Gene 551 (2014) 49–54

Contents lists available at ScienceDirect

Gene

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

# The complete mitochondrial genome of Temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904)

Zelda Du Toit [a,*], J. Paul Grobler [a], Antoinette Kotzé [a,b], Raymond Jansen [c], Helene Brettschneider [b], Desiré L. Dalton [a,b]

[a] Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa
[b] National Zoological Gardens of South Africa, P.O. Box 754, Pretoria 0001, South Africa
[c] Department of Environmental, Water and Earth Sciences, Tshwane University of Technology, Private Bag X680, Pretoria 0001, South Africa

ARTICLE INFO

*Article history:*
Received 6 January 2014
Received in revised form 14 August 2014
Accepted 22 August 2014
Available online 23 August 2014

*Keywords:*
Mitochondrial DNA
Temminck's ground pangolin
*Smutsia temminckii*
Phylogeny

ABSTRACT

Temmincki's ground pangolin is primarily a nocturnal mammal belonging to the order Pholidota. The body is covered in hard overlapping scales and these animals find refuge in burrows, feeding only on termites and ants. In this study, the whole mtDNA of Temmincki's ground pangolin was sequenced and the phylogenetic position of Pholidota determined within Eutheria, using whole mtDNA sequences from various representative species. The results indicate that the whole mtDNA of Temmincki's ground pangolin is 16,559 bp long and shared some similarities with the whole mtDNA of the back-bellied tree pangolin and the Chinese pangolin. Phylogenetic analysis indicate that the order Pholidota is closely related and share a recent common ancestor with the order Carnivora rather than with the ant/insect eating order Xenarthra and the group Afrotheria. A time measured phylogeny of Pholidota estimated a split from Carnivora at around 87 mya, followed by a split of the African pangolins from their Asian counterparts such as the Chinese pangolin at around 47 mya. This suggests a Laurasian origin and convergent evolution of the Pholidota with respect to Xenarthra and Afrotheria.



## 1. Introduction

Mitochondrial DNA (mtDNA) in mammals is a circular structure consisting of approximately 14.5 to 19.5 kb which codes for 37 gene regions. These gene regions include between 20 to 30 tRNA regions, two rRNA regions, 13 protein coding regions, various nucleotide gene regions which comprise of cytochrome *c* oxidase (COI, COII, COIII), ATPase 6 and cytochrome *b* as well as eight unidentified reading frames (Castro et al., 1998; Freeland, 2005; Ki et al., 2010; Roe et al., 1985). Mitochondrial DNA is well suited for the analysis of ancient or degraded samples as it is preserved for longer compared to nuclear DNA and only a small amount of DNA is necessary for mtDNA sequencing (Freeland, 2005). Mitochondrial genome sequence and structure is widely used to provide information on phylogenetic relationships (Ermakov et al., 2006), to determine the genetic structure of populations and to assess patterns of gene flow (Avise, 1994).

Temminck's ground pangolin (*Smutsia temminckii*; Smuts 1832); is a mammal belonging to the family Pholidota (Weber, 1904). This is the only mammal species where the entire body is covered in hard overlapping scales (Dollens, 2010; Herklots, 1937) and they will roll into a ball when threatened to protect the inner softer parts of their bodies that are more vulnerable. These animals are predominantly nocturnal and find refuge in abandoned burrows made by other species, where they hide during the day (Heath, 1992). These animals' only food source is ants and termites (Heath, 1992). They generally use their sharp claws to break open the termite nests (Herklots, 1937) and use their long proboscis tongue to remove termites from their tunnels (Heath, 1992). There are eight extant pangolin species, four of these occur in Asia and four in Africa (Herklots, 1937). All pangolin species are considered critically endangered, vulnerable or endangered by the International Union for the conservation of Nature (IUCN, 2014). The four African species are further divided into two arboreal and two ground-dwelling species. Temminck's ground pangolin (*S. temminckii*) has the largest distribution range of all four African species, ranging from southern Africa all the way to north-east Chad (Heath, 1992). To date, the only complete

*Abbreviations:* %, percentage; µl, microlitre; °C, degree Celsius; A, adenine; AIC, Akaike Information Criterion; BI, Bayesian inferences; bp, base pair; C, cytosine; $ddH_2O$, double distilled water; DNA, deoxyribonucleic acid; G, guanine; G, gamma parameter; GTR + I + G, General Time Reversal model with invariant sites and Gamma distribution; kb, kilobase pair; M, mole; mg, milligram; ML, maximum likelihood; mtDNA, mitochondrial DNA; Mya, millions of years ago; NCBI, National Centre for Biotechnology Information; NJ, Neighbour-Joining; NRF, National Research Foundation; NZG, National Zoological Gardens of South Africa; PCR, Polymerase Chain Reaction; R, transition:transversion ratio; rRNA, ribosomal RNA; RNA, ribonucleic acid; sec, seconds; T, thymine; tRNA, transfer RNA.

* Corresponding author.
*E-mail address:* zdutoit@gmail.com (Z. Du Toit).

http://dx.doi.org/10.1016/j.gene.2014.08.040
0378-1119/

mitochondrial genome sequences available in the order Pholidota (Weber, 1904), is that of the Chinese pangolin (*Manis pentadactyla*; Linnaeus, 1758) (Qin et al., 2012) and the black-bellied tree pangolin (*Phataginus tetradactyla*; Linnaeus, 1766) (Arnason et al., 2002). Considering that full mtDNA sequencing is useful for obtaining more reliable phylogenetic data (Freeland, 2005; Roe et al., 1985), the current study aimed to sequence the whole mtDNA genome of Temminck's ground pangolin. We discuss results with respect to genome structure, gene arrangement, nucleotide composition, and codon usage. Further, we report on phylogenetic analysis and molecular dating of Temminck's ground pangolin with the two pangolin species mentioned above in order to describe the phylogenetic relationships within the order Pholidota. We also describe the phylogenetic relationships between the order Pholidota with representatives from other ant-eating mammals from the order Eulipotyphla and Marsupialia as well as the closest related order Carnivora that has been identified previously in phylogenetic studies to be closely related to the order Pholidota (Arnason et al., 2002; Murphy et al., 2007). The out-groups used during the phylogenetic analysis were representatives from the order Monotremata.

## 2. Materials and Methods

### 2.1. Sample Collection and DNA Extraction

The specimen used to sequence the whole mtDNA of Temminck's ground pangolin (*S. temminckii*) originated from a wild deceased individual, collected by the African Pangolin Working Group (APWG) in the Kalahari region of South Africa. The cause of death was unknown, but suspected to be a road killing accident. A total of three muscle tissue samples were collected from various parts of the pangolin's body. Collected samples were placed in absolute ethanol for preservation and were stored at −20 °C until further usage.

DNA extraction was performed by using the ZR Genomic DNA ™–Tissue MiniPrep Kit (Zymo Research Corporation) using the protocol for solid tissue and following the manufacturers' protocol. The quantity of the extracted DNA was determined by using a NanoDrop ND-1000 Spectrophotometer and the extracted DNA was stored at −20 °C.

### 2.2. PCR Amplification and Sequencing of Full mtDNA Sequence

A total of fifteen primer pairs were designed based on the 16,571 bp reads identified in a previous study on the black-bellied tree pangolin (Arnason et al., 2002), to amplify fragments of approximately 1000 bp in length. Sequences obtained in this study were aligned to the full mtDNA sequence of *P. tetradactyla* in order to identify missing gaps. The missing regions were then filled using the primer walking method with species-specific markers, designed from the sequences obtained from Temminck's ground pangolin. Polymerase Chain Reaction (PCR) was performed using 9.5 μl $ddH_2O$, 12.5 μl 2× Dream Taq™ Mastermix, 1 μl of each primer pair and 1 μl DNA to obtain a PCR reaction of 25 μl. The thermal cycling was conducted as follows: initial denaturation at 95 °C for 5 min, 45 cycles of denaturation at 95 °C for 30 s, annealing at 50–55 °C for 30 s, extension at 72 °C for 45 s, followed by final extension at 72 °C for 7 min. Resulting amplicons were inspected on 2% agarose gels followed by purification using Exosap (Thermo Scientific, Lithuania). Purified templates were sequenced using the Big Dye V3.1 Terminator Kit (Applied Biosystems, Foster City, CA), used according to the manufacturer's instructions and run with the ABI 3130 genetic analyser (Applied Biosystems, Foster City, CA). The ZR DNA Sequencing Clean-up™ Kit (Zymo Research Corporation) was used to purify the sequences and remove excess products (BigDye and buffer) prior to genetic analysis.

### 2.3. Sequence Assembly and Phylogenetic Analysis

Sequences were viewed and edited using the Chromas programme embedded in MEGA v5.2 (Tamura et al., 2011). A sequence blast was done on the National Centre for Biotechnology Information (NCBI) website (www.ncbi.nm.nih.gov/blast) to verify sequence identity. Assemblies, mapping and primer design were performed in CLC Bio Genomics Work Bench 5.0 (CLC Bio, Aarhus, Denmark) to obtain an overlapping view of the full mtDNA sequence from Temminck's ground pangolin in relation to the full mtDNA of *P. tetradactyla* (AJ421454; Arnason et al., 2002) and *M. pentadactyla* (NC016008; Qin et al., 2012) for comparison and verification (Table 1). Final alignment of the whole mtDNA of Temminck's ground pangolin with that of the other species mentioned in Table 2 was performed using MAFFT 7 (Standley, 2013).

Phylogenetic analysis was performed using MrBayes v3.1 (Ronquist et al., 2011) to establish relationships based on Bayesian inferences (BI), while a Maximum Likelihood (ML) approach was implemented in PhyML3.1 (Guindon et al., 2010). MEGA v5.2 software was used to infer a Neighbour-Joining (NJ) tree while BEAST v1.7 (Drummond and Rambaut, 2007) software was used to infer a time-measured phylogeny and obtain estimates of divergence times for this dataset. The Akaike Information Criterion (AIC) in jModeltest v2.1.3 (Darriba et al., 2012) was employed to determine the best fit model of sequence evolution. The same model was selected for BI, ML, NJ and BEAST analysis as it proved to be the best fit model calculated for each analyses. Nodal support for the NJ tree was evaluated through 10,000 non-parametric bootstrap replications, while ML analyses were carried out with 1000 bootstrap replications. The BI and BEAST calculations were run over two million generations, after which 25% of the trees were discarded as burn-in. The divergence time used to calibrate the BEAST tree was estimated by BEAST during the final run and the out-groups were both the orders Monotremata and Marsupialia. Tracer v1.5 (Drummond and Rambaut, 2007) software was used to asses trace files generated by MrBayes v3.2 and BEAST v1.7, in order to assess whether mixing was achieved and to choose a suitable percentage burn-in.

## 3. Results

### 3.1. Whole Mitochondrial DNA (mtDNA) Comparison

The mtDNA genome of *S. temminckii* consists of 16,559 bp (Fig. 1). Compared to *P. tetradactyla* and *M. pentadactyla* which are 16,571 bp and 16,578 bp in length, respectively. As illustrated in Table 1, the mtDNA genome of *S. temminckii* varied from *P. tetradactyla* and *M. pentadactyla* in terms of gene region size at several genes. Two

**Table 1**
List of mtDNA gene region sizes located in *S. temminckii*, *P. tetradactyla* and *M. pentadactyla*.

| Gene regions | *Smutsia temminckii* | *Phataginus tetradactyla* (Arnason et al., 2002) | *Manis pentadactyla* (Qin et al., 2012) |
|---|---|---|---|
| Entire genome (bp) | 16,559 | 16,571 | 16,578 |
| 12S rRNA | 957 | 956 | 960 |
| 16S rRNA | 1553 | 1553 | 1570 |
| ND1 | 955 | 956 | 955 |
| ND2 | 1041 | 1043 | 1038 |
| COX1 | 1549 | 1553 | 1550 |
| COX2 | 683 | 683 | 683 |
| ATP8 | 200 | 206 | 200 |
| ATP6 | 680 | 680 | 680 |
| COX3 | 783 | 783 | 783 |
| ND3 | 345 | 345 | 345 |
| ND4I | 296 | 296 | 296 |
| ND4 | 1377 | 1377 | 1377 |
| ND5 | 1820 | 1826 | 1820 |
| ND6 | 527 | 527 | 524 |
| Cyt*b* | 1139 | 1134 | 1134 |
| D-loop | 1155 | 1164 | 1164 |