Article type: Original Article
Community assembly and diversification in a species-rich radiation of island weevils (Coleoptera: Cratopini)

Running title: Community assembly in Mascarene weevils.

James J. N. Kitson[1,6], Ben H. Warren[2], Christophe Thébaud[3], Dominique Strasberg[4] and Brent C. Emerson[5,6]

[1] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom.
[2] Institut de Systématique, Evolution, Biodiversité, UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, CP 51, 57 Rue Cuvier, 75231 PARIS Cedex 05, France.
[3] Laboratoire Evolution & Diversité Biologique, UMR 5174 CNRS-Université Paul Sabatier-ENFA-IRD, 31062 Toulouse Cedex 9, France.
[4] UMR PVBMT, Université de La Réunion - Faculté des Sciences et Technologies, CS 92003, 97744 Saint-Denis, La Réunion, France.
[5] Island Ecology and Evolution Research Group, IPNA-CSIC , C/Astrofísico Francisco Sánchez 3, 38206 La Laguna, Santa Cruz de Tenerife, Spain.
[6] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

Correspondence author: James Kitson (James.Kitson@newcastle.ac.uk)

ABSTRACT

Aim To test a prediction derived from island biogeographic theory that *in situ* speciation should make an increasingly important contribution to community assembly as islands age. This prediction is tested on estimated biogeographic histories from Mauritius (approximately 9 Ma) and Reunion (approximately 5 Ma). We additionally investigate the evolutionary dynamics of insect flight loss, since the loss of flight in island lineages can influence patterns of diversification.
Location Mascarene Islands; Southwest Indian Ocean.
Taxon Weevils.
Methods Up to five individuals of each taxonomically described species sampled within each sampling site were sequenced for the mitochondrial gene Cytochrome Oxidase II to delimit operational taxonomic units (OTUs). OTUs were further sequenced for the nuclear genes Arginine Kinase, Histone 3, and ribosomal 28s, to reconstruct the phylogenetic history of the group. Timings of colonisation and in situ speciation events were estimated with BEAST2.
Results Our results support the hypothesis that present-day species richness on the older island of Mauritius is largely the result of *in situ* speciation, with few colonisation events, of which all but the most basal are recent. In contrast, Reunion presents a more uniform temporal spectrum of colonisation times. Flight loss has evolved convergently at least five times, and speciation events associated with flight loss are significantly younger than speciation events that have not resulted in flight loss.
Main conclusions Patterns of community assembly on the islands of Mauritius and Reunion fit a model where the addition of new species and species turnover is increasingly dominated by *in situ* speciation as an island community matures. Repeated flight loss indicates selection for flightlessness, with the young age of flightless lineages suggesting higher extinction rates over longer evolutionary time scales and little influence on present-day species richness.

Key words
Colonisation, speciation, extinction, island, invertebrate, diversification, Mascarene

INTRODUCTION

Island species assemblages result from the accumulation of species through time by colonisation and establishment from outside areas, anagenetic change, *in situ* speciation and extinction. The relative probabilities of these processes are predicted to be functionally related to each other (e.g. MacArthur & Wilson, 1963, 1967; Emerson & Kolm, 2005; Emerson & Oromí, 2005; Emerson & Gillespie, 2008). A time lag before *in situ* speciation begins to contribute to island community assembly can be expected, given that: (i) speciation takes time; (ii) speciation between island and mainland source populations may be needed before *in situ* speciation can occur, and; (iii) an initially low number of established lineages means there are fewer lineages across which speciation can occur. It has been further hypothesised that community assembly on young islands should initially be dominated by colonisation, but that *in situ* speciation should later make an important contribution, giving rise to a dynamic equilibrium of species turnover involving colonisation, extinction and speciation (e.g. Emerson & Oromí, 2005; Emerson & Gillespie, 2008). Under this model one would expect that for islands belonging to the same regional species pool, and with a comparable size and species richness, but of different ages, species origin by *in situ* speciation would be more predominant on older islands, assuming that younger islands had not reached a turnover equilibrium among colonisation, speciation and extinction.

The Indian Ocean islands of Mauritius and Reunion offer a near-ideal opportunity to examine the temporal dynamics of community assembly, as both islands are of a similar size and isolation from source areas of colonisation, but are of markedly different ages (Thébaud et al., 2009; Casquet et al., 2015). The subaerial age of Mauritius is estimated to be approximately 8.9 Ma (million years) (Moore et al., 2011), while the age of Reunion is estimated to be up to 5 Ma (Bonneville et al., 1988). The closest long-term source area for both islands is Madagascar, from which both can be considered to be similarly isolated, being 870 km and 665 km away, respectively. Connectivity of both islands with the Seychelles and India has been promoted by the appearance of stepping stone islands along the Mascarene plateau during cycles of sea-level low stands (Warren et al., 2010).

One of the most species rich invertebrate groups distributed across the islands of the Southwest Indian Ocean is the weevil tribe Cratopini, which is represented by five of the nine recognised genera (Alonso-Zarazaga & Lyal, 1999). Species richness is concentrated on the islands of Mauritius and Reunion, with 75 of the approximately 107 described species of *Cratopus* (abbreviated to *C.*), 12 of the 14 described species of *Cratopopsis* (abbreviated to *Cr.*) and both described species of *Scaevinus*. Cratopine weevils of the Southwest Indian Ocean Islands are unusual in that a large proportion of species are flighted, contrasting with other known large radiations of oceanic island weevils where all species are flightless (e.g. the genus *Laparocerus* in Macaronesia and *Rhyncogonus* in Polynesia). While *Cratopus* species are capable of flight, with the exception of one flightless species (*C. triangularis*) and one flight polymorphic species, *C. murinus* [Kitson et al. (2013)], all *Cratopopsis* and *Scaevinus* species are exclusively flightless.

The equilibrium model of island biogeography (MacArthur & Wilson, 1963, 1967) emphasizes island biotas as being typified by rapid turnover through ongoing colonisation and extinction (Heaney 2007). While the equilibrium model does not explicitly include speciation, it is expected that species turnover will be increasingly dominated by *in situ* speciation as an island community matures (addressed by

MacArthur & Wilson, 1963; e.g. Emerson & Oromí, 2005; Emerson & Gillespie, 2008). It thus follows that colonisation would be expected to contribute more to species richness on the younger island of Reunion, because it is in a younger stage of the community assembly/species turnover spectrum (sensu Whittaker et al., 2008), and less to the more mature island of Mauritius, because it is at a later stage. Here we reconstruct phylogenetic relationships among cratopine weevils of the Mascarene Islands to infer the biogeographic history of the group, and to test the hypothesis that *in situ* speciation contributes proportionally more to species richness on mature islands (Mauritius) than on younger islands (Reunion). We explore this dynamic by distinguishing between divergence events that have added species to an island by colonisation from those that have added species through *in situ* speciation, and by testing the following three predictions: (i) the probability of establishment and persistence of a colonising species decreases as an island matures, (ii) excluding the earliest colonisation event, *in situ* speciation events on a mature island are more likely to be older than colonisation events, reflecting a dynamic dominated more by species turnover within the local species pool rather than colonisation of empty niche space by the regional species pool; and (iii) *in situ* speciation events on young islands are more likely to be of comparable age or even younger than colonisation events, reflecting a dynamic dominated more by colonisation of unoccupied ecological niches by the regional species pool, rather than species turnover within the local species pool.

An important part of community assembly is *in situ* evolution following colonisation, as colonist lineages adapt to their new niche. A recurring feature of this process in many insular insect and vertebrate groups is the loss of flight (Wright et al., 2016) with multiple hypotheses for this trend being proposed such as; reduced temperature at high altitude (Medeiros & Gillespie, 2011), strong winds displacing individuals from restricted environments (Sattler, 1991) and increased fecundity due to rebalancing an organism's energy budget (Roff, 1990). Loss of flight is particularly relevant in the context of community assembly within islands or groups of islands, as it can accelerate allopatric speciation and promote diversification, as shown in carrion beetles (Ikeda et al., 2012). Given that many weevil genera present on archipelagos are flightless (see Williams, 2000; Samuelson, 2003; Machado et al., 2008 for examples), the occurrence of just two flightless species within the otherwise flight capable genus *Cratopus* is surprising. However, the number of flight loss events within the Cratopini of the Southwest Indian Ocean Islands may be underestimated by taxonomy. The monophyly of the genus *Cratopopsis* in particular is defined by morphological characteristics that could be the consequence of convergent evolution through the repeated evolution of flight loss (partially fused elytra and reduced prothoracic anatomy). Thus it is possible that cratopine taxonomy is a poor predictor of the evolutionary history within the tribe. We used our phylogenetic data to re-examine the evolutionary dynamics of flight loss in Mauritius and Reunion Cratopini to assess the frequency and timing of speciation events associated with flight loss, and their potential influence on patterns of community assembly.

METHODS

Beetle sampling
Specimens were collected from the islands of the Southwest Indian Ocean by foliage beating during wet seasons between June 2007 and December 2011. Samples were

placed directly in 99% ethanol and sampling sites were recorded with a handheld GPS unit. Samples were identified to taxonomic species using morphological keys for *Cratopus* and *Cratopopsis* provided by Jaques Poussereau (unpublished work) and the original descriptions of most species (Hustache, 1919; Williams & Cox, 2003 - but see Appendix S1.2 in Supplementary Materials for more references).

Defining operational taxonomic units for the multilocus phylogeny
While there is a rich taxonomic literature for the three Cratopine genera of the Southwest Indian Ocean (see Appendix S1.2), the validity of several Mascarene species has been questioned (Voisin & Poussereau, 2009). To address the taxonomic uncertainty within the group, we used a phylogenetic approach to help define operational taxonomic units (OTUs). It was beyond the scope of this project to sequence all samples for multiple loci so we first selected up to five individuals per species per Mascarene sampling locality for genomic DNA extraction, PCR, and sequencing of the mitochondrial locus Cytochrome Oxidase II (COII). All COII sequences locus were aligned using MAFFT v6.814b and model testing was undertaken to determine the best phylogenetic model in jModeltest2 (Darriba et al., 2012). A maximum likelihood tree was generated using the GTRGAMMA model in RAxML 7.7.1 (Stamatakis et al., 2008) using the RAxML BlackBox (https://embnet.vital-it.ch/raxml-bb/index.php). Operational taxonomic units (OTUs) were then defined as the minimal monophyletic group containing all samples belonging to one or more taxonomic species. If an OTU occurred on multiple islands, then a representative from each island was included in the multilocus phylogeny. Single individuals of species from islands outside the Mascarenes were also included in the final set of OTUs. Full details of the procedure for OTU definition (including PCR conditions and sequencing) can be found in Appendix S1.

Molecular protocols for phylogenetic analyses
To reconstruct phylogenetic relationships among the defined OTUs, individuals were chosen for the amplification and DNA sequencing of three additional loci: Arginine Kinase (ArgK), Histone 3 and the ribosomal gene 28S. Sequences for all primers used can be found in Table S1.6.2 and detailed molecular methods including PCR conditions are available in Appendix S1.6.

All sequences were checked, and in the case of reverse sequencing, consensus sequences generated with Geneious Pro version 5.6. Sequences were aligned using MAFFT v6.814b (Katoh et al., 2002) with the following parameter values: scoring matrix 1PAM/k = 2, Gap open penalty = 1.53, Offset value = 0.123, and then checked by eye. The aligned sequences were tested for saturation using the entropy-based index of substitution saturation as implemented in DAMBE v5.2.78 (Xia et al., 2003). This was performed on two data sets, one comprised of the first and second codon positions and the second comprised of the third codon positions, with the exception of 28S which was not partitioned.

Phylogenetic analyses
Trees for OTUs defined from the mtDNA COII analysis were constructed from individual alignments of each locus using MrBayes 3.2 (Huelsenbeck & Ronquist, 2001). Four MrBayes replicate analyses were performed on each alignment for 20 million generations using eight MCMC chains, discarding 25% of the samples as burnin. Model testing was undertaken to determine the best model for each locus in jModeltest2 (Darriba et al., 2012). The alignments for COII, ArgK and H3 were assigned the GTR+G

substitution model and 28S was assigned the HKY+I+G model, all parameters permitted under these models were estimated. The outputs of all MrBayes analyses were assessed for stationarity and convergence in Tracer v1.5.0 (Rambaut & Drummond, 2007) and AWTY (Nylander et al., 2007), and only runs with ESS scores greater than 200 for all parameters were accepted. Consensus trees were plotted and formatted in R (R Development Core Team, 2017) using the 'ggtree' R package (Yu et al., 2017). Before performing multilocus analyses, congruence between gene trees was assessed by comparing consensus tree for each gene to check for conflicting well supported clades (>0.95 Bayesian posterior probability) present in one partition but not the others as (as in Meseguer et al., 2013). In addition to this, a sensitivity analysis was performed as in Meseguer et al. (2013) to examine the effects of missing data. Four separate concatenated alignments were produced: (1) all-samples; (2) samples with at least two loci; (3) samples with at least three loci; (4) samples which have all four loci. Each concatenated alignment was analysed using the same parameters as above with all substitution model parameters unlinked for each partition.

Divergence time estimation

Timings of colonisation and divergence events were estimated with BEAST2 v2.4.5 (Bouckaert et al., 2014) using all four loci. Each partition was given the substitution model previously determined by jModeltest2 and all substitution models were unlinked to allow each partition to have its own parameters. Ten replicate analyses each of 100 million generations were performed. The COII partition was assigned a relaxed lognormal clock model with a mean rate of 0.0154 substitutions/site/Ma with a standard deviation of 0.06, taken from (Cicconardi et al., 2009). A single nuclear clock, estimated relative to the COII clock, was applied to all nuclear partitions. Stationarity and convergence were assessed as with the individual locus MrBayes trees. A maximum clade credibility (MCC) tree was generated from the BEAST2 output by combining tree files in Logcombiner v2.4.5 (Bouckaert et al., 2014) and then generating the MCC tree in TreeAnnotator v2.4.5 (Bouckaert et al., 2014). An additional *BEAST analysis (Heled & Drummond, 2009) was attempted using BEAST2 with the same parameters as above but with each locus permitted to have a separate unlinked tree topology. This analysis failed to converge even after 1 billion generations and will not be considered further. This is likely due to the use of single samples per OTU for each locus resulting in over-parameterisation of the analysis but it should be noted that the unconverged species MCC tree produced was topologically compatible with the standard BEAST2 analysis.

Biogeographic analyses

For all analyses, the Seychelles were treated as a single island as they have formed one connected landmass at least six times due to changing sea levels over the last 500,000 years (Warren et al., 2010). All other islands were treated as independent entities as they are surrounded by deep water and have never been connected. Ancestral ranges for each node in the multilocus phylogeny were estimated using the BioGeoBEARS package (Matzke, 2013) in R. BioGeoBEARS has frequently been used to choose among variants of the dispersal-extinction-cladogenesis model using AIC. However, Ree and Sanmartín (2018) recently emphasise that the likelihoods of variants are not statistically comparable, casting serious doubt on such approaches. Following Ree and Sanmartín (2018) recommendations, we therefore use biological considerations to make our selection. In our oceanic island system, our focal islands have never been connected to each other or to any other landmass. Furthermore, geographic distances

between islands far exceed the regular dispersal ability of Cratopini. As a consequence, the frequency of dispersal between biogeographic areas is expected to be much lower than in continental contexts, with gene flow between areas being nil or extremely rare. In such a system, range expansion that is strictly anagenetic (i.e. in which the daughter population never diverges from the parent population) is expected to be rare or absent. Rather, we expect speciation and inter-island dispersal (range expansion) to be tightly associated, with almost all dispersal events leading to speciation within a restricted time interval. We therefore selected the DEC+J+X (Ree & Smith, 2008; Matzke, 2014; Van Dam & Matzke, 2016) model for biogeographic reconstruction, which most closely models our island scenario. Notwithstanding some recent critique (Ree & Sanmartín, 2018), we find this model highly feasible for our island scenario, in which long periods of biogeographic stasis are expected to be punctuated by periods in which biogeographic change and cladogenesis are closely linked.

Statistical analysis of node ages

Statistical tests comparing nodal ages were performed to test if: nodes representing colonisations of Mauritius are younger than nodes representing colonisations of Reunion (prediction 1); nodes representing colonisations of Mauritius are younger than nodes representing *in situ* speciation on Mauritius (prediction 2); nodes representing *in situ* speciation events on Reunion are younger than nodes representing colonisation events of Reunion (prediction 3). Nodes leading to flight loss were also tested to evaluate if they are significantly younger than other nodes, as might be expected if there is high evolutionary turnover of flightless lineages. To account for uncertainty in the estimated phylogenetic history and node ages, the above analysis was replicated across 1000 random trees drawn from the post burnin BEAST2 .trees files. For each node in each randomly selected tree, the most probable biogeographic range for that node and the ancestor of that node were extracted from the BioGeoBEARS output along with the median age of each node. These were then used to identify the node types listed above and calculate the maximum possible age for each colonisation or *in situ* speciation event. The mean age of each node was then used in a series of t-tests that were paired to account for overall variation in tree age. R scripts for subsampling of BEAST2 trees and performing the BioGeoBEARS analysis are available as part of our supplementary GitHub repository.

RESULTS

Beetle sampling

A total of 5565 beetles were collected across 19 islands, representing 77 described species, accounting for approximately 70% of the recognised species richness. A complete list of species and islands is provided in Table S2.3 in Appendix S2.

Operational taxonomic units

Subsampling five individuals, where possible, of each taxonomic species sampled within each Mascarene sampling site resulted in a dataset of 915 individuals for the definition of OTUs. A full list of Mascarene sites and GPS coordinates can be found in Table S1.2.1 in Appendix S1. A total of 32 OTUs were identified across the Mascarenes (See Figures S1.5.1 and S1.5.2 in Appendix S1 for OTUs and the taxonomic species they contain). There are an additional 19 species from non-Mascarene islands. Consistent with the

biogeographic expectations presented in our choice of biogeographic model, nearly all OTUs in the Mascarenes and Comoros are endemic to a single island. Only elsewhere (in the Seychelles and Iles Eparses) are OTUs found on multiple islands. There were also two species that would not amplify for COII and one outgroup (*Polyclaeis equestris*), making a total of 58 OTUs in the final dataset.

DNA sequencing for the OTU phylogeny

A full summary of sequencing success for each locus can be found in Table S2.4 in Appendix S2 while Table S2.5 in Appendix S2 contains the corrected and uncorrected genetic distances among individuals for each locus. The sequencing success for each OTU partitioned by locus is provided in Table S2.6 in Appendix S2. All sequences are available on GenBank (accession numbers will be provided upon acceptance). The results of Xia's test for saturation for each locus divided into first and second codon position vs. third codon position (except 28S which was not partitioned) is provided in Table S2.7 in Appendix S2. Results indicate that sequences are not saturated with the possible exception of the third codon position for COII which had a test statistic that was not significantly less than the critical value.

Phylogenetic analyses of individual genes

Among the four partitions, the COII (Fig S2.3 in Appendix S2) and ArgK (Fig. S2.4 in Appendix S2) trees are better supported, with ArgK providing greater resolution among basal nodes. The Histone 3 (Fig. S2.5 in Appendix S2) and 28S (Fig. S2.6 in Appendix S2) trees are less well supported. Histone 3 contains much missing data (39.0% of the alignment is missing or ambiguous compared to 6.8% for COII, no missing data for ArgK and 2.1% for 28S). The lack of support in the 28S tree is consistent with the low number of variable sites (90.2% of aligned sites are identical compared to 59.4% for COII, 65% for ArgK and 64.3% for Histone 3). The individual gene trees are in broad agreement with one notable exception. *Cratopopsis bistigma* from Reunion and the Mauritius *Cratopopsis* OTU are recovered as sister OTUs with very high support for all nuclear loci. However *Cr. bistigma* is recovered with very high support as a sister species to *C. frappieri*, also from Reunion, in the COII tree. The sequences for the Mauritius *Cratopopsis* OTU and *Cr. bistigma* were confirmed by repeat PCR and resequencing. The incongruent mtDNA relationships among the three OTUs was also evident within the broader mtDNA analysis to define OTUs, indicating that the incongruence between nuclear and mitochondrial loci is not due to error. An explanation of incomplete lineage sorting can be excluded due to both the phylogenetic depth involved and limited opportunity in the context of island colonisation (see Faria et al., 2016), indicating mtDNA introgression from *C. frappieri* to *Cr. bistigma.* Because of this, the COII partition for *Cr. bistigma* was excluded from the multilocus phylogenetic alignment. As with individual gene analyses, the trees produced by the sensitivity analysis were in broad agreement. Overall we found no large changes in tree support when including missing data with: 71.2% of nodes were well supported (≥0.95 Bayesian posterior probability) in the all sample analysis (58 OTUs); 70.8% with a minimum of two loci (54 OTUs); 71.1% with a minimum of three loci (52 OTUs); and, 74.2% with a minimum of four loci (37 OTUs). Subsequent analyses were thus performed using the four locus matrix for all 58 OTUs.

Multilocus phylogenetic analysis

The concatenated multilocus alignment of the 58 OTUs was 2,173 bp long. Overall, the MCC chronogram (Fig. 2. See Fig. S2.7 for HPD intervals on node ages) is generally well supported and can be divided into two main clades. Clade one contains much of the diversity from the Seychelles and the Comoros with some additional species from Rodrigues, Mauritius and Reunion. Clade 2 includes most species from Mauritius and Reunion and all of the species strictly associated with coralline or coastal habitats in the smaller islets of the Indian Ocean and the Seychelles (*C. gloriosus*, *C. adspersus*, *C. viridisparsus* and *C. griseovestitus*). Clade two can be further divided into three subclades (Fig. 2, Clades 2a, 2b and 2c). Clades 2b and 2c are well supported but the arrangement of relationships among them and clade 2a is poorly resolved, largely due to the unresolved position of *C. nigrogranatus*. *Cratopus* and *Cratopopsis* are not recovered as reciprocally monophyletic, with *Cratopopsis* OTUs occurring in both clades and two of the three subclades of clade two. Rodrigues and Reunion each have their own *Cratopopsis* lineage and a third is represented on both Mauritius and Reunion. *Scaevinus* is recovered within clade 1 as a sister lineage to *C. bernei*.

Biogeographic and dating analyses

Figure 2 shows the results of the DEC+J+X analysis performed on the BEAST2 MCC chronogram. All general biogeographic discussion refers to this analysis while the statistical analysis of node ages (next section) employs the same biogeographic model but iterates across a random sample of 1000 dated trees from the BEAST2 posterior.

Overall, the biogeographic history estimated on the BEAST2 MCC tree (Fig. 2, see also comparison to unconverged starBEAST analysis Fig. S2.8) suggests that the sampled diversity arose from an ancestor on Mauritius. The DEC+J+X analysis infers nine colonisations from Mauritius to Reunion and three colonisations from Reunion to Mauritius. There are two inferred colonisations to Rodrigues, two to the Seychelles and two to the Comoros. One of the two colonisations to each of the Seychelles and the Comoros involves a group of species that have colonised coastal habitats in Mauritius, Reunion and the Comoros and eventually northwards to the Seychelles.

Of the nine colonisations from Mauritius to Reunion inferred by BioGeoBEARS, four have a median estimated divergence time of two million years or less. Most of these have resulted in taxonomically defined species that are extremely similar (*Scaevinus dombayae*, Fig.2 clade 1, *Cr. bistigma* and *C. brunnipes,* Fig. 2, clade 2b) to their counterparts on Mauritius while the remaining colonisation (the *C. nigridorsis* species group, Fig. 2, clade 2b) is represented by a single OTU containing three taxonomically defined species. A fifth colonisation of Reunion is estimated to have occurred approximately 4.8 Mya (Million years ago) (95% HPD 3.6 - 6.2 Mya) , giving rise to the remaining *Cratopopsis* from Reunion (seven taxonomically defined species), *C. nanus*, *C. ditissimus* and the *C. humeralis* species group (four taxonomically defined species). Two further colonisations of Reunion occurred 4.2 Mya (95% HPD 2.8 - 5.7 Mya) and 2.6 Mya (95% HPD 1.7 - 3.7 Mya) respectively with one colonisation resulting in *C. frappieri* (Fig 2, clade 2b) and another resulting in *C. murinus* (Fig2, clade 2c).

The remaining two colonisations of Reunion were followed by back colonisations to Mauritius from Reunion. One occurred approximately 6.4 Mya (95% HPD 4.1 - 8.8 Mya), giving rise to *C. septemvittatus* and *C. exquisitus*, followed by a recolonisation of Mauritius giving rise to *C. melanocephalus* no more than 3.2 Mya (95% HPD 1.9 - 4.6 Mya), while the other is estimated have occurred approximately 8.2 Mya (95% HPD 6.3 - 10.3 Mya) giving rise to *C. bernei* and *S. dombayae*, with a back colonisation to Mauritius

no more than 1.5 Mya (95% HPD 0.8 - 2.2 Mya) giving rise to *S. subtruncatus*. The 95% HPD interval of this event does not include the maximum estimated age of Reunion, invoking either (i) an older age for Reunion, (ii) an underestimation of molecular rate, or (iii) colonisation associated with a more recent node, but this is masked by unsampled or extinct Mauritian taxa within this clade. The third colonisation from Reunion to Mauritius is inferred to have occurred at least 1.2 Mya (95% HPD 0.6 - 1.8 Mya) leading to the Mauritian population of *C. punctum*.

Lineage 2c contains the exclusively coastal species *C. punctum*, *C. griseovestitus*, *C. viridisparsus*, *C. adspersus*, *C. gloriosus*, *Cratopus GrandeComore* sp. 1 and *Cratopus Moheli* sp. 1. Dating analyses suggest the diversification among these taxa began approximately 2.1 Mya (95% CI 1.5 - 2.7 Mya), radiating from Reunion to the low lying islands of Juan de Nova, Europa, Grande Glorieuse and Aldabra not less than 1.1 Mya (95% HPD 0.8 - 1.5 Mya), followed more recently by a colonisation of Seychelles not less than less than 0.5 Mya (95% HPD 0.3 - 0.8 Mya).

Rodrigues is inferred to have been colonised twice. A colonisation giving rise to *C. virescens* and *C. viridipunctatus* (Fig. 2, clade 2b) is inferred to have its source on Mauritius and to have occurred at least 3.8 Mya (95% HPD 2.4 - 5.4 Mya). The second colonisation gave rise to *C. inornatus*, *C. rocki* and *Cr. pauliani* (Fig. 2, clade 1), with Reunion inferred as the most probable source between 7.2 Mya (95% HPD 5.6 - 9.2 Mya) and 8.1 Mya (95% HPD 6.3 - 10.3 Mya). As with the genus *Scaevinus*, the 95% HPD interval of this event does not include the maximum estimated age of Reunion. Given the remoteness of Rodrigues from Reunion, it seems likely that unsampled or extinct species on other islands, especially Mauritius, would be the most probable cause for this discrepancy.

The remaining species in clade 1 are recovered in two well supported groups. The first consists of species restricted to the granitic Seychelles while the second exclusively contains species from the Comoros. The granitic Seychelles appear to have been colonised by the ancestor of *Cratopus La Digue* sp. 1, *Cratopus Mahe* sp. 1, *C. segregatus* and *C. aurostriatus* at least 4.3 Mya (95% HPD 3.0 - 5.8 Mya) while the Comoros are estimated to have been colonised by the ancestor of *C. subdenudatus*, *Cratopus Anjouan* sp. 1 and *Cratopus Anjouan* sp. 2 at least 3.8 Mya (95% HPD 2.8 - 5.1 Mya).

Statistical analysis of node ages

Across all 1000 trees, there was a mean of 8.97 colonisations from Mauritius to Reunion and 3.03 colonisations from Reunion to Mauritius (mean difference 5.94 colonisations, 95% CI = 5.92 to 5.95, t = 665.9, df = 999, p < 0.0001). Colonisations of Mauritius from Reunion are inferred to be significantly younger than those of Reunion from Mauritius (mean difference = -1.07 Ma, 95% CI = -1.10 to -0.1.04 Ma, t = -66.0, df = 999, p < 0.0001). Colonisation events to Mauritius are inferred to be significantly younger than *in situ* speciation events (mean difference = -2.19 Ma, 95% CI = -2.22 to -2.16 Ma, t = -120.3, df = 999, p < 0.0001). *In situ* speciation events on Reunion are inferred to be significantly younger than colonisation events (mean difference = -0.097 Ma, 95% CI = -0.116 to -0.078 Ma, t = -10.275, df = 999, p < 0.0001). Statistical analyses of node ages indicate that nodes leading to flight loss are significantly younger than other nodes (mean difference = -1.327 Ma, 95% CI = -1.344 to -1.1.308 Ma, t = -146.7, df = 999, p < 0.0001) indicating that flight loss is a relatively recent phenomenon within this group, or alternatively that flightless lineages show high extinction rates.

DISCUSSION

Biogeography of Southwest Indian Ocean cratopine weevils

In addition to addressing our hypotheses, the phylogenetic analyses provide an overview of the colonisation dynamics that has given rise to the diversity of cratopine weevils in the islands of the Southwest Indian Ocean. Rodrigues has been colonised twice, as have both the Seychelles and Comoros archipelagos. The more recent colonisation of both the Seychelles and Comoros archipelagos also involves colonisations of the islands to the northeast (Aldabra and Grande Glorieuse) and southwest (Juan de Nova and Europa) of the Comoros archipelago with an apparent source on Reunion. The older colonisation of both archipelagos is inferred to be from Rodrigues. Warren et al. (2010) have described how sea level changes in the Indian Ocean over at least the last five million years would have exposed many now submerged islands of the Mascarene plateau. Thus stepping stone islands are likely to have been intermittently present over much of the evolutionary history of the cratopine weevils, potentially providing routes for dispersal between the granitic Seychelles, the Comoros and Rodrigues.

The estimated colonisation times for Rodrigues are significantly older than its presumed age. The 1.5 Ma age for Rodrigues suggested by McDougall et al. (1965) and based on the oldest exposed lavas falls outside the 95% HPD interval for both colonisation events. Additionally, all three speciation events within Rodrigues are estimated to significantly predate 1.5 Ma. It may be that our mutation rate overestimates divergence times within the Cratopini, but it has previously been suggested that Rodrigues is much older than 1.5 Ma (Thébaud et al., 2009; Warren et al., 2013; Strijk et al., 2014), and the estimates of McDougal et al. (1965) do not sample older buried strata, nor take into account those removed by extensive erosion (Montaggioni & Nativel, 1988).

All nodes inferred as having their most probable location on Mauritius are either younger than or have confidence intervals that include the 8.9 Ma provided by Moore et al. (2011). While the oldest nodes involving Reunion are much older than the 2.0 Ma age for the oldest lavas sampled on Reunion (McDougall, 1971), all but one are consistent with the maximum 5 Ma inferred using methods other than radiometric dating of lavas (Bonneville et al., 1988). While estimates of molecular rate and island age may hold some explanation for this single difference, no speciation events within Reunion exceed its 5 Ma age estimate. An alternative explanation is that colonisation time is overestimated due to unsampled or extinct taxa from Mauritius.

Island age, community assembly and species turnover

Biogeographic inference of ancestral areas supports the hypothesis that *in situ* speciation has contributed more to cratopine species richness on the older island of Mauritius than it has on the younger island of Reunion. Our inferences regarding the origins of species diversity are dependent on our OTUs being equivalent to species. Comparison of our OTUs with the most recent taxonomic treatise (e.g. Williams & Cox, 2003) support this inference, with recently-defined species corresponding with our OTUs. Our first prediction that colonising species to Mauritius should have lower persistence times, compared to those colonising Reunion, is supported by the statistical analysis of node ages. Mauritius is characterised by older *in situ* speciation events, with younger colonising lineages, supporting prediction two. With regard to Reunion, likelihood inferences suggest that species origin by colonisation has been more

predominant in the history of the island, with the contribution of *in situ* speciation appearing in only a small subset of lineages present on the island, supporting prediction three.

Results are largely consistent with the expectation that the contribution of colonisation to community assembly is more important in the early stages of assembly, with speciation assuming a more dominant role as community assembly gives way to a dynamic of species turnover (Emerson & Oromí, 2005; Emerson & Gillespie, 2008), which is expected under the equilibrium model (Heaney, 2007). Over evolutionary time, it appears that niche space within an island becomes occupied by colonising species. Newly arriving species are less able to establish and remaining niche space becomes increasingly likely to be colonised by resident species through *in situ* speciation, as are niches that may become vacant through extinction (Emerson & Oromí, 2005; Emerson & Gillespie, 2008).

Polyphyly of Cratopopsis and loss of flight capability
Our data demonstrate that the genera *Cratopus* and *Cratopopsis* are not monophyletic (Fig. 2), with at least three independent instances of the *Cratopopsis* morphology evolving convergently within *Cratopus*. *Cratopopsis* species mainly differ from *Cratopus* by a lack of wings and a characteristically-shaped reduced thorax that is probably associated with a loss of flight musculature. This pattern is seen more widely in the tribe with at least two further flight loss events (in *C. murinus* and *Scaevinus*) and the number of events may be higher, as two flightless Indian Ocean cratopine weevils were not sampled (*C. triangularis* from Mauritius and *Cr. matilei* from Anjouan). A number of authors have argued that there are benefits to flightlessness, such as energetic savings and increased reproductive output (Roff, 1990), and selection against dispersal has also been hypothesised for organisms inhabiting limited habitat patches (Harrison, 1980). A tendency toward flightlessness is part of a wider pattern seen in beetles on islands, with examples from many archipelagos [e.g *Rhyncogonus* of Polynesia (Samuelson, 2003), *Laparocerus* of Macaronesia (Machado et al., 2008)]. While there has been repeated selection for flightlessness within the Cratopini of the Indian Ocean, our statistical analysis of node ages indicates that flightless lineages are significantly younger than flighted lineages. Rather than flight loss being a recent phenomenon within this group, we speculate that it is more likely to be a recurrent phenomenon, but characterised by a higher per species extinction probability for flightless species than flighted species.

CONCLUSIONS

Phylogenetic reconstruction and ancestral area inference indicate that the extensive radiation of Cratopine weevils across the islands of the Indian Ocean has occurred within the last 10 million years, with the earliest diversification events inferred to have occurred on the island of Mauritius. Geographic inferences for colonisation and speciation on the species rich islands of Mauritius and Reunion are consistent with predictions from equilibrium theory regarding species turnover, when extended to incorporate speciation. Compared to Mauritius, colonisation is found to have contributed more to species richness on Reunion, consistent with its younger geological age and earlier stage of community assembly. In contrast, much of the species richness on the older island of Mauritius is consistent with a more mature community dynamic where turnover is dominated by *in situ* speciation of already established lineages. These

data offer support for a model where island community assembly is initially dominated by colonisation, but where in the later stages of an island life-cycle, turnover is more likely to involve species replacement by speciation within the island, rather than colonisation from an external source.

REFERENCES

Alonso-Zarazaga, M.A. & Lyal, C.H.C. (1999) *A world catalogue of families and genera of Curculionoidea (Insecta: Coleoptera)(excepting Scolytidae and Platypodidae).* Entomopraxis, Barcelona.

Bonneville, A., Barriot, J.P., & Bayer, R. (1988) Evidence from geoid data of a hotspot origin for the southern Mascarene Plateau and Mascarene Islands (Indian Ocean). *Journal of Geophysical Research, [Solid Earth]*, **93**, 4199–4212.

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**, e1003537.

Casquet, J., Bourgeois, Y.X.C., Cruaud, C., Gavory, F., Gillespie, R.G., & Thébaud, C. (2015) Community assembly on remote islands: a comparison of Hawaiian and Mascarene spiders. *Journal of biogeography*, **42**, 39–50.

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., & Fanciulli, P.P. (2009) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular ecology*, **19**, 386–400.

Darriba, D., Taboada, G.L., Doallo, R., & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, **9**, 772.

Emerson, B.C. & Gillespie, R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in ecology & evolution*, **23**, 619–630.

Emerson, B.C. & Kolm, N. (2005) Species diversity can drive speciation. *Nature*, **434**, 1015–1017.

Emerson, B.C. & Oromí, P. (2005) Diversification of the forest beetle genus Tarphius on the Canary Islands, and the evolutionary origins of island endemics. *Evolution; international journal of organic evolution*, **59**, 586–598.

Faria, C., Machado, A., Amorim, I.R., Gage, M.J.G., Borges, P.A.V., & Emerson, B.C. (2016) Evidence for multiple founding lineages and genetic admixture in the evolution of species within an oceanic island weevil (Coleoptera, Curculionidae) super-radiation. *Journal of biogeography*, **43**, 178–191.

Harrison, R.G. (1980) Dispersal Polymorphisms in Insects. *Annual review of ecology and systematics*, **11**, 95–118.

Heaney, L.R. (2007) Is a new paradigm emerging for oceanic island biogeography? *Journal of biogeography*, **34**, 753–757.

Heled, J. & Drummond, A.J. (2009) Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, **27**, 570–580.

Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* , **17**, 754–755.

Hustache, A. (1919) Synopsis des Curculionides de la faune malgache. I. Brachydérides et Otiorrhynchides. *Annales de la Societe entomologique de France. Societe entomologique de France*, **87**, 441–520.

Ikeda, H., Nishikawa, M., & Sota, T. (2012) Loss of flight promotes beetle diversification. *Nature communications*, **3**, 648.

Katoh, K., Misawa, K., Kuma, K.-I., & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, **30**, 3059–3066.

Kitson, J.J.N., Warren, B.H., Florens, F.B.V., Baider, C., Strasberg, D., & Emerson, B.C. (2013) Molecular characterization of trophic ecology within an island radiation of insect herbivores (Curculionidae: Entiminae: Cratopus). *Molecular ecology*, **22**, 5441–5455.

MacArthur, R.H. & Wilson, E.O. (1963) An Equilibrium Theory of Insular Zoogeography. *Evolution; international journal of organic evolution*, **17**, 373–387.

MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography.* Princeton University Press,

Machado, A., López, M., Almeida, T., & Hernández, M. (2008) Mitochondrial DNA phylogenetic analysis of the genus *Laparocerus* (Coleoptera: Curculionidae: Entiminae). I. The Madeiran clade. *Zoologica scripta*, **37**, 415–427.

Matzke, N.J. (2013) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, **5**, 242–248.

Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Systematic biology*, **63**, 951–970.

McDougall, I. (1971) Geochronology and evolution of the young volcanic island of Reunion, Indian Ocean. *Geochimica et cosmochimica acta*, **35**, 261–&.

McDougall, I., Upton, B.G.J., & Wadsworth, W.J. (1965) A geological reconnaissance of Rodriguez Island, Indian Ocean. *Nature*, **206**, 26–27.

Medeiros, M.J. & Gillespie, R.G. (2011) Biogeography and the evolution of flightlessness in a radiation of Hawaiian moths (Xyloryctidae: *Thyrocopa*). *Journal of biogeography*, **38**, 101–111.

Meseguer, A.S., Aldasoro, J.J., & Sanmartín, I. (2013) Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St. John's wort (Hypericum). *Molecular phylogenetics and evolution*, **67**, 379–403.

Montaggioni, L.F. & Nativel, P. (1988) *La Reunion, Ile Maurice: Geologie et apercus biologiques, plantes et animaux.* Masson, France.

Moore, J., White, W.M., Paul, D., Duncan, R.A., Abouchami, W., & Galer, S.J.G. (2011) Evolution of shield-building and rejuvenescent volcanism of Mauritius. *Journal of Volcanology and Geothermal Research*, **207**, 47–66.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., & Swofford, D.L. (2007) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* , **24**, 581–583.

Rambaut, A. & Drummond, A.J. (2007) *Tracer v1.5.0 available from http://tree.bio.ed.ac.uk/software/tracer/.*

R Development Core Team (2017) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.,

Ree, R.H. & Sanmartín, I. (2018) Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of biogeography*, **45**, 1–9.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, **57**, 4–14.

Roff, D.A. (1990) The evolution of flightlessness in insects. *Ecological monographs*, **60**, 389–421.

Samuelson, G.A. (2003) Review of *Rhyncogonus* of the Hawaiian Islands. *Bishop Museum Bulletins in Entomology*, **11**, 1–107.

Sattler, K. (1991) A review of wing reduction in Lepidoptera. *Bulletin of the British Museum*, **60**, .

Stamatakis, A., Hoover, P., & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Systematic biology*, **57**, 758–771.

Strijk, J.S., Bone, R.E., Thébaud, C., Buerki, S., Fritsch, P.W., Hodkinson, T.R., & Strasberg, D. (2014) Timing and tempo of evolutionary diversification in a biodiversity hotspot: Primulaceae on Indian Ocean islands. *Journal of biogeography*, **41**, 810–822.

Thébaud, C., Warren, B.H., Strasberg, D., & Cheke, A. (2009) Mascarene Islands, biology. *Encyclopedia of Islands* (ed. by R.G. Gillespie and D.A. Clague), pp. 612–619. University of California Press, Berkeley.

Van Dam, M.H. & Matzke, N.J. (2016) Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of biogeography*, **43**, 1514–1532.

Voisin, J.-F. & Poussereau, J. (2009) A propos de *Cratopus nanus* et de *Cratopus parvus* de l'île de la Réunion (Coleoptera, Curculionidae). *Nouvelle Revue d'Entomologie*, **26**, 59–65.

Warren, B.H., Safford, R.J., Strasberg, D., & Thébaud, C. (2013) Bird biogeography and evolution in the Malagasy region. *The Birds of Africa: Volume VIII: The Malagasy Region: Madagascar, Seychelles, Comoros, Mascarenes* (ed. by R. Safford, F. Hawkins, and J. Gale), pp. 35–36. A&C Black, London.

Warren, B.H., Strasberg, D., Bruggemann, J.H., Prys-Jones, R.P., & Thébaud, C. (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics: the international journal of the Willi Hennig Society*, **26**, 526–538.

Whittaker, R.J., Triantis, K.A., & Ladle, R.J. (2008) A general dynamic theory of oceanic island biogeography. *Journal of biogeography*, **35**, 977–994.

Williams, J.R. (2000) A revision of the Mascarene weevil genus Syzygops Schonherr (Coleoptera : Curculionidae : Entiminae). *Invertebrate Taxonomy*, **14**, 411–432.

Williams, J.R. & Cox, M.L. (2003) A contribution to the study of Mascarene Weevils of the genus *Cratopus* Schonherr (Coleoptera: Curculionidae: Entiminae: Cratopini): The species of Mauritius and Rodrigues. *The Mauritius Institute Bulletin*, **12**, 1–67.

Wright, N.A., Steadman, D.W., & Witt, C.C. (2016) Predictable evolution toward flightlessness in volant island birds. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 4765–4770.

Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003) An index of substitution saturation and its application. *Molecular phylogenetics and evolution*, **26**, 1–7.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., & Lam, T.T.-Y. (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in ecology and evolution*, **8**, 28–36.

BIOSKETCH

JJNK undertook most of the beetle sampling and performed all the laboratory work and data analyses. BHW and DS contributed to the fieldwork. JJNK and BCE wrote the manuscript. All authors discussed the results and commented on the manuscript. This project was conceived by BCE, DS and CT. It was supervised primarily by BCE and secondarily by BHW.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:
Appendix S1: Molecular methods and results for definition of OTUs.
Appendix S2: Gene trees and summary information for each locus used in the main phylogeny.
GitHub repository: https://github.com/James-Kitson/Biogeography <doi will be provided upon acceptance>. This repository includes all scripts used for plotting and analysing trees.

DATA ACCESSIBILITY

All sequences used in this manuscript are available under GenBank accession numbers <These will be provided upon acceptance>. All plotting scripts are available in our GitHub repository.
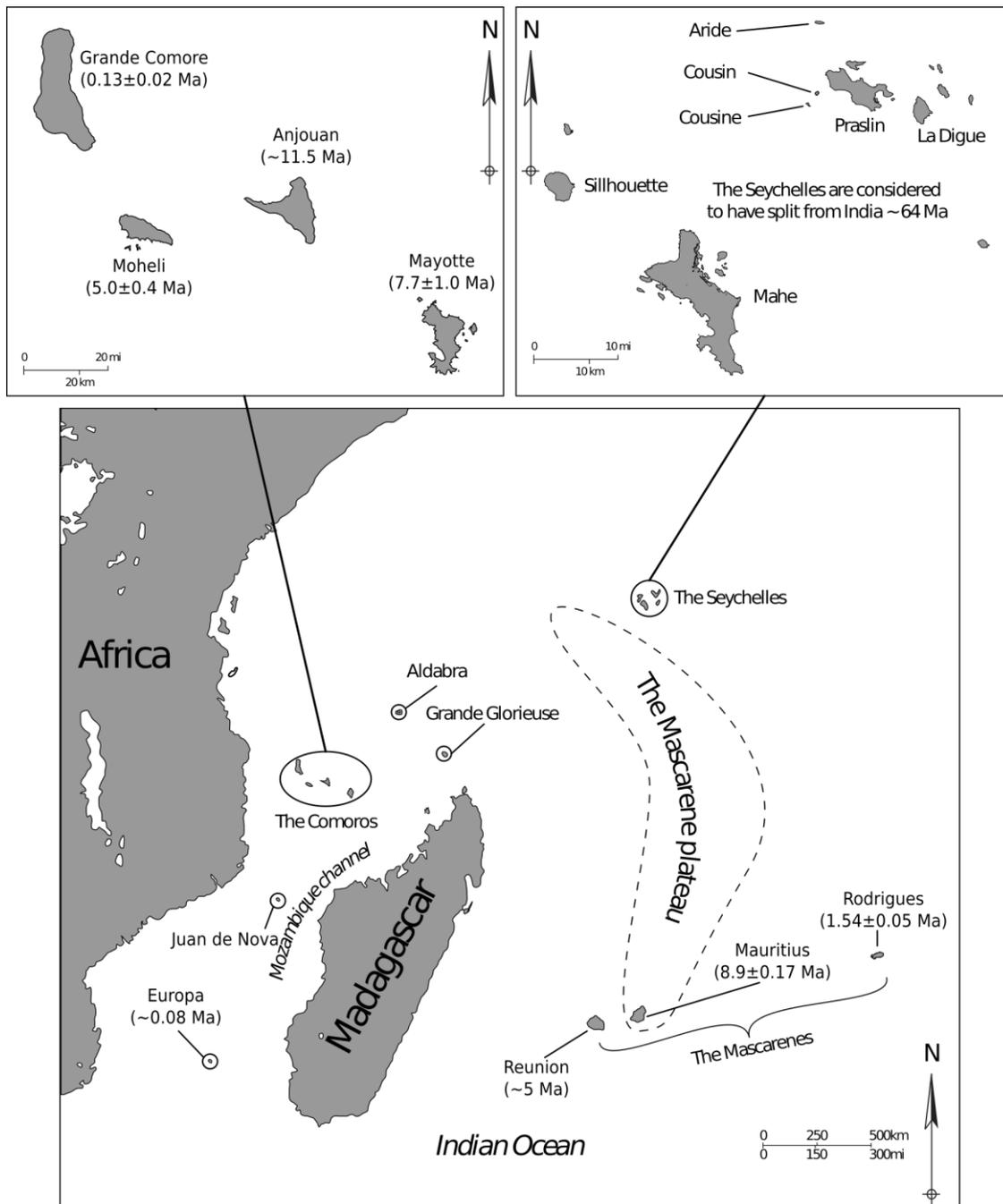
FIGURE LEGENDS

Fig. 1. The islands of the Southwest Indian Ocean area, with known ages. Additionally marked is the Mascarene plateau, an area of shallow water that is likely to have been exposed multiple times during glacial periods.
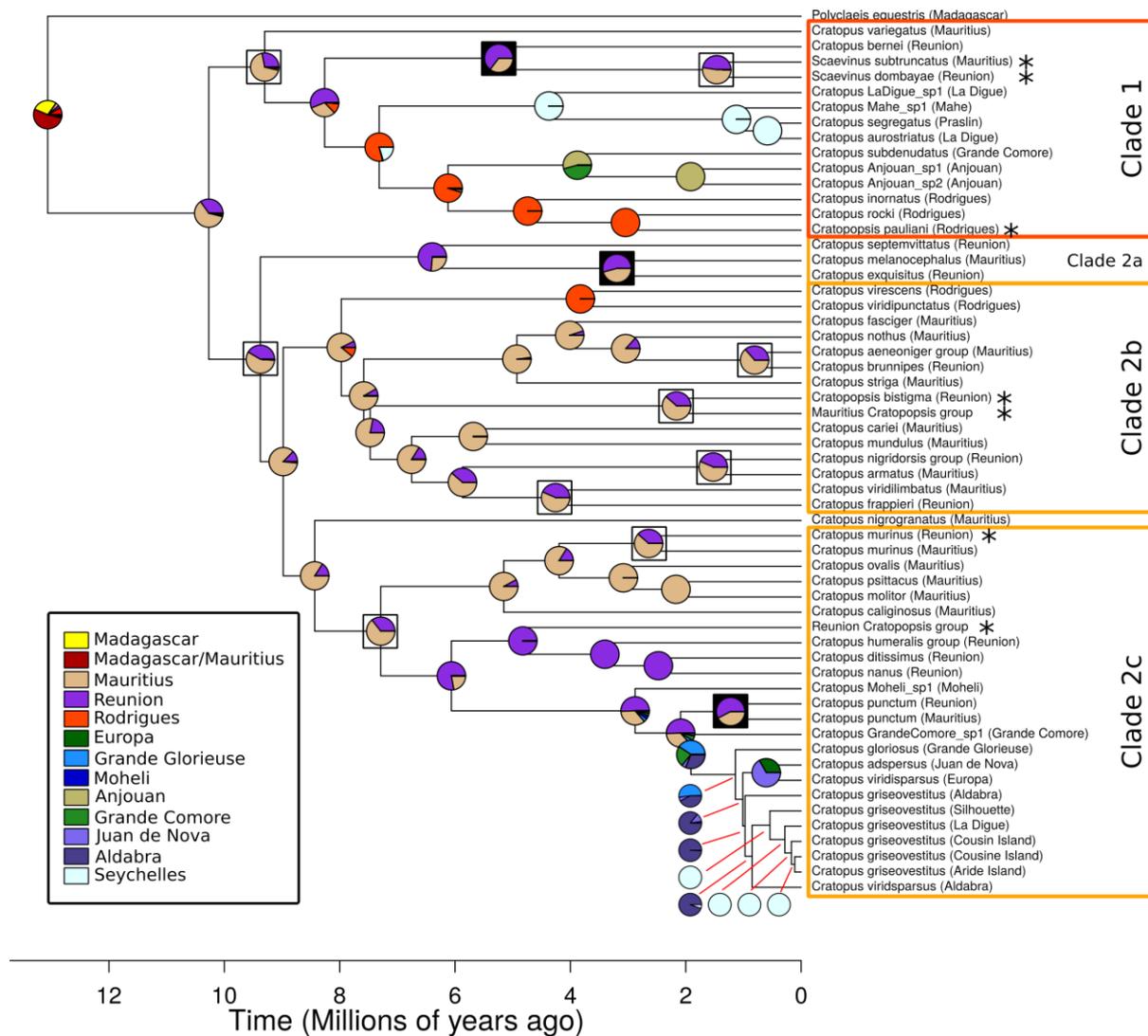
Fig. 2. The estimated biogeographic history of Cratopine weevils as inferred by BioGeoBEARS under the DEC+J+X model using the BEAST2 MCC chronogram. Pie charts on nodes represent the relative probabilities of the inferred geographic range of the ancestral taxon represented by the node. Empty boxes on nodes represent inferred colonisations of Reunion from Mauritius while black boxes on nodes represent inferred colonisations of Mauritius from Reunion. Taxa with * after their label are flightless.