2012

When and Where: Divergence times and colonization history of Darwin's darkling beetles in the Galápagos Archipelago

Julia Wucherpfennig
Wellesley College, jwucherp@wellesley.edu

When and where: Divergence times and colonization history of Darwin's darkling beetles in the Galápagos Archipelago

Julia Wucherpfennig
Advisor: Andrea Sequeira

Department of Biological Sciences

May 2012

Submitted in Partial Fulfillment of the Prerequisite for Honors in the Biological Sciences Department of Wellesley College

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Professor Andrea Sequeira for teaching me a love of evolutionary biology and for her support for the past three years in the lab. Without her, this project would never have been possible.

I would like to thank Weixia Guo and Christina Tran for all the work they put into the data collection on this project and for being wonderful friends and lab partners. I would also like to thank Courtney Stepien for teaching me everything when I first started working in the lab.

Thank you to Kara Lu, Adrienne Cheng, Blair Uhlig, and Katie Chen for their company in lab this year.

Many thanks to those in the Galápagos, at the Charles Darwin Research Station and the Galápagos National Park, for their work in making this project possible.

I would like to thank my thesis committee: Professor Hughes, Professor Königer, and Professor Hearn. You have been wonderful teachers, and I am grateful for all of your advice. Also, thank you to Elaine Igo for maintaining the sequencing machine.

Thanks to all of my friends for their moral support. Finally, I would like to thank, my parents, Kai and Anne Wucherpfennig without whom I would not be at Wellesley.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABSTRACT

Island archipelagos are the focus for many studies of species formation because they provide a simplified environment in which the processes of speciation, dispersal, and extinction can be elucidated. Of importance in understanding the evolutionary and dispersal history of an organism are the colonization histories, divergence times, and ancestral areas of groups of organisms. This study will consider these three components to understand the radiation of endemic beetle genus *Stomion* in the Galápagos Islands. In order to generate phylogenies and date the divergences, we sequenced two nuclear and five mitochondrial genes in 25 individuals from within the genus of *Stomion*. We found that *Stomion*'s presence in the archipelago is the product of a single colonization that occurred between 1 and 2.3 million years ago. While younger than the age of the islands, the colonization is older than the age of the islands included in this study. The generation of species diversity in this radiation is the result of multiple instances of inter-island speciation, and despite the old geological age of the archipelago and the large area of some the islands, the diversity associated with the radiation does not appear to be the result of intra-island speciation. The pattern of colonization in *Stomion* is peculiar in comparison to other groups in several ways. First, the reconstructed ancestral *Stomion* divergence took place in Darwin and Isabela, both young islands that are not in the path of the major oceanic currents. Second, the colonization order does not complying with the progression rule, nor does the order reflect a geographical pattern because the islands closer to the potential source (South America) were not the first colonization platforms. In order to understand why *Stomion* has unique pattern and elucidate more general patterns within the archipelago, the colonization patterns were compared to other invertebrate groups in the Galápagos. Based on these comparisons, it appears that the geography and other abiotic factors impact the first colonization point of the species, which is generally seen in the southeast of the archipelago. Due to its large size, Isabela also plays a central role in determining the colonization pattern. While the study of each group of organisms is unique, understanding each genus is important to assembling a broader picture about the evolution of groups in island archipelagos.

INTRODUCTION

Island archipelagos, such as the Galápagos Islands, provide an ideal environment to study the mechanism of species formation and colonization patterns of organisms. Despite the fact that the islands can be small, their geologic activity and diverse environments produce a dynamic system that contributes to the formation of new species by creating many barriers to gene flow (Emerson 2002, Whittaker and Fernández-Palacios 2007). Even though islands are considered to be species poor, their fauna contribute disproportionately to global biodiversity because many species are endemic to the archipelago, meaning that they are found nowhere else on earth (Whittaker and Fernández-Palacios 2007). Their endemism also means that the species are among the most threatened. As a result of these two factors, islands are considered biodiversity hotspots and are the focus of many conservation efforts (Myers et al. 2000, Whittaker and Fernández-Palacios 2007). The Galápagos, in particular, are home to many endemic species of birds, reptiles, and invertebrates, and efforts have been made to identify their individual patterns of colonization and speciation (Finston and Peck 1995, 1997, Caccone et al. 2002, Ciofi et al. 2002, Finston and Peck 2004, Parent and Crespi 2006, Peck 2006, Parent et al. 2008, Sequeira et al. 2008a, Sequeira et al. 2008b). This study aims to understand these patterns in the endemic beetle genus *Stomion*, by estimating the time of divergence and elucidating a general trend of colonization across different species of invertebrates in the Galápagos Islands.

Galápagos Islands

The Galápagos Archipelago consists of a group of 127 volcanic islands and islets 1,000 km off the western coast of South America. The islands lie on the Nazca plate and were formed when the plate moved over a hotspot that is thought to have developed between 80 and 90 million years ago (Figure 1a) (Peck 2006). As a result of the plate's movement to the southeast,

the young islands, Wolf and Darwin, are found in the northwest and the oldest islands, Española and San Cristóbal, in the southeast (Table 1) (Geist 1996, Peck 2006). Some of the islands, including Isabela and Fernandina, are still volcanically active. In contrast to the linear chains of the Canary and Hawaiian Islands, the Galápagos Archipelago has clusters of islands that are approximately the same age and are the product of a larger hotspot (Figure 1b, Table 1) (White et al. 1993). Nevertheless they are similar to the Hawaiian and Canary Islands in that none of the island systems were ever connected to the mainland, and all include seamounts (sunken islands) older than the extant islands (Price and Clague 2002). In the Galápagos, at least eight seamounts have been located in the eastern and southeastern part of the archipelago. These seamounts extend the time available for diversification on the islands to more than 11 million years (my) (White et al. 1993).

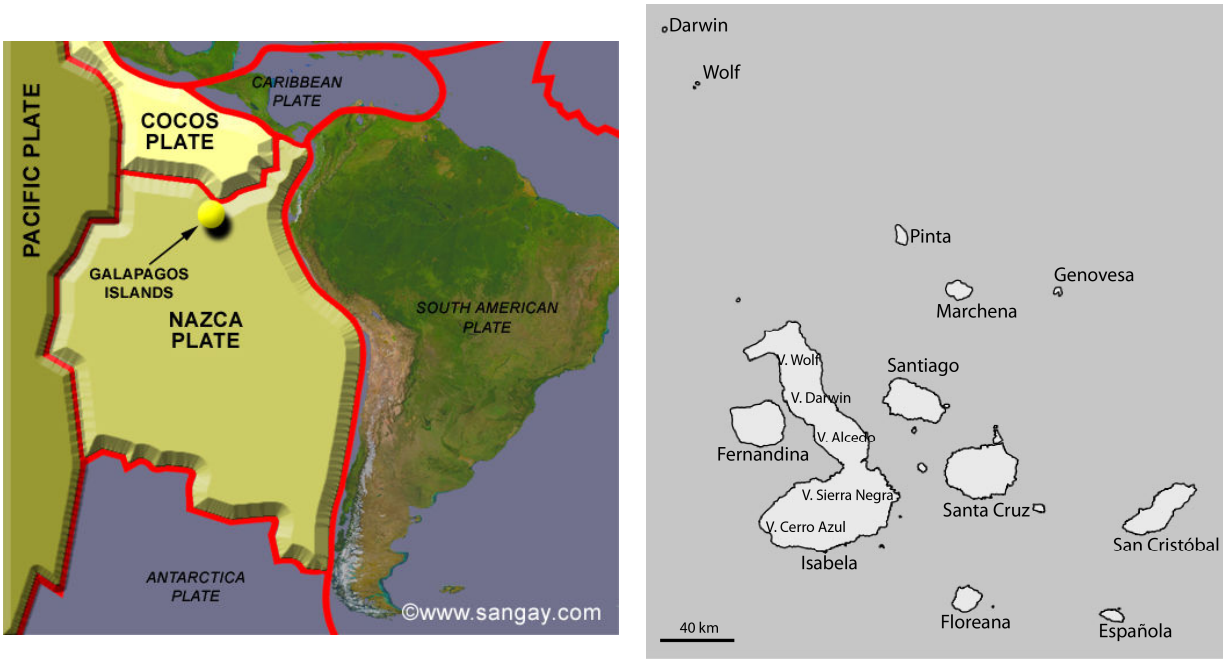

Figure 1. Location and map of the Galápagos Archipelago A. Nazca plate and neighboring plates (Source: http://www.sangay.com/). **B.** Map of Galápagos Archipelago with islands and main volcanoes (Source: http://en.wikipedia.org/).

Table 1. Ages of the Galápagos Islands and volcanoes in millions of years. The minimum age of the island is based on the potassium-argon age of the oldest exposed subaerial lava. The maximum age is based on the hotspot model. It assumes that all islands originated where Fernandina is now and moved to their current position at a rate of 37 mm per year (Geist 1996, Peck 2006).

Island/Volcano	Minimum Age (million years)	Maximum Age (million years)
San Cristóbal	2.3	6.3
Española	2.8	5.6
Santa Fé	2.8	4.6
Floreana	1.5	3.3
Santa Cruz	2.2	3.6
Isabela	0.06	0.7
Volcán Wolf	0.06	0.3
Volcán Darwin	0.07	0.7
Volcán Alcedo	0.15	0.30
Volcán Sierra Negra	0.07	0.30
Volcán Cerro Azul	0.06	0.30
Santiago	0.77	2.4
Genovesa	<0.7	-
Pinta	<0.7	-
Ferandina	0.006	0.3

Colonization Patterns

Due to geography, wind patterns, and ocean currents, it has been hypothesized that the individuals colonizing the archipelago originated from the South American mainland (Peck 2006). Those species with wings could have flown or been blown out to sea by a storm, but most probably reached the islands on vegetation rafts. These means of transport favor birds,

invertebrates, and reptiles and have led to the "unbalanced" distribution of taxa in the archipelago (Peck 2006, Whittaker and Fernández-Palacios 2007).

The species that are seen on the islands today are not necessarily the species that colonized the islands but are their descendants (Peck 2006, Whittaker and Fernández-Palacios 2007). Since the first individuals colonized the archipelago, micro-evolutionary processes have been on-going, resulting in the differentiation of the colonizers from their ancestors. In beetles, the most conspicuous phenotypic change that often occurs after colonization is the loss of wings. This loss of long-distance dispersal ability has been seen in insects, birds, and other organisms that inhabit island archipelagos (Whittaker and Fernández-Palacios 2007). While *Stomion* beetles are thought to have colonized the islands in a wingless form, other changes have occurred that distinguish them from the colonizing and mainland ancestors (Peck 2006). Many of these changes are influenced by the small size of the colonizing population. When a population colonizes the archipelago it undergoes a type of bottleneck, a drastic reduction in population size, known as a founder effect. Because only a subset of the mainland population reaches the archipelago, the survivors have lower genetic variation than their parent population and some alleles are lost. Genetic drift then acts on the remaining variation, and due to its small size, the population is more susceptible to its effects (Whittaker and Fernández-Palacios 2007).

Stomion

Beetles are among the most well represented and species-rich taxonomic groups in the archipelago. In total, the fauna includes 56 families, 297 genera and 486 species of beetles, and of these, 266 species are endemic to the islands (Peck 2006). The Galápagos is also home to three genera known as Darwin's darkling beetles, *Stomion*, *Ammophorus*, and *Blapstinus*. They are a part of the subfamily Pimeliinae in the family Tenebrionidae (Finston and Peck 2004).

Stomion, in particular, consists of 11 recognized species that were designated based on morphology and allozyme genotypes. Even though there is no known ancestor, the genus is thought to be the result of a single colonization from the South American mainland that radiated throughout the archipelago (Figure 2) (Finston and Peck 1995, 1997, 2004, Peck 2006). As a result, *Stomion* inhabits at least 25 islands or islets, yet very little is known about their biology (Peck 2006). Because of the conditions found on newly emerged volcanic islands, *Stomion* and the majority of the beetle fauna are generalists, scavengers and predators, as opposed to herbivores (Peck and Kukalová-Peck 1990, Peck 2006). Furthermore, generalists, such as *Stomion*, which feeds on detritus or debris, are more likely to colonize other islands and differentiate throughout the archipelago because they are not dependent on the presence of a specific food source (Becker 1975, Peck 2006, Whittaker and Fernández-Palacios 2007). They are often found in sand with vegetation in the supra-littoral zone but also live in litter under bushes and rocks in other zones. *Stomion* beetles are known to be more active at night and after rain, but their activity also varies according to temperature and number of hours of sunlight. In spite of the fact that their life history is so poorly understood, studies have been done to understand their morphology and genetic diversity (Peck and Kukalová-Peck 1990, Finston and Peck 1995, 1997, 2004, Peck 2006).

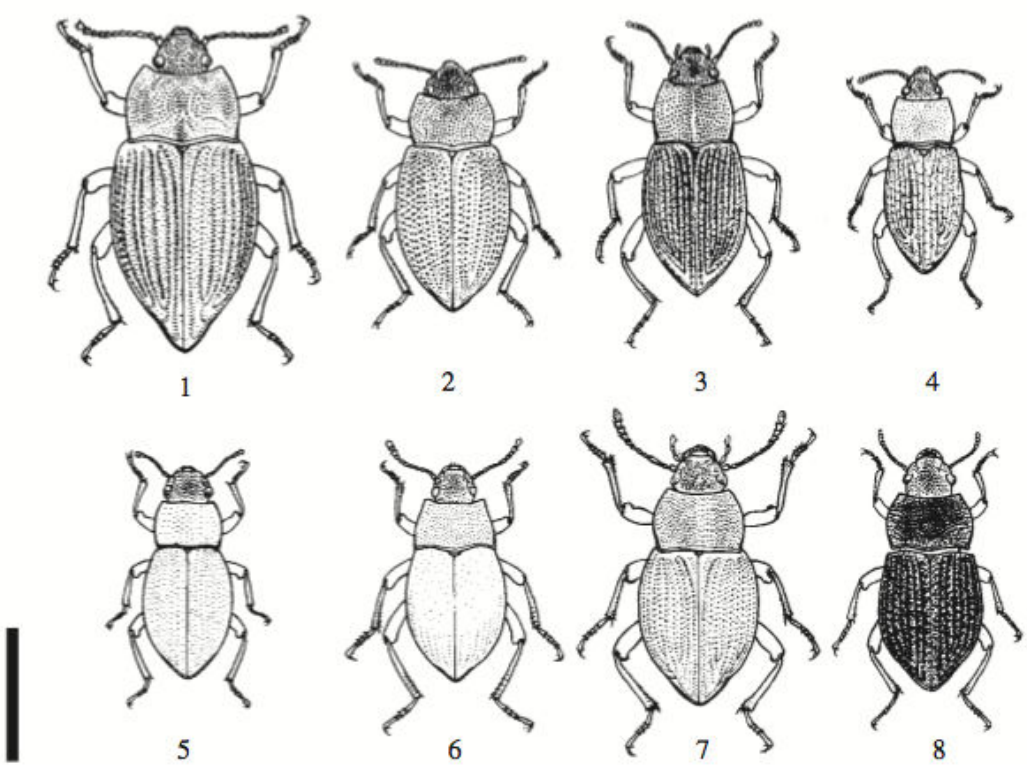

Figure 2. Drawings of eight species of *Stomion*. 1. *S. galapagoensis*, 2. *S. helopoides*, 3. *S. cribicollis*, 4. *S. longulum*, 5. *S. laevigatum*, 6. *S. linelli*, 7. *S. longicornis*, 8. *S. rugosum*. Scale bar =5 mm. Image source: (Finston and Peck 2004)

Previous studies of *Stomion*'s radiation

The first study of *Stomion* was done by Finston and Peck (1995) in order to analyze the population structure and gene flow among 35 populations representing nine of the then 13 recognized species. Based on chi-square contingency analysis of allele frequencies in eight polymorphic enzyme loci (allozymes), they showed that there were significant differences in the gene frequencies of populations of four species. The genetic subdivision was especially high between populations on different islands. Furthermore, the genotypic frequencies deviated from those expected by Hardy-Weinberg because there were fewer heterozygotes. Based on these results, they concluded that the populations are small and that there is little gene flow between

them (Finston and Peck 1995). These results were further confirmed in an expanded study that included more taxa and loci (Finston and Peck 1997).

Stomion has also been analyzed based on its morphological variation. The taxa can be divided into three groups based on "external sculpturing of the elytra" (Finston and Peck 1997). Building on these morphological groupings, Finston and Peck reevaluated the seventeen previously defined species to determine the number of morphologically distinct species. Principle coordinate analysis was conducted using 23 qualitative characteristics to sort the individuals, representing all named species, into morphological groups (Finston and Peck 2004). Based on the analysis, they identified eleven species, one of which has two subspecies. Their cladograms (a phylogenetic tree that shows the patterns of inheritance through a branching scheme) supported a single colonization by *Stomion* and yielded two weakly supported groups, but their findings confirmed neither the original species groups defined by Van Dyke based on Darwin's collection nor Finston and Peck's allozyme analysis (Van Dyke 1953, Finston and Peck 1995, 1997, 2004). The poorly resolved groups were attributed to the variations within and between populations of the same species. More broadly, Finston and Peck determined that the speciation of *Stomion* is not an example of adaptive radiation but is an example of allopatric speciation and subsequent morphological divergence. They also hypothesized that species divergences within the genus are relatively young; however, they had little evidence to support their hypothesis (Finston and Peck 2004).

Estimation of divergence times

Phylogenetic trees (phylogenies) are a useful tool to elucidate the geographic pattern of colonization in an archipelago, and with additional information can be used to establish a time frame for colonization and divergence within a group. They are visual representations of the

evolutionary relationships between species or taxa and are generated based on morphological character traits and/or molecular data that differ between the individual taxa. The nodes within the tree represent the most recent common ancestor of the taxa. In order to be able to determine which divergences are the most ancient and which species are most derived, the tree must be rooted, often using an outgroup. This outgroup is a taxon or set of taxa that is closely related to the group being studied but more distantly related than any of the other taxa in the group are to each other. Without an outgroup, the tree is an unrooted phylogeny from which only the branching pattern can be determined, and the divergence time cannot be estimated. After rooting the tree, a molecular clock or other models can be applied to a tree based on sequence data in order to approximate the ages of important nodes (Drummond et al. 2006). As a result of this process, the branch lengths are correlated to the length of time since the divergence.

Multiple models have been proposed to estimate both phylogenies and divergence times based on molecular sequences, but they have frequently been rejected because they make assumptions that are unrealistic (Drummond et al. 2006, Drummond and Rambaut 2007). Unrooted models of phylogeny assume a different and independent rate of mutation for each branch, but due to the lack of information about these rates, they cannot be used to estimate divergences. Strict molecular clocks, on the other hand, assume a constant rate of mutation over millions of years. While they can be used to estimate the age of divergence, the assumption of a constant mutation rate is unrealistic in many large datasets (Drummond et al. 2006). The "intermediate" model between these two extremes is known as a Bayesian relaxed molecular clock (Drummond et al. 2002, Drummond et al. 2006, Drummond and Rambaut 2007). While each branch has a different rate of mutation, those rates are derived from a parametric distribution, which is determined based on the rate of the parent branch. That distribution can

either be log-normal or exponential. A log-normal distribution assumes a changing rate along each branch relative to the length of the branch, while an exponential distribution assumes that the rate is independent of branch length and that changes in the rate occur at the nodes (Drummond et al. 2006). The computer program BEAST (Bayesian evolutionary analysis by sampling trees) implements a version of a relaxed clock model, using the Bayesian Markov Chain Monte Carlo method. It estimates both the phylogeny and divergence times under a relaxed clock, while fixing neither of the two elements. It can be used when the rate of evolution and ages of the nodes in the tree are not well known (Drummond et al. 2006, Drummond and Rambaut 2007, Whittaker and Fernández-Palacios 2007).

Nevertheless, to obtain a realistic estimate for the ages of divergence, either the age range of a node or the mutation rate must be set. This information is known as an informed prior, and along with the sequence data is the input needed to run the basic form of the model. The most commonly used prior is the age of a fossil; however, in the case of *Stomion*, no fossil records are known (Finston and Peck 2004, Drummond and Rambaut 2007). Instead, three different types of priors were used to calibrate the tree: 1. the age of the Galápagos Islands, 2. known rate of mitochondrial DNA mutation in arthropods, and 3. ages of the outgroups.

Geologic estimates have established a wide age range for each of the islands (Geist 1996) (Table 1). The minimum age was determined based on the potassium-argon dating of the oldest exposed subaerial lavas, while the maximum age was estimated based on the hotspot model. The latter method assumes that every volcano in the Galápagos emerged in the west where Fernandina is now and then traveled on the Nazca plate to its current location at a rate of 37 mm per year (the estimated rate of movement of the plate). Many of the ages obtained using this method are probably overestimates because the hotspot is large, and the islands could have

emerged much closer to their current locations (Geist 1996). Due to the fact that it is unclear exactly how accurate these models are with respect to the actual age of the islands and how close the divergence of *Stomion* is to the age of the islands, these values can only be used to define the distribution of a prior and must be considered a starting point of our estimations of beetle divergence time (method 1).

Because of the uncertainties relating to the use of island ages as priors described above, two other types of priors were also implemented. First, instead of constraining the age of the node, as was done using the age of the islands, the rate of divergence was set (method 2). Between two sequences, a certain percentage of the bases are different, and studies have determined the average percentage of bases that changes in arthropods over the course of a million years. This rate, known as the sequence divergence per million years, can be determined when the age of the divergence and the percentage of sequence divergence can be identified independently (Brower 1994, Farrell 2001). In seven closely related arthropods, the rate of pairwise sequence divergence per million years has been estimated to be 2.3%. This result means that every million years one sequence has a mutation rate of around 1.15% (Brower 1994).

The previously described percentage of sequence divergence has also been employed in other studies to estimate the divergence of *Pimelia*, a genus of darkling beetle found in the Canary Islands. The sequences of the COII gene region from six species within the genus were used to generate a phylogeny and approximate the divergence times (Juan et al. 1995b, Contreras-Diaz et al. 2003b). Because there are no fossil records of *Pimelia*, Brower's estimate for mtDNA mutation rate was used to calibrate the molecular clock. Because *Pimelia* is an outgroup used in the generation of the *Stomion* phylogeny presented in this study, the age

estimations for the divergences within *Pimelia* were also used to calibrate the divergence of *Stomion* externally (method 3).

Biogeography of Galápagos Species

While many radiations in the Galápagos have been studied in detail, few efforts have been made to find commonalities between these groups and elucidate a general colonization pattern (Sato et al. 2001, Caccone et al. 2002, Ciofi et al. 2002, Finston and Peck 2004, Arbogast et al. 2006, Parent and Crespi 2006, Schmitz et al. 2007, Parent et al. 2008, Sequeira et al. 2008a, Sequeira et al. 2008b, De Busschere et al. 2010). Due to the isolation and arid climate of the islands very few species have successfully colonized the islands, and even fewer have diversified within the archipelago. Most of the taxa found in the archipelago are thought to have ancestors on the South American continent, but the island fauna has less diversity than the mainland because many either do not survive the trip or cannot sustain themselves on the islands (Peck 2006, Whittaker and Fernández-Palacios 2007, Parent et al. 2008). Even among those that have thrived on the islands only some have diversified; most are represented by one species that may or may not have diverged from its mainland ancestor. Of the seven reptile lineages that have colonized the islands, four have diversified, and of the approximately 1,000 genera of insects, only five percent have diversified (Parent et al. 2008). Among the insect genera that have diversified, none have generated more than twenty species (Peck 2006, Parent et al. 2008). In contrast, the land snail genus, *Bulimulus*, has differentiated into 71 different species, but the other nine genera of land snails have diversified into a maximum of four species (Parent and Crespi 2006, Parent et al. 2008). Why some diversify and others do not is not well understood.

Among the most remarkable and well-studied diversifications are *Bulimulus*, tortoises, mocking birds, *Galapaganus* (weevils), *Galagete* (microlepidoptera), and finches (Sato et al.

2001, Caccone et al. 2002, Ciofi et al. 2002, Schmitz et al. 2007, Parent et al. 2008, Sequeira et al. 2008a, Sequeira et al. 2008b). The first three (*Bulimulus,* tortoises, and mockingbirds) generally follow the progression rule, which says that the most ancient species are found on the oldest islands, while the most derived species are found on the youngest islands (Parent et al. 2008). The oldest species in these groups are found in the southeast while the youngest species are found in the Northwest (Darwin and Wolf) or Central West (Isabela and Fernandina). Because the Galápagos do not form a straight chain of islands and are found in clusters with islands of similar age, these species do not follow the progression rule as strictly as other species in more linear archipelagos, such as Hawaii (White et al. 1993, Funk and Wagner 1995, Parent and Crespi 2006, Parent et al. 2008). The other three groups (*Galapaganus, Galagete,* and finches) do not follow the progression rule, and these differences cannot be attributed to phylum, vagility, or age of divergence (Table 2).

Table 2. Diversifications of groups in the Galápagos Archipelago. (Parent et al. 2008)

Taxa	Number of species or subspecies	Time of divergence within Galápagos	Number of Colonizations	Follows progression rule
Bulimulus (1)	71	?	1	Yes
Rice Rats (2)	8	?	3	?
Galápagos giant tortoise (3)	15	1.5-2.0	1	Yes
Darwin's finches (4)	14	1.6	1	No
Mockingbirds (5)	4	?	1	Yes
Galagete (6)	12	2.9-3.7	1	No
Galapaganus (7)	10	10.7-12.1	1	No

1: (Parent and Crespi 2006); 2: (Clark 1984); 3: (Caccone et al. 2002, Ciofi et al. 2002); 4: (Sato et al. 2001); 5: (Arbogast et al. 2006); 6: (Schmitz et al. 2007); 7: (Sequeira et al. 2008a, Sequeira et al. 2008b)

The divergences are due to two types of speciation: inter-island and intra-island. Inter-island speciation occurs when individuals from a species colonize another island and over time accumulate enough changes to be considered a separate species. Intra-island speciation occurs when populations diverge on the same island either in allopatry or sympatry. While many of the divergences are due to inter-island speciation, some of the less vagile species, such as *Bulimulus*, have also diverged *in-situ* within the island they inhabit (Parent and Crespi 2006, Parent et al. 2008). This is especially true on Isabela because it is made up of multiple volcanoes, is large in comparison to the other islands, and contains a variety of habitats. On an island, the populations of less motile species are often isolated by small barriers, such as lava flows, while the more vagile species often display gene flow between populations on the same island (Parent et al. 2008). Potential instances of intra-island speciation have been found in *Galapaganus* and *Bulimulus*, but in some cases, sympatry could also be secondary contact (Finston and Peck 1995, Parent and Crespi 2006, Parent et al. 2008, Sequeira et al. 2008a, Sequeira et al. 2008b).

Methods of Biogeographic Analysis

To be able to answer questions about the dispersal, extinction, and speciation of organisms in the Galápagos or anywhere, the biogeography of the area needs to be considered. While in biology, biogeography is commonly defined as the study of the distribution of species, the methods employed in the study use the information about the species to understand the relationships between different areas. This information can then be used for many different purposes, including better understanding the movement of continents and the establishment of conservation areas (Morrone 2009). In this study, the methods of evolutionary biogeography are used to understand how the genus *Stomion* and other groups came to be distributed in the patterns seen today.

Within the field of evolutionary biogeography, there are a wide variety of methods due in part to the origins of the study of biogeography. It is an interdisciplinary field because of the variety of data that can be used, including geology, ecology, and molecular biology. Furthermore, it can be considered at many different levels from continents to populations to organisms (Morrone 2009). Because of the differences in scale, aims and subjects of biogeographic studies the methods developed are often unique to each level of inquiry and attempts to generalize the methods are often unsuccessful. The two approaches that seem appropriate to understand the evolutionary biogeography of species in the Galápagos and will be used in this study are ancestral areas and area cladograms.

To be able to interpret the colonization path of *Stomion*, the ancestral area or range was reconstructed at each node. This information can be used to determine on which island the radiation started and from where and to where the members of the genus dispersed at each node in the tree. The first model used is a character model implemented by Mesquite (Maddison and Maddison 2011). It treats the areas as character states and calculates which of the areas could be reconstructed as an ancestral state at each node. To accomplish this, the program produces the proportional likelihood of the group being in each region at each node. While it is useful to establish the likelihoods of each region, it does not account for the possibility of a widespread range. At each node, the group can only be found in one area and to be found in another it must go extinct in the original area. This model is most effective when the characters are mutually exclusive (e.g. winged and wingless) but not when they are compatible (e.g. a species that is found on two islands).

To correct this problem, the dispersal extinction cladogenesis (DEC) model was created and later implemented in the software Lagrange (Ree and Smith 2008). It is designed to use

geographic ranges instead of geographic locations as characters and conducts "a likelihood analysis of geographic range evolution" (Ree and Smith 2008). Instead of assuming that derived taxa must have the same state as the ancestor, it allows the range to be divided and allows the branching taxa to inherit non-identical ranges. The model also incorporates information about how the geographic areas are distributed. Nevertheless, the model is still based on unrealistic assumptions. For example, it assumes that the geographic range evolution and speciation are separate. This is especially problematic on islands where the dispersal to another island will often result in allopatric speciation and thus a new lineage not a range expansion. Under a better model (yet to be developed), dispersal would occur only at the nodes not over branches as in the current model so that they represent lineage divergence. Additionally, ancestral ranges would be limited to one area so that the ancestral lineage would retain the original area, and the new lineage would colonize a new area (Ree and Smith 2008, Ree and Sanmartin 2009).

In the second approach to evolutionary biogeography, area cladograms are obtained by using the phylogeny of a group and the areas of endemism. First, a taxon-area cladogram is produced by replacing the taxa on the phylogeny with the areas they inhabit. This is also the area cladogram for a group when each species is found only in one area and none are found in the same areas. If these criteria are not met and there are widespread taxa, redundant distribution or missing areas, then the taxon-area cladogram must be "resolved" to produce an area cladogram. Within the field of cladistic biogeography, there are multiple different ways of "resolving" the cladograms in order to extract a general trend (Morrone 2009). One method is known as 3item analysis (Ducasse et al. 2008). It uses two methods, one to resolve the presence of multiple areas of endemism for a single terminal taxa (MASTs) and one to resolve the repetition of areas of endemism in two or more taxa in the cladogram (paralogy). MASTs are

common when a species is widely distributed, and paralogy is common when species occur in sympatry and thus have overlapping ranges (Ducasse et al. 2008). The area cladograms of many different groups can be combined to produce one general area cladogram; however, due to the differences in the cladograms, too much information would have been lost in the process.

Objectives, questions, and predictions specific to this study

In order to explore the geographic pattern of species formation within *Stomion* and provide a time scale for speciation events, we considered the questions detailed below. Furthermore, to place *Stomion*'s radiation in a broader context, we also explored the geographic origins of other invertebrate groups in the Galápagos Islands.

Table 3. Questions and Predictions.

Colonization History	
Does *Stomion*'s colonization history, follow the progression rule?	If the islands were colonized according to their geologic age, then individuals from Santiago would be basally located on the tree and younger islands would be colonized from that source.
Did *Stomion* colonize the islands more than once?	If all members of the genus *Stomion* are the product of a radiation following a single colonization of the islands, then the genus would be reconstructed as monophyletic (sharing a single ancestor). If the genus is the product of multiple colonizations, then the genus would be reconstructed as polyphyletic (not sharing a single ancestor).
Timing and Mode of Speciation	
Is *Stomion*'s diversification older or younger than the age of the islands?	If the initial divergence is much younger than the islands, then all the divergence time estimates would fall within the geologic age of the islands and colonization and divergence patterns on the tree would probably only be due to geographic proximity.
Did *Stomion* diversify due to inter or intra island speciation?	If some of the species diversity is the result of intra-island speciation, then species on the same island would be each other's closest relatives on the phylogenetic tree.
Areas of Ancestry in *Stomion* and other Galápagos groups	
Is the origin of the colonizers of the islands South America?	If the origin of the colonizers is South America then the southeastern portion of the archipelago should harbor the most ancient divergences.

Our study of *Stomion* is based on the analysis of five mitochondrial and two nuclear gene regions from 25 individuals in the genus *Stomion*. Representing six of the eleven putative species, the beetles were collected from five different islands, seven different volcanoes. Based on maximum parsimony, the oldest species was found on Darwin, one of the youngest islands. In order to explain this odd colonization pattern, the age of divergence of the species has been estimated using a relaxed molecular clock and compared to the ages of the islands they inhabit. To understand the radiation of *Stomion* across the archipelago (biogeography), the ancestral states were reconstructed at each of the nodes using two methods, character reconstruction and

the dispersal-extinction-cladogenesis model. *Stomion*'s pattern of colonization was compared to that of other invertebrate species found in the Galápagos (*Bulimulus*, *Galapaganus*, *Galagete*, and *Hogna*) using area cladograms.

MATERIALS AND METHODS

Sample Collection

Beetle samples of *Stomion* were collected from the islands of Darwin, Pinta, Genovesa, Santiago, and Isabela (on three different volcanoes, Alcedo, Sierra Negra, and Darwin) in 2006 and 2007 by other lab members (Figure 3). Specimens were collected by hand from the soil or off of the host plant. The beetles were preserved in 100% ethanol in vials, one for each locality. Each individual beetle was labeled with "S", a two letter code from the island on which they were collected (DA = Darwin; PI = Pinta; GA = Genovesa; SA = Santiago; IS = Isabela), the number of the locality, and a letter for each individual. Between one and four specimens were analyzed from each locality (Table 4). Specimens were preserved at -20° C in 100% ethanol until they were processed.

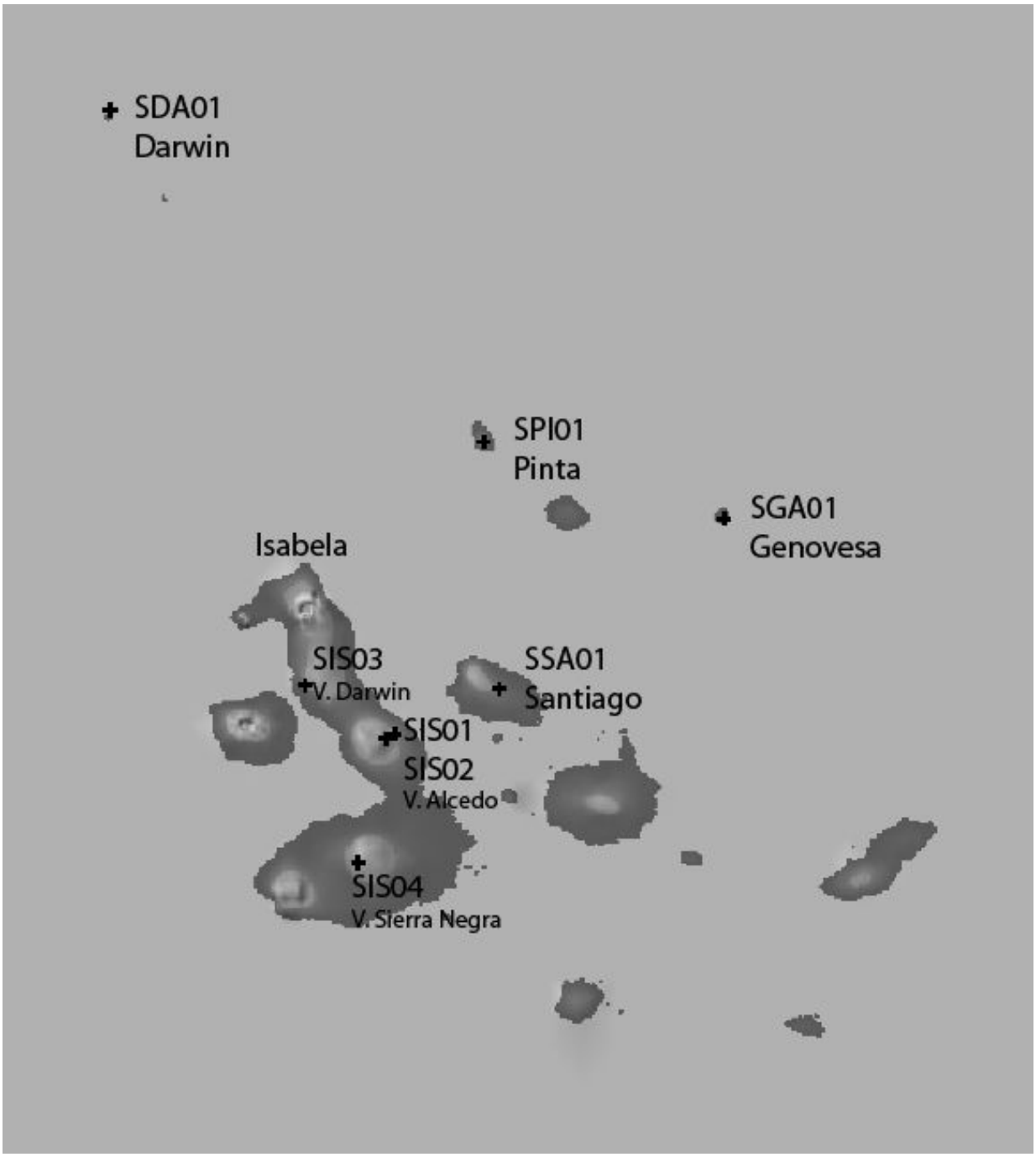

Figure 3. Digital Elevation Map of the Galápagos Archipelago with *Stomion* collection localities. The map was obtained from the Oak Ridge National Laboratory Distributed Active Archive. The localities were mapped using GPS data in GenGIS (Parks et al. 2009).

Table 4. *Stomion* collection localities. Island, population code, locality, altitude and number of specimens collected are given for each species.

Species	Island	Population Code, Locality, and Altitude	N
S. cribicollis	Darwin	SDA01, Darwin, 14m	4
S. rugosum	Pinta	SPI01, Pinta, 308m	4
S. linelli	Genovesa	SGA01, Genovesa, 200m	2
	Santiago	SSA01, La Bomba, 1m	3
S. genovesa	Genovesa	SGA01, Genovesa, 200m	2
S. laevigatum	Isabela, Volcan Darwin	SIS03, Campamento, 306m	4
	Isabela, Volcan Alcedo	SIS02, Los Guayabillos, 872m	3
Stomion sp.	Isabela, Volcan Sierra Negra	SIS04, Crater, 995m	2
	Isabela, Volcan Alcedo	SIS01, Los Pega Pega, 493m	1

DNA extraction and sequencing

DNA was extracted alternatively from two legs, three legs, head, or thorax using the DNeasy blood & tissue kit (QIAGEN). Polymerase chain reaction was used to amplify two nuclear gene regions, internal transcribed spacer (ITS) and the D2–D3 segment of the nuclear large ribosomal subunit (28S), and five mitochondrial gene regions, Cytochrome c Oxidase I (COI), Cytochrome c Oxidase II (COII), small mitochondrial ribosomal subunit (12S), Cytochrome B (CytB), and large mitochondrial ribosomal subunit (16S). Amplification reactions included 0.2μM of each primer (Table 5), 0.8 mM of dNTPs, between 0.5 and 2.5 mM of magnesium chloride, buffer, and 0.025 U Taq polymerase (NEB). The temperature profile for COI, COII, 12S, CytB consisted of 33 cycles: three cycle at 94°C for 1 min, 47°C for 30 s and 72°C for 1 min 30 s, followed by 30 cycles of 94°C for 1 min 50°C for 30 s 72°C 1 min 30 s. For the final extension step, the temperature was held at 72 °C for 5 min. The temperature profile for 16S consisted of 39 cycles: 95°C for 1 min followed by 39 cycles of 95°C for 30 s, 47°C for 2 min, 72°C for 2 min 30 s. For the final extension step, the temperature was held at 72 °C for 3 min. The temperature profile for 28S consisted of 39 cycles: 95°C for 10 minutes,

followed by 39 cycle of 95°C for 30s, 47°C for 1 min, 72°C for 1 min 30s. The temperature profile for ITS consisted of 34 cycles: one cycle of 94°C for 30 s, 60°C for 1 min, and 72°C for 40 s; the annealing temperature was dropped by two degrees for each of the next 6 cycles, and the last 27 cycles consisted of 94°C for 30 s, 46°C for 1 min, and 72 for 40 s. Amplification success was confirmed by running 10% of the reaction mix on a 1.5% agarose gel with 1% ethidium bromide at 100v, 60mA, for 20 minutes and visualized under UV light. The amplified DNA was purified using a MinElute PCR purification kit (QIAGEN). Purified products were sent out for sequencing, using the PCR primers, to University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility.

Table 5. Primers used for PCR amplification and sequences of mitochondrial and nuclear gene regions.

Gene Region	Primer Name	Sequence
ITS (1)	ITS6F	5'-TAATTGCGCGTCAACTTGTG-3'
	ITS29R	5'-CCGCTACTGAGGGAATCCTA-3'
28S (2)	S3660	5'-GAGAGTTMAASAGTACGTGAAAC-3'
	A247	5'-CCTGACTTCGTCCTGACCAGGC-3'
COI (3)	S2183	5′-CAACATTTATTTTGATTTTTTGG-3′
	A2771	5′-GGATARTCAGARTAACGTCGWGGTATWC-3′
COII (4)	J3038	5'-TAATATGGCAGATTAGTGCATTGGA-3'
	N3668	5'-GCTCCACAAATTTCTGAGCA-3'
12S (5)	J14233	5'-AAGAGCGACGGGCGATGTGT-3'
	N14588	5'-AAACTAGGATTAGATACCCTATTAT-3'
Cyt B (6)	CB1	5'-TATGTACTACCATGAGGACAAATATC-3'
	CB2	5'-AATACACCTCCTAATTTATTAGGAAT-3'
16S (7)	J12585	5'-GGTCCCTTACGAATTTGAATATATCCT-3'
	N13398	5'-CGCCTGTTTAACAAAAACAT-3'

1: Modified from (Hillis and Dixon 1991); 2: Modified from (Hillis and Dixon 1991); 3: Farrell laboratory; 4: C. Linnen, Farrell Laboratory; 5: (Simon et al. 1994); 6: Modified from (Crozier and Crozier 1992, Vogler and Welsh 1997, Cryan et al. 2001); 7: (Simon et al. 1994)

Alignment and editing

Sequences were compiled, edited, and aligned in SEQUENCHER version 4.6 (Gene Codes Corporation, Ann Arbor, MI, USA). 16S, 12S, ITS, and 28S were aligned with ClustalW (Larkin et al. 2007). A matrix of 4683 characters was compiled in MacClade (Maddison and Maddison 2000) from all five mtDNA regions, 644 from COI, 770 from COII, 841 from 16S, 419 from 12S, and 616 from CytB, and two nDNA regions, 561 from ITS and 832 from 28S for 25 individuals from 8 populations within *Stomion*. The dataset was fairly complete with one exception, only 10 sequences were included for ITS. One individual from the genus Carabidae was also sequenced to serve as an outgroup. In order to find sequences from more closely related species to use as an outgroup, public databases were searched for individuals within the same clade as *Stomion*, the Eurymetopine group within Tenebrionidae (Finston and Peck 2004). Because there are no published sequences within the clade, sequences were used from individuals in two closely related clades, Pimeliine and Tentryriine (Doyen 1993). These included sequences from the genera *Onymacris* (COII, CytB) (Steckel et al. 2010), *Anatolica* (COII, CytB) (unpublished source NCBI), *Physadesmia* (COII, CytB) (Steckel et al. 2010) within Tentryriine and *Pimelia* (COI, COII, CytB, 16S) (Contreras-Diaz et al. 2003b) within Pimeliine.

Incongruence length difference (ILD) tests were done to test for the agreement between the gene regions in the dataset, one between the five mitochondrial gene regions and two nuclear gene regions, and another between the protein coding (COI, COII, CytB) and non-protein coding gene regions (Farris et al. 1994). To estimate the average sequence divergence among members of particular groups within the *Stomion* group, a pairwise divergence test was conducted in PAUP (Swofford 2002) under the uncorrected and Kimura two parameter (K2P) model (in K2P,

transitions and transversions occur at different rates and all nucleotides have the same frequencies). From this, the maximum divergence between any two individuals and the average divergence could be determined.

Evolutionary relationships among endemic taxa

Maximum Parsimony

In order to establish the relationships between the taxa, a phylogenetic tree was built using maximum parsimony. It is a character-based method, which assumes that individuals share a characteristic because they inherited it from a common ancestor and that a simpler hypothesis with fewer steps is better. In this case, the characters are bases in the sequences. The trees that require the least number of changes to explain the differences in the sequences are the results of the analysis. The simplest trees were obtained using heuristic searches completed in PAUP (Swofford 2002). A heuristic search is done using step-wise addition, rearrangement of taxa, and branch swapping. While it is commonly used, it does not guarantee the best tree because it is not possible to include every variation. The search started with a random tree to which rearrangements were made to find the shortest, optimal, minimal or most parsimonious tree. This process was repeated 100 times, each time with 1,000 rearrangements of the starting tree. All of the genes, positions, and types of changes were flatweighted meaning that they were all treated equally. The trees were summarized in a consensus tree. To estimate the reliability of the tree, bootstrapping analysis was done, with 100 replicates and 20 random addition sequences. Bootstrapping uses resampling to create pseudoreplicate datasets that can be used to see if stochastic effects are influencing the distribution of characters. The bootstrap value is produced for each branch on the tree and is a measure of the frequency at which a branch is found in the replicates (Figure 5).

Bayesian Inference

Maximum likelihood models consider the probability that a given model of evolution will generate the observed sequences and choose the tree with the highest probability. The Bayesian approach to generating phylogenies is similar to maximum likelihood because it uses a likelihood function and a predetermined model of substitution (Yang and Rannala 1997, Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003). It produces a topology, and the uncertainty for the groups on the tree. Bayesian statistics are unique in that they use prior information to inform the result by constructing a prior from the posterior distribution of the previous run. In most cases, the original prior information used to generate the trees is uninformative (flat priors). The ability to define the original priors is not used because if the priors are incorrect, the resulting phylogeny is also incorrect or could at the very least skew the results to what is expected and not what is "true." The tree is based on the hypothesis that has the highest posterior probability (final probability based on the data). In theory, this posterior probability is generated based on the likelihood and the prior probability of the hypothesis being tested, but in reality due to the uninformative nature of the prior, the probability mainly reflects the likelihood (Huelsenbeck et al. 2001, Archibald et al. 2003). Bayesian statistics also allow for the use of more complex models of sequence evolution than likelihood models.

The implementation of Bayesian model is made possible by the use of Markov Chain Monte Carlo (MCMC) simulation in the software MRBAYES (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003). The MCMC simulation, with the DNA substitution model, uses the data to generate a posterior probability distribution of trees. This probability distribution can then be summarized into a consensus tree. The substitution model chosen was

based on the results of jModelTest (yet another likelihood procedure described in the next section) (Posada 2008), which recommended the model GTR+I+G.

Estimation of divergence time

In order to estimate the colonization time of *Stomion* in the Galápagos and other divergences within the genus, the sequence data were used to co-estimate the phylogeny and time estimates under a relaxed molecular clock implemented by BEAST (Drummond et al. 2012). Before the model was run, the best model of DNA substitution and type of clock were chosen in order to set the parameters of the model.

To determine the best DNA substitution model for each gene region and the whole dataset, jModelTest was run with its default settings (Posada 2008). It tests 88 models; these include 11 substitution schemes and allow for equal or unequal base frequencies, invariable sites (+I) and rate variation among sites (+G). The best fit is chosen using likelihood scores, which are based on the maximum likelihood optimized tree generated by the program. The simplest model that fits the data is selected (Posada 2009). For ITS, the model chosen was JC (Jukes Cantor) under which all substitutions are equally likely, there is one rate of substitution, and all nucleotides occur with the same frequency. For COI and CytB, the model chosen was GTR+I+G (General Time Reversible). The GTR model allows each type of nucleotide change (n=6) to have a different rate of substitution and nucleotides to occur at different frequencies. For 28S, the TPM2uf model was chosen; under this model, there are two substitution rates for transversions and one for transitions, and the base frequencies are unequal. For 12S and 16S, the model chosen was TIM2+G, and for COII, TIM2+I+G model was chosen. The TIM2 model uses two substitution rates for transitions and two for transversions, and permits unequal base frequencies.

To determine what type of clock (strict or relaxed) should be used in the analysis, the dataset was tested to see if it fit the molecular clock hypothesis. In other words, the sequences were examined to see if they have a constant rate of mutation over time as opposed to a variable one. If the test fails to reject the null hypothesis (molecular clock hypothesis), then a strict clock that assumes a constant rate of evolution can be used. If, on the other hand, the test rejects the null hypothesis, then a relaxed clock should be used because it does not assume a constant rate of evolution over the whole phylogeny and allows for independent rates on different branches. In PAUP (Swofford 2002), a likelihood ratio test was done in which the likelihood was determined with and without a clock. The clock model chosen corresponds to the best model for DNA substitution for the whole dataset. The model chosen by jModelTest (Posada 2008) was GTR+I+G. It is the most complex DNA substitution model, where none of the rates are equal, there are invariant sites, and rates are allowed to vary between sites (Posada 2009). Using the two likelihood values to estimate a chi square statistic (2(likelihood with clock – likelihood without clock)) and the number of taxa included in the analysis (n-2) as the degrees of freedom, a p–value was determined (Posada 2003). The same was done for the entire dataset and for the *Stomion* taxa without the outgroups. The use of an outgroup allows taxa to be compared to a more distantly related species to determine if individuals have accumulated the same number of changes. At the same time, if the outgroup is too distant, then the test in not as accurate (Posada 2003). The molecular clock was rejected for the whole data set ($p = 0.001$) but was not rejected without the outgroups ($p = 0.25$). Because the strict clock was rejected for the whole dataset, a relaxed clock was used for the remaining analysis.

To estimate the time of the original colonization and times of divergence within *Stomion*, a Bayesian relaxed molecular clock was implemented in BEAST (Drummond et al. 2012). The

run file was generated in Beauti (Drummond et al. 2012). The seven gene regions were imported separately, and each used the DNA substitution model that was closest to the best fit determined by jModelTest (Posada 2008) (not all models in jModelTest are options in Beauti). The clock models and tree parameters were the same for each region. Each of the labeled nodes was specified as a taxa set so that the median and 95% confidence intervals were calculated and included in the results file. All other operators were left at their default settings.

To ensure that the ages estimated under the model are supported by a sufficient amount of data, the program also calculates the effective sample size (ESS) of the prior and posterior probability distributions. These values are representative of 'the number of independent draws from the posterior distribution of the markov chain' (Drummond et al. 2007). Ideally, to be considered well supported the ESS needs to be greater than 200 (Drummond et al. 2007). The sample size is influenced in part by the number of generations for which the model is run. In this case, it was run for 100 million generations.

To be able to estimate the divergence times on the tree structure, priors are required to estimate the parameters of the model. The starting tree was randomly generated using a Yule tree prior. This type of prior is used for species level phylogenies, and assumes a constant speciation rate since the species are closely related (Drummond et al. 2007). The ucld.mean prior was set as uniform from 0 to MAX because there was no information to make a more accurate estimate about the mean of the branch rates. In addition to setting the priors for the tree and branches, the model also needs an approximate age for at least one node in the tree as a starting point to estimate the ages of other node in the tree. Generally, the most reliable and accurate source for this information is the age of a fossil. As previously noted, there are no known *Stomion* fossils; therefore, three different dating schemes were used to set the priors for

different nodes in the trees as mentioned in the previous section: 1. Ages of the islands; 2. Known rate of mitochondrial DNA mutation in arthropods; 3. Age of the outgroup. Details of the implementation are provided below.

1. The ages of the islands were used as an approximation for the ages of the clades, and were used as priors for three nodes (A, H, and I) (Figure 7). At A, the origin of *Stomion*, the maximum (6.3 my) and minimum (2.3 my) ages of the oldest extant island were used. At H, the minimum (0.7 my) and maximum (2.4 my) ages of Santiago were used. At I, the maximum (0.7 my) and minimum ages of Isabela (0.07 my) were used (Table 6). A total of 21 permutations were run using either a normal or uniform distribution for the prior. For those with a normal distribution of the ages, either minimum or maximum ages of the islands was used. For those with a uniform distribution of the ages, the maximum age and zero were used as the extremes meaning that the divergences could not be older than the islands but could be younger. For all of the methods, the analysis was run three times using one, two, or all of the three of the nodes to calibrate the age (Table 6).

2. The rate of pairwise sequence divergence per million years has been estimated to be 2.3% for the mitochondrial DNA of arthropods (Brower 1994). This means that the rate of mutation for one lineage is 1.15% per million years. This was used as the mean for a normal prior to describe the mean rate of divergence with a relaxed clock model. In a second run, it was also used as the mean rate under a strict clock model.

3. The age of the *Pimelia* genus, which was used as an outgroup to root the tree, has been estimated (Juan et al. 1995b, Contreras-Diaz et al. 2003b). The origin of *Pimelia* (9.3-10.7 my) and the divergence of *P. granulicollis* (3.6 - 4.2my) were used to externally calibrate the *Stomion* divergences.

A maximum clade credibility tree with branch and node positions representing the median estimates and excluding the first 1,000 trees was constructed in TreeAnnotator v1.6.2 (Drummond et al. 2012) and visualized in FigTree v.1.3.1 (Rambaut 2009). In Tracer v1.5 (Rambaut and Drummond 2007), the medians and 95% confidence intervals were recorded for nodes A-L with a burn-in of 10% of the length of run, 10 million generations (Tables 8 - 11).

Table 6. Ages of the islands used to calibrate BEAST runs.

Nodes Constrained	Maximum Age	Standard Deviation	Minimum Age	Standard Deviation	Uniform Range
A	6.3	0.1	2.3	0.1	0-6.3
H	2.4	0.1	0.7	0.1	0-2.4
I	0.7	0.1	0.07	0.03	0-0.7
All *Pimelia*	--	--	--	--	9.3-10.7
Divergence of *P. granulicollis*	--	--	--	--	3.6-4.2

Ancestral Areas in the Archipelago

To determine the most likely location of *Stomion* at the nodes in the tree, the ancestral areas of the taxa were reconstructed using two different methods, a character model and a dispersal extinction and cladogenesis model, also referred to as the DEC model.

The character model was implemented in Mesquite (Maddison and Maddison 2011), and the characters were mapped onto the three topologies (Bayes tree, two trees from BEAST runs, BEAST1 and BEAST2). Using the characters (area of the archipelago) of the extant taxa and the given topology, the states were reconstructed using maximum likelihood methods under the mk1 model. This model is the simplest and has one rate of change and assumes that any change is equally probable. The states mapped onto the tree correspond to the regions in the archipelago where the taxa are found and the outgroup (Figure 4). The archipelago was divided into areas grouping islands based on age and geography. Northwest contained Darwin and Wolf; Central North contained Pinta, Marchena, and Genovesa; Central West contained Fernandina and

Isabela; Central contained Santiago and Santa Cruz; Southeast contained Española, Floreana, and San Cristóbal (No taxa from the Southeast were available for this study) (Figure 4). The proportional likelihoods of each area were calculated at all nodes and are displayed in Table 12 for six nodes of interest.

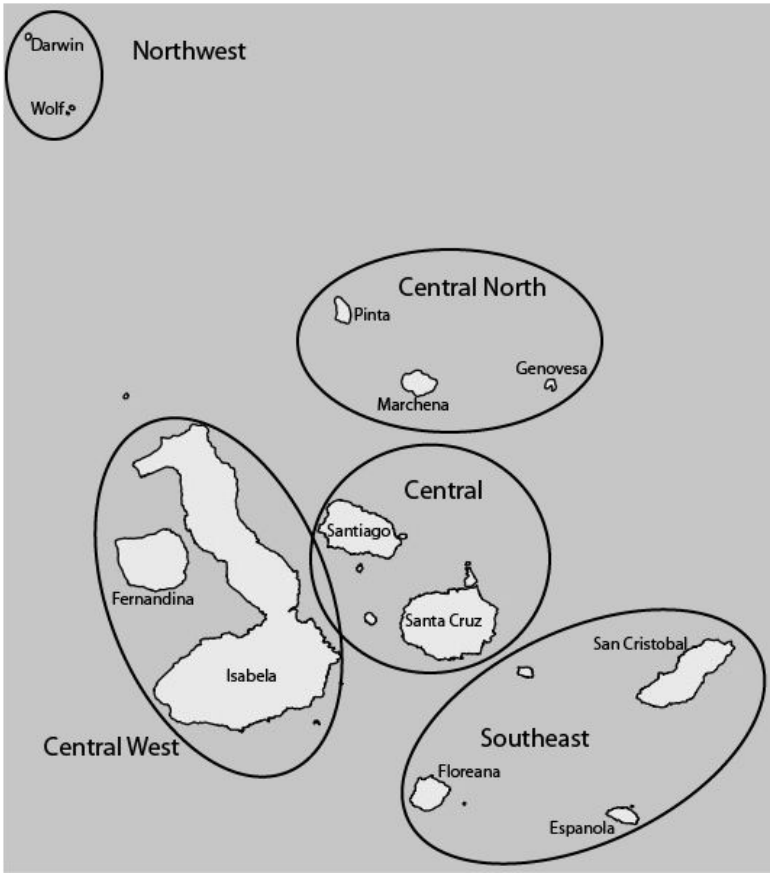

Figure 4. Regions of the Galápagos Archipelago. Islands were divided into five regions based on their age and geographic locations and named according to their location in the archipelago.

While the DEC model and the character model can both be used to reconstruct the ancestral areas, the DEC model is designed specifically to consider the ancestral ranges as opposed to being a general character model. The likelihood analysis of geographic range evolution implemented by Lagrange-2011 reconstructs a range of areas at each node instead of a single area (Ree and Smith 2008, Ree and Sanmartin 2009). Furthermore, Lagrange allows for

the inheritance of a widespread ancestral range and for range evolution by dispersal, extinction, and cladogenesis. The areas were defined following Figure 4. The model constrained the ancestral ranges to a maximum of four areas (all regions of the archipelago from which *Stomion* taxa were included in this study) and excluded the possibility of having a species found only in two non-adjacent regions. In the first permutation of the analysis, dispersal rates between all of the islands were assumed to be equal. In the second permutation of the analysis, dispersal rates between adjacent areas were doubled to 2.0, rates between non-adjacent areas were left at 1.0, rates from archipelago to the outgroup were set at 0, and rates from the outgroup to Central North and Central were set to 2.0 (Table 7). The baseline rates of dispersal and local extinction were estimated by maximum likelihood. The program required an ultrametric tree (a topology with branch length proportional to their age but all tips are equidistant from the root of the tree), which was obtained from a BEAST run.

Table 7. Dispersal rate matrix for Lagrange analysis. The rate of dispersal is show as being from the areas in the top row to the areas in the columns. 0 represents no dispersal from one area to another, and increasing numbers represent increasing rates of dispersal.

	Outgroup	Northwest	Central North	Central	Central West
Outgroup	--	1.0	2.0	2.0	1.0
Northwest	0.0	--	2.0	1.0	2.0
Central North	0.0	2.0	--	2.0	2.0
Central	0.0	1.0	2.0	--	2.0
Central West	0.0	2.0	2.0	2.0	--

Biogeographic Analysis

In order to determine whether there are commonalities in the colonization histories of similar species, the four most thoroughly studied invertebrate groups (with available

phylogenies) and *Stomion* were compared. These groups include the weevil genus, *Galapaganus*, (Sequeira et al. 2008a, Sequeira et al. 2008b), the snail genus, *Bulimulus* (Parent and Crespi 2006), the spider genus, *Hogna* (De Busschere et al. 2010), and the microlepidopteran genus, *Galagete* (Schmitz et al. 2007). To be able to effectively compare the patterns, taxon-area cladograms were constructed for each of the five groups using 3area, a three-item analysis biogeography program that is a part of the Lisbeth package (Ducasse et al. 2008). 3area uses the cladograms of the taxa and replaces the terminal taxa with areas of endemism. The islands were divided into the same five regions used for both Mesquite and Lagrange (Figure 4). No specimens from the Southeast are available for *Stomion*. While multiple taxa can be found in the same location, the program cannot effectively interpret widespread taxa (taxa found in many different areas of endemism, also known as MASTs) (Ebach et al. 2005). Generally, all of the islands to which a species is endemic would be defined as one area (Morrone 1994). In these genera, this could not be done because the widespread distributions of many species would have resulted in the loss of most of the resolution. Taxic paralogies (different taxa in a cladogram that are found in the same location) are identified because they are considered uninformative (Ebach et al. 2005). Paralogy-free subtree analysis is then applied and MAST-free, paralogy-free subtrees are produced (Ducasse et al. 2008). To produce these subtrees, the program starts at the tips of the trees and takes out parts of the tree that do not have taxa found in the same areas. These trees were analyzed in 3ia (another part of the Lisbeth package) to produce one intersection tree (Ducasse et al. 2008).

RESULTS

Agreement between mitochondrial and nuclear phylogenetic signal

No significant conflict was found between the mitochondrial and nuclear gene regions or between the protein and non-protein coding regions. As a result, the regions were concatenated for the phylogenetic analysis. The average pairwise genetic difference within *Stomion* calculated under an uncorrected and K2P-model was 2.7% and 2.8% respectively. The maximum divergence was approximately 5% within the genus. The average divergence between *Stomion* individuals and members of the outgroup was much larger (17.7% and 20.4% under the two models).

Phylogenetic Relationships and the Colonization History of *Stomion*

Based on the outgroups and taxa included in the analysis, the genus is monophyletic, which suggests that the genus is the result of one introduction into the islands as predicted (Table 3). This is seen in all models used to construct the trees, including maximum parsimony (Figure 5), BEAST, which produced two topologies BEAST1 and BEAST2 (Figure 7), and Bayesian (model: GTR+I+G) (Figure 10). The maximum parsimony analysis produced 36 most parsimonious trees (L=2165 steps). Out of the 4683 characters, 590 were parsimoniously informative. While *Stomion* is monophyletic in all of the topologies, there are slight differences in the branching order of some groups. In the maximum parsimony model, the basally located species (*S. cribicollis*) is found on one of the younger islands, Darwin (Figure 5). Darwin is also one of the furthest islands from the South American mainland, which is thought to be the source of *Stomion*'s common ancestor. In both the Bayesian model and one of the trees obtained using BEAST (BEAST 1), the taxa from Isabela are the basally located species (Figure 10; Figure 7). The two topologies (BAYES and BEAST1) do not agree as to whether the groups are mono or

paraphyletic. The last alternative is presented in the BEAST 2, which shows the taxa on Darwin and Isabela as sister groups along the first branch to diverge within *Stomion* (Figure 7).

With the exception of the disagreement between the trees about the basal location of taxa from Darwin or Isabela and some unresolved nodes in the maximum parsimony tree, the topologies are identical. Due to the relatively young age of both Darwin and Isabela, the age of *Stomion*'s divergence in the archipelago is needed to fully understand the colonization history of *Stomion* in the Galápagos. Nevertheless, the colonization pattern is already in contrast to that expected under the progression rule because the earliest branchings are to the younger islands, Darwin and Isabela, instead of the older island, Santiago.

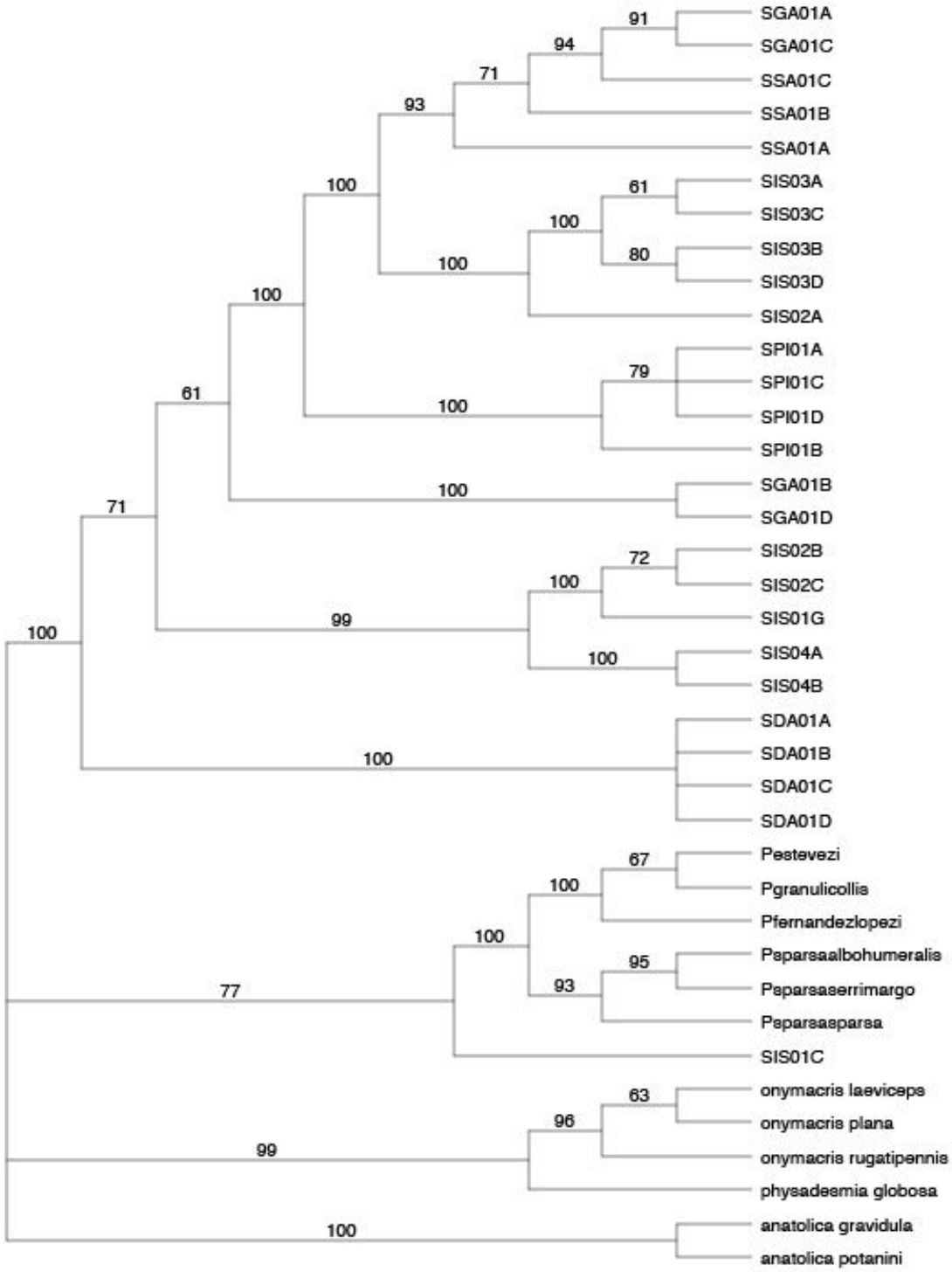

Figure 5. Strict consensus of 36 most parsimonious trees of all *Stomion* taxa and outgroups. All gene regions were used and the run was unweighted. *Pimelia*, *Onymacris*, *Physadesmia*, *Anatolica* and SIS01C (Carabidae) were used as outgroups. Bootstrap values are given above the branches.

Taxonomic and Phylogenetic Agreement

All the specimens identified as members of the each of the five species group together on trees except for *S. laevigatum*. The specimens identified as *S. laevigatum* are found on two different positions on the tree even though some were collected from the same locality (Figure 6). *S. linelli* is found on Santiago and Genovesa and all of the taxa from that species are each other's closest relatives. The specimens labeled SIS01G, SIS04A, and SIS04B do not have species designations, and at this time, they cannot be assigned to a species based on their placement on the tree because of disagreements in the topologies (Figure 6). With the exception of *S. laevigatum* and the undesignated specimens, our molecular phylogenies indicate a strong agreement between phylogenetic placement and taxonomic designations (Finston and Peck 2004). The clear species boundaries recovered from mitochondrial and nuclear data argue against the possibility of inter-species hybridization within the genus.

Modes of Speciation

To determine the most common type of speciation (inter v. intra), the placement of taxa from the same islands was considered. The taxa from Genovesa are not grouped together on any of the topologies. They belong to two different species, *S. linelli* and *S. genovesa*. Because the species are not each other's closest relatives, their sympatry is probably the product of two different colonizations of the island (Figure 10). Similarly, the phylogeny indicates that the island of Isabela has been colonized at least twice (Figures 10-12) because the inhabitants of Isabela are not each other's closest relatives within *Stomion*. There are no instances of intra-island speciation, and as a result, it appears that all the diversity has been generated through dispersal and colonization of neighboring islands (inter-island speciation).

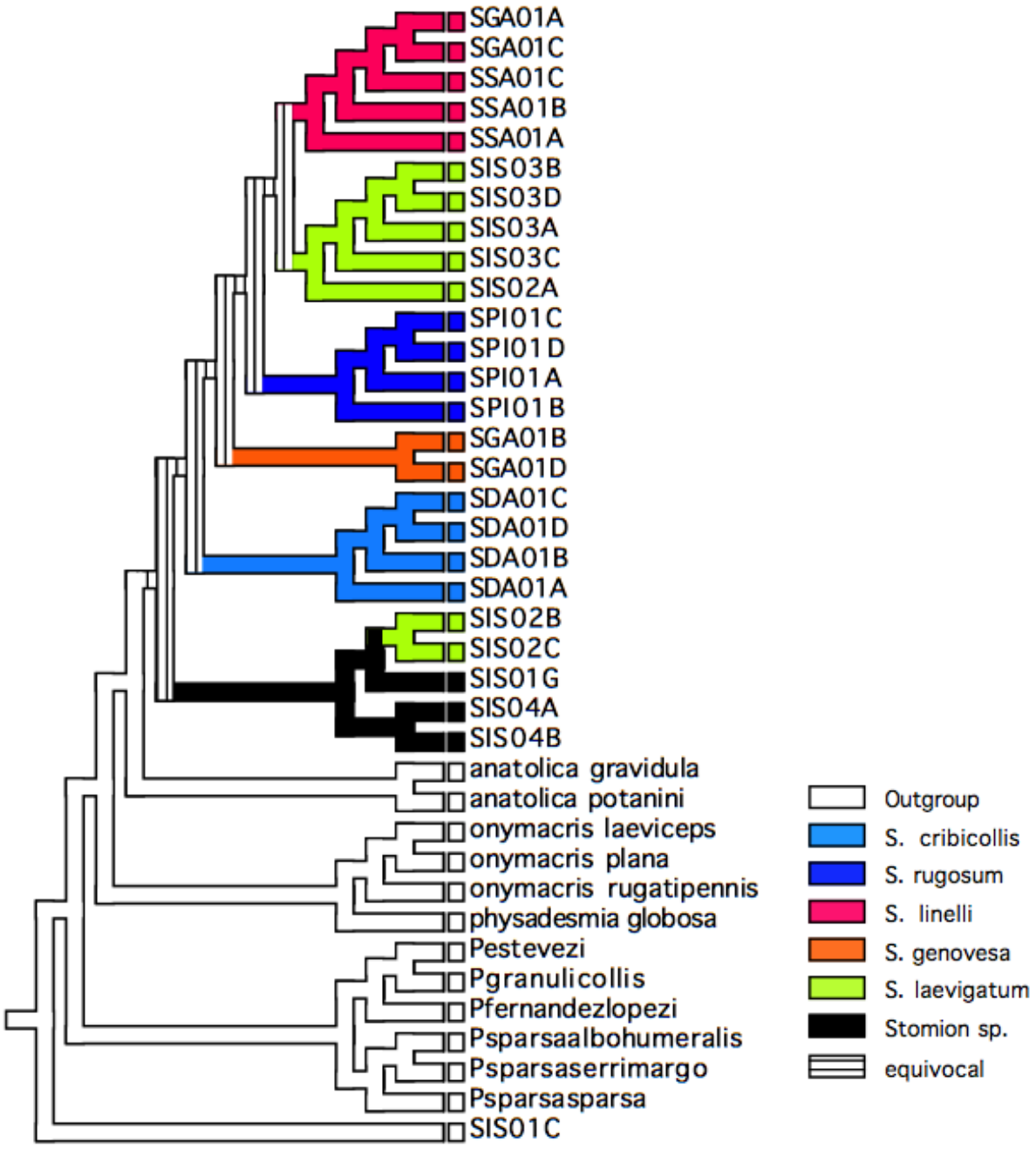

Figure 6. One of the maximum likelihood topologies with species designations. Details of the model and run specifications are provided in the materials and methods section.

Estimation of Divergence Times

Assessment of and Confidence in age estimates

The 25 permutations with different constraints led to widely divergent time estimates for all nodes (Tables 8-11; Figure 7-9). We will argue that despite the higher estimates, *Stomion*'s divergence in the Galápagos Archipelago is well within the age range of the extant islands.

When the tree is calibrated so that the nodes must be younger than the islands using uniform priors, the age of the first divergence of *Stomion* is approximately 0.015 my (15,00 years) (Table 8). 15,000 years is a very recent estimate in evolutionary terms and considering the average percentage of sequence divergence within the genus (2.7%) and the well defined morphological differences in the species (Figure 2), this scenario is considered very unlikely and attributed to an unrealistic prior.

The estimates obtained using the maximum age in a normal distribution all resulted in ages of approximately 6 my for the origin of *Stomion*, except for that in which only the youngest node (I) was used to calibrate the divergence (Figure 7; Table 9). In that run, the age of *Stomion* was approximately 2 my (Table 9). 6 my is the maximum age of the oldest extant island; however, no individuals from those islands are included in this study; the oldest island included in this analysis is 2.4 my old (Table 1). Even the divergences within the islands are older than the maximum age of the islands (Table 9). This prior is thus thought to overestimate the ages of the divergences. Furthermore, the maximum age estimates, on which the calibrations were based, are considered to be an overestimation of the age even by geologists (Geist 1996). The run that constrained only node I is consistently much younger than any of the other estimates (both when the maximum and minimum ages were used) implying that the age used to constrain that node may not be consistent with the ages used at other nodes (Tables 9-10). Because it is a

divergence found at the base of the tree and found on an island younger than the colonization, it is possible that the divergence is older than the island that the taxa currently inhabit (Figures 7-8). The run using the calibration of node I was thus not considered to be an accurate estimate and will not be included in further discussion.

The ages for node A (colonization of archipelago) obtained using the minimum age as the mean for a normal prior, were between 1 and 2.3my (Table 8; Figure 8). Considering the other information about the morphological and genetic divergence, these estimates were considered to be the most representative. Furthermore they were consistent with the runs using the rate of mtDNA mutation, which produced ages of between 1.4 and 1.8 my for node A (Table 9B; Figure 9B).

The other runs with external calibrations used the ages of divergences within the *Pimelia* outgroup (Table 11A; Figure 9A.). These two runs produced ages of either 4.9 or 8.3 my for the origin of *Stomion* depending on which node within the divergence of *Pimelia* was incorporated (Table 11A). The later is older than the age of the oldest extant island. These estimates are based on the calibration of *Pimelia* based on the age of the Canary Islands and the rate of mtDNA mutation since there are no fossils available. Thus, calibrating the tree based on these data is introducing the error of this dating in addition to the error in the dating of our tree.

As a result, the two estimation schemes that are most consistent with other sources of information about *Stomion* and that will be used to draw conclusions from are: the dating using the rate of mutation in mtDNA of arthropods and the minimum age of the islands. Furthermore, to avoid oversimplification, the interpretation of the results is based on the medians but includes the 95% confidence intervals (Tables 8-11).

To see whether the values for each node converge and the run provides accurate estimates through thorough sampling, the program also estimates the effective sample size (ESS) of the posterior and prior probability distribution. For runs of 100 million generations, the ESS values are between 25 and 45, which is well below the desired threshold of 200. When a run was extended to 1E9 generations, the ESS values were above 200 and the ages of the nodes had changed only in the third decimal place (data not shown). Due to limits in computational power, the runs were not extended to 1E9 generations for all of the permutations of the analysis. Given the small variation in the estimates generated by increased sampling we are confident that the estimates provide an accurate picture.

Age Estimates with external calibrations (mtDNA)

Because the tree calibrated using the rate of mtDNA mutation is much younger than the age of the oldest extant island, many of the nodes were younger than the ages of the islands that the taxa inhabit (Figure 9; Table 11). All of the nodes that represent divergences within an island (C, E, G, I, J, L) have median ages below the maximum age of the islands (Table 9). The divergences between taxa on different islands either have median ages within the ages of the older islands (H,K) or have median ages that are older than the islands that the taxa inhabit, but the maximum age of the islands is still found in the lower end 95% confidence interval (A, B, D, F) (Table 11). Under this divergence time estimate, it is possible for species to have diverged in the islands that they inhabit without having to invoke a scenario of extinction in the source area and later colonization of their present island.

Age estimates with minimum age of the islands

1) Divergences to Darwin and Isabela (nodes B and B1)

The colonization and first divergences within the archipelago is reconstructed in two different ways. Either, upon colonization the first divergence is to Isabela and the second is to Darwin or the first divergence is to the sister groups in Isabela and Darwin. The divergence to Darwin in BEAST1 (node B) is dated to be between 0.8 and 1.8 my (Table 10; Figure 8A). The minimum age of Darwin is 0.4 my, so the divergence to Darwin is older than the age of the island. The first divergence in BEAST2 to Isabela and Darwin (node B1) is estimated to have a median age between 1.4 and 2.0 my (Table 10; Figure 8B). This is older than the maximum ages of both islands. The divergence within Darwin (node C) has a median age of between 0.1 and 0.3 my, which is younger than the island (Table 10; Figure 8). The divergence within Isabela (node I) has a median age of between 0.1 and 0.7 my (Table 10; Figure 8); the maximum age of the island is 0.7 my so the divergence within the island is younger than the maximum age of Isabela. In sum, while the divergences within Darwin and Isabela are younger than the age of the islands, the divergences to the islands (or colonizations) are older than the islands.

2) Divergences to Pinta and Genovesa (nodes D and F)

The first colonization of Genovesa (node D) has an estimated age between 0.8 and 2.2 my. The divergence within Genovesa (node E) has a median age between 0.1 and 0.3 my (Table 8; Figure 8). The age of the island is less than 0.7 my, so while the divergence within the island is younger than the maximum age of the island, the divergence from the ancestor (generally assumed to be the colonization) is older than the age of the island.

The colonization of Pinta (node F) has an estimated median age between 0.6 and 1.4 my. The diversification within Pinta (node G) has an estimated median age between 0.1 and 0.3 my

(Table 10; Figure 8). Again, while the divergence within the island is within the age range of the island (younger than 0.7 my), the colonization of the island or divergence from the ancestor could be older than the age of the island. A general pattern is becoming apparent in which the divergences within the islands are younger than the age of the islands but the divergences between islands are older than the maximum age of the islands

3) Divergence to Santiago and Isabela (node H)

At node H, *Stomion* colonized Santiago for the first time and Isabela for the second time. The median age for node H is between 0.5 and 0.9 my (Table 10; Figure 8). Santiago has a maximum age of 2.4 my, and Isabela has a maximum age of 0.7 my, so the divergence is at least younger than Santiago and possibly younger than Isabela. The divergence within Isabela has a median age between 0.2 and 0.5 my (Table 10; Figure 8). This is within the age range of Isabela. The divergence within Santiago has a median age between 0.3 and 0.6 my. From Santiago, *Stomion* recolonized Genovesa. The timing of the colonization cannot be determined based on the topology but the divergence within the island has an estimated median age of 0.07 and 0.2my (Table 10; Figure 8). This divergence is within the geological age range of Genovesa.

In sum, while the divergences within each island are estimated to have occurred within the age range of the island, the ages of the colonizations of Darwin and Pinta, and the first colonizations of Isabela and Genovesa are suggested to be older than the age of the extant islands based on the dating of the phylogenetic tree. Only the colonization of Santiago and second colonizations of Isabela and Genovesa are within the age of those islands. We will consider aspects of *Stomion*'s dispersal abilities and ecology to explain these dynamic colonization patterns.

Table 8. Age estimates for twelve of the nodes (A-L) for all permutations of BEAST analysis with uniform priors on internal nodes. Median values and 95% confidence intervals are displayed. Effective Sample Size (ESS) values are also shown for each run.

	Age of Nodes												ESS Values	
	A	B	C	D	E	F	G	H	I	J	K	L	Posterior	Prior
AHI	0.015 (0.005-0.051)	0.015 (0.004-0.044)	0.002 (0.001-0.007)	0.012 (0.004-0.039)	0.002 (0.0003-0.007)	0.008 (0.002-0.026)	0.002 (0.0004-0.006)	0.005 (0.002-0.017)	0.005 (0.001-0.017)	0.003 (0.001-0.011)	0.004 (0.001-0.013)	0.002 (0.0003-0.005)	35.7	34.8
AH	0.015 (0.005-0.051)	0.013 (0.004-0.044)	0.002 (0.001-0.007)	0.012 (0.004-0.039)	0.002 (0.0003-0.007)	0.008 (0.002-0.026)	0.002 (0.0004-0.006)	0.006 (0.002-0.018)	0.005 (0.001-0.017)	0.003 (0.001-0.011)	0.004 (0.001-0.013)	0.001 (0.0003-0.005)	32.9	32.0
AI	0.016 (0.005-0.051)	0.015 (0.004-0.045)	0.002 (0.001-0.008)	0.012 (0.004-0.039)	0.002 (0.0003-0.007)	0.008 (0.002-0.026)	0.002 (0.0004-0.006)	0.005 (0.002-0.018)	0.005 (0.001-0.017)	0.003 (0.001-0.011)	0.004 (0.001-0.013)	0.001 (0.0003-0.005)	28.7	28.2
HI	0.017 (0.005-0.050)	0.014 (0.004-0.043)	0.002 (0.001-0.007)	0.012 (0.004-0.038)	0.002 (0.0003-0.007)	0.008 (0.002-0.026)	0.002 (0.0004-0.006)	0.005 (0.002-0.017)	0.005 (0.001-0.017)	0.003 (0.001-0.010)	0.004 (0.001-0.013)	0.001 (0.0003-0.005)	45.3	42.9
A	0.016 (0.005-0.051)	0.014 (0.004-0.045)	0.002 (0.001-0.008)	0.012 (0.004-0.040)	0.002 (0.0004-0.007)	0.008 (0.003-0.027)	0.002 (0.0004-0.007)	0.005 (0.002-0.018)	0.005 (0.001-0.017)	0.003 (0.001-0.011)	0.004 (0.001-0.013)	0.001 (0.0003-0.005)	38.9	37.0
H	0.016 (0.005-0.052)	0.014 (0.004-0.045)	0.002 (0.0005-0.008)	0.012 (0.004-0.040)	0.002 (0.0003-0.007)	0.008 (0.002-0.027)	0.002 (0.0004-0.007)	0.005 (0.002-0.018)	0.005 (0.001-0.017)	0.003 (0.001-0.011)	0.004 (0.001-0.013)	0.001 (0.0003-0.005)	39.0	38.1
I	0.015 (0.004-0.051)	0.013 (0.004-0.045)	0.002 (0.001-0.008)	0.012 (0.004-0.040)	0.002 (0.0003-0.007)	0.008 (0.002-0.027)	0.002 (0.0004-0.006)	0.005 (0.002-0.018)	0.005 (0.001-0.017)	0.003 (0.001-0.011)	0.004 (0.001-0.013)	0.001 (0.0003-0.005)	39.3	37.5

Table 9. Age estimates for twelve of the nodes (A-L) for BEAST runs using the maximum ages of the islands as a prior. Runs produced two different topologies depending on the nodes that were used to calibrate the analysis. All nodes were the same except for B, which in the alternate topology is labeled B1. Topology containing B is shown in Figure 7a (BEAST 1), and the topology containing B1 is shown in Figure 7b (BEAST 2). Median values and 95% confidence intervals are displayed for each node. Effective Sample Size (ESS) values are also shown for each run.

	Ages of the Nodes													ESS Values	
	A	B	B1	C	D	E	F	G	H	I	J	K	L	posterior	prior
A H I	6.290 (6.092-6.484)	5.687 (4.501-6.499)	--	0.806 (0.355-1.409)	4.994 (3.819-6.474)	0.718 (0.186-1.493)	3.386 (2.617-4.393)	0.701 (0.275-1.290)	2.375 (2.184-2.569)	0.782 (0.598-0.977)	1.349 (0.731-1.985)	1.699 (1.077-2.247)	0.564 (0.174-1.009)	27.8	27.0
A H	6.294 (6.102-6.490)	--	5.342 (4.481-6.453)	0.870 (0.550-1.266)	6.205 (4.846-6.532)	0.849 (0.435-1.348)	3.785 (3.077-4.511)	0.790 (0.465-1.175)	2.399 (2.218-2.587)	1.814 (0.140-2.609)	1.423 (1.033-1.830)	1.787 (1.378-2.168)	0.735 (0.409-1.116)	31.2	30.1
A I	6.285 (6.090-6.482)	5.395 (3.856-6.499)	--	0.745 (0.289-1.367)	4.626 (3.091-6.438)	0.646 (0.550-1.422)	2.945 (1.717-4.263)	0.630 (0.234-1.206)	1.896 (1.076-2.817)	0.767 (0.582-0.957)	1.090 (0.504-1.812)	1.363 (0.696-2.127)	0.462 (0.128-0.887)	26.5	26.0
H I	5.315 (3.591-7.582)	4.966 (3.450-6.756)	--	0.726 (0.305-1.315)	4.630 (3.129-6.318)	0.658 (0.161-1.352)	3.217 (2.503-4.194)	0.650 (0.255-1.218)	2.354 (2.155-2.553)	0.785 (0.594-0.977)	1.286 (0.686-1.950)	1.632 (0.991-2.198)	0.545 (0.169-0.967)	28.2	27.3
A	6.296 (6.102-6.495)	--	5.380 (4.438-6.477)	0.874 (0.523-1.287)	6.195 (4.572-6.551)	0.838 (0.394-1.346)	3.780 (2.804-4.674)	0.788 (0.450-1.193)	2.395 (1.760-3.016)	1.808 (1.086-2.655)	1.411 (0.932-1.921)	1.768 (1.226-2.325)	0.725 (0.355-1.117)	26.1	25.6
H	6.097 (4.561-7.963)	--	5.166 (3.550-7.622)	0.850 (0.487-1.301)	5.811 (4.132-7.538)	0.818 (0.385-1.375)	3.681 (2.860-4.617)	0.770 (0.423-1.197)	2.388 (2.194-2.589)	1.761 (1.001-2.732)	1.406 (0.984-1.848)	1.764 (1.321-2.186)	0.714 (0.699-1.096)	32.8	31.5
I	2.051 (0.841-3.293)	--	1.758 (0.808-2.896)	0.284 (0.104-0.508)	1.947 (0.590-3.172)	0.272 (0.061-0.512)	1.228 (0.412-2.022)	0.256 (0.086-0.474)	0.781 (0.296-1.309)	0.653 (0.448-0.856)	0.459 (0.163-0.791)	0.578 (0.205-0.975)	0.235 (0.050-0.428)	39.1	38.1

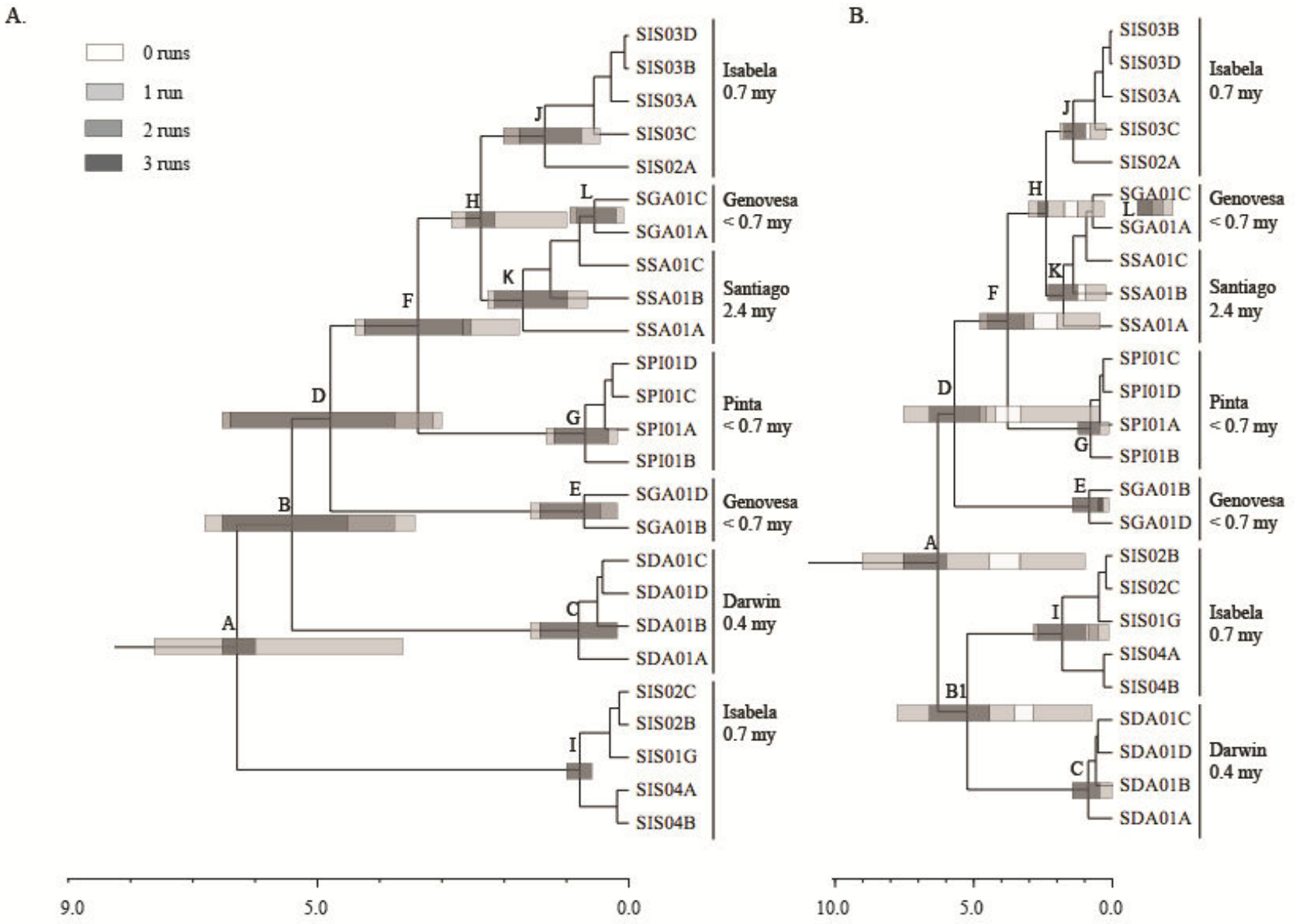
A.
0 runs
1 run
2 runs
3 runs
SIS03D
SIS03B
SIS03A
SIS03C
SIS02A
Isabela
0.7 my
SGA01C
SGA01A
Genovesa
< 0.7 my
SSA01C
SSA01B
SSA01A
Santiago
2.4 my
SPI01D
SPI01C
SPI01A
SPI01B
Pinta
< 0.7 my
SGA01D
SGA01B
Genovesa
< 0.7 my
SDA01C
SDA01D
SDA01B
SDA01A
Darwin
0.4 my
SIS02C
SIS02B
SIS01G
SIS04A
SIS04B
Isabela
0.7 my
J
H
L
K
F
D
G
E
B
A
C
I
9.0
5.0
0.0
B.
SIS03B
SIS03D
SIS03A
SIS03C
SIS02A
Isabela
0.7 my
SGA01C
SGA01A
Genovesa
< 0.7 my
SSA01C
SSA01B
SSA01A
Santiago
2.4 my
SPI01C
SPI01D
SPI01A
SPI01B
Pinta
< 0.7 my
SGA01B
SGA01D
Genovesa
< 0.7 my
SIS02B
SIS02C
SIS01G
SIS04A
SIS04B
Isabela
0.7 my
SDA01C
SDA01D
SDA01B
SDA01A
Darwin
0.4 my
J
H
L
K
F
D
G
E
A
I
B1
C
10.0
5.0
0.0

Figure 7. Ultrametric trees and ages estimates obtained with BEAST using priors based on the maximum age of the islands. The maximum age of the islands was used as the mean of a normal distribution for the priors of the nodes A, H and/or I. Depending on which nodes were calibrated two different topologies were obtained. **A.** The topology (BEAST 1) was obtained three times by calibrating the run using [A,H, and I], [A and I], and [H and I]. **B.** The topology (BEAST 2) was obtained four times by calibrating the run using nodes [A and H], [A], [H], and [I]. The 95% confidence intervals, shown in gray, were obtained for nodes A-L in A., and A, B1, C-L in B. Because multiple runs (three in A. and four in B.) were condensed into one tree, the tree with the longest branch lengths is displayed above and the confidence intervals for all runs are shown. The confidence intervals that overlap are shown with progressively darker shades of grey. The white regions in the intervals are the age ranges that are not included in confidence intervals of any runs. The ages are given in millions of years. The outgroups were included for the analysis and were removed afterwards.

Table 10. Age estimates for twelve of the nodes (A-L) for BEAST runs using the minimum ages of the islands as a prior. Runs produced two different topologies depending on the nodes that were used to calibrate the analysis. All nodes were the same except for B, which in the alternate topology is labeled B1. Topology containing B is shown in Figure 8a (BEAST 1), and the topology containing B1 is shown in Figure 8b (BEAST 2). Median values and 95% confidence intervals are displayed for each node. Effective Sample Size (ESS) values are also shown for each run.

	Age of Nodes													ESS Values	
	A	B	B1	C	D	E	F	G	H	I	J	K	L	posterior	prior
A H I	2.262 (2.067-2.458)	1.765 (1.146-2.314)		0.237 (0.072-0.529)	1.458 (0.853-2.013)	0.197 (0.026-0.552)	0.945 (0.615-1.397)	0.203 (0.054-0.462)	0.660 (0.483-0.838)	0.121 (0.075-0.169)	0.366 (0.148-0.613)	0.455 (0.229-0.680)	0.123 (0.025-0.270)	28.2	27.7
A H	2.260 (2.069-2.455)		1.958 (1.524-2.430)	0.304 (0.176-0.454)	2.156 (1.528-2.453)	0.292 (0.130-0.477)	1.276 (0.929-1.594)	0.271 (0.146-0.412)	0.774 (0.617-0.922)	0.633 (0.352-0.969)	0.468 (0.313-0.619)	0.584 (0.417-0.745)	0.241 (0.119-0.374)	43.1	40.1
A I	2.255 (2.060-2.457)	1.692 (0.979-2.302)		0.225 (0.067-0.498)	1.373 (0.689-2.032)	0.183 (0.0267-0.517)	0.855 (0.404-1.373)	0.187 (0.048-0.419)	0.547 (0.264-0.905)	0.119 (0.073-0.168)	0.311 (0.112-0.585)	0.377 (0.161-0.677)	0.106 (0.0213-0.249)	28.2	27.4
H I	0.956 (0.185-1.703)	0.829 (0.201-1.423)		0.119 (0.020-0.303)	0.745 (0.192-1.253)	0.100 (0.012-0.321)	0.577 (0.126-0.913)	0.102 (0.017-0.274)	0.461 (0.086-0.689)	0.078 (0.059-0.096)	0.214 (0.042-0.457)	0.273 (0.057-0.517)	0.069 (0.012-0.180)	25.2	24.9
A	2.286 (2.091-2.484)		1.944 (1.561-2.412)	0.316 (0.191-0.466)	2.23 (1.665-2.507)	0.304 (0.147-0.489)	1.376 (1.009-1.706)	0.287 (0.162-0.435)	0.871 (0.637-1.110)	0.656 (0.391-0.960)	0.512 (0.341-0.707)	0.642 (0.446-0.856)	0.264 (0.130-0.408)	31.4	30.2
H	1.665 (1.016-2.396)		1.411 (0.806-2.210)	0.231 (0.118-0.381)	1.581 (0.945-2.262)	0.221 (0.092-0.393)	1.007 (0.637-1.415)	0.201 (0.010-0.349)	0.653 (0.445-0.855)	0.477 (0.236-0.794)	0.381 (0.227-0.557)	0.478 (0.294-0.671)	0.193 (0.085-0.317)	32.8	31.4
I	0.206 (0.087-0.329)		0.176 (0.085-0.291)	0.029 (0.010-0.051)	0.195 (0.063-0.314)	0.027 (0.007-0.052)	0.123 (0.045-0.203)	0.026 (0.008-0.047)	0.078 (0.030-0.130)	0.065 (0.045-0.086)	0.046 (0.016-0.079)	0.058 (0.021-0.098)	0.023 (0.005-0.043)	25.9	25.4

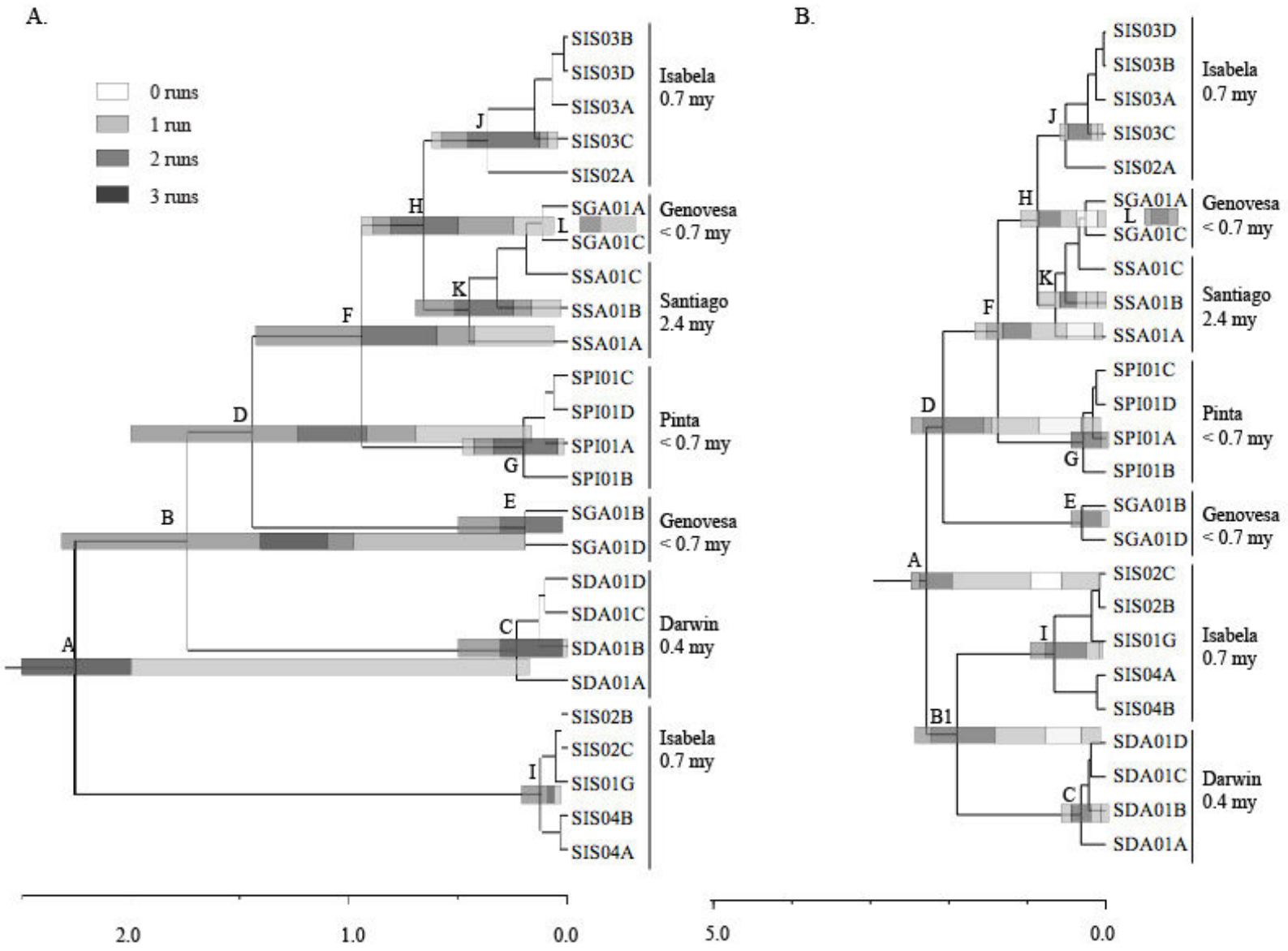
A.
0 runs
1 run
2 runs
3 runs
SIS03B
SIS03D
SIS03A
SIS03C
SIS02A
Isabela
0.7 my
SGA01A
SGA01C
Genovesa
< 0.7 my
SSA01C
SSA01B
SSA01A
Santiago
2.4 my
SPI01C
SPI01D
SPI01A
SPI01B
Pinta
< 0.7 my
SGA01B
SGA01D
Genovesa
< 0.7 my
SDA01D
SDA01C
SDA01B
SDA01A
Darwin
0.4 my
SIS02B
SIS02C
SIS01G
SIS04B
SIS04A
Isabela
0.7 my
A
B
C
D
E
F
G
H
I
J
K
2.0
1.0
0.0
B.
SIS03D
SIS03B
SIS03A
SIS03C
SIS02A
Isabela
0.7 my
SGA01A
SGA01C
Genovesa
< 0.7 my
SSA01C
SSA01B
SSA01A
Santiago
2.4 my
SPI01C
SPI01D
SPI01A
SPI01B
Pinta
< 0.7 my
SGA01B
SGA01D
Genovesa
< 0.7 my
SIS02C
SIS02B
SIS01G
SIS04A
SIS04B
Isabela
0.7 my
SDA01D
SDA01C
SDA01B
SDA01A
Darwin
0.4 my
A
B1
C
D
E
F
G
H
I
J
K
L
5.0
0.0

Figure 8. Ultrametric trees and ages estimates obtained with BEAST with priors based on the minimum age of the islands. The minimum age of the islands was used for the prior, as the mean of a normal distribution, to calibrate nodes A, H and/or I. Depending on which nodes were calibrated two different topologies were obtained. **A.** The topology (BEAST 1) was obtained three times by calibrating the run using [A,H, and I], [A and I], and [H and I]. **B.** The topology (BEAST 2) was obtained four times by calibrating the run using nodes [A and H], [A], [H], and [I]. The 95% confidence intervals, shown in gray, were obtained for nodes A-L, and B1 in the second topology. Because multiple runs (three in A. and four in B.) were condensed into one tree, the tree with the longest branch lengths is shown, but the confidence intervals for all runs were used. The confidence intervals that overlap are shown with progressively darker shades of grey. The white regions in the intervals are the age ranges that are not included in any of the runs. The outgroups were included in the run and were cut from tree after the analysis. The ages are given in millions of years. The outgroups were included for the analysis and were removed afterwards.

Table 11. Age estimates for 12 nodes (A-L) for BEAST analysis with external calibrations. **A.** The divergence of the Gran Canarian *Pimelia* and the diversification of the *P. granulicollis* species group were used to calibrate two runs with a relaxed molecular clock. The corresponding tree is shown in Figure 9a (BEAST 1). **B.** The rate of mtDNA mutation in arthropods (1.15%) was used under a relaxed molecular clock and under a strict molecular clock. The corresponding tree is shown in Figure 9b (BEAST 2). The nodes in the two trees are identical except for B, which are alternately named B1. Median and 95% confidence intervals are shown. Effective Sample Size (ESS) values are also shown for each run.

A.	Age of Nodes												ESS Values	
	A	B	C	D	E	F	G	H	I	J	K	L	posterior	prior
Origin of *Pimelia*	4.901 (2.853-7.770)	4.266 (2.488-6.668)	0.672 (0.230-1.246)	3.764 (2.171-5.857)	0.591 (0.198-1.234)	2.522 (1.427-4.001)	0.576 (0.237-1.080)	1.676 (0.967-2.696)	1.437 (0.555-2.943)	0.996 (0.499-1.701)	1.239 (0.655-2.023)	0.456 (0.173-0.839)	26.4	26.0
Divergence of *P. granulicollis*	8.302 (4.014-13.849)	7.210 (3.357-12.024)	1.152 (0.442-2.197)	6.356 (2.959-10.775)	1.001 (0.274-2.186)	4.280 (1.998-7.238)	0.972 (0.351-1.917)	2.847 (1.333-4.791)	2.444 (0.780-5.150)	1.686 (0.703-3.055)	2.102 (0.934-3.656)	0.762 (0.219-1.505)	31.5	29.9

B.	Age of Nodes												ESS Values	
	A	B1	C	D	E	F	G	H	I	J	K	L	posterior	prior
Relaxed Clock	1.482 (0.617-3.518)	1.350 (0.536-3.279)	0.197 (0.070-0.494)	1.362 (0.575-3.164)	0.181 (0.054-0.472)	0.852 (0.359-1.999)	0.163 (0.056-0.421)	0.519 (0.208-1.229)	0.449 (0.157-1.156)	0.292 (0.111-0.714)	0.371 (0.143-0.882)	0.152 (0.050-0.376)	25.1	24.6
Strict Clock	1.842 (1.611-2.094)	1.588 (1.304-1.917)	0.244 (0.163-0.337)	1.783 (1.497-2.071)	0.235 (0.131-0.349)	1.108 (0.921-1.317)	0.217 (0.137-0.309)	0.685 (0.557-0.827)	0.553 (0.379-0.751)	0.401 (0.298-0.515)	0.484 (0.372-0.609)	0.208 (0.124-0.307)	31.8	31.0

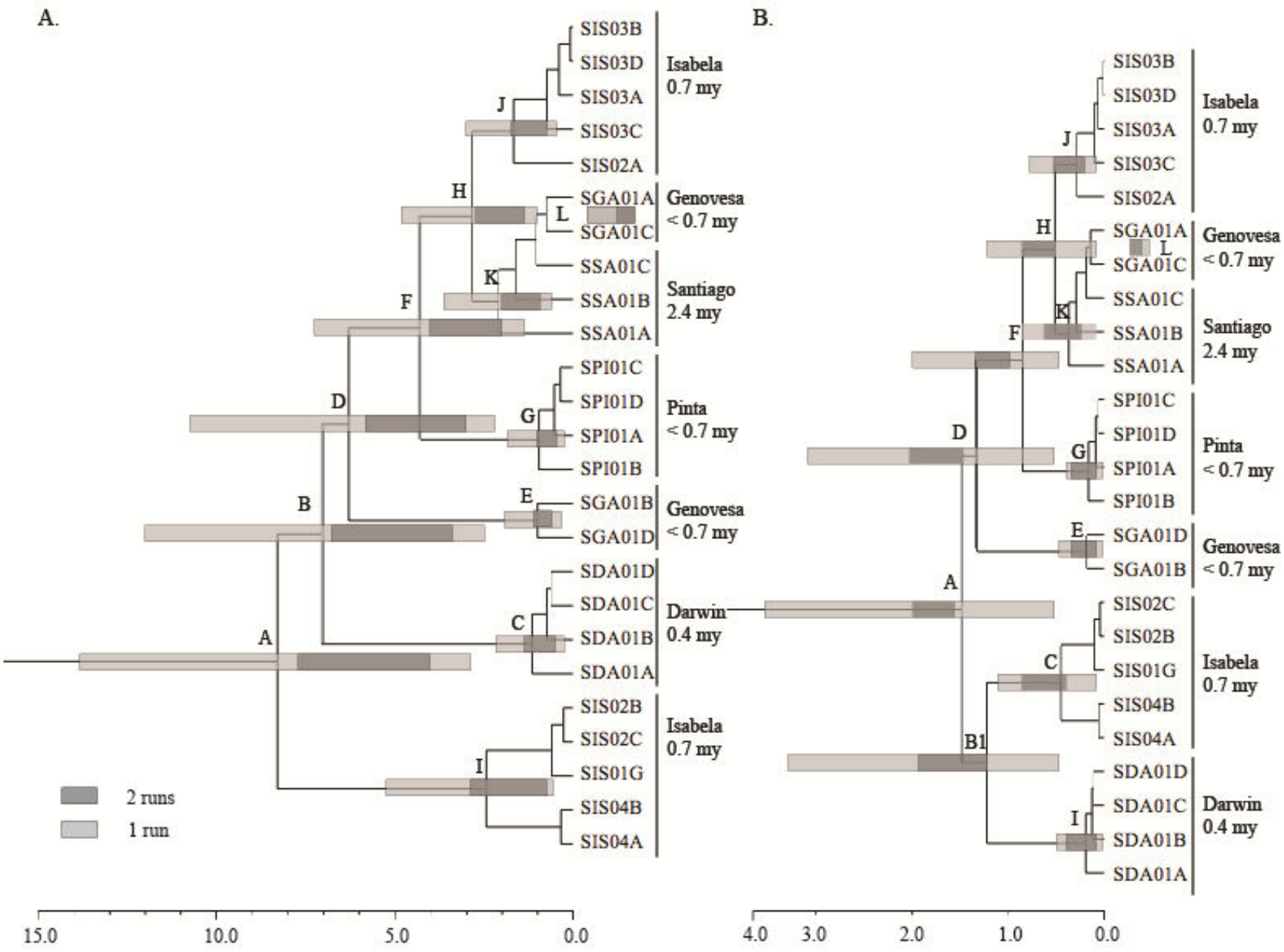
A.
SIS03B
SIS03D
SIS03A
SIS03C
SIS02A
Isabela
0.7 my
J
H
SGA01A
L
SGA01C
Genovesa
< 0.7 my
SSA01C
K
SSA01B
F
SSA01A
Santiago
2.4 my
SPI01C
D
SPI01D
G
SPI01A
SPI01B
Pinta
< 0.7 my
B
E
SGA01B
SGA01D
Genovesa
< 0.7 my
SDA01D
SDA01C
C
A
SDA01B
SDA01A
Darwin
0.4 my
SIS02B
SIS02C
I
SIS01G
SIS04B
SIS04A
Isabela
0.7 my
2 runs
1 run
15.0
10.0
5.0
0.0
B.
SIS03B
SIS03D
SIS03A
SIS03C
SIS02A
Isabela
0.7 my
J
H
SGA01A
L
SGA01C
Genovesa
< 0.7 my
SSA01C
K
F
SSA01B
SSA01A
Santiago
2.4 my
SPI01C
D
SPI01D
G
SPI01A
SPI01B
Pinta
< 0.7 my
E
SGA01D
SGA01B
Genovesa
< 0.7 my
A
SIS02C
SIS02B
C
SIS01G
SIS04B
SIS04A
Isabela
0.7 my
B1
SDA01D
SDA01C
I
SDA01B
SDA01A
Darwin
0.4 my
4.0
3.0
2.0
1.0
0.0

Figure 9. Ultrametric trees and ages estimates obtained with BEAST using external calibrations. **A.** The toplogy (BEAST 1) obtained when the age of the *Pimelia* outgroup was used to calibrate the tree. **B.** The topology (BEAST 2) was obtained when the rate of mtDNA mutation in arthropods was used under a relaxed molecular clock and under a strict molecular clock. In runs with identical topologies were obtained, the tree with longer branch lengths is shown. The 95% confidence intervals for all runs are shown in gray and were obtained for nodes A-L, B1. The regions where the confidence intervals overlap for the two runs are shown in a darker grey. The ages are in millions of years. The outgroups were included for the analysis and were removed afterwards.

Ancestral Areas

Character model (Mesquite)

To explore the history of colonization of *Stomion*, the ancestral characters (regions of the Galápagos) were reconstructed onto the three different topologies (Bayes, BEAST1, and BEAST2) (Figures 10-12). Optimization on the Bayesian topology indicates that the most likely platform for colonization was Central West because it is the only area with a significant proportional likelihood (Figure 10; Table 12). Central West also had the largest proportional likelihood in another topology, BEAST1 but all the other areas (except for Central) also display significant likelihoods (Figure 11-12; Table 12).

The "first branching" (labeled in Figures 10-12) was different in all three topologies either to Northern Isabela, to all of Isabela, or to both Isabela and Darwin. Central West was the only area with a significant likelihood in Bayes. In the BEAST1 topology, the first branching went to different parts of Isabela, and the most likely area of ancestry was also Central West (Figures 10-11; Table 12). In BEAST 2, all areas of ancestry displayed had significant likelihoods except for Central (Figure 12; Table 12). These results support an important role for the Central West area as the point of first colonization in the archipelago.

The "second branching" was either to Isabela or Darwin. In the Bayes tree, the second branching was to Darwin and the most probable areas of ancestry were Central North and Central West (Figure 10; Table 12). Interestingly, the Northwest is not reconstructed to have a significant proportional likelihood. BEAST1 had the same branching to Darwin, but all of the areas except for Central had a significant likelihood as ancestral sources (Figure 11; Table 12). The second branching in BEAST2 was to Isabela and the most probable area was Central West (Figure 12; Table 12). Because Central West was again reconstructed as having a significant

likelihood, this branching further suggests that Central West was an important area for the first colonization.

The third branching was to Genovesa and was the same in all three topologies (Figures 10-12). In Bayes, Central North and Central West were the most probable ancestral areas (Figure 10). In BEAST1 and BEAST2, Central North was the most probable (Figure 11-12). The same pattern was seen in the fourth branching to Pinta (Table 12).

In the fifth branching, all of the central regions (Central, Central North and Central West) had significant likelihoods. In the two BEAST topologies, the likelihoods were approximately equal, but in the Bayes topology, Central West was more probable than the other two central regions (Figures 10-12). The general pattern of colonization starting in the Central West is reconstructed in Figure 13.

Overall, we found that the age of colonization is between 1 and 2.3 my and that the ancestral platform for the colonization is in the Central West area. These findings must be reconciled with the fact that the maximum age of any Central West platform (0.7 my) is younger than our age of divergence. To do this, we will consider both the variation in the geologic estimates as well as in the calculation of our age estimates in the reconstruction of the evolutionary history of *Stomion* in the Galápagos.

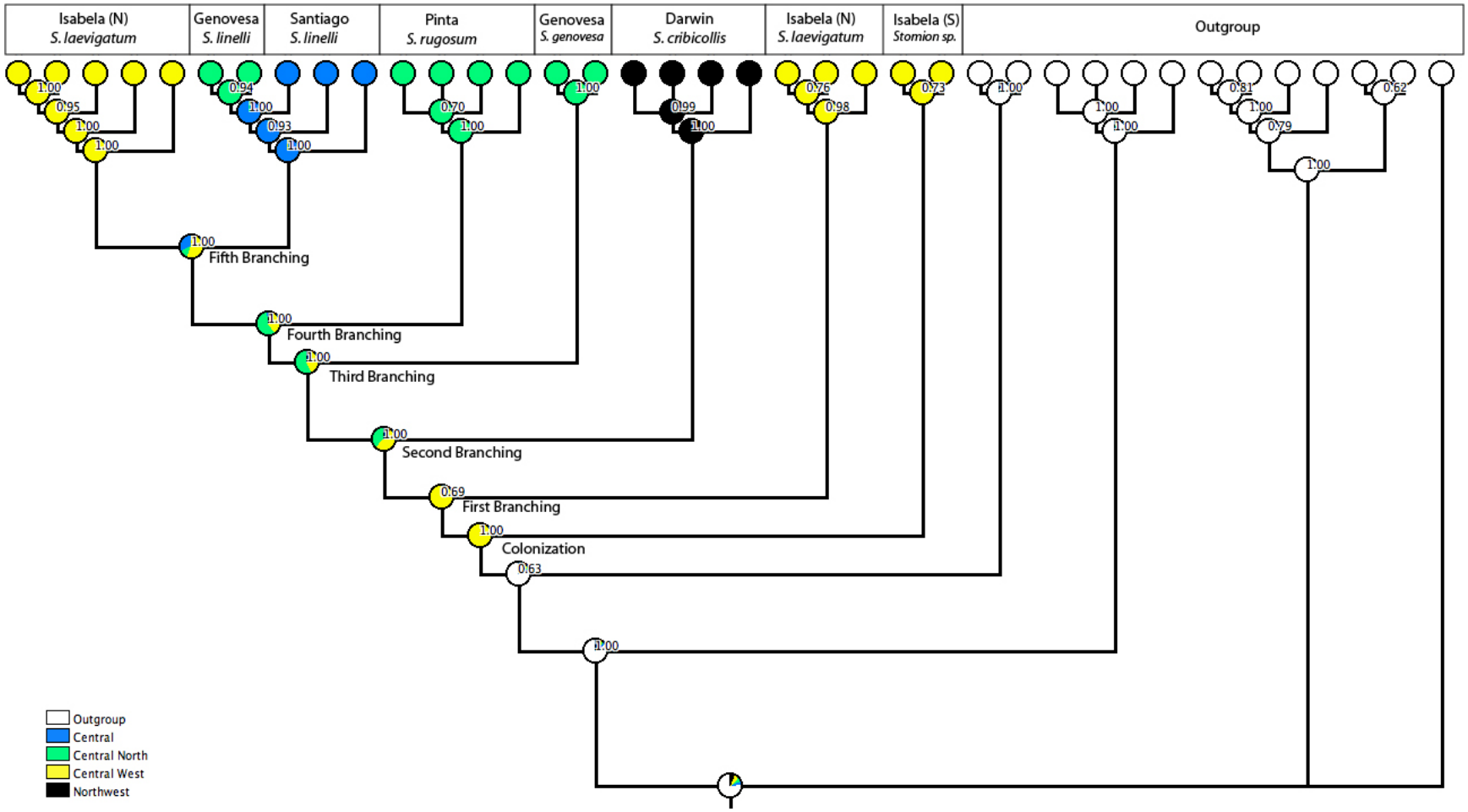

Figure 10. Bayesian tree with the proportional likelihoods of the ancestral areas within the Galápagos Archipelago. Pie charts at each node show the proportional likelihood of the ancestor being in a region of the archipelago. The islands and the species are found on the tips. The reliability of the Bayesian reconstruction of the tree is reflected in the values displayed on each of the nodes. Proportional likelihoods of each area on six of the nodes are found in Table 12 where 1 is the maximum.

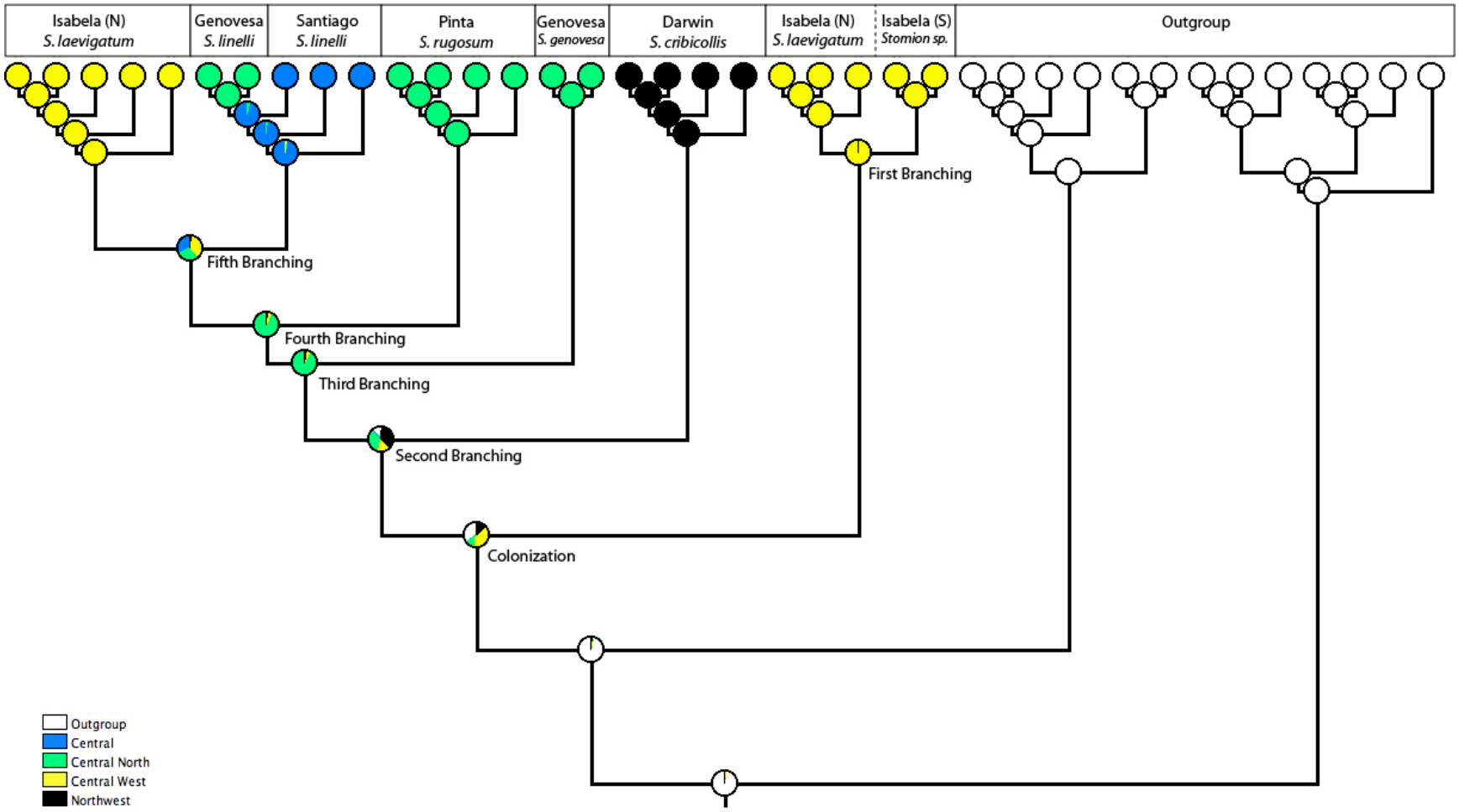

Figure 11. BEAST 1 tree with proportional likelihoods of the ancestral areas within the Galápagos Archipelago. Pie charts at each node show the proportional likelihood of the ancestor being in a region of the archipelago. The islands and the species are found on the tips. The reliability of the Bayesian reconstruction of the tree is reflected in the values displayed on each of the nodes. Proportional likelihoods of each area on six of the nodes are found in Table 12 where 1 is the maximum.

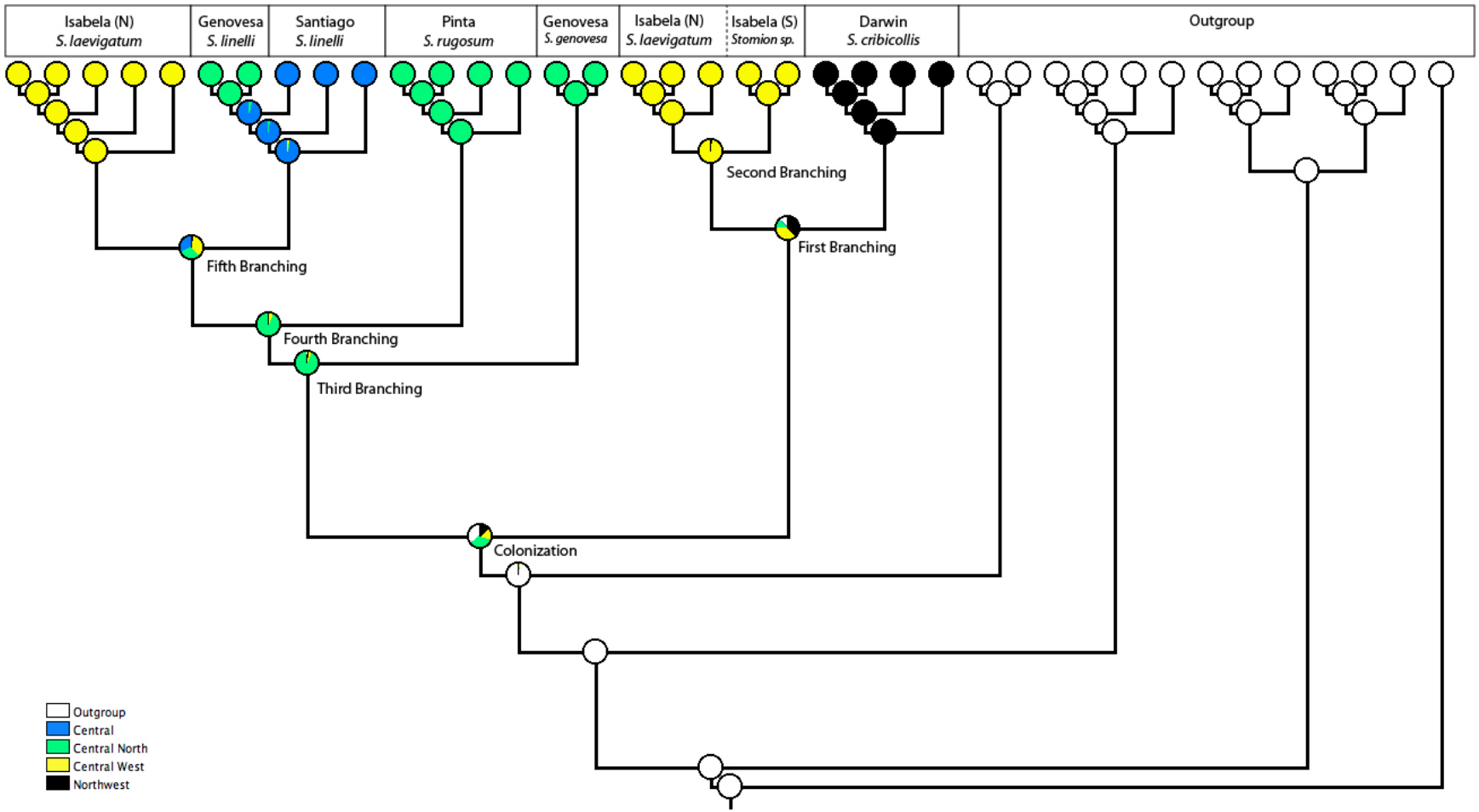

Figure 12. BEAST 2 tree with the proportional likelihoods of the ancestral areas within the Galápagos Archipelago. Pie charts at each node show the proportional likelihood of the ancestor being in a region of the archipelago. The islands and the species are found on the tips. The reliability of the Bayesian reconstruction of the tree is reflected in the values displayed on each of the nodes. Proportional likelihoods of each area on six of the nodes are found in Table 12 where 1 is the maximum.

Table 12. Proportional likelihoods of each reconstructed area at six nodes within the divergence of *Stomion*. The colonization, third, fourth and fifth branching are found at the same nodes in all three topologies. First and Second Branching nodes contain different taxa due to differences between the topologies. NW=Northwest; CN=Central North; C=Central; CW=Central West.

		Bayes	BEAST1	BEAST2
Colonization of Archipelago (Node A)	Outgroup	0.0002	0.352*	0.362*
	Central	0.0007	0.009	0.011
	Central North	0.001	0.119*	0.344*
	Central West	0.999*	0.395*	0.159*
	Northwest	0.0002	0.125*	0.124*
First Branching in Bayes (To Southern Isabela)	Outgroup	0.0001	n/a	n/a
	Central	0.0001	n/a	n/a
	Central North	0.002	n/a	n/a
	Central West	0.997*	n/a	n/a
	Northwest	0.0004	n/a	n/a
First Branching in BEAST1 (Within Isabela)	Outgroup	n/a	0.009	n/a
	Central	n/a	0.001	n/a
	Central North	n/a	0.003	n/a
	Central West	n/a	0.985*	n/a
	Northwest	n/a	0.003	n/a
First Branching in BEAST2 (To Isabela and Darwin)	Outgroup	n/a	n/a	0.124*
	Central	n/a	n/a	0.009
	Central North	n/a	n/a	0.118*
	Central West	n/a	n/a	0.391*
	Northwest	n/a	n/a	0.358*
Second Branching in Bayes and BEAST1 (To Darwin)	Outgroup	0.002	0.122*	n/a
	Central	0.013	0.011	n/a
	Central North	0.382*	0.344*	n/a
	Central West	0.515*	0.160*	n/a
	Northwest	0.088	0.363*	n/a
Second Branching in BEAST2 (To Isabela)	Outgroup	n/a	n/a	0.003
	Central	n/a	n/a	0.001
	Central North	n/a	n/a	0.003
	Central West	n/a	n/a	0.985*
	Northwest	n/a	n/a	0.009
Third Branching (To Genovesa)	Outgroup	0.001	0.008	0.022
	Central	0.016	0.008	0.008
	Central North	0.563*	0.915*	0.916*

	Central West	0.400*	0.047	0.047
	Northwest	0.020	0.021	0.008
Fourth Branching (To Pinta)	Outgroup	0.001	0.004	0.009
	Central	0.037	0.019	0.019
	Central North	0.566*	0.914*	0.915*
	Central West	0.389*	0.054	0.054
	Northwest	0.007	0.009	0.004
Fifth Branching (To Isabela, Sanitago, Genovesa)	Outgroup	0.001	0.008	0.010
	Central	0.329*	0.319*	0.319*
	Central North	0.121*	0.309*	0.309*
	Central West	0.546*	0.353*	0.353*
	Northwest	0.002	0.010	0.008

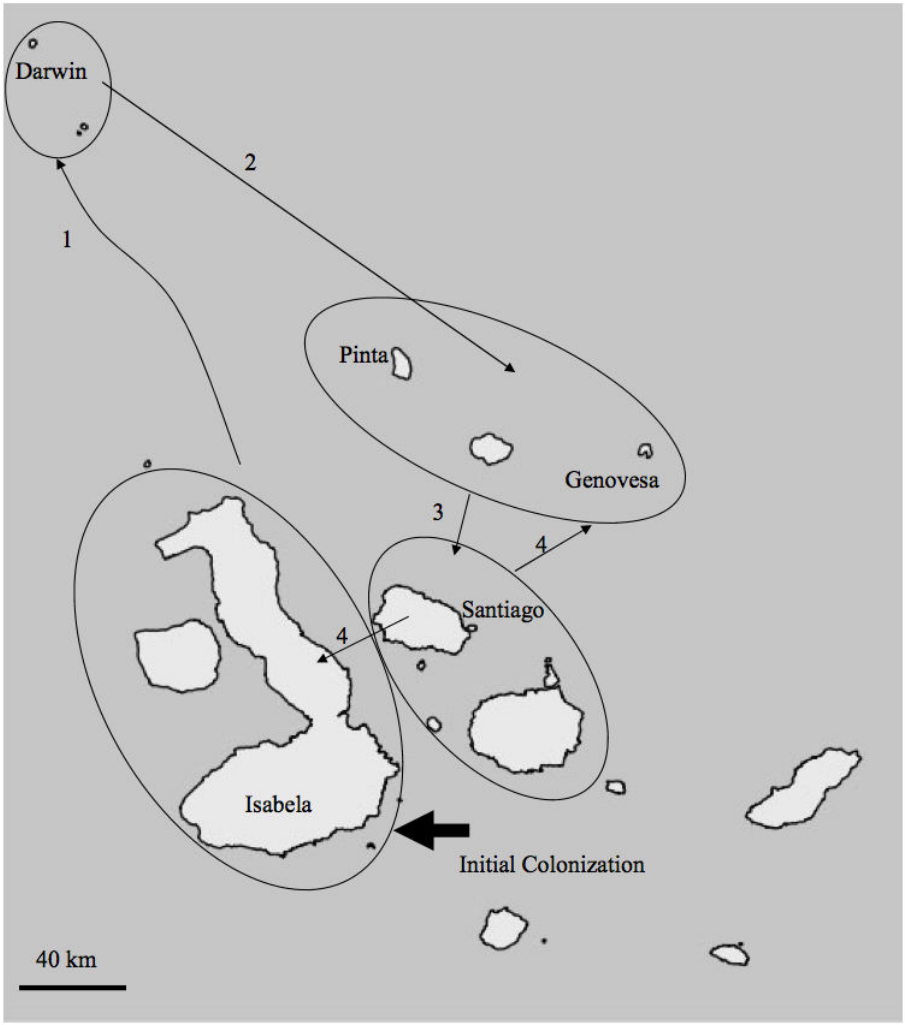

Figure 13. Reconstructed colonization history of *Stomion* on a map of the Galápagos Islands. The colonization patterns are based on the branching order on most of the topologies (Bayes and BEAST1) and the ancestral state reconstructions.

Dispersal, Colonization and Extinction Model (Lagrange)

While both Mesquite and Lagrange reconstruct the ancestral areas at the nodes of each tree, Mesquite reconstructs characters meaning that a node can only be in one place whereas Lagrange reconstructs the geographic range (potentially in multiple locations). At each node, the scenario with the highest relative probability was chosen. BEAST1 with short branch lengths from the uniform age run was successfully used to estimate the ancestral areas. All other topologies could not be used due to a convergence error and for other genera the ultrametric trees needed were not available.

The reconstruction is flawed in many ways including that the implied vicariant, or range splitting events at many of the nodes, A, B, D, F, H, K, and M, are unrealistic for an island archipelago. Even though Lagrange implements a model specifically designed for geographic ranges, the types of events it allows for are not compatible with the island system. The range reconstructions performed in Lagrange are described below for the sake of completeness however the results will not be considered in the final analysis. Instead, the results from Mesquite will be considered because even though it is a character model, it does not rely on a type of event and thus scenario that is almost impossible in the Galápagos.

In BEAST 1, node A was reconstructed to be in Central West and Northwest. The bottom branch containing taxa from Isabela was reconstructed to have originated in the Central West. The rest of the tree was reconstructed to have originated in both Central West and Northwest regions (Table 13; Figure 14). No range at this node was reconstructed as only one area, which in an island archipelago would be the point of colonization. Since vicariance is not realistic in an island archipelago, the implication is that there is not enough information to conclude which area the group colonized first. At node B, the geographic range was also

reconstructed to be in the Central West and Central North regions. The taxa from Darwin were reconstructed as having originated in the Northwest range and the rest of the tree from the Central West range. Between node B and D there was a dispersal event to the Central North (Figure 14). This event is implied by the fact that node D has a range of Central North and Central West and the geographic area of Central North was not "inherited" from the ancestor at node B. The taxa from Genovesa were reconstructed as having originated from Central North and the rest of the tree from Central West and Central North. At node F, the range was Central West and Central North (Figure 14). The taxa in Pinta were reconstructed as having originated in the Central North range, and the taxa in the rest of the tree from Central North and Central West. The node H has a range of Central North and Central West and the taxa in Isabela are reconstructed as having originated in the Central West and the taxa in Genovesa and Santiago are reconstructed as having originated in the Central North. Between nodes H and K there was a dispersal event to the Central region (Figure 14). The range at node K was Central and Central North. The individuals in Santiago are reconstructed as having originated in the Central region and the rest of the group are reconstructed as having originated in the Central and Central North range. The same happened again at M. At the next derived node the group in Genovesa are reconstructed as having originated in the range in the Central North and the taxa in Santiago are reconstructed as having originated in the Central range.

Table 13. Ancestral range reconstruction. Relative probabilities are shown for five nodes in Fig. 14 inferred using maximum likelihood in Lagrange. NW=Northwest; CN=Central North; C=Central; CW=Central West.

Node	Area Reconstruction	Relative Probability Run 1	Relative Probability Run 2
A	NW CW | CW	0.7069	0.6518
	NW | CW	0.1235	0.1528
	CN | CW	0.1216	0.152
B	CW | NW	0.7566	0.6972
	CN | NW	0.2431	0.2971
D	CN CW | CN	0.7566	0.6974
	CN | CN	0.2391	0.2971
F	CN | CN CW	0.7298	0.6654
	CN |CN	0.1852	0.2268
	CN | CW	0.0796	0.09906
H	CW | CN	0.9175	0.8953
	CW | C	0.0821	0.1047
K	CN C | C	0.9191	0.8973
	C | C	0.07986	0.1013
M	CN C | C	0.9289	0.9097
	C | C	0.06984	0.08864

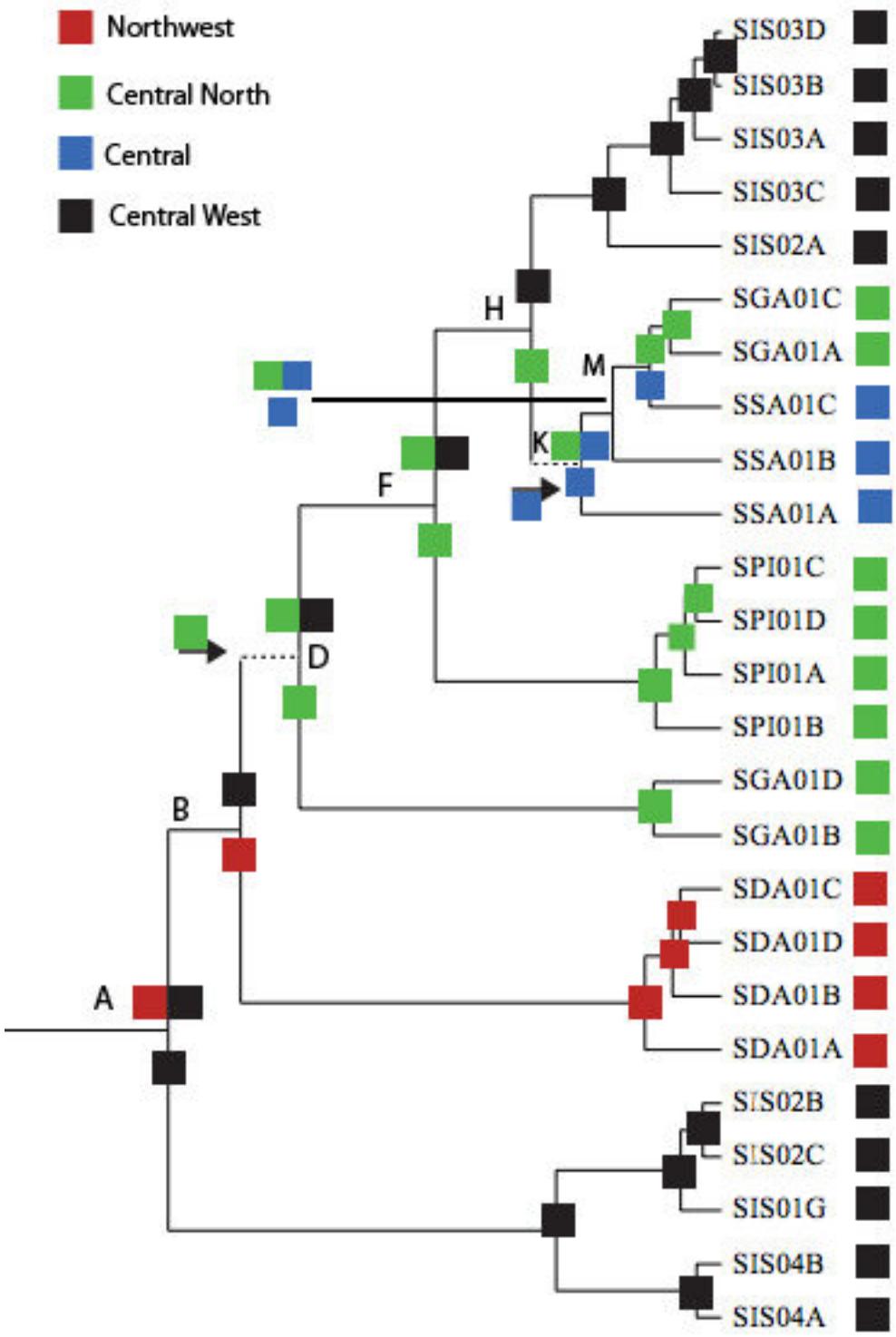

Figure 14. BEAST 1 tree with biogeographic analysis in Lagrange. Colored squares at the tips indicate the current distribution of the individual's species, and those on the nodes represent the range inferred using Lagrange. The branches with dotted lines and arrows indicate dispersal to explain the current distribution of taxa. The inferred range with the highest proportional likelihood is shown but those nodes with alternative ranges and their respective likelihoods are shown in Table 13.

Area Cladograms

To determine if there is a common pattern in the colonization histories of the invertebrates in the Galápagos Islands, area cladograms were constructed from the phylogenetic trees of five invertebrate genera including *Stomion*. While not all of the genera have the same geographic ranges and vary substantially in other ways, there are still conclusions that can be drawn from the cladograms. For each tree, one outgroup was included to root the tree. In *Bulimulus*, the first branching is to the Southeast (Figure 15C). In *Galagete*, and *Galapaganus*, the Southeast is also one of the first branchings but it is unresolved (Figure 15A-B). For *Hogna*, the Southeast was found at the most derived position in the tree (Figure 15D). For *Stomion*, no taxa from the Southeast were included in the taxa tree therefore it cannot be compared (Figure 16).

The central islands of the Galápagos were divided into three regions for the cladogram (Central, Central West and Central North). In all cladograms, Central contained one of the youngest species (Figures 15-16). Central North and Central West were found in very different positions for the genera analyzed. *Stomion*, *Galapaganus*, and *Galagete* are the three genera for which taxa from Central North were included. In both *Stomion* and *Galagete*, the taxa from Central North are the youngest but for *Galapaganus* they are found in an intermediate position in the tree (Figures 15-16). There are taxa from Central West in all of the genera. *Bulimulus* and *Galapaganus* have the youngest species in Central West, but for all the other genera, they are found in intermediate or unresolved basal positions (Figures 15-16).

Hogna contains one species that is widespread in the Galápagos (De Busschere et al. 2010). Since 3area does not accept taxa that are found in more than one locality, the colonization pattern for this species could not be resolved (Figure 15D).

The two BEAST trees were used to produce two separate cladograms. They are identical except for the branching of the northwest is only resolved in one of the topologies (Figure 16).

With the exception of *Hogna*, all of the groups that have samples from the Southeast show the first colonization in the Southeast. These are the islands closest to South America, thus supporting the prediction that the groups came from the South American mainland. This study did not include samples of *Stomion* from the Southeast so there is no way to know where it would fall in the pattern. The Southeast also contains the oldest islands. The next oldest islands are found in the Central region, but the region is one of the last to be colonized by all of the groups.

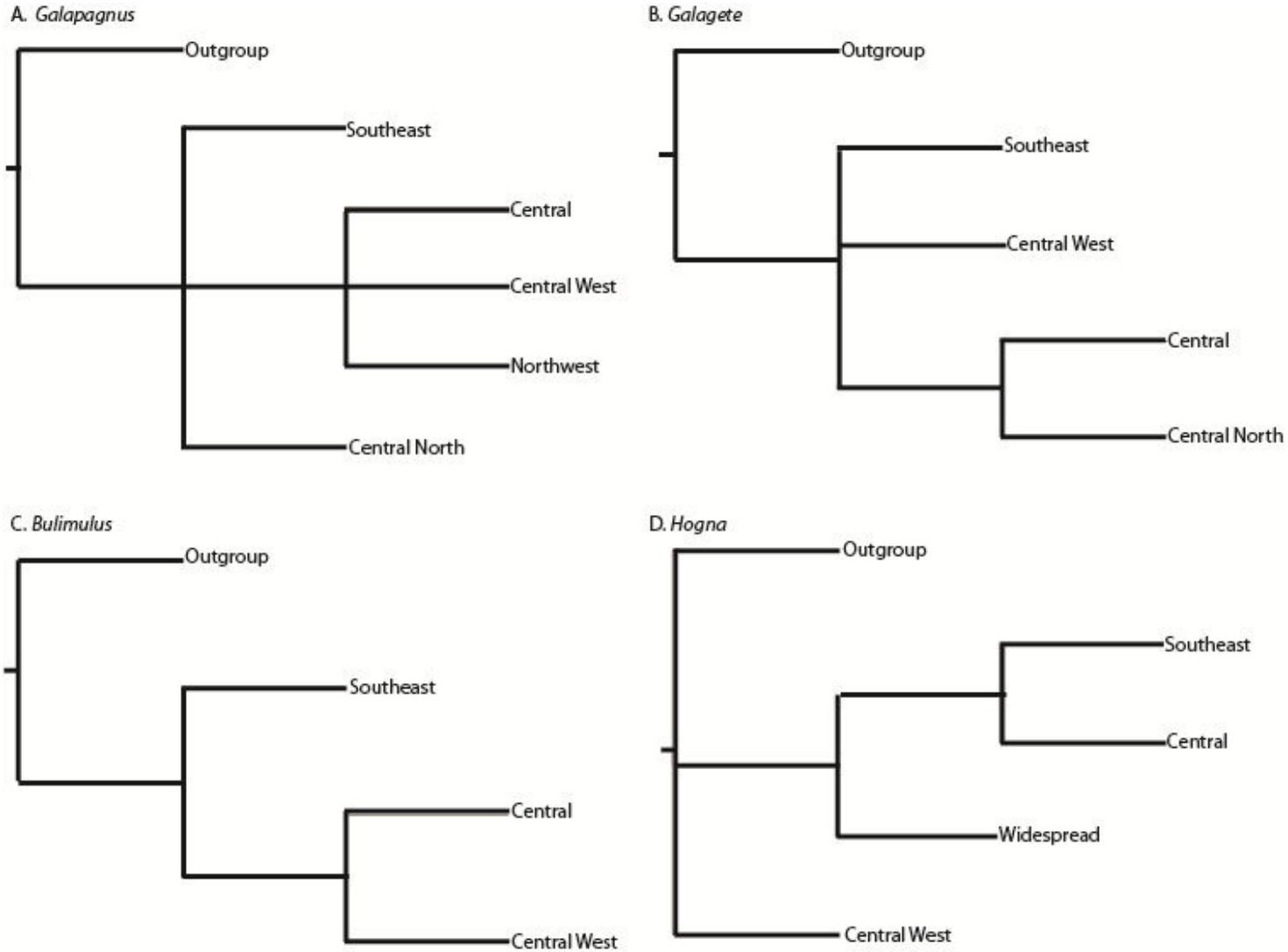

Figure 15. Area cladograms constructed using 3area for four different genera in the Galápagos Archipelago. **A**. *Galapaganus* **B.** *Galagete* **C.** *Bulimulus* **D.** *Hogna*. The Galápagos was divided into five regions (Northwest, Central North, Central, Central West, and Southeast). Intersection trees are shown.

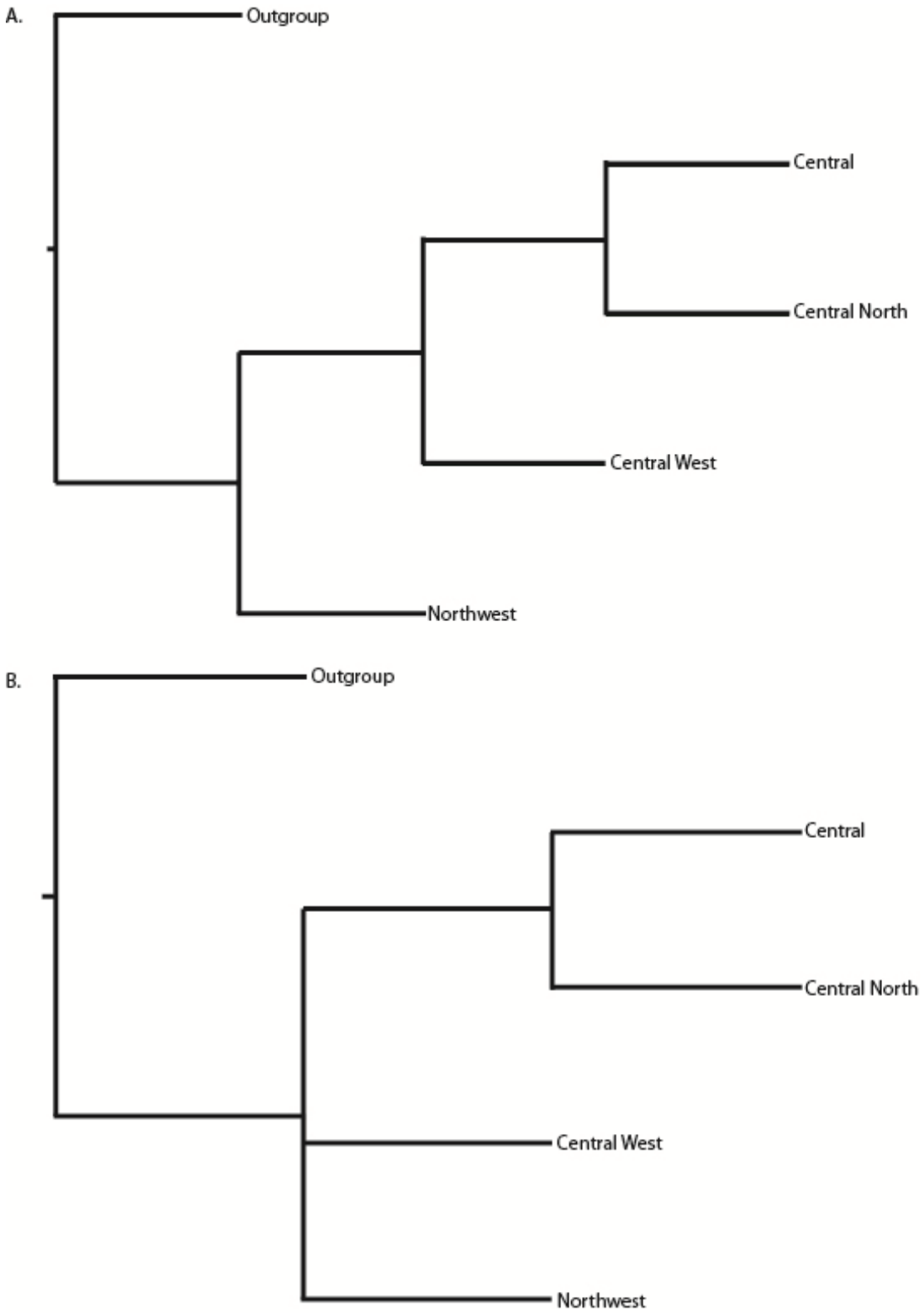

Figure 16 Area cladograms constructed using 3area for *Stomion*. A. The cladogram was built using BEAST 1 **B.** The cladogram was built using BEAST 2. Galápagos was divided into five regions (Northwest, Central North, Central, Central West, and Southeast). Intersections trees are shown.

DISCUSSION

Overall, this study found that the divergence of the Galápagos endemic genus *Stomion* is between 1 and 2.3 my old, well within the geologic age range of the islands. Our results indicate that it is the product of one colonization event and that inter-island speciation led to the radiation of the group across the archipelago. Additional sampling from the older islands would help us construct a more accurate picture of the colonization history. Based on the existing data and models, the Central West area of the archipelago seems to play an important role in this young radiation. To be able to fully understand these findings, they must be considered in the larger context of the findings on other groups in the Galápagos Archipelago. Our larger analysis suggests a different scenario for other groups where the Southeast is the first point of colonization in the archipelago.

Colonization History

Progression Rule

Comparisons are often drawn between the Canary, Hawaiian and Galápagos Islands because of their unique status as volcanic island archipelagos with well-studied examples of radiation and speciation (MacArthur and Wilson 1967, Funk and Wagner 1995, Juan et al. 1996, Sato et al. 2001, Caccone et al. 2002, Ciofi et al. 2002, Price and Clague 2002, Contreras-Diaz et al. 2003a, Finston and Peck 2004, Bonacum et al. 2005, Arbogast et al. 2006, Parent and Crespi 2006, Peck 2006, Parent et al. 2008, Sequeira et al. 2008a, Sequeira et al. 2008b, De Busschere et al. 2010). One of the keys to studying these radiations is an understanding of the order in which the groups colonized the islands. One often observed pattern is known as the progression rule under which the oldest species are found on the oldest islands and the youngest species are found on the youngest islands (Funk and Wagner 1995). In a phylogenetic tree, this is seen

when the organisms endemic to or present on the oldest islands are basally located on the tree and the youngest islands contain the most derived species. This pattern is not seen in *Stomion* because the base of the tree has taxa from some of the youngest islands. Even though this study does not include samples from the oldest islands, the lack of progression rule cannot be explained by incomplete sampling because Santiago, the oldest island that was sampled, is in a derived position on the tree.

The first extensive study of the progression rule by Funk and Wagner (1995) was based on the Hawaiian Island groups. The pattern is clearest in this archipelago because the oldest islands are furthest from the mainland, and thus, it is not obscured by a pattern of geographic proximity. Some of the groups that follow the progression rule, such as *Drosophila*, are also among the oldest lineages in the islands and colonized the older islands before the formation of the younger islands (Funk and Wagner 1995, Price and Clague 2002, Bonacum et al. 2005). There are also younger groups, such as the plant genus, *Hesperomannia* (Asteraceae), that follow the progression rule even though some of the younger islands had already formed and were geographically closer to the source of the colonization (Funk and Wagner 1995, Price and Clague 2002). In the Canary Islands, the oldest islands are closest to the mainland, and as a result, a pattern from oldest to youngest could also be caused by the geographic proximity to the mainland source (Whittaker and Fernández-Palacios 2007). There are many groups that appear to follow the progression/geographic pattern in the Canary Islands, including *Pimelia* and *Hegeter,* which are both genera of flightless darkling beetles from the same family as *Stomion* (Juan et al. 1995a, Juan et al. 1996, Juan et al. 2000, Contreras-Diaz et al. 2003a). Regardless of the archipelago, the pattern can be obscured by extinction, back colonization (when the same species or closely related species disperses back to an island on which it is already established),

and within-island speciation (Funk and Wagner 1995, Juan et al. 2000). In *Stomion*, no within-island speciation is seen, but extinction and back colonization are both possible and should be taken into account when considering the underlying processes.

In the Galápagos, the patterns are further confounded by the fact that the islands are not arranged in a linear chain as seen in the Canary and the Hawaiian Islands. The Galápagos Islands instead consist of groups of islands with similar ages. Some researchers have applied a looser interpretation of the progression rule in the Galápagos by considering islands that are geographically proximal and are approximately the same age in groups (Parent and Crespi 2006). Even then, many groups in the Galápagos do not follow the progression rule, and *Stomion* is no exception (Parent et al. 2008).

In addition to *Stomion,* the finches, *Galagete* (microlepidoptera), *Hogna* (spiders) and *Galapaganus* (weevils) do not strictly follow the progression rule (Sato et al. 2001, Schmitz et al. 2007, Parent et al. 2008, Sequeira et al. 2008a, Sequeira et al. 2008b, De Busschere et al. 2010). The older genera are expected to follow the progression rule since the younger islands were not above water when the genus colonized the archipelago. This is not observed in the Galápagos; most notably *Galapaganus* is older (10.7-12.1 my) than the oldest extant island (6.3 my), but it does not strictly follow the progression rule, and aspects of the pattern must be explained by dispersal and extinction, as will be discussed later (Sequeira et al. 2008a, Sequeira et al. 2008b). In younger groups, all of the islands were extant at the time of colonization so a pattern of geographic proximity to the source given by the geography of the islands could be the main driver of the colonization pattern. That is also not seen in *Stomion* because the two basally located groups are from Darwin and Isabela, which are the farthest from the mainland.

In spite of these complications with the interpretation of the progression rule, there are still groups in the Galápagos that roughly follow the age of the islands. These groups include the Galápagos giant tortoises, lava lizards, marine and land iguanas, and the mocking birds (Parent et al. 2008). It is unclear what makes the groups that conform and do not conform to the rule different, but it does not seem to be related to age or vagility of the groups. We also need to consider that some authors have argued that deviations from the rule could be explained by the use of sunken seamounts as platforms for older colonizations (Sequeira et al. 2008a).

Number of colonizations

Island archipelagos are central to the study of evolution mainly because they offer the simplest context in which to study phenomena (dispersal, competition, adaptation, and extinction), that are affected by a variety of different variables (environmental, geological etc.) (MacArthur and Wilson 1967, Losos et al. 2010). Their relative simplicity comes from their insularity. Most lineages, such as *Stomion*, only colonized the islands once because it is so unlikely for a group of organisms to arrive in a distant archipelago, such as the Galápagos, survive, and then radiate throughout the islands more than once. Of the insect lineages that have survived in the Galápagos, only 5% have radiated, and *Stomion* is one of them (Parent et al. 2008). Furthermore, because each island archipelago contains multiple islands, the islands then serve as replicates of each other (MacArthur and Wilson 1967). This allows genera to radiate on these platforms and allows us to understand which island properties may lead to species diversity and how the aforementioned phenomena interact to produce this diversity (MacArthur and Wilson 1967, Losos et al. 2010). In *Stomion*, it is difficult to consider the different island factors that produce within-island diversity because most islands only support one species, and only two islands have been colonized by two species.

The first attempt to quantify the processes associated with biogeography was completed by MacArthur and Wilson in *The Theory of Island Biogeography* (1967). Their main question was: Why do some islands have more species than others? Central to their theory was an attempt to quantitatively explain the effect of distance from the source, age, area, and habitat diversity. Since then, many different studies have analyzed one or more than one of these factors (MacArthur and Wilson 1967, Losos and Schluter 2000, Parent and Crespi 2006, Ricklefs and Bermingham 2007, Losos et al. 2010). They have yielded variable but not entirely conflicting results. Generally, one can conclude that many of the aforementioned factors have some form of effect but the degree to which each is important, depends on the species and island archipelago.

One form of distance, island insularity (quantified as the distance to the next closest island or distance from the source population) is thought to limit the species diversity because distance decreases the chances of individuals reaching the island (MacArthur and Wilson 1967, Losos et al. 2010). In *Stomion*, even the two most distant islands, Wolf and Darwin, have been colonized by one species. Three out of the four islands that Finston and Peck (2004) found to be colonized by more than one species, Isabela, Santiago, and Santa Cruz are found in the center of the archipelago suggesting that the centrality of their location might contribute to the frequency of inter-island colonization and thus species diversity. Other studies, such as that on *Bulimulus* in the Galápagos, have found no link between insularity and the number of species.

Additionally, it has been suggested that with increasing island age there should be more species on an island because there has been more time for colonization of the island and divergence within the island (MacArthur and Wilson 1967, Losos et al. 2010). This trend is also not seen in *Stomion*. Genovesa and Isabela are two of the youngest islands, both less than 0.7 my old and the only ones with more than one lineage (Table 1) (Geist 1996). The oldest islands

were not included in this study, but Finston and Peck (2004) only found one species on San Cristóbal and two on Española, which is not greater than the number on the younger islands (Finston and Peck 1995, 1997, 2004).

While linked to age, the area of an island also has an independent effect on species diversity (Parent and Crespi 2006). MacArthur and Wilson argued that while island area does not have a direct impact on the number of species, it has an indirect effect and can be used in place of habitat diversity when no other information is available (MacArthur and Wilson 1967). Furthermore, island area is also believed to be linked to species diversity because they are larger targets for colonization. In this study of *Stomion*, there are two islands, Isabela and Genovesa, that harbor two species, while all other islands only contain one species. Isabela is four times the size of the next largest island in the archipelago and Genovesa is one of the smallest, thus based on this study alone there seems to be no relationship between area and number of species. Finston and Peck (2004) included eleven species in their study, three species on Santa Cruz, two on Isabela, Santiago, and Española, and one on San Cristóbal, Darwin, Wolf, Pinta, Marchena, Santa Fé, and Fernandina. Even when their findings are considered there is no correlation between the number of species and the area of the island. Habitat diversity has also been identified as a contributor to the generation of diversity, especially in *Bulimulus* (MacArthur and Wilson 1967, Parent and Crespi 2006).

While most of the factors thought to contribute to speciation are environmental or geologic, Emerson and Kolm (2005) proposed the idea that the diversity of one species contributes to the diversity of another. It is not a new idea because it was part of MacArthur and Wilson's theory, and Parent and Crespi (2006) used plant diversity as a measure of habitat diversity. In an area with more species, the extinction rate is expected to increase because of the

competition between species. At the same time if the species does not succumb to extinction, the pressures can lead it to adapt, thus speeding the rate of speciation. Furthermore, when the species diversity increases, the average population size should decrease. A smaller population is more susceptible to the effects of genetic drift, and there is a higher probability that it will diverge from the source population (Emerson and Kolm 2005). In a young radiation like that of *Stomion*, it remains to be seen if species diversity can really drive speciation. Furthermore, *Stomion* beetles are generalists and thus are not host specific making it harder to share a niche even when there is a high level of diversity.

Each of these factors have both positive and negative effects on the species diversity, but the overall effect is that with increasing area, species diversity, age, habitat diversity, and lower insularity, there is a greater species diversity. With regards to *Stomion*, only one island contains three species, only four islands harbor two species, and all other islands inhabited by *Stomion* only display a single species. This is probably due in large part to the young divergence of the group and the fact that the beetles are generalists, thus geographic isolation (through intra-island speciation) rather than specialization on plant hosts might drive their diversity. As a result, it is difficult to use this radiation to explain the relative importance of each of these factors except to say that there is no evidence to discount the importance of any of them.

TIMING AND MODE OF SPECIATION

Mode

The process of inter and intra island speciation have both been documented as generators of diversity in genera in the Galápagos; however, *Stomion* only displays instances of inter-island speciation. This may be due in part to the relatively young age of *Stomion*'s divergence in the

Galápagos. Additionally, there are other island properties that influence the occurrence of the two modes of speciation.

Emerson and Oromí (2005) considered the effect of island age on the mode of generating species diversity in *Tarphius* and other arthropod genera in the Canary Islands. They showed that in the older Canary Islands more of the species were generated as the result of in situ speciation than in the younger islands. While the age of the divergences themselves did not differ between intra-island speciation among the older islands or inter-island speciation among the younger islands, a difference due to the increasing age contributed to the higher proportion of intra-island speciation (Emerson and Oromí 2005). They did not speculate as to what factor related to island age could have contributed to this phenomenon. The contribution of intra-island events to biodiversity of older islands is not seen in *Stomion* in part because the genus is much younger than those in the Canary Islands, and the oldest island in the Galápagos is younger than the older islands in the Canary Islands (6.3 my compared to 10-16 my) (Geist 1996, Emerson and Oromí 2005).

Building on the discussion of the effect of area, Losos and Schluter (2000) considered not just inter-island speciation but also within-island speciation. They contend that there are more species on larger islands because the larger area provides a greater potential for intra-island speciation due to the fact that geographic isolation increases with area. To support this contention they used the *Anolis* lizards of the Caribbean. They found that on islands larger than 3,000 km^2 the contribution of within-island speciation was higher than that of immigration (inter-island speciation) (Losos and Schluter 2000). While a decrease in extinction rate is probably also associated with increasing island size, they concluded that it did not account for all of the differences. Furthermore, even though they acknowledge a link between habitat diversity and

island area, they point to some of the smaller islands that have equally diverse climates and are old, but do not have any instances of within-island speciation (Losos and Schluter 2000). In the Galápagos, only Isabela is larger than 3,000 km^2 and since there are not instances of intra-island speciation in *Stomion*, this prediction cannot be evaluated.

Intra-island speciation has been seen extensively in the diversification of *Bulimulus* in the Galápagos (Parent and Crespi 2006). In part, this can be attributed to the low vagility of the snails and thus lower gene flow among populations and higher potential for isolation within an island. It has been observed in other more or less vagile groups (*Hogna* and *Galapaganus*) in the older islands of the archipelago (San Cristóbal and Santa Cruz) (Sequeira et al. 2008a, Sequeira et al. 2008b, De Busschere et al. 2010). Additionally, they concluded that habitat diversity and the indirect contributions to habitat diversity as the result of increased age, size and location, contributed to the observed amount of within-island speciation (Parent and Crespi 2006). While the pattern they observed was consistent with the observations by Emerson and Oromí (2005) with respect to age, they considered more variables. Many of these factors and variables cannot easily be separated from each other, and while they conclude that different factors are important, their findings are not in direct opposition to each other.

Timing

Divergences that are between 1 and 2 million years are not uncommon for endemic groups in either the Galápagos or the Hawaiian Islands. In fact, they are more common than older divergences (Price and Clague 2002, Parent et al. 2008).

The Hawaiian archipelago contains islands that are up to 30 million years old and thus are potentially home to groups with very old divergences. Despite studies that have looked at many of the largest divergences in the islands most divergences appear to be less than 5 million

years old. Price and Clague (2002) estimated the ages of 15 multi-species lineages and found only four that are older than 5 my. While these are some of the largest species groups (between 28 and 1000 species), the bulk of their diversity may be more recent. In the Hawaiian fruit flies (*Drosophilidae*), there are an estimated 1,000 species, and their divergence is estimated to be 26 my old; however, most of the diversity is thought to have occurred in the last 5 my (Russo et al. 1995, Price and Clague 2002, Bonacum et al. 2005). Price and Clague (2002) attribute this to the dynamic nature of the landscape. In reconstructing the life history of the islands, they found bottlenecks in which many habitats, especially the higher elevation and cooler habitats, shrunk or ceased to exist leading to the extinction of many species. To have a large amount of diversity, the groups must speciate quickly during "peak periods" when the islands are larger, taller, and closer to each other (Price and Clague 2002). Many of the plant and bird species in Hawaii have been estimated to have ages between 1 and 5 my, which is the same approximate age as *Stomion* in the Galápagos. The Galápagos have more lineages that are older than the age of the extant islands possibly due to differences in climatic patterns, geographic elements, and geologic variables, such as the subsidence rates of the islands (White et al. 1993, Geist 1996, Price and Clague 2002, Parent et al. 2008).

The elements of the geologic history, such as changes in sea level, must also be considered to evaluate how they affect the colonization history of one group before they can be compared. The change in sea level affects the emergence and submergence of islands but could also have lead to connections between the islands. These connections could have increased the rates of immigration between the islands. While precise data were not available due to the many factors affecting sea level, estimates (based on data from Bermuda) suggest that the sea level was only 130 m lower than its current level in the last 10 million years (older than the age of the

extant islands) (Geist 1996). This means that the formation of a bridge between Isabela and Fernandina could have fostered the exchange of organisms between the islands. Even if the islands are not directly connected, a decreased distance between Santa Cruz and Isabela could still increase the rate of immigration between the islands.

Dispersal and Extinction

In some groups, the ages of specific divergences within the archipelago suggested by the phylogenetic tree are older than the age of the islands they are thought to have colonized. In the case of *Stomion*, this occurs at the basal nodes of the tree where the islands of Isabela and Darwin are younger than the colonization calculated from the molecular data. Because the islands did not exist it must be suggested that the group was on another island at that time, dispersed to the island where they are now found, and then became extinct in the first island. This pattern cannot be explained by lack of sampling because additional samples would not decrease the estimates of colonization times to each of the islands. The dispersal and extinction explanation also had to be invoked to explain the colonization history of *Galapaganus* since Isabela's endemic weevils are placed at the base of the tree and the divergence is estimated to be between 10.7-12.1 my old (Sequeira et al. 2008a, Sequeira et al. 2008b). Furthermore, the divergence is older than the age of any of the extant islands meaning that the first colonizers must have been living on an island that is now submerged (Christie et al. 1992, White et al. 1993, Geist 1996).

At least four divergences within the Galápagos or with the closest extant lineage have been estimated to be older than the ages of the extant islands, with the maximum being around 12 my (Parent et al. 2008, Sequeira et al. 2008a). Seamounts located in the east of the archipelago have been estimated to be at least 9.5 my old and because the hotspot that has

resulted in the formation of these islands is between 80 and 90 my old, it is thought that there could be sunken islands that are as old as the hotspot (Christie et al. 1992). The divergence of *Stomion* is much younger than even the extant islands; therefore, these are now submerged islands are not required to support our proposed scenario of colonization for *Stomion.*

Sources of Colonization and Ancestral Areas

The fauna of the Galápagos are generally assumed to have originated from the Central or South American mainland because of the many phylogenetic relationships that have been documented in a variety of different species (Sato et al. 2001, Peck 2006). An additional source that has been proposed is the Cocos Islands, but there is little evidence to support this as the original source of any Galápagos groups (Sato et al. 2001, Peck 2006). Assuming that they did reach the archipelago from South America, four main dispersal methods have been proposed: 1. flying or passively being carried by the winds, 2. floating, rafting or swimming in or on the water, 3. hitchhiking on other organisms (method for parasites), and 4. introduction by humans (Peck 2006). *Stomion* is flightless, not parasitic, and has a colonization age of at least 1 my, so the most plausible method of introduction is floating on water or rafting on vegetation. Since no mainland ancestor is known at present, the origin of *Stomion* cannot be definitively determined, but there is no reason to believe that it has an origin different from that of other species in the Galápagos.

Even if the groups under consideration (*Hogna, Bulimulus, Galagete, Galapaganus,* and *Stomion*) colonized the islands from the South American mainland, they do not necessarily need to have the same arrival site. Following the progression rule or geographic proximity, the point of colonization would be in the Southeast, on San Cristóbal or Española. The clearest example of this pattern is seen in *Bulimulus* where the first colonization point is indeed in the Southeast

followed expansion to the Central and Central West regions. In *Galagete* and *Galapaganus*, the southeast is in an unresolved position close to the base reflecting that some details of the pattern are probably lost in the construction of area cladograms when each area can only be displayed once. The species phylogenies show that for both *Galapaganus* and *Galagete* the most basal branch contains inhabitants of the Central West region (Fernandina or Isabela). In *Stomion*, the Central West (Isabela) seems to be the point of origin considering the available data. It is possible that this is due to lack of sampling from the Southeast, but it is a pattern that has been seen in *Galapaganus* even when individuals from the Southeast were included. In *Hogna*, the Southeast is found at the most derived position of the tree and the basal node divergence to Isabela (De Busschere et al. 2010).

In summary, the area cladograms support the notion that the relative closeness of the southeastern islands to the continental source played a role in the colonization patterns of fairly divergent groups (weevils, microlepidoptera and snails). One aspect of their location relevant to colonization might be their presence in the trajectory of major wind and ocean currents (Humboldt) after they have passed the Ecuadorian and Peruvian coasts (Peck 2006).

Future

We interpret these results as evidence in support of the idea that *Stomion* diversity is the product of a single colonization with subsequent radiation across the archipelago. Our results indicate that the diversity of species originated mostly through colonization of neighboring islands in a pattern of inter-island speciation that does not follow the progression rule. The calculated age of the initial *Stomion* divergence (and of the initial colonization) is within the geological age range of the islands, but is older than its first colonization target, Isabela. While the discordance between the age of divergence and the geological age of the first target is not

unique to *Stomion* and has been observed in other groups in the Galápagos, the colonization history is unlike any pattern observed in other groups. This study has been relatively exhaustive in terms of the number and types of genes analyzed; however, it could benefit from the inclusion of populations from the older islands and from close relatives in the continent. The addition would allow us to determine how much the group has diverged from its continental ancestor and to construct a more complete comparison with other Galápagos groups.

REFERENCES

Sequencher® version 5.0 sequence analysis software. Gene Codes Corporation, Ann Arbor, MI USA.

Arbogast, B. S., S. V. Drovetski, R. L. Curry, P. T. Boag, G. Seutin, P. R. Grant, B. R. Grant, and D. J. Anderson. 2006. The origin and diversification of Galapagos mockingbirds. Evolution **60**:370-382.

Archibald, J. K., M. E. Mort, and D. J. Crawford. 2003. Bayesian inference of phylogeny: a non-technical primer. Taxon **52**:187-191.

Becker, P. 1975. Island Colonization by Carnivorous and Herbivorous Coleoptera. Journal of Animal Ecology **44**:893-906.

Bonacum, J., P. M. O'Grady, M. Kambysellis, and R. DeSalle. 2005. Phylogeny and age of diversification of the *planitibia* species group of the Hawaiian *Drosophila*. Molecular phylogenetics and evolution **37**:73-82.

Brower, A. 1994. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences **91**:6491.

Caccone, A., G. Gentile, J. P. Gibbs, T. H. Fritts, H. L. Snell, J. Betts, and J. R. Powell. 2002. PHYLOGEOGRAPHY AND HISTORY OF GIANT GALAPAGOS TORTOISES. Evolution **56**:2052-2066.

Christie, D., R. Duncan, A. McBirney, M. Richards, W. M. White, K. Harpp, and C. Fox. 1992. Drowned islands downstream from the Galapagos hotspot imply extended speciation times.

Ciofi, C., M. C. Milinkovitch, J. P. Gibbs, A. Caccone, and J. R. Powell. 2002. Microsatellite analysis of genetic divergence among populations of giant Galápagos tortoises. Molecular ecology **11**:2265-2283.

Contreras-Diaz, H. G., O. Moya, P. Oromi, and C. Juan. 2003a. Phylogeography of the endangered darkling beetle species of Pimelia endemic to Gran Canaria (Canary Islands). Molecular ecology **12**:2131-2143.

Contreras-Diaz, H. G., O. Moya, P. Oromi, and C. Juan. 2003b. Phylogeography of the endangered darkling beetle species of *Pimelia* endemic to Gran Canaria (Canary Islands). Journal of Molecular Ecology:12131-12143.

Crozier, R. and Y. Crozier. 1992. The cytochrome b and ATPase genes of honeybee mitochondrial DNA. Molecular Biology and Evolution **9**:474-482.

Cryan, J. R., J. K. Liebherr, J. W. Fetzner Jr, and M. F. Whiting. 2001. Evaluation of relationships within the endemic Hawaiian Platynini (Coleoptera: Carabidae) based on molecular and morphological evidence. Molecular phylogenetics and evolution **21**:72-85.

De Busschere, C., F. Hendrickx, S. Van Belleghem, T. Backeljau, L. Lens, and L. Baert. 2010. Parallel habitat specialization within the wolf spider genus Hogna from the Galápagos. Molecular ecology **19**:4029-4045.

Doyen, J. T. 1993. Cladistic Relationships among Pimeliine Tenebrionidae (Coleoptera). Journal of the New York Entomological Society **101**:443-514.

Drummond, A. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology **7**:214.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology **4**:e88.

Drummond, A. J., S. Y. W. Ho, N. Rawlence, and A. Rambaut. 2007. A rough guide to BEAST 1.4. University of Auckland, Aukland, New Zealand. Website http://code. google. com/p/beast-mcmc/downloads/list.

Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics **161**:1307.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution.

Ducasse, J., N. Cao, and R. Zaragüeta-Bagils. 2008. LisBeth. Three-item analysis software package. Laboratoire Informatique et Systématique, UPMC Univ Paris 06, UMR 7207 (CR2P) CNRS MNHN UPMC. .

Ebach, M. C., C. J. Humphries, R. A. Newman, D. M. Williams, and S. A. Walsh. 2005. Assumption 2: opaque to intuition? Journal of Biogeography **32**:781-787.

Emerson, B. 2002. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. Molecular ecology **11**:951-966.

Emerson, B. C. and N. Kolm. 2005. Species diversity can drive speciation. Nature **434**:1015-1017.

Emerson, B. C. and P. Oromí. 2005. DIVERSIFICATION OF THE FOREST BEETLE GENUS TARPHIUS ON THE CANARY ISLAND, AND THE EVOLUTIONARY ORIGINS OF ISLAND ENDEMICS. Evolution **59**:586-598.

Farrell, B. D. 2001. Evolutionary Assembly of the Milkweed Fauna: Cytochrome Oxidase I and the Age of Tetraopes Beetles. Molecular phylogenetics and evolution **18**:467-478.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing significance of incongruence. Cladistics **10**:315-319.

Finston, T. L. and S. B. Peck. 1995. Population structure and gene flow in Stomion: A species swam of flightless beetles of the Galapagos Islands. Heredity **75**:390-397.

Finston, T. L. and S. B. Peck. 1997. Genetic differentiation and speciation in Stomion (Coleoptera: Tenebrionidae): flightless beetles of the Galapagos Islands, Ecuador. Biological Journal of the Linnean Society **61**:183-200.

Finston, T. L. and S. B. Peck. 2004. Speciation in Darwin's darklings: taxonomy and evolution of Stomion beetles in the Galapagos Islands, Ecuador (Insecta: Coleoptera: Tenebrionidae). Zoological Journal of the Linnean Society **141**:135-152.

Funk, V. A. and W. L. Wagner. 1995. Biogeographic patterns in the Hawaiian Islands.

Geist, D. 1996. On the emergence and submergence of the Galapagos Islands. Noticias de Galapagos **56**:5-9.

Hillis, D. M. and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quarterly Review of Biology:411-453.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics:754-755.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294**:2310-2314.

Juan, C., B. C. Emerson, P. Oromí, and G. M. Hewitt. 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends in Ecology & Evolution **15**:104-109.

Juan, C., P. Oromi, and G. M. Hewitt. 1995a. Mitochondrial DNA phylogeny and sequential colonization of Canary Islands by darkling beetles of the genus Pimelia (Tenebrionidae). Proceedings of the Royal Society of London. Series B: Biological Sciences **261**:173-180.
Juan, C., P. Oromi, and G. M. Hewitt. 1995b. Mitochondrial DNA Phylogeny and Sequential Colonization of Canary Islands by Darkling Beetles of the Genus Pimelia (Tenebrionidae). Proceedings: Biological Sciences **261**:173-180.
Juan, C., P. Oromi, and G. M. Hewitt. 1996. Phylogeny of the genus Hegeter (Tenebrionidae, Coleoptera) and its colonization of the Canary Islands deduced from Cytochrome Oxidase I mitochondrial DNA sequences. Heredity **76**:392-403.
Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947-2948.
Losos, J. B., R. E. Ricklefs, and R. H. MacArthur. 2010. The theory of island biogeography revisited. Princeton Univ Pr.
Losos, J. B. and D. Schluter. 2000. Analysis of an evolutionary species, area relationship. Nature **408**:847-850.
MacArthur, R. H. and E. O. Wilson. 1967. The theory of island biogeography. Princeton Univ Pr.
Maddison, D. R. and W. P. Maddison. 2000. MacClade 4. Sinauer Associates.
Maddison, W. P. and D. R. Maddison. 2011. Mesquite: a modular system for evolutionary analysis.
Morrone, J. J. 1994. On the identification of areas of endemism. Systematic Biology **43**:438-441.
Morrone, J. J. 2009. Evolutionary biogeography: An integrative approach with case studies. Columbia Univ Pr.
Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. Da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.
Parent, C. E., A. Caccone, and K. Petren. 2008. Colonization and diversification of Galapagos terrestrial fauna: a phylogenetic and biogeographical synthesis. Philosophical Transactions of the Royal Society B: Biological Sciences **363**:3347.
Parent, C. E. and B. J. Crespi. 2006. Sequential colonization and diversification of Galapagos endemic land snail genus Bulimulus (Gastropoda, Stylommatophora). Evolution **60**:2311-2328.
Parks, D. H., M. Porter, S. Churcher, S. Wang, C. Blouin, J. Whalley, S. Brooks, and R. G. Beiko. 2009. GenGIS: A geospatial information system for genomic data. Genome research **19**:1896-1904.
Peck, S. B. 2006. The Beetles of the Galápagos Islands, Ecuador: Evolution, Ecology, and Diversity (Insecta: Coleoptera). NRC Research Press, Ottawa.
Peck, S. B. and J. Kukalová-Peck. 1990. Origin and biogeography of the beetles (Coleoptera) of the Galápagos Archipelago, Ecuador. Canadian Journal of Zoology **68**:1617-1638.
Posada, D. 2003. Selecting models of evolution. The phylogenetic handbook:256-282.
Posada, D. 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution **25**:1253-1256.
Posada, D. 2009. Selection of models of DNA evolution with jModelTest. Methods Mol Biol **537**:112.

Price, J. P. and D. A. Clague. 2002. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. Proceedings of the Royal Society of London. Series B: Biological Sciences **269**:2429-2435.
Rambaut, A. 2009. FigTree v1.3.1. Available from http://tree/bio.ed.ac.uk/software/figtree/.
Rambaut, A. and A. Drummond. 2007. Tracer v1.5. Available from http://beast.bio.ed.ac.uk/Tracer
Ree, R. H. and I. Sanmartin. 2009. Prospects and challenges for parametric models in historical biogeographical inference. Journal of Biogeography **36**:1211-1220.
Ree, R. H. and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology **57**:4-14.
Ricklefs, R. E. and E. Bermingham. 2007. The causes of evolutionary radiations in archipelagoes: passerine birds in the Lesser Antilles. The American Naturalist **169**:285-297.
Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics:1572-1574.
Russo, C., N. Takezaki, and M. Nei. 1995. Molecular phylogeny and divergence times of drosophilid species. Molecular Biology and Evolution **12**:391-404.
Sato, A., H. Tichy, C. O'hUigin, P. R. Grant, B. R. Grant, and J. Klein. 2001. On the origin of Darwin's finches. Molecular Biology and Evolution **18**:299-311.
Schmitz, P., A. Cibois, and B. Landry. 2007. Molecular phylogeny and dating of an insular endemic moth radiation inferred from mitochondrial and nuclear genes: the genus Galagete (Lepidoptera: Autostichidae) of the Galápagos Islands. Molecular phylogenetics and evolution **45**:180-192.
Sequeira, A., A. Lanteri, L. R. Albelo, S. Bhattacharya, and M. Sijapati. 2008a. Colonization history, ecological shifts and diversification in the evolution of endemic Galápagos weevils. Molecular ecology **17**:1089-1107.
Sequeira, A., M. Sijapati, A. Lanteri, and L. R. Albelo. 2008b. Nuclear and mitochondrial sequences confirm complex colonization patterns and clear species boundaries for flightless weevils in the Galapagos archipelago. Philosophical Transactions of the Royal Society B: Biological Sciences **363**:3439-3451.
Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Ent. Soc. Am **87**:651-701.
Steckel, J., M. L. Penrith, J. Henschel, R. Brandl, and J. Meyer. 2010. A preliminary molecular phylogeny of the Namib Desert darkling beetles (Tenebrionidae). African Zoology **45**:107-114.
Swofford, D. 2002. PAUP* Phylogenetic Analysis Using Parsimony (*and other methods) version 4. Sinauer Associates, Sunderland, Massachusetts.
Van Dyke, E. C. 1953. The Coleoptera of the Galapagos Islands. California Academy of Sciences.
Vogler, A. P. and A. Welsh. 1997. Phylogeny of North AmericanCicindelaTiger Beetles Inferred from Multiple Mitochondrial DNA Sequences. Molecular phylogenetics and evolution **8**:225-235.
White, W. M., A. R. McBirney, and R. A. Duncan. 1993. Petrology and geochemistry of the Galapagos Islands: Portrait of a pathological mantle plume. Journal of Geophysical Research **98**:19533-19519,19563.

Whittaker, R. J. and J. M. Fernández-Palacios. 2007. Island biogeography: ecology, evolution, and conservation. Oxford University Press, USA.
Yang, Z. and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Molecular Biology and Evolution **14**:717-724.