Large-scale phylogeny of chameleons suggests African origins and Eocene diversification

Krystal A. Tolley1,4, Ted M. Townsend2 and Miguel Vences3

1 Applied Biodiversity Research Division, South African National Biodiversity Institute, Claremont 7735, Cape Town, South Africa
2 Department of Biology, San Diego State University, San Diego, CA 92182, USA
3 Zoological Institute, Technical University of Braunschweig, Mendelssohnstrasse 4, 38106 Braunschweig, Germany
4 Department of Botany and Zoology, University of Stellenbosch, Matieland 7602, South Africa

Oceanic dispersal has emerged as an important factor contributing to biogeographic patterns in numerous taxa. Chameleons are a clear example of this, as they are primarily found in Africa and Madagascar, but the age of the family is post-Gondwanan break-up. A Malagasy origin for the family has been suggested, yet this hypothesis has not been tested using modern biogeographic methods with a dated phylogeny. To examine competing hypotheses of African and Malagasy origins, we generated a dated phylogeny using between six and 13 genetic markers, for up to 174 taxa representing greater than 90 per cent of all named species. Using three different ancestral-state reconstruction methods (Bayesian and likelihood approaches), we show that the family most probably originated in Africa, with two separate oceanic dispersals to Madagascar during the Palaeocene and the Oligocene, when prevailing oceanic currents would have favoured eastward dispersal. Diversification of genus-level clades took place in the Eocene, and species-level diversification occurred primarily in the Oligocene. Plio-Pleistocene speciation is rare, resulting in a phylogeny dominated by palaeo-endemic species. We suggest that contraction and fragmentation of the Pan-African forest coupled to an increase in open habitats (savannah, grassland, heathland), since the Oligocene played a key role in diversification of this group through vicariance.

1. Introduction

Most lizard families are of ancient origin, with diversification between families occurring in the Jurassic when arid conditions prevailed globally [1–3]. However, chameleons (family Chamaeleonidae) are comparative newcomers, having diverged as recently as the Mid- to Late-Cretaceous [4]. The overwhelming majority of extant chameleon species are found in Africa and Madagascar, both of which are fragments of the ancient super-continent Gondwana. A few species presumably of relatively recent origin are distributed on the southern fringes of Europe and Asia, and some others occur on islands in the Indian Ocean (e.g. the Comoros and Seychelles archipelagos, and Socotra). The sister group to chameleons (Agamidae) is completely absent from Madagascar, Comoros and Seychelles, although it is distributed across most other East Gondwanan and Laurasian landmasses, which is a pattern common to multiple faunal groups, particularly within mammals and reptiles [5]. The fossil record suggests that the probable common ancestors of chameleons and agamids were distributed across Laurasia in the Mesozoic period [6–9]. Chameleons themselves are not abundant in the fossil record (and stem chameleons are completely absent), but they do have a presence in Africa and Europe in the Miocene, and on Madagascar in the Holocene [10–13].

Molecular divergence estimates strongly support a role for oceanic dispersal in shaping historical biogeographic patterns of chameleons, as the earliest divergence events between African and Malagasy species are estimated at 47–90 Myr ago [14,15]. This is substantially later than the fragmentation event.
of the Gondwanan supercontinent that has kept these landmasses separated by ca 400–1200 km of open ocean for the last 120 Myr (million years) [16–18]. Although it is therefore evident that oceanic dispersal has played a pivotal role in the current distribution pattern of chameleons, the geographical origin of the family and the direction of dispersal are still uncertain. The hypothesis that chameleons originated in Madagascar and later dispersed to Africa where they subsequently radiated [14] is oft-cited [18–21], despite some conspicuous inconsistencies in that scenario. Although multiple major chameleon clades are found on Madagascar, the chameleon sister-group Agamidae does not occur there at present, nor is it in the Malagasy fossil record. A Malagasy origin for crown-group chameleons would require either (i) that agamids were present on Madagascar in the Late Cretaceous (at the latest), and subsequently became extinct, or (ii) that the original agamid-chameleon split occurred in Africa, and then the common ancestor to crown chameleons dispersed from Africa to Madagascar in the Late Cretaceous, but became extinct in Africa. However, given the occurrence of extinct and extant chameleons and agamids in Africa, and high potential for their interchange between Africa and Laurasia throughout the Cretaceous [5,22], an equally plausible scenario is that of an ancestral chamaeleonid lineage in Africa that both dispersed to Madagascar and also radiated within Africa. Previous biogeographic scenarios derived from mitochondrial markers and rough molecular dating suggested that chameleons dispersed over water between 47 and 90 Myr ago [14,15]. Dispersal would therefore have occurred when Africa and Madagascar were located about 10° more southerly. However, according to explicit palaeogeographic reconstructions, prevailing oceanic currents promoted dispersal from Africa to Madagascar, and not in the inverse direction, at that time [23,24]. Direct observations of long-distance oceanic dispersal demonstrate that direction of currents may be key to successful dispersal among terrestrial vertebrates [25–28]. Given these considerations, an African origin for chameleons is not only plausible, but is also consistent with the fossil record, present day distributions and oceanographic processes. Indeed, dated phylogenetic analyses using multiple markers and complete generic-level taxon sampling strongly suggest an Eocene out-of-Africa dispersal of one chameleon genus to the Seychelles [29], rather than the out-of-Madagascar dispersal previously proposed [14].

The initial divergence of Chamaeleonidae into the Malagasy genus Brookesia and a clade composed of all other genera is supported in most molecular phylogenetic analyses to date [15,29,30]. However, relationships among the remaining chameleon genera still lack a well-supported phylogenetic hypothesis, possibly owing to a lack of extensive taxon sampling and inclusion of too few markers, or possibly because a rapid initial radiation caused a ‘true’ polytomy [14,15,29–31]. Analyses with extensive taxon sampling are limited to single genera [30,32–35], and these have usually recovered Oligocene/Miocene origins for genera. This in turn suggests the possibility of a collective explanation for generic-level diversification, i.e. vicariance through the break-up of Pan-African forests which began in the Oligocene [36,37].

The objective of this study was to investigate competing hypotheses of Malagasy and African origins for the family Chamaeleonidae by examining phylogenetic relationships and the timing of diversification events in the light of contemporaneous and palaeo-oceanic currents. We postulated that oceanic dispersal out-of-Africa should involve clades that diverged prior to the Oligocene (ca 35 Ma), when currents favoured easterly dispersal across the Mozambique Channel [23,24]. To examine this hypothesis, multiple nuclear and mitochondrial markers were used to generate a fossil-calibrated dated phylogeny, incorporating greater than 90 per cent of chameleon species. Using this dated phylogeny, ancestral areas and habitats were then inferred and examined for fit with hypothesized biogeographic scenarios.

2. Material and methods
Phylogenetic analyses were conducted using two datasets. Given the historical difficulty in resolving deeper relationships within chameleons, we assembled a dataset with reduced-taxon sampling but containing data from 13 nuclear and mitochondrial markers, for a total of 14 089 base pairs (see the electronic supplementary material, tables S1 and S2). Twenty-eight out-group squamate reptiles and Sphenodon were included, plus 25 chameleon taxa spanning the deepest splits within all major clades as identified in previous phylogenetic analyses [4] and in our full-taxon analyses (see below). This dataset included three mitochondrial markers (16S rRNA: 16S; NADH-dehydrogenase, subunits 2 and 4: ND2 and ND4) and 10 nuclear markers (recombination-activating gene 1: RAG1; oocyte maturation factor: CMOS; prolactin receptor: PRLR; A kinase anchor protein 9: AKAP9; BTB and CNC homology 1, basic leucine zipper transcription factor 1: BACH1; BTB and CNC homology 1, basic leucine zipper transcription factor 2: BACH2; brain-derived neurotrophic factor: BDNF; mutS protein 6: MSH6; natural killer-tumour recognition protein: NKTR; and REV3-like, polymerase (DNA directed), zeta, catalytic subunit: REV3L). The second dataset included greater taxon sampling, with 174 chameleons representing greater than 90 per cent of described species, plus the same out-group taxa as above, and used three mitochondrial and three nuclear markers totalling 5054 base pairs (16S, ND2, ND4, RAG1, CMOS, PRLR).

For both datasets, MrBayes v. 3.1.2 [38] was used to estimate time-free (unrooted) topologies in a Bayesian framework, BEAST v. 1.6.2 [39] was used to simultaneously estimate phylogeny and divergence times under a Bayesian relaxed-clock, and RAxML HPC v. 7.2.8 [40,41] was used to produce maximum-likelihood tree topologies. Squamate monophyly was constrained in all analyses. Additional details of these analyses are provided in the electronic supplementary material. Sequences were accessioned in EMBL (accession nos. HF570393–HF570918, full details in the electronic supplementary material, table S1). Alignments and tree files were deposited in Dryad (accession no. doi:10.5061/dryad.11350). Analyses were performed on the CIPRES Science Gateway (www.phylo.org/sub_sections/portal/).

To test the robustness of our biogeographic inferences, multiple runs of the 6-gene, full-taxon analysis were made for each phylogenetic method, under three constraint scenarios: (i) completely unconstrained, (ii) monophyly of Malagasy large chameleons (i.e. Calumma and Furcifer) constrained and, (iii) the genus Rhampholeon constrained to be the sister taxon to Calumma + Furcifer. These constraints were imposed because the unconstrained analysis of the 6-gene dataset lacked support at some deep nodes, whereas the 13-gene BEAST analysis showed significant support at these same nodes (see §3 and the electronic supplementary material, methods).

Divergence-time estimates were constrained by placing age calibrations on a total of eight nodes in the tree corresponding to the oldest known fossils of various lepidosaurian taxa (seven nodes within squamates, as well as the lepidosaurian root; see the electronic supplementary material, table S3). For each calibration, we used BEAUti (part of the BEAST package) to construct a translated-lognormal (TL) distribution (i.e.
‘lognormally distributed’, with an offset from zero roughly equal to the age of the fossil, [42–44]). Specific values of the zero-offset, mean and standard deviation for each TL distribution were chosen such that the more recent bound of the 95% highest posterior density (HPD) was placed at a point approximately 1 per cent more recent than the estimated fossil age, and the bulk of the probability was placed near the estimated fossil age. We left (somewhat arbitrarily) long probability tails for the soft maxima of node ages. To examine whether heterogeneous rates of molecular evolution would have an effect on our dating estimates, analyses were run with the dataset partitioned (i) by locus, (ii) by codon position and (iii) by codon position with first and second positions combined + third position.

(a) Inference of ancestral area and habitat
Multiple ancestral area analyses were carried out to infer the geographical origins of the Chamaeleonidae, and to establish direction of oceanic dispersal using (i) dispersal-extinction-cladogenesis (DEC) analysis in LAGRANGE v. 20110117 [45,46], (ii) likelihood optimization in MESQUITE v. 2.74 [47] and (iii) Bayesian binary ‘Markov-chain-Monte-Carlo’ (BBM) in RASP v. 2.0 [38,48,49] using the ultrametric trees generated in BEAST from the complete taxonomy. To ensure that topological differences between the analyses would not influence the interpretation of biogeographic patterns, the biogeographic analyses were run on each of the constraint scenarios: completely unconstrained, monophyly of Malagasy large chameleons (i.e. Calumma + Furcifer) constrained and the genus Rhampholeon constrained to be the sister taxon to Calumma + Furcifer. Taxa were coded for area based on present day occurrence (Africa, Madagascar, Europe, Asia, Seychelles, Socotra). Because we only examined large-scale geographical patterns, taxa which occur on coastal Atlantic islands in the Gulf of Guinea or islands proximal to Madagascar (including the Comoros) were coded as Africa or Madagascar, respectively.

For the DEC analysis, temporal constraints were placed on dispersal (see the electronic supplementary material, methods) which took into account the shifts in the prevailing direction of oceanic currents. Out-group taxa were restricted to agamids, and these were coded as occurring in Africa and Asia. The online ‘LAGRANGE configurator’ (http://www.reelab.net/lagrange/configurator/index) was used to produce the Python v. 2.6 script that was run locally [46]. Likelihood optimization in MESQUITE used the same data matrix as the DEC analysis using the Markov k-state, one-parameter model (all states equally probable).

The BBM analysis was carried out with the same in-group coding, but also allowed us to examine the effect of coding at the root by applying six different distribution scenarios for the out-group: (i) African, (ii) Asian, (iii) African and Asian, (iv) Laurasian (Europe and Asia), (v) widespread (occur in all areas), and (vi) null (occur in none of the areas). These analyses were run under a variable model (P8I + G) for 2,000,000 generations, with 10 chains, a sampling frequency of 100 generations and a burn-in of 10 per cent.

Ancestral habitats were reconstructed using likelihood optimization in MESQUITE 2.74 [47], using the ultrametric tree generated in BEAST from the 6-gene analysis, under all three differing constraint scenarios. Because there is a functional basis for micro-habitat choice in all species of chameleons studied to date [50,51], each taxon was first coded for micro-habitat (substrate) preference: terrestrial (forest or desert), arboreal-small perch (as found in grassland, montane heath and small bushes), arboreal-large perch (as found in closed canopy forest and open canopy woodland) and mixed (from both small and large perch habitats). Second, the analysis was run to optimize macro-habitat type on the phylogeny: closed canopy forest, open canopy woodland, heathland, grassland, desert and mixed (species recorded in several vegetation types). Coding was based on multiple sources from the literature [52–54].

3. Results

(a) Phylogeny and dating
The 13-gene, reduced-taxon analyses (MrBayes, BEAST and RAxML) produced trees featuring monophyly of the large Malagasy species (i.e. Calumma + Furcifer; electronic supplementary material, figures S1–S3). In the BEAST analysis, posterior probability (PP) support for this finding was 1.0, but the node was not strongly supported in the MrBayes or RAxML analyses. As taxon sampling in this 13-gene dataset was designed to span the deepest split within each major clade (mostly genera), we consider it relevant to the question of intergeneric relationships. We therefore used these results to justify constraining monophyly of the large Malagasy species in the main 6-gene, full-taxon BEAST analysis (see the electronic supplementary material, methods), which we consider by virtue of its much larger taxon sampling (174 versus 25 chameleon species) to provide the most reliable estimates of branch lengths (and thus divergence times). A less intuitive finding, because it has never been found for any other analyses for this family and cannot be readily explained biogeographically, was that the 13-gene analyses also recovered the genus Rhampholeon as the sister taxon to the Calumma/Furcifer clade (PP = 0.95 in the BEAST analysis, but node not present in other analyses).

The 6-gene unrooted Bayesian (MrBayes) and maximum-likelihood (RAxML) analyses each produced more poorly resolved/supported topologies that differed from the BEAST topology at some deeper nodes, although none of these conflicting nodes were well supported in either analysis (see the electronic supplementary material, figures S5 and S6). All analyses strongly supported the Malagasy leaf chameleon genus Brookesia as the sister group to a large clade comprising all other chameleon genera, consistent with other recent studies that have incorporated nuclear DNA data [29,30,55].

The 6-gene BEAST relaxed-clock analysis performed with Calumma + Furcifer constrained to monophyly produced a topology with significant support for all clades at the genus level as well as for some deep nodes (figure 1). It is this tree that will form the main basis for our discussions of phylogenetic relationships and biogeography, but it should be noted that our biogeographic inferences were unchanged when we considered topologies obtained under our two alternative constraint scenarios (i.e. no topological constraints, and Rhampholeon + Calumma + Furcifer constrained to monophyly). The dated phylogeny (figure 1) suggests that the initial split between Brookesia and all other clades occurred approximately 65 (+10) Ma, followed by the divergence of the Rieppeleon/Archaeis clade (52 ± 8 Ma). The remaining genera diverged in rapid succession 40–50 Myr ago. Heterogeneous rates of molecular evolution among sites did not appear to affect on our dating estimates, as we obtained similar estimates of divergence with the dataset partitioned (i) by locus, (ii) by codon position, and (iii) by codon position with first and second positions combined + third position (see the electronic supplementary material, table S4).

(b) Ancestral areas and habitats
The DEC analysis, likelihood optimization and BBM analyses all showed an unequivocal African origin for Chamaeleonidae
Bayesian binary analyses showed generally high probability that the ancestral area for chameleons is Africa, regardless of which coding scenario was used for the out-group. The exception was that if the out-group was constrained to Asia or Laurasia, the result was equivocal. This analysis thus suggests three separate dispersals out of Africa: a dispersal ca 65 (±8) Ma to Madagascar (Brookesia), followed by a dispersal ca 47
Transitions to open habitats such as grassland and heathland and transitions to other types only occur late in the history of chameleons (figure 1). Not considered here is the dispersal of two species of Furcifer westward to the Comoros, which doubtless occurred out of Madagascar in the Miocene when sea currents were conducive to rafting in this direction.

An ancestral terrestrial habitat received unequivocal support in the likelihood optimization, regardless of topology. There were two transitions to arboreal habitats with large perches (i.e. either forest or woodland): the first ca 50 (+7) Ma for most of the genera, and the second ca 34 (+8) Ma with the dispersal of Archaius to the Seychelles (see figure 2a and electronic supplementary material, table S9). Transition out of habitats with large perches to habitats with small perches (i.e. grassland and heathland) is invariably more recent, typically near the start of the Pliocene. For the broader scale biome-level analysis, closed canopy forest shows the greatest likelihood of being the ancestral habitat type (see figure 2b and electronic supplementary material, table S10). Transitions to other types only occur late in the history of chameleons, in particular with the shift by multiple lineages of Chamaeleo and Furcifer to open canopy (ca Early Miocene). Transitions to open habitats such as grassland and heathland occur primarily since the Pliocene in Brookesia and Trioceros.

4. Discussion
Chamaeleonidae originated in the Late Cretaceous, around 90 Ma (figure 1), post-dating the separation of Africa from Madagascar/India (ca 120 Ma). This agrees with previous higher-level phylogenies that show vicariance owing to Gondwanan break-up was not a factor influencing chameleon diversification [3,4]. The initial divergence within the family, ca 65 (+10) Ma, divides chameleons into one large clade containing 10 of the 11 extant genera (ca 155 species), and another comprising only the Malagasy leaf chameleons, Brookesia (ca 30 species). Thus, Brookesia is not a paraphyletic group as suggested previously [14], but instead is the monophyletic sister clade to all other chameleons [30]. A series of very short, poorly supported branches are found deep in the tree, despite near complete taxon sampling and a considerable number of nuclear and mitochondrial markers applied to this analysis. Our results therefore suggest rapid origins of the genus-level clades within in the Eocene followed by cladogenesis within genera primarily within the Oligocene and the Miocene.

The most likely biogeographic scenario points to an African origin for chameleons, with two separate dispersals to Madagascar, the first by the ancestor to the Brookesia clade ca 65 Ma, and the second ca 47 Ma in the Mid-Eocene by the common ancestor of Furcifer/Calumma. A Malagasy origin for the family is not favoured in any of the analyses, even when alternative topologies are considered (see the electronic supplementary material, table S7). The 95% HPD associated with the divergence dates of these dispersal events overlap with the timing for the prevailing west to east (Africa to Madagascar) palaeo-currents [23], further supporting an African origin and two separate dispersals to Madagascar.

Extant European and Asian chameleons are all relatively young species originating in the Late Miocene or Pliocene, and all are nested within the African genus Chamaeleo. This supports the suggestion that chameleons only recently occupied Laurasian landmasses [14], in accordance with an African origin for chameleons. Of these peripherally distributed species, Chamaeleo monachus from Socotra Island is the oldest (ca 20 ± 5 Myr). Although sea floor spreading in the Western portion of the Gulf of Aden is recent, Socotra is located substantially eastward, where rifting was complete ca 20 Ma [56–59]. The timing of divergence for C. monachus overlaps with the timing of separation between Socotra and Africa, and therefore, neither the isolation owing to vicariance nor oceanic dispersal [60] can be ruled out. The remaining three Asian taxa (Chamaeleo zeylanicus, Chamaeleo arbicus and Chamaeleo calyptratus) probably dispersed once from Africa to Arabia/Asia (ca 13 Ma), and thereafter diversified in situ (ca 6–8 Myr ago).

The oldest extant chameleon clades comprise primarily terrestrial, closed canopy forest chameleons (Brookesia, Rhampholeon and Rieppeleon according to our favoured tree). By contrast, clades that correspond to all other extant genera diverged in the Early to Mid-Eocene (ca 49–40 Myr ago), and these contain primarily arboreal species (figure 2). The latter could point to an adaptive radiation, owing to increased ecological opportunity [61], associated with expansion of forest environments in Africa and Madagascar connected to greenhouse Earth conditions during the Early Eocene Climatic Optimum ca 51–53 Myr ago [62–67]. Post-Palaecocene Madagascan left a subtropical arid belt as it drifted northward towards the equator, which led to a significant increase in moisture levels on the island. Furthermore, owing to the opening of the Indian Ocean and the movement of India northward to Asia, trade winds started hitting the island from the east, leading to the full extension of the eastern rainforest in the Oligocene [37]. The expansion of forest environments would have allowed chameleons to take full advantage of the arboreal niche in both regions (figure 2).

Within genera, species-level divergences occurred primarily within the last 35 Myr, since the start of the Oligocene, during which time Icehouse Earth conditions prevailed [66,68]. Aside from short episodic climatic optima (ca 15 and 24 Myr), the climate has remained comparatively cool, which has induced dry conditions, resulting in fragmentation and reduction of forests [36]. Given that most chameleon species (both terrestrial and arboreal) are closely associated with forested/wooded habitats [32], fragmentation of forests could have led to vicariance causing the species formation observed within most genera during this period. Some species that use large perches were also apparently able to make the transition from closed canopy forests to open woodlands (savannah, dry forests of Madagascar) allowing colonization of that niche as closed canopy forests diminished. Transition out of habitats with large perches into habitats with fine structure (i.e. grassland and heathland) has only occurred for a few clades [32,34], and primarily since the Pliocene (figure 2) when these habitats became present in the African landscape [69–72].

Oceanic dispersal has become fundamental to many biogeographic and diversification hypotheses [28,73–78] even for taxa which were traditionally considered poor dispersers, such as amphibians [79–81]. Consideration of oceanic palaeo-currents and the timing of diversification events are both necessary components of many biogeographic scenarios. Ancestral area reconstruction for the Chamaeleonidae shows an African origin, consistent with the physical processes that would have promoted oceanic dispersal during the Eocene. Initial diversification from terrestrial to arboreal lifestyle
Figure 2. (a) Ancestral habitat reconstruction using the 6-gene constrained phylogeny. Species are colour-coded according to ancestral habitat structure (terrestrial, brown; arboreal thick perch, green; arboreal fine perch, purple; mixed, red) for unequivocal character states (i.e. $\geq 0.80$ likelihood). Species with equivocal states are coloured grey, and those where habitat information was not available are coded with thin black lines. (b) Broad-scale ancestral habitat reconstruction using the 6-gene constrained phylogeny. Species are colour-coded according to ancestral habitat (closed canopy forest, green; open canopy woodland, orange; heathland, purple; grassland, light blue; desert, black; mixed, grey) if the character state was unequivocal (i.e. $\geq 0.80$ likelihood). Species with equivocal states are coloured grey, and those where habitat information was not available are coded with thin black lines. Supported nodes denoted as in figure 1.
corresponds with Greenhouse Earth conditions and is probably the result of ecological opportunity.

We thank the Port Elizabeth Museum (Bayworld), California Academy of Sciences, Trento Science Museum, Zoologisches Forschungsmuseum, A. Koenig, Wolfgang Böhme, Frank Glaw, John Measey, Michele Menegon, Fanomezana Ratsoavina, Colin Tilbury and David R. Vieites for assistance with sampling. Shelley Edwards, Zoe Davids, Gabriele Keunecke and Meike Kondermann provided assistance in the laboratory, and we thank two reviewers for providing very insightful comments. This study was supported in part by Deutsche Forschungsgemeinschaft grant no. VE247/3–1 to M.V. and the South African National Biodiversity Institute.

References


26. Couvreur TLP, Chatrou LW, Sosef MSM, Richardson JE. 2008 Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain
forest trees. BMC Biol. 6, 54. (doi:10.1186/1741-7007-6-54)


49. Yu Y, Harris AJ, He XJ. 2011 RASP (reconstruct ancestral state in phylogenies) 2.0b. See http://imnh.sce.edu.cn/soft/blog/RASP.


53. Tilbury CR. 2010 Chameleons of Africa—an Atlas. Including the chameleons of Europe, the Middle East and Asia. Frankfurt am Main, Germany: Edition Chimaira.


