., Dupont, L., & Schefuß, E. (2011). Miocene to Pliocene development of surface and subsurface temperatures in the Benguela Current system. *Paleoceanography*, *26*, 3. https://doi.org/10.1029/2010PA002074

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*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Scherer, S. (1989). The relative-rate test of the molecular clock hypothesis: A note of caution. *Molecular Biology and Evolution*, *6*(4), 436–441. https://doi.org/10.1093/oxfordjournals.molbev.a040561

Silva, F. R., Almeida-Neto, M., do Prado, V. H. M., Haddad, C. F. B., & de Cerqueira Rossa-Feres, D. (2012). Humidity levels drive reproductive modes and phylogenetic diversity of amphibians in the Brazilian Atlantic Forest. *Journal of Biogeography*, *39*(9), 1720–1732. https://doi.org/10.1111/j.1365-2699.2012.02726.x

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Swofford, D. L. (1998): PAUP* version 4.0. Phylogenetic analysis using parsimony (and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Verboom, G. A., Archibald, J. K., Bakker, F. T., Bellstedt, D. U., Conrad, F., Dreyer, L. L., & Mummenhoff, K. (2009). Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution*, *51*(1), 44–53. https://doi.org/10.1016/j.ympev.2008.01.037

Xia, X. (1998). How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics*, *149*, 37–44.

Xia, X. (2013). DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, *30*(7), 1720–1728. https://doi.org/10.1093/molbev/mst064

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686–693. https://doi.org/10.1126/science.1059412

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*, 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

Zhao, Z., Heideman, N., Bester, P., Jordaan, A., & Hofmeyr, M. D. (2020). Climatic and topographic changes since the Miocene influenced the diversification and biogeography of the tent tortoise (*Psammobates tentorius*) species complex in Southern Africa. *BMC Evolutionary Biology*, *20*(1), 1–33. https://doi.org/10.1186/s12862-020-01717-1

Zhao, Z., Heideman, N., Grobler, P., Jordaan, A., Bester, P., & Hofmeyr, M. D. (2020). Unraveling the diversification and systematic puzzle of the highly polymorphic *Psammobates tentorius* (Bell, 1828) complex (Reptilia: Testudinidae) through phylogenetic analyses and species delimitation approaches. *Journal of Zoological Systematics and Evolutionary Research*, *58*(1), 308–326. https://doi.org/10.1111/jzs.12338

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zhao Z, Heideman N, Hofmeyr MD. Codon-based analysis of selection pressure and genetic structure in the *Psammobates tentorius* (Bell, 1828) species complex, and phylogeny inferred from both codons and amino acid sequences. *Afr J Ecol*. 2021;59:497–509. https://doi.org/10.1111/aje.12840