Historical biogeography of *Pelargonium*:

Evidence of a Cape to Cairo scenario using explicit and non-explicit species ranges.

Merlijn Schram

961216748050

Supervisors: Sara van de Kerke & Dr. Freek T Bakker

Biosystematics

19 – 07 - 2019

Summary

The purpose of historical biogeography is estimating ancestral species ranges. Modern programs, such as BioGeoBEARS, require pre-set areas constructed by the user. GEM, implemented in the program evs, is a promising new method that allows the use of explicit species ranges. Both BioGeoBEARS and evs were used to study the historical biogeography of the mainly South-African plant genus *Pelargonium* (Geraniaceae). The divergence of *Pelargonium* and its sister *Hypseocharis,* and the origin of the *Pelargonium* genus were investigated. The possibility of a Cape to Cairo scenario was explored, as were several disjunct distributions within the genus. These disjunctions include *P. karooicum* in the South-African cape, and its sister species *P. caylae, P. endlicherianum* and *P. quercetorum* on Madagascar and in the Middle-East respectively. They also include *P. insularis* on Socotra, with its close sister groups in Ethiopia and the South-African Cape.

While BioGeoBEARS can perform a statistical comparison of six different models, evs requires the user to pick a fitting model based on pre-existing biological knowledge of the studied group. Due to *Pelargonium's* young age, vicariance caused by continental drift was excluded from this model, while sympatry, point sympatry, and jump dispersal were included. BioGeoBEARS found the BAYAREALIKE+j model, which excludes vicariance and point sympatry, to be the most fitting.

Both the BAYAREALIKE+j model implemented in BioGeoBEARS, and the NOVIC model implemented in evs found a South-African origin for *Pelargonium*. The evs output, combined with known divergence times, suggests that *Pelargonium* and *Hypseocharis* diverged via long-distance dispersal. While the ancestral range of the *Pelargonium* and *Hypseocharis* ancestor remains unclear, we suggest that it may have included Antarctica. Both the BAYAREALIKE+j and NOVIC models support a Cape to Cairo scenario as an explanation for the eastern- and northern-African *Pelargonium* species. Abiotic factors such as the Pleistocene uplift, arid corridors, and the formation of the East-African Rift may have mitigated the dispersal of these eastern- and northern-African species.

Long-distance dispersal and not vicariance is behind the occurrences of *Pelargonium* on Australia and St. Helena. The exact order in which *P. caylae, P. endlicherianum,* and *P. quercetorum* colonized Madagascar and the Middle-East is unknown, but long-distance dispersal seems to be behind all of the disjunct distributions.

BioGeoBEARS and evs each have clear advantages and flaws, but complement each other well. When studying historical biogeography, it may be beneficial to use them both.

Contents

Introduction

Historical biogeography

A brief history

Historical biogeography is the study of species' distributions in time and space, and of the estimation of ancestral areas. There are several biological processes that play a role in speciation events. Vicariance, or allopatric speciation, means that a geological barrier is formed which separates a population into multiple factions that then become separate species (Crisci, 2001; Sanmartín, 2012). Since these new barriers often affect more than one species, vicariance can lead to congruent patterns of speciation among clades. Jump dispersal, or founder-event speciation, happens when a few members of a population cross an existing barrier such as a mountain range or an ocean, and then eventually form a new species (Crisci, 2001; Sanmartín, 2012). This does not necessarily lead to congruent patterns within clades. Other types of dispersal also exist: dispersion happens when an existing barrier disappears, leading to species range expansion. Sometimes multiple species are affected by the disappearance of these barriers, leading to congruent dispersal patterns (Lieberman, 2003). This is called "geodispersal" or "predicted dispersal". Jump dispersal can also cause congruent patterns when it is mitigated by abiotic factors. This is called "concerted dispersal" (Sanmartín & Ronquist, 2004). Vicariance and dispersal are not the only processes involved in speciation, as extinction and sympatric speciation also take place. Point sympatry, or subset sympatry, means that one descendant inherits only a small section of the ancestors range, while the other descendant inherits the full range. With full sympatry, both descendants occupy the entire ancestral range.

For a long time, biogeographers have disagreed on whether vicariance or dispersal is behind the current distribution of species, and the processes were often seen as mutually exclusive (Sanmartín, 2012). Nowadays we believe that all the previously described processes played a role in the geographic distribution of species on earth (Crisci, 2001). Sanmartín and Ronquist looked at general patterns of species distribution in the southern hemisphere, and while it was often assumed that vicariance would explain almost all these patterns, their study shows that dispersal has greatly influenced plant diversity (Sanmartín & Ronquist, 2004).

The evolution of species' distribution is inherently tied to the evolution of the earth itself (Crisci, 2001; Sanmartín, 2012). Plate tectonics often cause vicariance, and in some cases also dispersal. Tectonic patterns can be reticulate: continents or islands drift apart, only to merge again later.

Geographic processes, however, are not all that we want to know about species histories. In order to truly place speciation events into context, timing is also important. In the case of vicariance, the barrier that forms and the allopatric speciation that follows are connected (Upchurch & Hunn, 2002), so the timing of both events needs to coincide. Sometimes lineages can display pseudo-congruence. This means they seem to have evolved in similar patterns, but at different times and therefore probably because of different events (Upchurch & Hunn, 2002). Due to reticulate patterns in plate tectonics, pseudo-congruence could occur quite often (Upchurch & Hunn, 2002). It is therefore important to infer the timing of speciation events. Nowadays, species divergence times can be approximated with molecular clock models, which use the DNA-based divergence between species to calculate how long ago they separated (Drummond *et al.*, 2006).

In the second half of the last century, historical biogeography became an increasingly popular topic among biologists and geologists. Over time, many different methods were considered, and often disputed later. The introduction of both plate tectonics and cladistics greatly influenced the field. Hennig (1965) introduced phylogenetic biogeography, the first technique to use cladograms to find ancestral areas. There are three general methods of studying historical biogeography that have been used frequently in the past decade: Pattern based (or cladistic) biogeography (Rosen, 1978), event-based biogeography (Sanmartin, 2007), and parametric biogeography (Ree & Sanmartín, 2009; Buerki *et al.*, 2011).

Cladistic biogeography

The process of cladistic biogeography starts with constructing an area cladogram. A species phylogenetic tree is made, and the species names at the terminal branches are replaced with their area of occurrence

(Sanmartín, 2012). Multiple area cladograms of species from the same regions can be combined into one "general area cladogram" (Sanmartín, 2012) (Figure 1). This cladogram can be used to make assumptions about the histories of these areas (Humphries, 2000), and the vicariance events that separated the species.

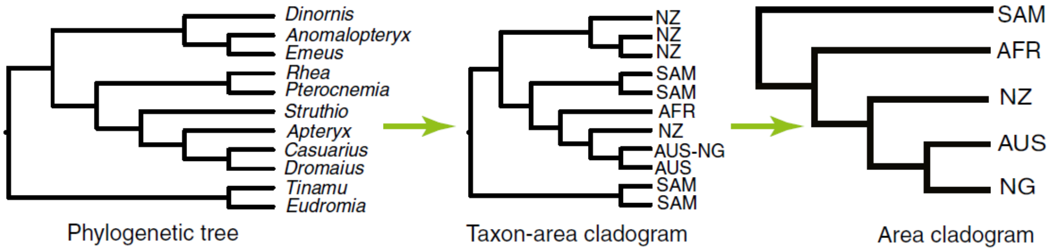

Figure 1: Cladistic biogeography. From Sanmartin, 2012.

Cladistic biogeography, however, only works when there are congruent patterns between the species in the region caused by vicariance events. If the species' distribution was affected by other processes, such as extinction and jump dispersal, then these processes will simply be 'noise' that disturbs the cladogram. Since we now know that dispersal events are not only frequent (Sanmartín & Ronquist, 2004), but can also cause congruent patterns of their own (Lieberman, 2003; Sanmartín & Ronquist, 2004; Sanmartin, 2007), cladistic biography seems too simplistic to be truly informative. New cladistic methods, such as Brooks parsimony analysis (Brooks & McLennan, 2001), can infer geodispersal and dispersion as they cause congruent patterns just like vicariance. However, these pattern based methods assign events *a posteriori (Ronquist, 2002)*, and therefore the possible events do not influence the analysis. In fact, the major limitation of cladistic biogeography is that the events that cause speciation are not considered when constructing the area cladogram, but only afterwards (Ronquist, 2002). The analysis itself focusses exclusively on patterns. (Sanmartin, 2007).

Event-based biogeography

Event-based methods of historical biogeography were designed to use both patterns and processes to find ancestral areas of species (Ronquist, 1994; Hovenkamp, 1997). Biographic processes such as vicariance, dispersal, extinction, and sympatry, are each assigned a cost based on how likely they are to take place. Unlikely events are assigned high costs and frequently occurring events are 'cheaper' . The least costly scenario is the best reconstruction of events (Sanmartin, 2007). While this seems simple, it does come with one complication: the way costs are assigned to events will influence the final result.

Oftentimes maximum parsimony (MP) is used in event based biogeography, meaning the most conserved solutions are the least costly. Vicariance and sympatry are given low costs, because they are constrained, meaning that the total descendants range is the same as the range of their shared ancestor (Sanmartin, 2007). Extinction and dispersal events do not leave conserved patterns, and are thus given higher costs (Sanmartin, 2007). The ideal costs of each process will depend on the specific history of the studied species. Additionally, when extinction or dispersal events leave no descendant in the ancestral location, event-based methods cannot place them in the reconstruction (Sanmartín, 2012).

While maximum parsimony is a popular method, there are some limitations. Because MP methods find the minimum amount of changes, the number of extinctions and dispersal events is often underestimated (Sanmartín, 2012; Sanmartin, 2016). Furthermore, while the timing of events can be inferred by using a parametric tree in the analysis, this timing cannot be used in the analysis itself, and will not influence the results*.* (Sanmartin, 2016). It is important to take the absolute timing of events into account, in order to distinguish between pseudocongruence and real congruent patterns (Upchurch & Hunn, 2002; Sanmartín, 2012).

Parametric biogeography

Throughout the years, the methods used in the field of historical biogeography have changed a great deal, especially with the introduction of new statistical methods and new software. Parametric methods are the newest addition in the field of historical biogeography.

In parametric historical biogeography the evolution of species' ranges is modelled along the phylogeny. This modelling can be done with Bayesian statistics, using Markov chains. The parameters in these models are the biogeographic events, for example vicariance and dispersal. The character states in the Markov chains are the species ranges (Ree & Sanmartín, 2009; Sanmartín, 2012; Sanmartin, 2016). Along the branches of the species' phylogeny, and over time, the character states (areas) can change between ancestors and their descendants, depending on how likely this is (Ree & Sanmartín, 2009). The likelihood of a species arriving at its current range through a possible speciation event, given a certain ancestral range, can be calculated for many possible ancestral ranges. This likelihood of range evolution is partly determined by the prior probabilities, the known descendant and ancestral ranges (Ree *et al.*, 2005).

Information, such as DNA divergence dates and (relaxed) molecular clocks, can be integrated into parametric models. Instead of assigning a set cost or likelihood to events, probabilities of them happening can be calculated directly from the data (Sanmartin, 2016). A major benefit of parametric biogeography is that with it, the timing of events can actually be used in the calculations. If more time passes, it is more likely that the species range will change in some way (Sanmartin, 2016). Another benefit is that parametric models can look at all the alternative hypotheses, and see which one best explains the data (Sanmartín, 2012). Furthermore, different versions of the model, with different parameters, can be compared to see which one fits better. Since the parameters represent biogeographical processes, this comparison shows which processes are most likely to have caused the visibly patterns (Sanmartin, 2016).

Geraniaceae and *Pelargonium*

Phylogeny and distributions

Pelargonium is a genus of plants in the family Geraniaceae and the order Geraniales which mostly occur in Africa and Australia. The genus consists of five main clades: A1, A2, B, C1, and C2 (Bakker *et al.*, 2004; Weng *et al.*, 2012), shown in a phylogenetic tree constructed by Bakker et al (2004). In their study, Bakker et al (2004) have described the geographic distributions of these different *Pelargonium* clades. Clade A is the biggest, and all of its species are found in South-Africa. Clade B contains mostly small annual weeds, and apart from the southern-African species it also contains species that are found on Australia, New Zealand, New Caledonia, and on St. Helena. Dispersal events are likely to have happened within this clade (Bakker *et al.*, 2004). Clade C2, mostly distributed in southern and eastern Africa, has some disjunct distribution patterns. The species P. endlicherianum and P. quercetorum are found in Asia Minor, and their closest sister, *P. caylae*, is found on Madagascar. The sister to these three is *P. karooicum* (Bakker et al., 2005), which occurs in Southern Africa. *P. insularis* is found on Socotra, and its close relatives *P. quinquelobatum* and *P. multibracteatum*, occur in Ethiopia, while most species in clade C remain in South-Africa. Clearly, some notable biological events are behind this distribution (Bakker et al., 2004). Another phylogeny of *Pelargonium* was reconstructed in 2012 using five markers: *rbcL, matK, ndhF, rpoC1*, and *trnL-F (Weng et al., 2012)*. It is similar to earlier reconstructions (Bakker *et al.*, 2004), except the modern methods have shown much better support of the clades. The newest *Pelargonium* phylogenetic tree was constructed by Kerke et al (2019). This phylogenetic tree still supports the five clades found by Bakker et al (2004), and does not contradict the clade C2 distributions. *Pelargonium* has a sister genus, *Hypseocharis,* which occurs in the South-American Andes.

Divergence times

Efforts to unravel the divergence times of Geraniaceae have been made before. The first dated *Pelargonium* phylogeny was reconstructed by Bakker et al (2005) using plastid DNA trnL-F and nuclear rDNA ITS. It estimated *Pelargonium* to be about 30 million years old. In 2008, Fiz et al. used molecular dating and Dispersal-Vicariance analysis (DIVA) to find out when the genera of Geraniaceae diverged and where they occurred. They found that Pelargonium diverged from its sister groups 38-47 Mya (Fiz *et al.*, 2008). This coincides with the findings of an earlier study that used DNA divergence data to find divergence times for the entire APG (Wikström *et al.*, 2001). According to Fiz et al, the *Pelargonium* crown age is approximately 10 My.

In 2012 Palazzesi et al (2012) performed another study including Geraniaceae, using molecular dating and focussing on discovering the timing of divergence events. They used two genes: *trnL-F* and ITS, and included five new fossils as calibration points. According to Palazzesi et al, *Pelargonium* split from the rest of the Geraniaceae 29 Mya, and the *Pelargonium* crown is 11 Million years old. They also found that Hypseocharitaceae and Geraniaceae diverged 37 Mya (Palazzesi *et al.*, 2012).

A more recent study, however, disagrees with these findings (Sytsma et al*.*, 2014). The authors found several problems with Palazzesi et al.'s methods. Firstly, the two genes they chose evolve relatively fast, and are therefore at risk of substitution saturation. This means the amount of divergence between the branches could be underestimated, making the splits seem more recent than they actually are. An earlier study specifically testing pitfalls in molecular dating already warned that using fast evolving genes by themselves is risky (Zeng *et al.*, 2011). Sytsma et al. also argue that while the five new fossils are useful, their placement is questionable. In their own study they mostly repeat all the steps of Palazzesi et al. and included the five fossils, but they added three additional priors, placed at deeper nodes. The addition of these priors changed the phylogenetic tree dramatically. The new reconstruction closely resembles an older version (Wang et al. 2009). According to this new phylogenetic tree the *Pelargonium* crown is around 20 My old, and the *Pelargonium* stem diverged from the other Geraniaceae around 50 Mya. The Hypseocharitaceae and Geraniaceae diverged 62-68 Mya (Sytsma *et al.*, 2014).

The most recent *Pelargonium* phylogeny, made using 78 plastome exons and nuclear DNA, estimates the *Pelargonium* crown to be around 9.8 My old (van de Kerke *et al.*, 2019). This is similar to the age found by Palazzesi et al. The phylogeny shows radiations happening from the late Miocene, and throughout the

Pliocene and Pleistocene. According to the phylogeny, *Hypseocharis* split off from the Geraniaceae 35.8 Mya (van de Kerke *et al.*, 2019).

Historical biogeography of *Pelargonium*

Hypseocharis

Fiz et al (2008) studied the historical biogeography of Geraniaceae using DIVA. DIVA placed the shared ancestor of Hypseocharitaceae and Geraniaceae in southern Africa and South America. The Andes, where *Hypseocharis* occurs, started to form in the Oligocene around 33 Mya. However, mountains continued to rise in the millions of years after (Taylor, 1991). The common ancestor of the Geraniaceae and the common ancestor of *Pelargonium* appear to have occurred in southern Africa (Fiz *et al.*, 2008). South America and Africa started to split between during the early cretaceous so 145 years ago. According to dated phylogenies the last common ancestor of *Hypseocharis* and *Pelargonium* is younger than 145 My (Palazzesi *et al.*, 2012; Sytsma *et al.*, 2014; van de Kerke *et al.*, 2019). It therefore seems unlikely that the last ancestor of Hypseocharis and Geraniaceae lived across both continents when they were connected.

Antarctica was connected to the south of South-America after Gondwana broke apart, until about 35-40 Mya (Livermore *et al.*, 2005). Since it is covered in ice, it is easily dismissed when studying extant plant distributions, but Antarctica only started cooling down in the late Eocene (around 34 Mya) (Zachos *et al.*, 2001; Galeotti *et al.*, 2016). Before that, it was covered with temperate flora, including angiosperms (Iglesias, 2016; Estrella *et al.*, 2018). Many of the plant lineages that used to cover Antarctica still exist (Estrella *et al.*, 2018). Due to its proximity to South-America, and it being in long-distance dispersal range of Africa, it is possible that the ancestor of *Pelargonium* and *Hypseocharis* occurred on Antarctica just before it froze, or that Antarctica mitigated the dispersal of *Pelargonium's* ancestor to Africa as a land-bridge. Recently, Estrella et al (2018) made a strong case as to why Antarctica should not be ignored in historical biogeography studies. They summarize the plant fossils found on Antarctica, and demonstrate that Antarctic presence is not limited to older plant lineages (Estrella *et al.*, 2018). In short, a possible role for Antarctica should be considered when studying the evolution of the Geraniaceae.

South-America and Antarctica were connected around the time the Geraniaceae + *Hypseocharis* ancestor existed, though they were no longer connected to Africa (Livermore *et al.*, 2005; Estrella *et al.*, 2018). While this means that vicariance is an unlikely explanation for the *Pelargonium* origin, it does mean that our potential species range for this ancestor should include Antarctica. Direct dispersal from South-America to Africa is also a definite possibility, for example by wind or by sea water (Renner, 2004). A more detailed review on trans-oceanic dispersal will follow in the paragraph on long-distance dispersal.

Clade A: Climate change and radiations in South-Africa

Climate changes

In the last 20 Million years, many environmental changes happened in Africa (Linder, 2003; Linder, 2017). Tectonic events formed mountain ranges, platforms and volcanoes, and climatic changes led to gradual shifts in vegetation type (Linder, 2017). There is still much that we do not know about the vegetation that covered East-Africa during the Cenozoic, but we do have some evidence that reveals the existence and timing of certain events, such as the uplift of the East-African Plateau (Wichura *et al.*, 2015; Linder, 2017). Once we have a solid estimation of the distribution patterns of *Pelargonium* in space and time, these events could provide context, telling us why and how speciation events might have happened. They could also provide evidence for or against our hypotheses.

Climate changes in southern Africa are likely to have had an effect on the distribution of *Pelargonium*. From the Miocene until the Pleistocene, so from 23 Mya until 2.6 Mya, the global climate changed: it became colder and drier (Zachos *et al.*, 2001). There are not many fossils from the CRF that can provide information about the local climate in this era, but the ones we do have suggest a summer-dry, Mediterranean climate from 10 Mya onwards (Linder, 2003). Before this, the cape was covered in mostly tropical vegetation (Linder 2003). On a global level, these climatic changes were driven by events such as the opening of the Drake passage (Scher & Martin, 2006) and worldwide tectonic forces (Zachos *et al.*, 2001). In the Cape, it is possibly that the Benguela upwelling system across the Namibian coast caused the temperature to go down (Linder, 2003). The aridification that took place in eastern Africa around this time is demonstrated by the increase of C4 over C3 plants, from their emergence around 8 Mya to their current dominance (Ségalen *et al.*, 2007). The late Miocene aridification of Africa seems to have contributed to the increase in diversity of *Pelargonium*, as both events happened around the same time, and many *Pelargonium* species possess some degree of drought tolerance (Fiz *et al.*, 2008).

Climate change can cause large areas to split-up into multiple smaller ones with different climates. The environments start to differ, and this can effectively result in vicariance of the species present there, leading to speciation (Linder, 2003; Martínez-Cabrera & Peres-Neto, 2013). The preferred niches of the *Pelargonium* species in the CFR have evolved relatively quickly. This could be caused by increased dryness and changes in rainfall patterns. For example, closely related species preferred different rainfall patterns along newly formed rainfall gradients (Martínez-Cabrera & Peres-Neto, 2013). The new Mediterranean climate could have supported the *Pelargonium* radiation as well.

Radiations

In Pelargonium multiple radiation events happened within the same clade, millions of years apart (Bakker *et al.*, 2005). These nested radiations appear to have occurred in clade A2, which is completely South-African, with its species occurring in the winter rainfall region (Bakker *et al.*, 2005). According to Bakker et al, the first radiation was that of the Xerophytic clade, around 18 Mya. The second radiation, nested within this clade, occurred in the section *Hoarea* around 10 Mya (Bakker *et al.*, 2005). According to new phylogenetic inferences, these radiations happened more recently (van de Kerke *et al.*, 2019), but they still appear to be nested. Since radiations happened multiple times, they may also have happened for different reasons. One possible explanation is that species rapidly evolved traits that improved drought tolerance, right as the area became more arid (Bakker *et al.*, 2005). It is likely that this caused the radiation of the Xerophytic clade (Bakker *et al.*, 2005). This is an example of an adaptive radiation, as the newly arid environment seems to have given rise to new niches for these species to occupy (Gittenberger, 1991). Non-adaptive radiation would be when the species of a rapidly radiated clade still occupy similar niches, and do not appear to be ecologically diverse (Gittenberger, 1991). Since the *Hoarea* species occupy similar niches, their radiation could have been non-adaptive (Bakker *et al.*, 2005). However, the *Hoarea* radiation does correlate with the *Hoarea* species' adaptations to a wide variety of pollinators. Therefore, the radiation of *Hoarea* flowers may have been adaptive after all. Since A2 is the youngest of the *Pelargonium* clades, these radiations may not provide evidence for a South-African origin of *Pelargonium*, but they do demonstrate the potential of the CFR as a centre of biodiversity. South-Africa has been shown to have a high floral carrying capacity when compared to the rest of Africa (Sanmartín *et al.*, 2010).

Clade B: vicariance or trans-oceanic long-distance dispersal

Vicariance

Sanmartín and Ronquist (2004) looked at the role of dispersal in southern hemisphere speciation, and found that dispersal had a large effect on the distribution of plants. They also found that the breaking-up of Gondwana did not really influence plant diversity (Sanmartín & Ronquist, 2004). It is possible that these studied plants are simply too young, and did not exist when Gondwana split up. It is also possible that vicariance did play a role in their distribution, but that years of dispersal and extinction erased those patterns. Evidence of this second option does exist, and fossils, for example from the Monimiaceae, show that some plants families are older than the break-up of Gondwana (Renner *et al.*, 2010). Thus, vicariance still needs to be considered when reconstructing the historical geography of plants.

A family with a similar distribution to *Pelargonium* is Monimiaceae. Renner et al (2010) researched whether or not the distribution of this family was influenced by the breaking up of Gondwana. They state that, since East-Gondwana broke up between 120 Mya and 40 Mya, the end of the break-up could have influenced the spread of the Monimiaceae. West-Gondwana was already separated 100 Mya, and back then the Monimiaceae did not exist (Renner *et al.*, 2010). A South-American and a Australian Monimiaceae clade seem to be connected, their dispersal aided by Antarctica acting as a land bridge around 57 Mya. Some long-distance dispersal from Australia to New Caledonia and New Zealand, where several Pelargonium species occur, also took place in the Monimiaceae, as did long-distance dispersal across the Indian ocean (Renner *et al.*, 2010).

Dispersal or vicariance in *Pelargonium* happened clearly in the section *Peristera* (clade B), to Australia and surrounding islands (Bakker *et al.*, 1998). *Peristera* includes most of the *Pelargonium* species that occur outside of Africa. They occur in Australia, New Zealand, New Caledonia, and Tasmania, but there are also *Peristera* species in South-Africa. The African and non-African *Peristera* species have similar habitats, and therefore a close relationship between the species always seemed likely (Bakker *et al.*, 1998).

Long-distance dispersal

Since Australia separated between 66 and 145 Mya, and *Pelargonium* appears to be around 10 My old (van de Kerke *et al.*, 2019), vicariance seems an unlikely explanation for the Australian *Pelargonium* species. Australia was last in contact with Antarctica 35 Mya (Sanmartín & Ronquist, 2004), long before the first Peristera species arose. Antarctica can therefore not have acted as a bridge. We could therefore hypothesise that the *Peristera* ancestor arrived in Australia after a long-distance dispersal event. Long distance dispersal of plants to Australia is not rare. Crisp and Cook (2013) studied 85 clades of Australian plants, and found that 48% arrived in Australia through long-distance dispersal. Many specific examples have been found, for example in Asteraceae (Bergh & Peter Linder, 2009). Concordant dispersal patterns, dispersal facilitated by various abiotic factors, are also present in plants (Sanmartín & Ronquist, 2004). The jump of *Pelargonium* from Africa to Australia could have happened as recently as 5 Mya (Bakker *et al.*, 1998; van de Kerke *et al.*, 2019). Trans-Tasman winds may explain the long distance dispersal events between Australia and New-Zealand (Pole, 2001).

A plant's general dispersal ability can often be linked to the shape of its seed (diaspore) (Heleno & Vargas, 2015). There are four basic diaspore's that match with four ways to disperse long distances: dispersal by wind, water, the intestines of animals, and the fur or feathers of animals (Heleno & Vargas, 2015). The seeds of *Pelargonium* (figure 2) appear to be best suited for wind-dispersal and epizoochory: sticking to fur or feathers.

Figure 2: Pelargonium Australe with seeds

Seed morphology, however, cannot be used to reliably predict the ways in which plants disperse over long distances. While their results were not fully conclusive, Higgins et al (2003) suggest that plants often rely on different dispersal method than their standard one when dispersing unusually long distances. It has since been shown that seeds of many different species can be dispersed by ways for which they are not adapted (Myers *et al.*, 2004; Heleno & Vargas, 2015), and that non-standard dispersal is not rare (Nathan, 2006). Thus, *Pelargonium* seeds could have gotten stuck to a bird (epizoochory), or have been carried by ocean or wind currents (Renner, 2004), regardless of seed morphology. In the family Urticaceae, research has shown that non-germinated seeds can survive in sea water long enough to travel long distances to isolated islands (Wu *et al.*, 2018).

Aside from normal wind, water, or animal facilitated dispersal, there are also more unpredictable and rare dispersal mechanisms (Nathan *et al.*, 2008). Extreme weather can carry seeds further than normally, perhaps even seeds that are not normally dispersed by wind, and natural rafts can be carried by ocean currents while preserving the seeds they carry (Waters & Roy, 2004; Nathan *et al.*, 2008; Heleno & Vargas, 2015). Extreme weather may even offer the benefit of disturbing the environment it deposits the seeds in, giving the invading species an advantage (Wu *et al.*, 2018).

One disadvantage of dispersal through extremely rare events is that it often takes multiple seeds for a species to settle, especially if plants are not self-compatible, or dioecious (de Waal *et al.*, 2014; Wu *et al.*, 2018). If dispersal is rare, the chance of successfully settling is very low. While the focus in historical biogeography studies tends to lie on the journey to different locations, the establishment of a species in a new location can be just as challenging. *Peristera* species have a high capacity for dispersal because they are small weeds that can colonize new regions relatively easily (Bakker et al., 1998).

Clade C: Long-distance dispersal across eastern Africa.

Cape to Cairo hypothesis

According to Bakker et al (2005), the species *P. karooicum, P. caylae, P. endlicherianum* and *P. quercetorum*, diverged around 19 Mya possibly through long-distance dispersal. Bakker *et al* also suggest that these species could provide evidence of *Pelargonium*'s history (Bakker *et al.*, 2005). Perhaps *Pelargonium* moved towards the Cape, and originated somewhere more northern. The northern *Pelargonium* species could be the remaining evidence of this movement (Bakker *et al.*, 2005). The most recent phylogenetic tree found that *P. karooicum, P. caylae*, *P. quercetorum* and *P. endlicherianum* diverged much earlier, approximately 5 Mya (van de Kerke *et al.*, 2019). This does not dispute the possibility that these species are traces left over from *Pelargonium's* migration across the African continent.

Galley et al (2007) studied historical biogeography in *Disa,* Iridae, *Pentaschistis* and Restionaceae, and discovered that their common ancestor occurred in the Cape. Migration from the Cape to other, more northern regions was relatively common, while migration towards the cape rarely happened. It seems that a Cape to Cairo scenario is more likely than a north to south migration scenario (Galley *et al.*, 2007). South to north plant dispersal was summarized by Sanmartin et al (2010). Other examples of plant species that originated and diversified in the CFR and then moved northwards include *Helichrysum* (Andrés-Sánchez *et al.*, 2018). Many CFR species are endemic to their highly specific ranges so it is likely that they first evolved in the CFR, and not somewhere else (Linder, 2003). The histories of other plant lineages suggest that Cape to Cairo migration is a relatively common occurrence. Therefore, it could have happened in *Pelargonium.*

Dispersal across eastern Africa

Long distance dispersal events of plants across entire continents are not unheard of. Jump dispersal is thought to be the cause behind the North-Eastern Euryops species (Mucina *et al.*, 2010). These species, however, form a monophyletic group (Mucina *et al.*, 2010), and this is not the case with the Ethiopian *Pelargonium* species. Aside from the two groups discussed in detail, there are several other northern and eastern African *Pelargonium* species, such as *P. rungvense*, *P. hararense*, and *P. erlangerianum*, that do not appear to form a monophyletic group (van de Kerke *et al.*, 2019). Several independent dispersal events in the same direction suggest a pattern, which means that these events were possibly facilitated in some way by external factors. One possible influence on *Pelargonium* dispersal is the existence of African Arid Corridors (AAC's) (Bellstedt *et al.*, 2012).

Bellstedt et al (2012) identified five migrations from south-west to north-east Africa in the plant family Zygophylloideae. Three of these happened in the last 5 million years, the *Pelargonium* dispersal timeframe. Bellstedt et al suggest that this pattern can be explained by a migration corridor in eastern Africa. This corridor, the AAC, has been identified before, but it may just be a long-distance dispersal route instead of an actual corridor. It may have existed multiple time, and its precise timing is unclear (Bellstedt *et al.*, 2012).

The possibility of migration corridors across eastern Africa towards Ethiopia was investigated by Chala et al (2017). They showed that drought-tolerant plant species inhabiting alpine environments could have used grassland and forest present in the Pleistocene to cross otherwise inhabitable mountain ridges (Chala *et al.*, 2017). More evidence of an Arid belt across eastern Africa was collected by Jurgens (1997), who demonstrated that floral habitats show patterns of fragmentation, suggesting that what was once an arid migration corridor has now broken up into several habitable ranges (Jürgens, 1997). It is possible that the trail of *Pelargonium* species across eastern and northern Africa have been similarly divided by habitat fragmentation in an Arid belt. Similar habitat fragmentation patterns were studied by (Fer *et al.*, 2017), who show that forest patches across eastern Africa could have been connected in the past, during the Miocene. They suggest that the formation of the East-African Rift could have caused some East-African areas to become more arid, causing the forest to fracture.

East-African rift and Pleistocene uplift

The East-African Rift is a mountain range in East Africa, that has formed over the past 17 My (Wichura *et al.*, 2010; Wichura *et al.*, 2015). New orographic events have been happening continuously and changing the local environment. Its formation is still ongoing (Sommerfeld *et al.*, 2016). Some of these orographic changes could have influenced the dispersal of Pelargonium ancestors, and even recent species. East-African Rift formation may also have caused some of the African aridification, starting around 15 Mya (Sommerfeld *et al.*, 2016). According to Wichura et al (Wichura *et al.*, 2010; Wichura *et al.*, 2015), the timing of the East-African uplift correlates with the formation and expansion of East-African grasslands and an increase in C4 plants, around 10 Mya.

It is suggested that the Pleistocene uplift in Africa might have formed a corridor through and over the tropical forests that acted as a barrier (Axelrod & Raven, 1978). This way, the non-tropical Pelargonium species *P. aridum*, *P. quinquelobatum*, and *P. insularis* could have dispersed from the cape to different locations. This dispersal appears to fit within the Pleistocene timeframe (van de Kerke *et al.*, 2019). Vicariance due to the separation of Socotra may have caused P. insularis current distribution, but long distance dispersal seems more likely. In the plant genus *Thamnosma*, which includes both African and Socotran species, long distance dispersal was inferred to be the reason behind its distribution (Thiv *et al.*, 2011). This was mainly because the latest possible vicariance event would have happened 18 Mya (Thiv *et al.*, 2011). Dispersal is the cause behind at least part of the Socotran flora (Thiv & Meve, 2007; Thiv *et al.*, 2011).

Migration patterns in Africa have been studied before. Sanmartín et al (2010) found that while exchange between northern and southern Africa appears low, the migration/dispersal that took place seems to have gone through the east, which matches the Pelargonium pattern. There is noticeable exchange between southern and eastern Africa (Sanmartín *et al.*, 2010). This matches results found by Ali et al (2013), which found several dispersal events in the plant subfamily Urgineoideae from South-Africa to eastern Africa, also within the past five years (Ali *et al.*, 2013).

Madagascar

Indian ocean dispersal could explain the occurrence of Pelargonium species in Asia minor and on Madagascar. Madagascar has been disconnected from Africa for 140 My, and it has been a separate island for 88 My (Reeves & De Wit, 2000; Yoder & Nowak, 2006). The speciation event that landed Pelargonium caylae on Madagascar happened within the last 10 My. Long-distance dispersal towards Madagascar from other continents such as Africa and Asia has been demonstrated and is not rare. At least 8 lineages of the plant family Celastraceae originate in Africa and arrived on Madagascar through independent jump-dispersals (Bacon *et al.*, 2016). Long distance dispersal from South-America and Africa towards Madagascar are behind the formation of two Madagascar clades of the Spermacoceae tribe (Janssens *et al.*, 2016). An earlier study summarized the origins of many species inhabiting Madagascar (Yoder & Nowak, 2006). Not only did they find that the vast majority of species arrived through long-distance dispersal sometime in the last 65 My, they also found that Africa was the largest contributor of Madagascar species, for plants, vertebrates, and invertebrates. Beside Africa, Asia was also a large contributor (Yoder & Nowak, 2006). Another example of plant dispersal towards Madagascar has been demonstrated in the plant genus *Plukenetia* (Cardinal-McTeague *et al.*, 2019).

Many species on Madagascar resemble Asian species, even though most of these species did not exist when Madagascar and Asia drifted apart (Warren *et al.*, 2010). Apparently, long-distance dispersal between Madagascar and Asia is not rare. There have occasionally been, and still are, islands in between these places mitigating this dispersal (Warren *et al.*, 2010). These islands were also present during the dispersal window of Pelargonium (Warren *et al.*, 2010). Hawaiian islands were used as stepping stones by *Melicope*, another plant genus (Harbaugh *et al.*, 2009). Similar stepping stone mechanisms might explain the distribution of the sister Pelargonium species *P. enlicherianum, P. quercetorum* and *P. caylae*. Perhaps a *Pelargonium* ancestor dispersed to Asia first, and then from there to Asia Minor.

P. Karooicum may have been created through hybridization (van de Kerke *et al.*, 2019). This would explain why it is difficult to place in the phylogenetic tree, and shares characteristics with multiple sectors. A more complete phylogenetic tree could improve our knowledge on the placement of several sections and unassigned species, such as *P. Caylae* and *P. Karooicum (van de Kerke et al., 2019).*

Historical biogeography: explicit vs non-explicit species ranges

BioGeoBEARS

Every method used in historical biogeography has to make some assumptions about geographical processes that have occurred. These assumptions can potentially have a large impact on the results of studies, even more so than the statistical method that is chosen (Matzke, 2013b). Most historical biogeography models don't use explicit geographic ranges of the species, but instead use area's that are defined before the analysis. The areas can be defined in many of different ways, but because the models usually cannot handle more than around ten of them they are often fairly large. This means that some species will get lumped together in an area even though they are actually separate (Arias, 2017). The way in which the areas are defined will inadvertently influence the outcome of the analysis, so it needs to be thoroughly thought through.

The R package BioGeoBEARS (Matzke, 2013b) implements several popular methods at once, such as DIVA, DEC and others, all of them are likelihood versions of the original (maximum parsimony-based) models. It combines al the possible parameters (processes) from these models, and adds a new one: founder-event speciation (jump dispersal). Matzke shows in his paper that the inclusion of founder-event speciation can have a large effect on the results (Matzke, 2013b). BioGeoBEARS also allows you to turn off parameters by setting them to zero. Each implemented model is made turning off the right parameters.

In essence, BioGeoBEARS offers a model selection method. It simulates many earlier models and allows the user to manually adjust them. Founder-events can be included in each model, creating "model+J" versions. Multiple models are run, and common statistical tests, such as the likelihood ratio test and the Akaike information criterion, can be used to see which model fits the best (Matzke, 2013b). It would be interesting to use BioGeoBEARS to see to what extent jump dispersal played a role in the history of *Pelargonium*, by comparing the +J versions with the normal models.

BioGeoBEARS does not use explicit species ranges, meaning the user has to manually assign ranges to every species. This means that the users choice areas will influence the outcome of the analysis.

GEM

GEM (Arias, 2017) is a historical biogeography program that uses an event-based method. Original event-based models don't use the exact geographic ranges of the species (Arias, 2017). There are newer event-based models that can use explicit geographical ranges, such as the DA-model, but these have another problem: they can only incorporate one biological process/event, for example vicariance. This means that nodes that do not seem to be caused by vicariance cannot be used (Arias, 2017). GEM promises to use both explicit geographical ranges and allow for multiple events (Arias, 2017).

GEM assigns events to internodes without looking at the actual size of the range. It decides which event fits best based on how the descendants inherit the range. The processes include vicariance, sympatry, point sympatry, and founder events. Events are decided on through cost/probability models. Many different reconstructions can be made, with different events and with different ancestral ranges, and the cheapest/most likely reconstructions are chosen (Arias, 2017). Ranges are modelled using pixel grids, where pixels are filled in if a taxa is present there. Usually, geographic data comes from collected specimens, and it is often incomplete. If the raster contains many of pixels it can be more accurate, but there will also be more gaps because of the incomplete data. To deal with this, a second pixel grid is made (beside the one with observed locations), where gaps around pixels are filled out by an algorithm. This also means that the recorded locations are a little less sensitive, as they are buffered by the pixels filled in around them. Range chance along the lineage is measured as the amount of pixels that are lost or gained.

Every ancestral pixel has to be included in one of the descendants ranges, and a pixel present in both descendants ranges has to be included in the ancestral range. These two rules are used to simplify the cost-matrix (Arias, 2017). Every descendant pixel lost or gained relative to the ancestral range brings an extra cost. In the default version of GEM all events are seen as equally likely, and they cost the same. GEM can also be used to emulate other methods such as DE and DIVA (Arias, 2017).

Reconstructing the ancestral range from two descendant ranges when both descendent ranges have several exclusive pixels leads to many of possible solutions (with dispersal and extinction) that all need to be stored. To reduce the amount of data, ancestral reconstructions are limited to: Descendant range 1 + descendant range 2, just descendant range 1, or just descendant range 2. If the two descendants are combined either vicariance or sympatry is assigned, and if just one descendant range is used sympatry, subset sympatry, or a founder event is assigned (Arias, 2017). Note that even with only 8 options per node there is still a large amount of possible reconstructions. To decrease the complexity of the analysis, the "flip algorithm" is used: Only one node of the tree is changed at the time, and only the changes that lead to lower costs are kept. This way the eventual reconstruction is the most parsimonious one (Arias, 2017).

GEM does have some possible shortcomings, described in the paper. Firstly, GEM does not give any information on the branch lengths. There is an implementation which can be used but Arias *et al* discourage the use of this as it can cause other problems with the analysis. Dispersal distances are also not given by GEM (Arias, 2017). Finally, evs does not have a model-comparison function, meaning the user has to choose the model themselves, which adds uncertainty.

A comparison

A big difference between GEM and BioGeoBEARS is that GEM uses explicit areas whereas with BioGeoBEARS The user will have to determine the areas, and there is a limit to how many they can use (Matzke, 2013b). This also means the determined areas will be quite large, and thus the program may not pick up on small-scale vicariance events caused by habitat fractioning, or shorter-ranged dispersal events.

Another difference is that BioGeoBEARS allows maximum-likelihood and Bayesian statistical methods, while GEM uses a cost-based method. BioGeoBEARS therefore provides useful options for model comparisons.

Evidently, both programs have advantages and drawbacks, and it will be interesting to see if the results are anything alike. Perhaps the results will be complementary, and together give a more complete picture of *Pelargonium* history, or perhaps they will be contradictory.

Both biogeographical reconstructions need a detailed parametric phylogeny as input, and the most recent one will be used in this study (van de Kerke *et al.*, 2019).

From this point, GEM will be referred to by its computer-implementation, evs.

Main problems, scientific and societal benefit, and expected results

Main problems

Three main questions about the history of *Pelargonium* will be the focus of this study. The first two concern the origin of the *Pelargonium* ancestor, and of its sister genus *Hypseocharis* which can be found in South-America*.* We want to find out where the last shared ancestor of *Pelargonium* and *Hypseocharis* occurred, and how the two genera diverged. We also want to see whether *Pelargonium* originates in South-Africa, or in a more northern region of Africa. Finally, we want to take a closer look at several closely related sister groups with unusual distributions. *P. karooicum, P. caylae, P. endlicherianum, P. quercetorum* occur in South-Africa, Madagascar and the Middle East (Asia Minor) respectively. *P. multibracteatum* and *p. quinquelobatum* occur in Ethiopia while *P. insularis* can be found on Socotra. Their closest sisters occur in South-Africa. We want to know which events took place that led these species to their current distribution.

Scientific and societal benefit

With climate change threatening different ecosystems around the world, it is important that we understand how similar climatic changes have influenced species in the past. Such changes likely fractured habitats and caused vicariance and extinction events. Using historical biogeography we can find out more about how environmental factors in the past influenced current species' distribution patterns. Our time and resources for conservation are limited, and therefore the best way to protect at-risk species is to anticipate how climate changes will influence them, so we can help the ecosystems that need it the most. While this study is about the history of *Pelargonium*, the distribution and speciation patterns found in this genus may resemble those of many others. It could therefore contribute evidence for or against longstanding hypotheses in the field.

Another way this study could prove useful is by illuminating new ways to improve biogeographical research. Historical biogeography programs that use explicit geographical ranges are still relatively new to the field, and evs (GEM) is the first that also allows for multiple biogeographical processes. To my knowledge, no large studies have tested evs at the time of publication, especially in comparison to a tried and true method such as BioGeoBEARS. It will be interesting to find out whether the explicit ranges have an influence on the outcome, and whether evs is superior to BioGeoBEARS because of it. If this is the case, explicit geographical ranges could become an essential part of historical biogeography in the future.

Expected results

I expect that the last common ancestor of *Hypseocharis* and *Pelargonium* occurred in South America, South-Africa, possibly Antarctica, or a combination of these. Currently, it seems as though *Hypseocharis* diverged from the Geraniaceae after the Gondwanan break-up. Therefore, long-distance dispersal seems the most likely cause of their divergence.

I expect that *Pelargonium* arose in South-Africa, and later dispersed to the north and east. Based on literature, I believe a Cape to Cairo scenario is the most likely explanation for *Pelargonium's* distribution in and around Africa.

The *Pelargonium* species on New Zealand and Australia likely arrived through long-distance dispersal.

Long-distance dispersal could have brought *P. caylae* to Madagascar, and long distance dispersal may also be behind the distribution of *P. multibracteatum, p. quinquelobatum* and *P. insularis*. The North-African *Pelargonium* species do not form a monophyletic group, and external factors may have mitigated dispersal from the south through the east to the north of Africa. Arid corridors, mountains formed by the Pleistocene uplift, and orogenic changes caused by East-African Rift formation are examples of such factors.

Stepping stone islands could have been involved in the trans-oceanic *Pelargonium* dispersals.

Method

Data collection

Compiling a coordinate dataset

A large dataset containing coordinates from 246 African *Pelargonium* and 7 Australian *Pelargonium* species was provided to me by Timothy Moore and Cynthia Jones from the University of Connecticut. This data was obtained from the South-African National Biodiversity Institute's PRECIS database (South African National Biodiversity Institute, 2018), the Harry Bolus Herbarium, BOL, at the University of Cape Town, and section *Hoarea* records from Marais (1994).

Additional *Pelargonium* location data was extracted from GBIF (GBIF.org, 2019a, d, c, b). The amount of coordinates per species is irrelevant, which is why potential overlap between the datasets is not important. During the data-search samples from herbariums and botanical gardens were excluded. Furthermore, only observations that included coordinates were used. Each dataset was carefully edited, and uncertain species occurrences that seemed out of place were removed if the locations could not be confirmed. *Hypseocharis bilobata* occurrence data was also extracted from GBIF.

To complete the dataset, a list of species was made from the combined original and GBIF data, and this list was compared to a list that includes information on all accepted and ambiguous *Pelargonium* species. All the species missing from our dataset were listed. The ones that are (partially) taxonomically accepted were kept while the rest was removed. Then, for these accepted species, JSTOR Plants was used to search for information on the location of some of the first submitted specimens of the species (for citations see supplementary material: table L). Every available detail was used to narrow down the location. Google earth was used to find the location based on this information, markers were placed, and coordinates were noted down.

Some *Pelargonium* species' names are synonymous with other *Pelargonium* species. Such synonyms were removed using a comprehensive *Pelargonium* species list, Rosenblëck et al's comprehensive review on *Pelargonium* taxonomy (Röschenbleck *et al.*, 2014), and with extra help from Timothy Moore and Cynthia Jones. The synonymous-species' location information was merged with the original species' location info.

Collecting phylogenetic data

A parametric phylogenetic tree including 106 recognized *Pelargonium* species was made by Kerke et al (2019). Both BioGeoBEARS and evs need a parametric tree that includes all the species in the dataset to work. Since the 2018 tree was incomplete, the missing species had to be manually added using R-package phytools (Revell, 2012). Older phylogenetic and taxonomic studies (Bakker *et al.*, 2005; Röschenbleck *et al.*, 2014) were used to find the best place to attach each species. The species were added section by section. The missing species with clear sister species present in the 2018 tree were attached first. Then the species that had a clade as a sister were attached. The Bakker et al (2005) tree used to attach species contains many polytomy's. Furthermore, every species not include in this tree needed to be added in a polytomy at the base of their entire section. Because BioGeoBEARS and evs do not allow for true polytomy's, these species were instead attached as sisters to the clades, with very short branch lengths.

The final full tree has 258 taxa excluding *Hypseocharis bilobata.* It can be seen in figure 3. A cleaned unaltered version of the tree was also prepared. This unaltered tree consists of 106 taxa excluding *H. bilobata.* Finally, a tree consisting of only clades B and C was prepared.

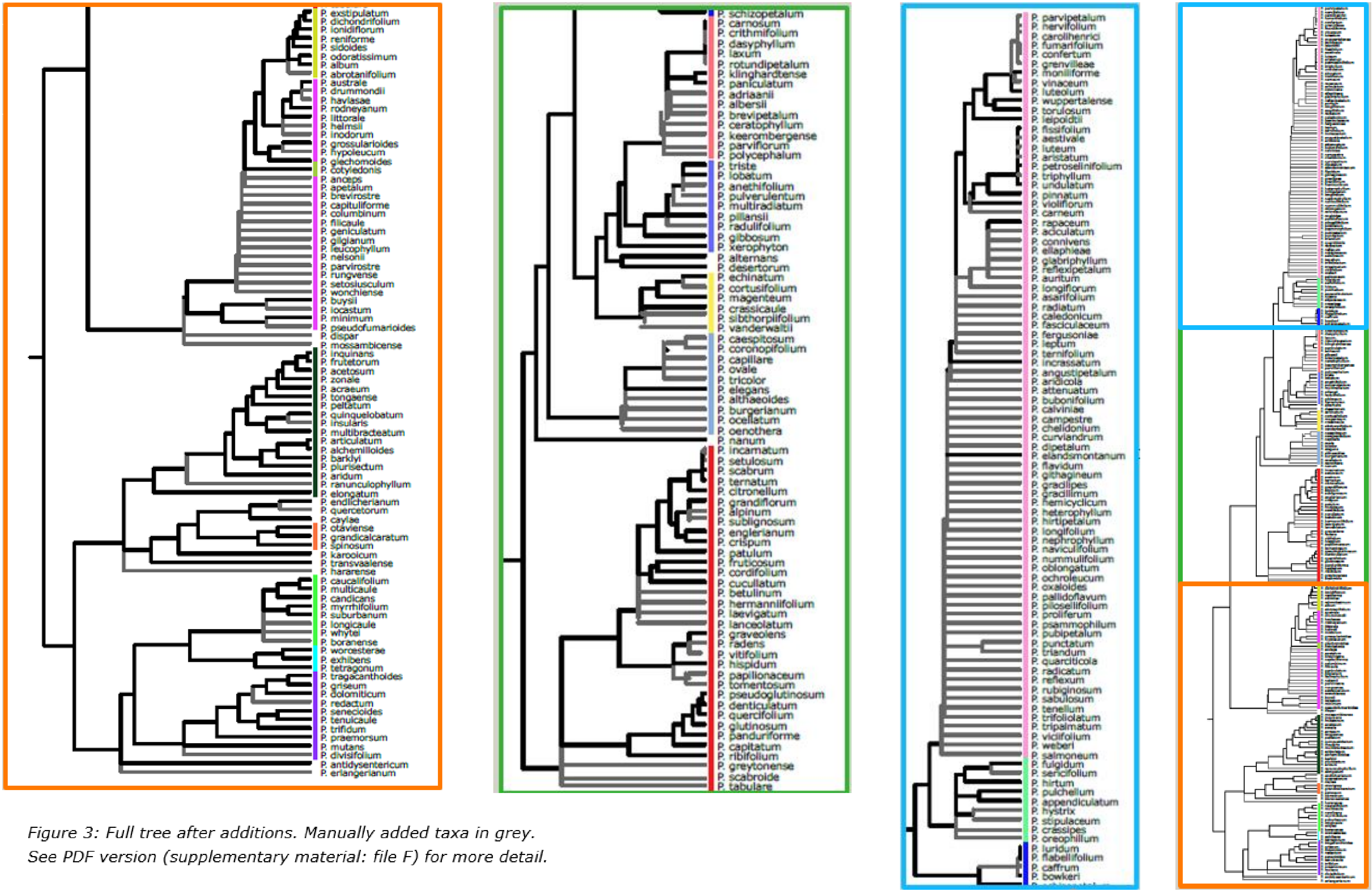

Figure 3: Full tree after additions. Manually added taxa in grey.
See PDF version (supplementary material: file F) for more detail.

BioGeoBEARS

Initial run

A preliminary BioGeoBEARS analysis was run before the definitive one. The coordinate data from both datasets ("original" and "GBIF"), were plotted in google earth pro, around 12 species at the time. Following recommendations in the BioGeoBEARS tutorial, the maximum number of areas for this preliminary run was set at 11. A division was made based on species' occurrence, and areas described in previous studies.

South-America was chosen because it is where *Hypseocharis* occurs. Australia and New Zealand were used together because there were no large differences in species occurrence between the two. St. Helena was taken as a separate area. Madagascar was chosen as an area because we are particularly interested in the ancestral range of the species that occurs there, *P. caylae*. Socotra and Asia Minor were used as areas for the same reason. Many of the non-South-African *Pelargonium* species occur on or alongside the mountains of the East-African Highway. This mountain range is believed to have possibly been an driving factor in the dispersal of *Pelargonium* across Africa. Therefore, the strip including all *Pelargonium* species in Ethiopia, Tanzania, Kenya, Malawi, and northern Mozambique was chosen as one single area. South-Africa and southern Mozambique were divided into four areas. The cape floristic region and area around it was divided into two areas based on rainfall patterns: winter rainfall and summer rainfall. I chose to use rainfall patterns as a deciding factor because of their alleged importance in the diversification of *Pelargonium* species. I used maps by Chase and Meadows (2007) to distinguish the winter and summer rainfall regions. The Eastern cape area starts where the summer rainfall region ends, and goes on into Mozambique, along the coastline. Central South-Africa is called the Karoo region in the analysis, based on the phytochoria evaluation by Linder et al (2005).

While placing species in areas using google earth, mountain ridges were used as borders. Species' distribution patterns were also taken into consideration. If one single observation differed drastically from the others it was discarded. The book "*Pelargoniums* of Southern Africa" was used to verify some of the data (Van der Walt *et al.*, 1977).

The R-package BioGeoBEARS was downloaded as per the instruction in the tutorial (Matzke, 2013a; Matzke, 2013b; Matzke, 2014). A detailed example script was run following some small adjustments.

Final analyses

After this initial analysis, the area division was tweaked based on our new understandings. South-America was excluded because we saw in our initial run that its inclusion did not help our understanding of *Pelargonium* history. Excluding it allowed us to subdivide the East-African countries into three different areas. This could give a more detailed picture of *Pelargonium* migrations across eastern Africa.

Three versions of the *Pelargonium* phylogenetic tree were used. These trees were: the full tree (258 taxa), a tree consisting of just clade B and C (88 taxa), and the unaltered tree (106 taxa). The final area division can be seen in table 1.

Region	Code
Winter rainfall	A
Summer rainfall	B
Karoo Region	C
Eastern Cape	D
Mozambique	E
Tanzania + Kenya + N Malawi	F
Ethiopia	G
Madagascar	H
Socotra	I
Asia Minor (Middle-East)	J
St. Helena	K
Australia + New Zealand	L

Table 1: Areas used in finalized BioGeoBEARS analysis.

BioGeoBEARS was also used to run a stochastic mapping analysis (Dupin *et al.*, 2017), which can be used to find out more about the events underlying the species' distribution. With stochastic mapping, the results of 50 repeated reconstructions of the *Pelargonium* historical biogeography under the BAYAREALIKE+j model are summarized. This gives an overview of how often events like sympatry and jump dispersal have taken place. It also provides information on dispersal rates between different areas. Again, an example script was provided by Matzke's BioGeoBEARS tutorial.

Evs

For the evs analysis, the 258-taxa tree will be used almost exclusively, because it is identical to the BioGeoBEARS tree, and therefore the results can be compared. This means that *Hypseocharis bilobata* is excluded by default. However, since evs could potentially provide more information on the ancestral ranges of the *Hypseocharis* + *Pelargonium* ancestor, another version of the tree with *H. bilobata* included was run as well. This version, the 259 taxa three, will only be used to study the history of the *Hypseocharis* and *Pelargonium* ancestor. Beside the unaltered tree with 106 taxa, an alternative version including *H. bilobata* was also tested.

For the evs analysis the program evs was installed. The tree-file and data file were stored as .tab files in a separate directory. The tree.tab file consisted of a newick tree-file adjusted with the evs program. The records.tab file consisted of three columns: species names, longitude, and latitude. Dr SJ Arias provided the necessary information to get the program to work (J. S. Arias, personal communication, 26-04-2019).

Evs allows the user to control a variety of parameters. These parameters are explained in the introduction and summarized below in table 2. A detailed description of each parameter can be found in Arias et al (2017).

Parameter	**Description**
C	Determines the size of the pixels
F	Determines the amount of filler pixels surrounding the occupied pixels.
Z	Determines the extra cost given to large ancestral areas.
R	The number of repetitions run by evs.

Table 2: 4 adjustable parameters in evs.

Some preliminary analyses were run to test the settings of these parameters. For the final analysis, the standard settings were used: c = 360, f = 2, z = 50, and r = 10000.

In addition to these parameters, evs allows the user to adjust the cost of four events: Sympatry, point sympatry, vicariance, and founder events. By allowing or disallowing certain events, previously described models such as DIVA, DEC and BAYAREA can be emulated. As such, several variants of these models and others were tested. To "turn an event off" the cost of the event was increased to 1000.

As with BioGeoBEARS, these models can be adjusted to include jump dispersal. The +j versions of the models are identical to the normal models, except that founder events are allowed, meaning the cost was lowered.

Model selection

Since statistical model comparison with evs is not possible yet, the recommended way of finding a preferred model is using biological and ecological information to determine what the cost of each event should be (J. S. Arias, personal communication, 06-06-2019). As a starting point of selecting a suitable evs model, the best fitting model from BioGeoBEARS was taken. This model was BAYAREALIKE+j. This model allows sympatry and founder events, but it does not allow vicariance and point sympatry. As vicariance does not appear to have had a big effect on the *Pelargonium* distribution, it can reasonably be excluded from our evs model, just as it is excluded from our BioGeoBEARS model. There is no reason, however, to exclude point sympatry from the evs analysis. Therefore, it was added to the evs model which will be called NOVIC (no vicariance). NOVIC is assumed to be the best fitting model based on biological knowledge of *Pelargonium*. The models Bayareaj, GEM, BAYAREA, and VIP were ran in comparison to NOVIC to confirm this. Table 3 shows the exact parameters making up each model.

	Vicariance	**Sympatry**	**Point Sympatry**	**Founder events**
GEM	-	-	-	-
VIP+j	-	-	1000	-
BAYAREA	1000	-	1000	1000
BAYAREA+j	1000	-	1000	-
NOVIC	1000	-	-	-

Table 3: Parameters of 5 different evs models. A dash means no cost has been assigned (making all allowed events equally expensive). All disallowed costs have been given a cost of 1000, as was recommended to me by Dr Arias.

Pollinators

In order to find correlations between environmental factors and *Pelargonium* morphology, pollinator and climate data had to be collected. Known pollinators were linked to *Pelargonium* species using literature (Struck, 1997; Goldblatt & Manning, 2000; Combs & Pauw, 2009; Newman *et al.*, 2014). Only the pollinators identified to species level could be used, the rest were removed. For every species in the dataset, a total of 27, coordinates were extracted from GBIF and iNaturalist (for citations see supplementary material – file G). The data was checked by plotting it in google earth and removing suspicious outliers. One butterfly species, *Cacyreus marshalli*, was introduced to Europe where it became a pest. The European/non-South-African coordinates of this species were removed.

Pollinators were then subdivided in the same areas used by BioGeoBEARS, and linked to their host-species. Using the R function aggregate (R-Core-Team, 2019) and R-package raster (Hijmans, 2019) the mean elevation of each species habitat was extracted from WorldClim.

Climate Data

If we want to know more about *Pelargonium* ancestral ranges, it may benefit us to know what the climate in these ranges was like. In order to reconstruct ancestral climates, elevation data, precipitation data, and temperature data was extracted for the habitat of each *Pelargonium* species. This data was extracted from WorldClim using species coordinates and R-package raster (Hijmans, 2019). When no data could be extracted from the given coordinates, surrounding species and additional literature were used (climate.data.org). The R function aggregate was used to calculate mean yearly and seasonal temperature, mean elevation, and total yearly and seasonal precipitation.

Then, ancestral elevation, precipitation and temperature were reconstructed using Mesquite (Maddison & Maddison, 2018), onto the full *Pelargonium* tree (258 terminals). The analysis was carried out with Mesquite's the Trace Character History function, using maximum parsimony.

It is important to note that several species' coordinates were estimated by me using literature describing *Pelargonium* species' distributions. The climate data extracted for these species is not informative and cannot be used to explain our results.

Results

Choosing the best models

BioGeoBEARS

Model comparison (comparing six models) with BioGeoBEARS found that BAYAREALIKE+j is the best fitting model for our data. BAYAREALIKE+j allows for founder events and sympatry, while disallowing vicariance and point sympatry. The BAYAREALIKE+j results for each tree can be found in table 4. In table A in the supplementary material the statistics for the winning and losing models can be found. Stochastic mapping analysis of the full reconstruction shows an average of 20.6 founder events**,** and an average of 236.4 sympatry events. Mean events and standard deviations of each tree can be found in table 4. The BioGeoBEARS BAYAREALIKE+j output of the full tree can be seen in figure 4, while the other models can be seen in the supplementary material (map B).

Tree	**Model**	**AICc**	**LnL Pval**	**Found Mean**	**Found Stdv**	**Vic**	**Symp Mean**	**Symp Stdv**	**Point**
Full Correct	BAYAREALIKE+j	1454	1.20E-10	20.6	2.41	0	236.4	2.41	0
BC	BAYAREALIKE+j	608.5	5.60E-11	18.52	1.97	0	68.48	1.97	0
Unaltered	BAYAREALIKE+j	648.3	5.30E-06	7.46	1.73	0	97.54	1.73	0
Morph Tree	BAYAREALIKE+j	393.7	1.00E-04	4.08	0.99	0	62.92	0.99	0

Table 4: Statistics and events for each tree.

The ratios of sympatry and founder events do not differ much between the full and unaltered tree. When comparing the two trees node by node, we can see that the only important difference is in the clade of *P. caylae, P. endlicherianum, and P. quercetorum,* where the addition of an extra species shifts the ancestral area towards Asia Minor*.* Beyond this, some ancestral areas are wider in the full tree, which can be expected because of the additional terminals with different locations. The unaltered tree reconstruction can be seen in the supplementary material (file C). From this point, the full tree will be used to discuss our results.

As expected, the ratio of founder events vs sympatry events was higher in the tree where clade A was excluded. Beyond this, there are no notable differences between the full tree and the clade BC-tree (supplementary material, file D). We can therefore conclude that the sheer amount of clade A2 species does not influence the way ancestral areas are assigned to the other clades.

Table E in the supplementary material shows dispersal rates from and to the 12 different areas, as found by the BioGeoBEARS stochastic mapping analysis. Overall, this information shows that dispersal rates from the south to the north are higher than dispersal rates from the north to the south. Some relatively high dispersal rates are found from South-Africa to Ethiopia, while dispersal rates in the other direction are close to zero. Dispersal rates are the highest among the four South-African regions, where most *Pelargonium* species occur.

The stochastic mapping results show that there was no subset sympatry or vicariance (supplementary material, table A). Neither of these are allowed under the BAYAREALIKE+j model. This means that of the six models, the best fitting one does not rely on these two events.

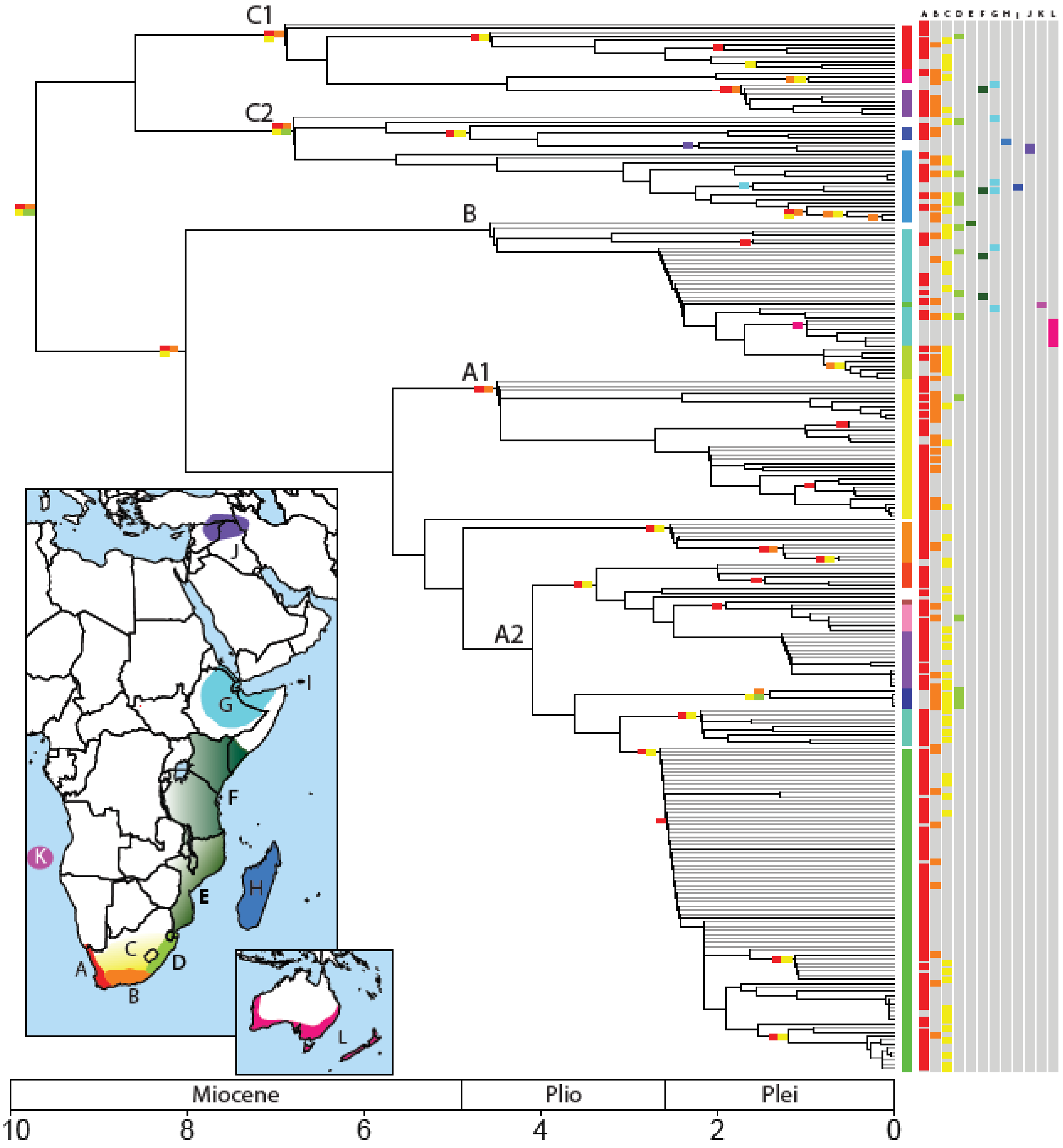

Figure 4: BioGeoBEARS analysis of Pelargonium under BAYAREALIKE+j model.

Evs

Using the NOVIC model, evs estimates that 55 founder events, 170 sympatry events, and 32 point sympatry events happened in the evolution of pelargonium. This reconstruction cost 961.480. The Bayareaj reconstruction is slightly more expensive (table 5a), but preliminary runs revealed that allowing fewer events tends to make the reconstructions more expensive. Therefore, I cannot use this cost alone to decide which model is better. The only difference between the Bayareaj model and the NOVIC model is that NOVIC allows point sympatry, and in the NOVIC reconstruction 32 point sympatry events were assigned, 12% of total events, showing that point sympatry is not rare in *Pelargonium* (supplementary material – table H).

	Bayareaj	**Bayarea**	**VIP**	**GEM**	**NOVIC**
Full Tree	1059.780	1307.480	1018.500	933.500	961.480
Clade BC	626.320	X	X	539.380	552.120
Unaltered Tree	619.600	X	X	539.260	556.620

Table 5a --> Total cost of the best reconstruction from each model.

	Founder events	**Sympatry**	**Point sympatry**	**Vicariance**
Bayareaj	65	192	0	0
Bayarea	0	257	0	0
NOVIC	55	170	32	0
GEM	50	168	34	5
VIP+j	58	192	0	7

Table 5b --> Events for all of the models, for the full tree.

Bayareaj in evs assigns more founder events and fewer sympatry events than BAYAREALIKE+j in BioGeoBEARS does.

VIP+j and NOVIC each allow 3 events, meaning that the cost difference could be more informative. The best VIP+j reconstruction is more expensive the NOVIC reconstruction, supporting our theory that point sympatry was more important in *Pelargonium* evolution than vicariance (table 5a).

The model GEM (not to be confused with the name of the program) allows all four events and gave the reconstruction with the lowest cost. This reconstruction assigned amounts of founder events, sympatry and point sympatry almost identical to the NOVIC model, while only assigning 3 vicariance events (table 5b). All of these vicariance events happened between species that occur closely together, and none of them in nodes we were actively investigating (supplementary material – map I). Again, this supports our choice of the NOVIC model.

Finally, BAYAREA was run to see the importance of founder events. In every model allowing founder events they make up a substantial part of the total events assigned, while in BAYAREA every event is sympatry (table 5b). Furthermore, the cost of the BAYAREA reconstruction is much higher (table 5a).

Based on these comparisons, and the biological knowledge we have of *Pelargonium*, the NOVIC reconstruction will be used to discuss our results. The NOVIC model was rerun 4 more times with the exact same parameters and data. The reconstruction with the lowest cost is the one used in this study .

When comparing the unaltered tree node by node, we can see that while there are differences, they do not appear in any of the nodes we will study in detail. The ratio of assigned events of each tree does not differ much (supplementary material – table H). We will therefore assume our full tree to be correct.

An output of the NOVIC reconstruction with the lowest cost, focussing on clade B and C, can be seen in figures 8 & 9. Figures 5-7 and 10-12 show maps of estimated ancestral area's for several nodes. Ancestral areas for every other node can be found in the supplementary material (map K). A table of assigned events from the best reconstruction of each model for each tree can be seen in table 5b.

Ancestral area's

Hypseocharis

Unfortunately BioGeoBEARS could not precisely estimate the ancestral area of the ancestor of *Pelargonium* and *Hypseocharis*. The output of the initial run in which we included *Hypseocharis bilobata* can be seen in the supplementary material (file J). Since this estimation could not get more precise than South-America + South-Africa, we decided not to include *H. bilobata* in further analysis. Excluding *H. bilobata* meant eliminating South-America as an area, which allowed us to make the division of the East-African areas more complex.

The 259 taxa-analysis was run in evs with the NOVIC and the GEM models. The ancestral area reconstructions from both of these models can be seen in figure 5. NOVIC does not allow vicariance, and as a result, the *Hypseocharis* ancestor is shown to have arrived in South-America through jump dispersal, while the *Pelargonium* ancestor stayed in South-Africa. Under the GEM model, which allows all four events, vicariance is shown to have caused the divergence between *Hypseocharis* and *Pelargonium.* The full evs reconstructions of the 259 tree for both models can be seen in the supplementary material (map M).

Figure 5a: Full tree with 259 terminals under the GEM model – Hypseocharis (blue) and Pelargonium (red) diverged through vicariance.

Figure 5b: Full tree with 259 terminals under the NOVIC model – Jump dispersal brought Hypseocharis' ancestor (white squares) to South-America

Pelargonium origin

Both the BioGeoBEARS analysis and the evs analysis show that *Pelargonium* originates in South-Africa. The exact BioGeoBEARS ancestral *Pelargonium* area is the winter rainfall region + the summer rainfall region + the Karoo region + the eastern cape (figure 4). Together these area's make up most of South-Africa. The *Pelargonium* ancestral range as estimated by evs can be seen in figure 6.

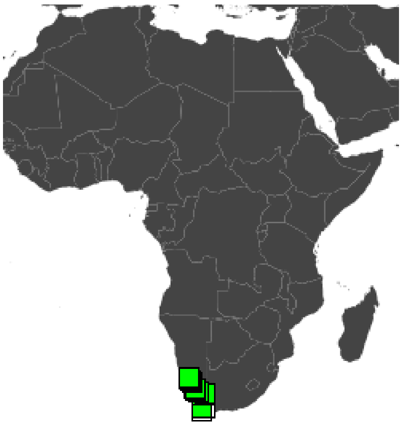

Figure 6: Full tree under the NOVIC model – distribution of the Pelargonium ancestor as estimated by evs.

Australia and St. Helena

In both the BioGeoBEARS and evs analyses, a single dispersal event to Australia caused the *Peristera* distribution (figures 7 & 8). Another jump-dispersal event brought *P. cotyledonis* to St. Helena (figures 7 & 8). Long distance dispersal is assigned in evs under both the NOVIC model, and the GEM model which allows vicariance (supplementary material – map I).

Figure 7: Jump dispersal from South-Africa to Australia and St. Helena of Peristera and P. cotyledonis respectively. Note that these are two separate dispersal events, shown in a single picture for convenience. Ancestral ranges shown as white squares. See pink triangles in figure 8.

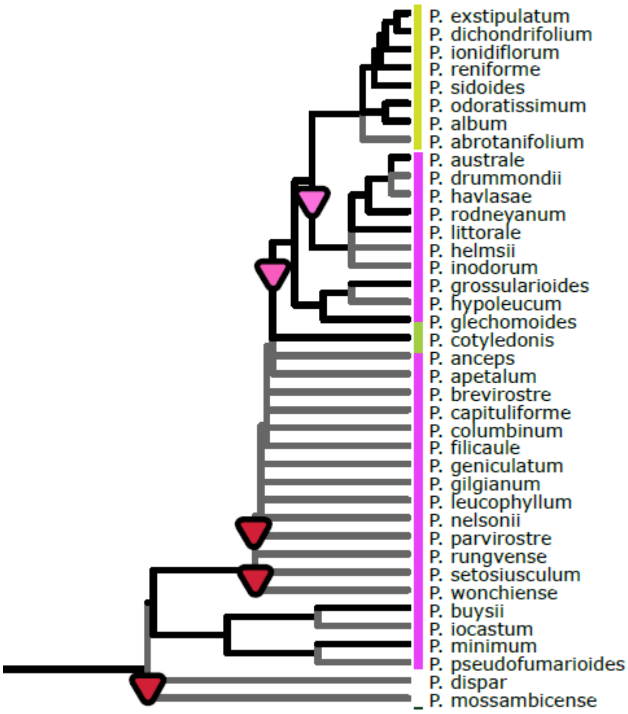

Figure 8: Clade B run by evs under NOVIC model. Pink: Peristera species. Pink Triangles: jump dispersal of Australian clade to Australia and of P. cotyledonis to St. Helena. Red triangles: additional jump dispersal towards the east or north across eastern Africa. Manually added taxa are coloured grey. Full tree can be seen in supplementary material (PDF N).

Clade C

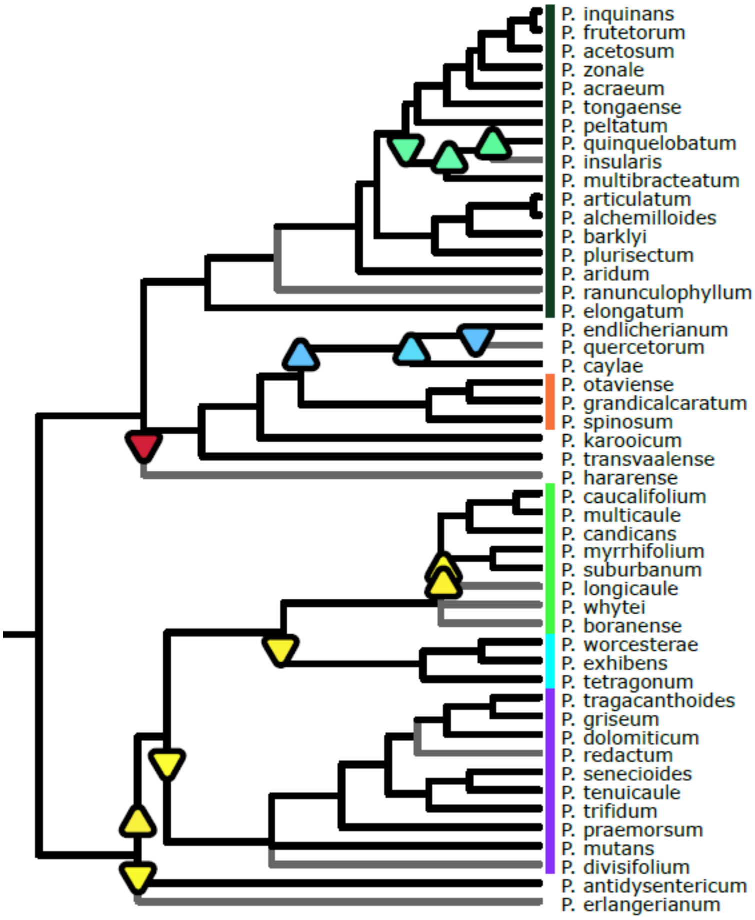

Figure 9: Evs output of NOVIC model: Clade C. Green: section Myrrhidium, blue: section Chorisma, purple: Jenkinsonia. Green triangles: dispersal of P. multibracteatum, P. insularis, and P. quinquelobatum. Blue triangles: dispersal of P. caylae, P. quercetorum, and P. endlicherianum. Yellow triangles: dispersal in clade C1. Red triangle: additional jump dispersal towards the east or north across eastern Africa. Manually added taxa are coloured grey. Full tree can be seen in supplementary material (PDF N).

According to BioGeoBEARS, the ancestor of *P. caylae, P. endlicherianum, and P. quercetorum* occurred in Asia Minor (Fig 4). Then, from there, *P. caylae* dispersed to Madagascar. Evs, however, shows their ancestor occurred on Madagascar, and from there the Middle-Eastern species dispersed to Asia Minor (Figure 10). It should be noted that in some of the slightly more expensive evs reconstructions under the same model (NOVIC), the history of this group was the same as it is in BioGeoBEARS. Under both the NOVIC and GEM models, jump dispersal is assigned.

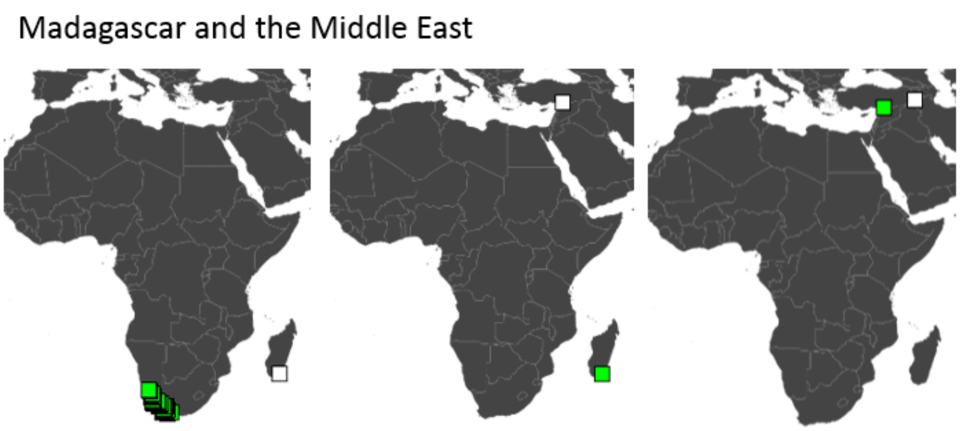
Madagascar and the Middle East

Figure 10a: Jump dispersal of Pelargonium lineage to Madagascar

Figure 10b: Jump dispersal of Pelargonium lineage to the Middle-East.

Figure 10c: Jump dispersal of Pelargonium quercetorum in the Middle-East.

Figure 10: Jump dispersal to Madagascar and the Middle east. See blue triangles in figure 9.

Under the NOVIC model, evs estimates the ancestral range of *P. insularis, p. multibracteatum,* and *P. quinquelobatum* to be in Ethiopia (Figure 11). From there, the ancestor of *P. insularis* dispersed to Socotra, and from there *P. quinquelobatum* dispersed back to Ethiopia and Tanzania. According to BioGeoBEARS the route is slightly different. The ancestor of *P. quinquelobatum* and *P. insularis* still occurs in Ethiopia and Tanzania. From there, *P. insularis* disperses to Socotra. In every analysis, jump dispersal was given as the event that caused the divergence. Even using the GEM model, vicariance was not assigned to the divergence of *P. insularis*


Ethiopia and Socotra

Figure 11a: Jump dispersal of pelargonium lineage to Ethiopia.

Figure 11b: Jump dispersal of Pelargonium lineage to Socotra.

Figure 11c: Jump dispersal of P. quinquelobatum.

Figure 11: Jump dispersal to Ethiopia and Socotra. See green triangles in figure 9.

Beyond these specific examples, there are multiple other jump-dispersal events from South-Africa to the north, across eastern Africa (figures 9 & 12).

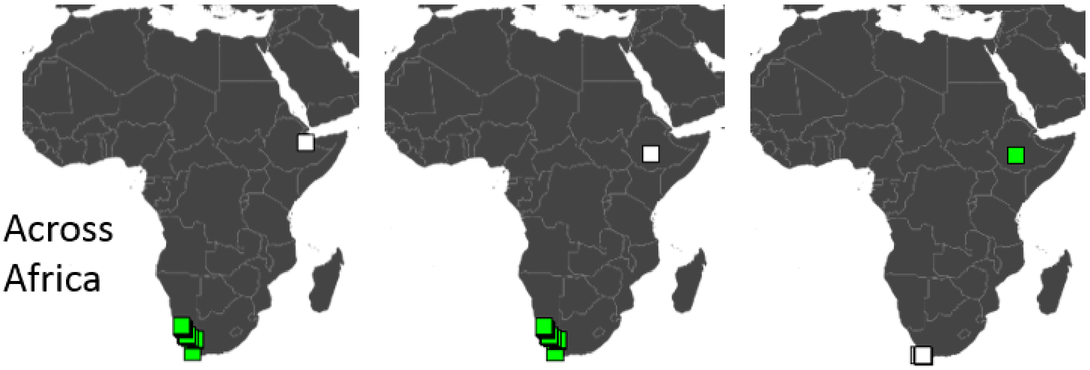

Figure 12a: Jump dispersal of Pelargonium erlangerianum.

Figure 12b: Jump dispersal of clade C1 minus P. erlangerianum and P. antidysentericum to Ethiopia.

Figure 12c: Jump dispersal of ancestor of section Jenkinsonia from Ethiopia to South Africa.

Figure 12d: Jump dispersal of ancestor of section Chorisma to south Africa

Figure 12e: Jump dispersal of section Myrrhidium minus P. boranense to Mid-Africa.

Figure 12f: Jump dispersal of section Myrrhidium minus P. whytei to South-Africa.

Figure 12: Dispersal across eastern Africa. See yellow triangles and color-coded clades in figure 9.

Climate Data

The predicted climate of the ancestral *Pelargonium* species had an average yearly temperature of 12.4 – 15.9 degrees Celsius. The estimated total annual rainfall is between 343 and 646 mm. The elevation is estimated between 605 and 899 meters (figure 13).

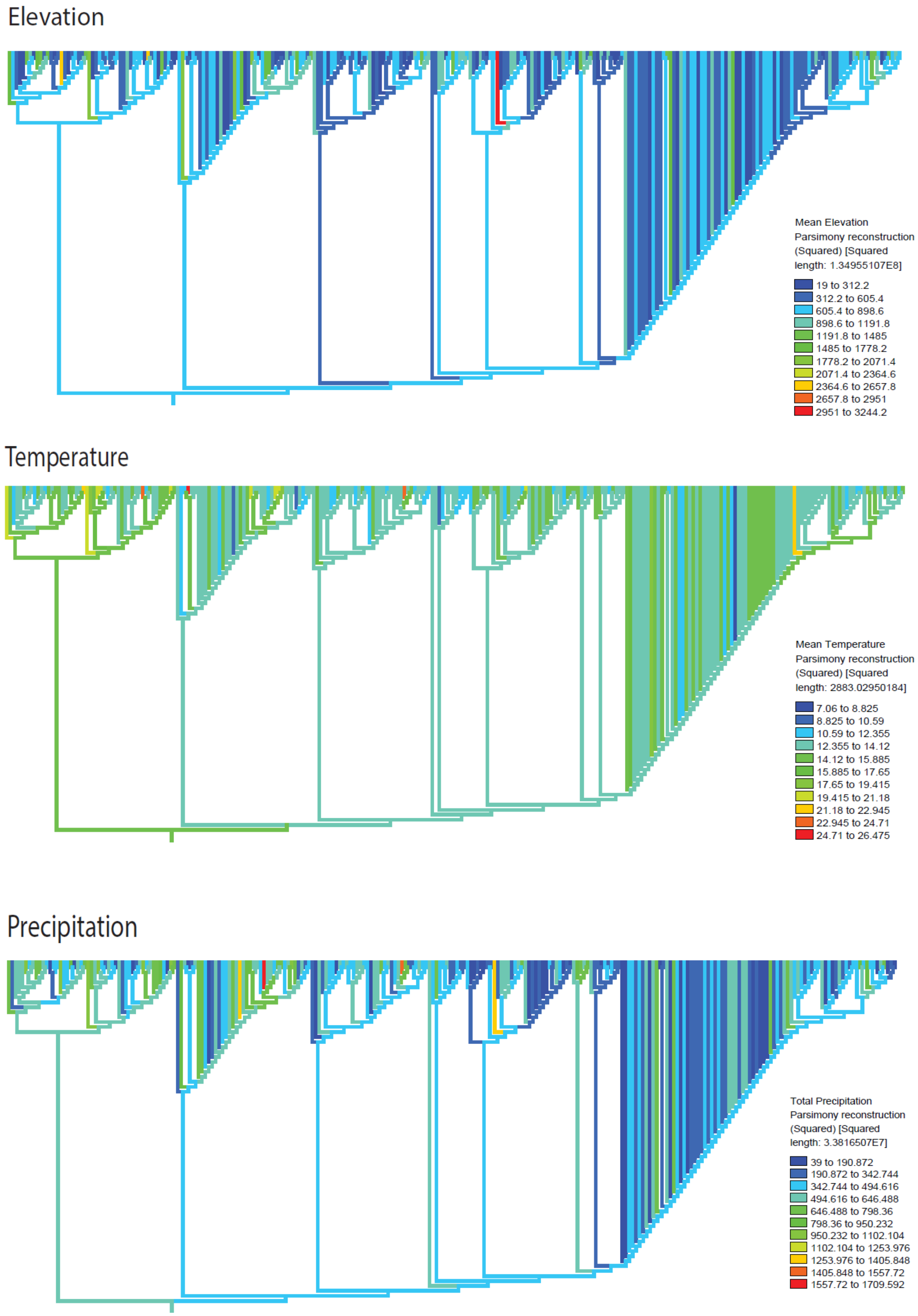

Figure 13: Ancestral climate reconstructions of Pelargonium.

Discussion

Hypseocharis ancestor

Pelargonium's ancestor first diverged from *Hypseocharis* 35.8 Mya. During this time Africa and South-America were no longer connected, not even via Antarctica.

Under the NOVIC model evs finds that jump dispersal brought *Hypseocharis* to South-America from South-Africa. This would suggest that the ancestor of *Pelargonium* and *Hypseocharis* already occurred in South-Africa. It is, however, also possible that the weight of 258 *Pelargonium* species versus one *Hypseocharis* species influenced this outcome and that the jump dispersal actually happened in the opposite direction: from South-America to Africa.

It would therefore be beneficial to do another historical biogeography study, this time focussing not only on one genus, but on the entire family of Geraniaceae and its sister lineages. One such study has been conducted by Fiz et al (2008), but with the new possibilities and methodology that have become available over the past ten years, new findings may be discovered. The usage of explicit species ranges, for example, seems promising, especially when used together with a method capable of model comparisons.

Under the GEM model, evs found that vicariance caused the divergence of *Pelargonium* and *Hypseocharis*. We know, however, that vicariance due to the continental drift of South-America and Africa is unlikely to be behind the *Pelargonium* and *Hypseocharis* distributions, unless the divergence times estimated in multiple studies are incorrect. Furthermore, the fact that BioGeoBEARS found a model without vicariance to be the most fitting also supports the exclusion of vicariance. Therefore, this presumably incorrect GEM result may demonstrate why (partially) basing the model choice on biological knowledge of a genus or family is a good idea.

Recent studies have suggested that trans–Atlantic jump dispersal between Africa and South-America may be more common than previously thought (Calviño *et al.*, 2016). One possible explanation for this is the presence of islands in the Atlantic Ocean, between the two continents (Ezcurra & Agnolín, 2011). Several Asteraceae lineages are thought to have dispersed across the Atlantic using an island chain (Katinas *et al.*, 2013).

These islands existed some millions of years into the Eocene (55.8 Mya) (Ezcurra & Agnolín, 2011), which means they might have been gone by the time *Pelargonium* and *Hypseocharis* diverged. *Pelargonium* is not the only group in which trans-Atlantic LDD events are younger than these island chains. In the Sapotaceae, long distance dispersal from South-America to Africa is inferred to have happened between 7 and 14 Mya (Bartish *et al.*, 2011).

Long distance dispersal between Africa and South-America was also suggested by Fiz et al (2008), Palazzesi et al (2012), and Sytsma et al (2014) as an explanation for the Geraniales distribution. The direction of these dispersals was unclear (Sytsma *et al.*, 2014). Renner (2004) found that water-mitigated dispersal across the Atlantic happens in both directions and is the most common, while wind-mitigated dispersal happened mainly from South-America towards Africa. *Pelargonium* seeds appear more suited for wind-dispersal, but we do know that non-standard dispersal events are often behind extreme long distance dispersals (Higgins *et al.*, 2003; Nathan, 2006). Abiotic factors are also thought to have caused the long distance dispersal of Spermacoceae from South-America to Madagascar (Janssens *et al.*, 2016). A dispersal route from Africa to South-America via Europe and North-America (Erkens *et al.*, 2009) seems unlikely, given the lack of pelargonium species in these continents.

There is, however, another possible explanation for the distribution of *Pelargonium* and *Hypseocharis*. Antarctica was still warm and non-glacial around the time their ancestor existed (Zachos *et al.*, 2001; Galeotti *et al.*, 2016). It is not unthinkable that the *Pelargonium* + *Hypseocharis* ancestor occurred on Antarctica back when it was still suitable for temperate vegetation. The ancestor may also have occurred on both Antarctica and South-America, as the two continents were connected before around 40 Mya (Livermore *et al.*, 2005). The *Hypseocharis* ancestor could have ended up in South-America through jump dispersal or vicariance, while the *Pelargonium* ancestor could have dispersed to South-Africa, where we know the species originates. There are multiple plant families with extant species in the Andes, of which fossils have been found on Antarctica (Taylor, 1991). Involvement of the connection between South-America and Antarctica in plant distributions has been suggested before (Gamerro & Barreda, 2008), as has the possibility of long-distance dispersal from South-America to Africa across Antarctica or its surrounding islands (Katinas *et al.*, 2013).

The biggest problem with this idea is that we cannot test it. In BioGeoBEARS, every included area needs to have at least one species in it. In evs, only the specific coordinates belonging to the extant species can be used in the analysis. Since there are no extant *Pelargonium* species on Antarctica, the whole continent is automatically excluded from the analysis. Considering that *Pelargonium* is not the only species with possible ties to Antarctica, this is a rather large shortcoming of current historical biogeography methods.

Recently, fossilised pollen grains belonging to the Asteraceae were found in ice sheets on Antarctica (Barreda *et al.*, 2015). The discovery of these fossils greatly pushed back the estimated Asteraceae origin, and it is believed that the Antarctica ice sheets still hold valuable information on the evolution of this family. While Geraniaceae fossils have not been found on Antarctica, this does not mean they do not exist, and the discovery of new evidence could greatly change our current perception of this group. Perhaps evidence supporting the idea that the ancestor of *Hypseocharis* and *Pelargonium* occurred on Antarctica is out there somewhere.

Pelargonium origin

The ancestral range of *Pelargonium* as found by evs is mostly limited to the winter rainfall region, and is therefore more narrow than the one found by BioGeoBEARS. This difference could be due the higher level of precision that is one of the benefits of using explicit geographical ranges. Evs is, however, still limited to coordinates that could be extracted from public databases. Parts of the habitat that *Pelargonium* occupies are quite hard to reach, and there is no doubt that the true species ranges differ at least a little from the ones shown by our coordinates. As a result, I expect that the ancestral ranges found by evs are likely smaller than the true ranges. BioGeoBEARS, however, is more likely to estimate wide ancestral ranges, since the program cannot perceive when a species only occupies part of a predetermined area.

Still, under both the BAYAREALIKE+j model in BioGeoBEARS, and the NOVIC model in evs, *Pelargonium* originates in South-Africa. These results strongly support the Cape-to-Cairo hypothesis (Galley *et al.*, 2007), and oppose the hypothesis that *Pelargonium* originates in the north and dispersed down to the Cape later.

There is still much of time between the *Pelargonium* crown 9.7 Mya and the divergence from *Hypseocharis* 35.8 Mya that is unaccounted for. We know *Pelargonium* likely dispersed to South-Africa from Antarctica or South-America, but it took many years before the genus started diversifying. Meanwhile, tropical vegetation is believed to have covered the South-African Cape regions during the Miocene, which is when these ancestral *Pelargonium* species were supposedly already there. Linder (2003) suggests that while the current CFR flora may have arrived after the cape region became drier and less tropical, it is also possible that some of the lineages were already present in South-Africa. These lineages may have occurred on mountains, were the environments would have been better suited to them.

According to the ancestral elevation reconstruction, the ancestral *Pelargonium* species occurred on habitats around 600-900 meters above sea level. These heights seem to match up with the South-African uplifts causing parts of the cape region to rise higher than 700M above sea level (Linder, 2003). Before this uplift, big parts of South-Africa were already over 500M above sea level (Linder, 2003). These ancestral heights support the idea that ancestral *Pelargonium* species took refuge in the mountains while the lower cape regions were tropical, and descended when the aridification started around 10 Mya (Zachos *et al.*, 2001; Linder, 2003).

Disjunct distributions

Australia & St Helena

As expected, long-distance dispersal events brought *Pelargonium* to Australia and St. Helena. While the estimated Australian species ranges in evs are more specific, the overall results are the same. Even the GEM model, which allows vicariance, assigned long-distance dispersal to these nodes. This, and the fact that these species diverged in the past 10 My means we can safely exclude vicariance as the reason behind *Pelargonium's* Australian distribution. How these long-distance dispersal events took place is unclear, though trans-oceanic plant dispersal is far from rare. Endozoochory appears unlikely given *Pelargonium* seed morphology, but water, wind and epizoochory (Katinas *et al.*, 2013) are all possibilities.

One way to narrow down the list of possibilities would be to test the capacity for survival of germinated and non-germinated *Pelargonium* seeds in sea water, in a study similar to Wu et al (2018). The distance from the South-African coast to Australia or St Helena is long, and *Pelargonium's* ability to survive in salt water would have to be extraordinarily high in order to disperse this way. Urticaceae seeds had the best survival rates when they landed in salt water without spending a long time in sweet water first. Thus, plants growing in coastal locations have a better change of successful oceanic dispersal than plants growing near rivers (Wu *et al.*, 2018). Investigating if this is the case with *Pelargonium*, which contains many species occurring relatively close to the coast, could be informative. Perhaps more costal *Pelargoniums* are more likely to disperse over long distances.

Several Australian *Pelargonium* species are self-compatible to a certain extent (Nicotra *et al.*, 2016). This self-compatibility could have made the colonisation of Australia easier, since fewer species would be needed for a successful settlement.

Caylae + endlicherianum + quercetorum

The ancestral areas assigned to the ancestors of *P. caylae, P. quercetorum,* and *P. endlicherianum* by evs were not consistent across different runs of the same model. We chose to discuss the reconstruction with the lowest cost, but the cost-differences are small. Apparently these nodes are somewhat uncertain, and a definitive conclusion on the order of colonisation of Madagascar and Asia Minor cannot be given based on these results. It does appear that jump dispersal caused this *Pelargonium* distribution. This fits with our hypothesis that this group is too young to have been influenced by vicariance.

There are multiple explanations for this jump dispersal across Africa and Arabia. Evidence of dispersal between eastern Africa and Madagascar has been collected in a wide array of species, including plants (Bartish *et al.*, 2011). Dispersal from South-Africa to the Middle-East could have happened in a similar pattern to the dispersal of multiple *Pelargonium* lineages towards Ethiopia, thus adding to the support for mitigated trans-African dispersal . It is also possible that this dispersal happened between Madagascar and the Middle-East directly, possibly through islands in the Indian ocean or the aforementioned long-distance dispersal mechanisms such as the wind and epizoochory.

If the disjunct *Pelargonium* distributions are going to be studied further in the future, it may be interesting to take the species' ecologies into account. *P. caylae*, for instance, has a liana-like appearance that is unusual compared to the rest of the genus. It is also notable for being the only known *Pelargonium* species on Madagascar, and differs phytochemically from the rest of the clade C2 species (Röschenbleck *et al.*, 2014). *P. caylae's* distinct ecology may be connected to its disjunct distribution, and studying both could lead to new insights about the history of *Pelargonium*.

Insularis + multibracteatum + quinquelobatum

It seems that the dispersal route of *P. insularis, p. multibracteatum,* and *P. quinquelobatum* cannot yet be solved completely, and though we show the reconstruction with the lowest cost here, we cannot say for certain that this route was really taken by *Pelargonium*. In each case, however, the last shared ancestor of the three species could be found in Ethiopia. We can also say that jump dispersal and not vicariance caused this distribution. This dispersal likely happened across East-Africa, in a similar pattern to the other Ethiopian and eastern-African *Pelargonium* lineages. *P. insularis* arrived on Socotra though jump dispersal, and not vicariance.

Cape-to-Cairo Dispersal

In the original version of the Kerke et al (2019) phylogeny, the northern-African *Pelargonium* species do not form a monophyletic group. In fact, northern-African *Pelargonium* species can be found across clades B, C1, and C2. Since we have strong evidence suggesting a South-African origin for *Pelargonium*, we see a pattern of multiple dispersal events from southern Africa towards northern Africa, with several species in eastern Africa. While the exact amount of separate dispersal events cannot be conclusively stated until we have a complete phylogeny based on genetic information, in our analysis the amount appears to lay between three and eight dispersal events, excluding the one to Madagascar and the Middle East. These dispersal events seem to have gone across eastern Africa, therefore adding to the growing body of evidence that there was a dispersal corridor of some sort stretching from South-Africa to Ethiopia.

The summarized dispersal rates between different areas as inferred by BioGeoBEARS stochastic mapping analysis strengthen the argument for a dispersal corridor, and show that the *Pelargonium* dispersal mainly happened from the south to the north, and rarely from the north to the south. The most dispersal happened between the four South-African regions. Apparently ancestral *Pelargonium* species were not specialized to their specific habitat to a degree where range-switching or -expanding to different rainfall-regions was no longer possible.

Ancestral climate reconstructions are useful for inferring how much the *Pelargonium* habitats in northern and southern Africa used to differ. Unfortunately, all the northern-African, Socotran, Madagascar, and Middle-Eastern coordinates are approximations based on primary literature, and can therefore not be used to infer the conditions these species faced after their dispersal events. Therefore, our climate reconstruction cannot be used to compare southern- and northern-African habitats. Coordinates were available for the other non-African *Pelargonium* species. Neither the Australian or St. Helena *Pelargonium* habitats differ much from the ancestral habitats in clade B, and both are variable.

BioGeoBEARS vs evs

While the best fitting model used in BioGeoBEARS does not allow for vicariance, several of our evs analyses did. In each of them, hardly any vicariance was assigned to the *Pelargonium* phylogeny. While we reasoned beforehand that vicariance due to the breaking up of continents is unlikely to have contributed to the *Pelargonium* distribution, it seemed possible that habitat fractioning due to climate changes could have caused speciation through vicariance. It is unsurprising that BioGeoBEARS did not assign a vicariance based model, given that the designated areas are too large for it to perceive vicariance between species that continue to exist closely together. Habitat fractioning can create a detailed patchwork that would get lost in any biogeography method allowing a limited number of areas.

It is possible that the few cases of vicariance assigned under the GEM model in evs are caused by this habitat fractioning, but no real patterns are distinguishable, and the few vicariance events are spread out over multiple clades. Perhaps if the original phylogeny was more complete, we would see clearer vicariance patterns, possibly in the rapidly radiated species in clade A2. Currently, the phylogeny of clade A2, which underwent multiple radiations, is heavily influenced by the manual phylogenetic reconstruction based solely on taxonomy and previous phylogenetic studies. Thus, this study cannot provide evidence for small-scale vicariance caused by habitat fractioning.

No subset sympatry in BioGeoBEARS is expected since the designated areas are too large to find most cases of subset sympatry. Many cases of subset sympatry will instead be seen as narrow sympatry since the ancestor and descendant seem to inhabit the same range even if they are miles apart. Some subset sympatry was visible in evs, showing that using the NOVIC model instead of the Bayareaj model was the correct choice. This does raise the question of what would change if a NOVIC model was used in BioGeoBEARS. Perhaps in the future, the evs results can be used to design some more BioGeoBEARS models.

Both BioGeoBEARS and evs have their own drawbacks. In BioGeoBEARS, the user has to decide upon different areas, areas which have to be quite large if the studied group occupies a wide range. I found out for myself that this process can be uncertain at best, and completely arbitrary at worst. I will say that the area division would be more straightforward in island-clades, which are particularly interesting for biogeographical research. Besides the very informative stochastic mapping function, one of BioGeoBEARS most useful features is the model comparison function. Not only does this compare the fit of previously described biogeography models, but it also tests the role of jump-dispersal.

The biggest drawback of evs that no such model comparisons are possible right now. While the program is versatile and allows the user to emulate many different models, the final model choice has to be based on previous biological insights into the species. The usage of explicit species ranges, however, is a promising addition to historical biogeography research, and there is no doubt that it can be expanded on in the future.

One drawback both programs share is that they view the world as static. Many plant and animal lineages are old enough to have been influenced by continental drift. While vicariance can be inferred by methods of historical biogeography, the actual continental drifts are not perceived. In fact, areas in BioGeoBEARS are just letters, and no information on geographical locations can be given. Evs plots coordinates on a map, without taking continental drift into account. Another shared drawback is that the user cannot include unoccupied areas into the analysis. While neither of these shortcomings have a clear solution right now, effort are always being made to improve upon our current biogeography methods.

Conclusions

Hypseocharis and *Pelargonium* possibly diverged through trans-Atlantic jump dispersal, but the hypothesis that the ancestor of *Hypseocharis* and *Pelargonium* occurred on Antarctica should be investigated. *Pelargonium* has a clear South-African origin, mostly concentrated in the winter-rainfall region. This, and several separate migrations towards the north, suggest a Cape to Cairo scenario. The pattern of jump dispersals across eastern Africa towards Ethiopia matches patterns found in other plants, and suggests that this dispersal is mitigated by abiotic factors. A single jump dispersal event brought an ancestral *Peristera* species from South-Africa to Australia, and another jump dispersal brought *P. cotyledonis* to St. Helena. A single jump dispersal from either Africa or the Middle-East brought *Pelargonium* to Madagascar. Another jump dispersal brought *Pelargonium* to Socotra from Ethiopia.

Both BioGeoBEARS and evs have clear advantages. BioGeoBEARS has the ability to compare different models, and thereby informs the user on the speciation events that were most important in shaping the species distribution seen today. Its biggest problem is that the user needs to divide species into manually constructed areas, which will influence the outcome of the analysis. Evs uses explicit species ranges, removing this problem. Explicit ranges are a promising feature that will hopefully be explored more in the future. Unfortunately, model comparison is not yet possible in evs, and as a result the user needs to find the best model based on biological knowledge of the group that is being studied. In an ideal world, we would have a program that is capable of using explicit ranges, and model comparisons. Until then, using multiple methods and allowing them to complement each other may be the most reliable way of studying historical biogeography.

Acknowledgements

I would like to thank Sara van de Kerke and Dr Freek Bakker for their guidance and for teaching me many new things. I want to thank Cynthia Jones and Timothy Moore for the coordinate data they collected and allowed me to use. I would like to thank Dr J. S. Arias for his help with running the program evs, and his elaborate explanations about the different parameters. Next, I want to thank Dr. Nicholas Matzke for writing extensive tutorials on the usage of BioGeoBEARS. I want to thank Sander Rodenburg for helping me set up a computer on the third floor after the analysis-computer I was using broke down, and for allowing me to use it for many weeks. I would also like to thank Eric Schrantz for checking in on me and providing new ideas, and Floris Breman for giving me tips and advice. Finally, I am grateful for everyone at Biosystematics for providing a nice environment to work in.

Literature

For the sources of pictures used: see supplementary material (file O).

Ali, S.S., Pfosser, M., Wetschnig, W., Martínez-Azorín, M., Crespo, M.B. & Yu, Y. (2013) Out of Africa: Miocene Dispersal, Vicariance, and Extinction within Hyacinthaceae Subfamily Urgineoideae. *Journal of Integrative Plant Biology*, **55**, 950-964.

Andrés-Sánchez, S., Verboom, G.A., Galbany-Casals, M. & Bergh, N.G. (2018) Evolutionary history of the arid climate-adapted Helichrysum (Asteraceae: Gnaphalieae): Cape origin and association between annual life-history and low chromosome numbers. *Journal of Systematics and Evolution*, **0**

Arias, J.S. (2017) An event model for phylogenetic biogeography using explicitly geographical ranges. *Journal of Biogeography*, **44**, 2225-2235.

Axelrod, D.I. & Raven, P.H. (1978) Late Cretaceous and Tertiary vegetation history of Africa. *Biogeography and Ecology of Southern Africa*, pp. 77-130. Springer Netherlands, Dordrecht.

Bacon, C.D., Simmons, M.P., Archer, R.H., Zhao, L.-C. & Andriantiana, J. (2016) Biogeography of the Malagasy Celastraceae: Multiple independent origins followed by widespread dispersal of genera from Madagascar. *Molecular Phylogenetics and Evolution*, **94**, 365-382.

Bakker, F., Culham, A., Marais & Gibby, M. (2005) *Nested radiation in Cape Pelargonium*.

Bakker, F.T., Hellbrügge, D., Culham, A. & Gibby, M. (1998) Phylogenetic relationships withinPelargonium sect.Peristera (Geraniaceae) inferred from nrDNA and cpDNA sequence comparisons. *Plant Systematics and Evolution*, **211**, 273-287.

Bakker, F.T., Culham, A., Priyani, H., Tasoula, T. & Gibby, M. (2004) Phylogeny of Pelargonium (Geraniaceae) Based on DNA Sequences from Three Genomes. *Taxon*, **53**, 17-28.

Barreda, V.D., Palazzesi, L., Tellería, M.C., Olivero, E.B., Raine, J.I. & Forest, F. (2015) Early evolution of the angiosperm clade Asteraceae in the Cretaceous of Antarctica. *Proceedings of the National Academy of Sciences*, **112**, 10989.

Bartish, I.V., Antonelli, A., Richardson, J.E. & Swenson, U. (2011) Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, **38**, 177-190.

Bellstedt, D.U., Galley, C., Pirie, M.D. & Linder, H.P. (2012) The Migration of the Palaeotropical Arid Flora: Zygophylloideae as an Example. *Systematic Botany*, **37**, 951-959.

Bergh, N.G. & Peter Linder, H. (2009) Cape diversification and repeated out-of-southern-Africa dispersal in paper daisies (Asteraceae–Gnaphalieae). *Molecular Phylogenetics and Evolution*, **51**, 5-18.

Brooks, D. & McLennan, D. (2001) *A comparison of a discovery-based and event-based method of historical biogeography*.

Buerki, S., Forest, F., Alvarez, N., Nylander, J., Arrigo, N. & Sanmartin, I. (2011) *An evaluation of new parsimony-based versus parametric inference methods in biogeography: A case study using the globally distributed plant family Sapindaceae*.

Calviño, C.I., Teruel, F.E. & Downie, S.R. (2016) The role of the Southern Hemisphere in the evolutionary history of Apiaceae, a mostly north temperate plant family. *Journal of Biogeography*, **43**, 398-409.

Cardinal-McTeague, W.M., Wurdack, K.J., Sigel, E.M. & Gillespie, L.J. (2019) Seed size evolution and biogeography of Plukenetia (Euphorbiaceae), a pantropical genus with traditionally cultivated oilseed species. *BMC Evolutionary Biology*, **19**, 29.

Chala, D., Zimmermann, N.E., Brochmann, C. & Bakkestuen, V. (2017) Migration corridors for alpine plants among the 'sky islands' of eastern Africa: do they, or did they exist? *Alpine Botany*, **127**, 133-144.

Chase, B.M. & Meadows, M.E. (2007) Late Quaternary dynamics of southern Africa's winter rainfall zone. *Earth-Science Reviews*, **84**, 103-138.

Combs, J.K. & Pauw, A. (2009) Preliminary evidence that the long-proboscid fly, Philoliche gulosa, pollinates Disa karooica and its proposed Batesian model Pelargonium stipulaceum. *South African Journal of Botany*, **75**, 757-761.

Crisci, J.V. (2001) The voice of historical biogeography. *Journal of Biogeography*, **28**, 157-168.

Crisp, M.D. & Cook, L.G. (2013) How Was the Australian Flora Assembled Over the Last 65 Million Years? A Molecular Phylogenetic Perspective. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 303-324.

de Waal, C., Rodger, J.G., Anderson, B. & Ellis, A.G. (2014) Selfing ability and dispersal are positively related, but not affected by range position: a multispecies study on southern African Asteraceae. *Journal of Evolutionary Biology*, **27**, 950-959.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. (2006) Relaxed Phylogenetics and Dating with Confidence. *PLOS Biology*, **4**, e88.

Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L. & Smith, S.D. (2017) Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, **44**, 887-899.

Erkens, R.H.J., Maas, J.W. & Couvreur, T.L.P. (2009) From Africa via Europe to South America: migrational route of a species-rich genus of Neotropical lowland rain forest trees (Guatteria, Annonaceae). *Journal of Biogeography*, **36**, 2338-2352.

Estrella, M.d.l., Buerki, S., Vasconcelos, T., Lucas, E.J. & Forest, F. (2018) The Role of Antarctica in Biogeographical Reconstruction: A Point of View. *International Journal of Plant Sciences*, **180**, 63-71.

Ezcurra, M.D. & Agnolín, F.L. (2011) A New Global Palaeobiogeographical Model for the Late Mesozoic and Early Tertiary. *Systematic Biology*, **61**, 553-566.

Fer, I., Tietjen, B., Jeltsch, F. & Trauth, M.H. (2017) Modelling vegetation change during Late Cenozoic uplift of the East African plateaus. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **467**, 120-130.

Fiz, O., Vargas, P., Alarcón, M., Aedo, C., García, J.L. & Aldasoro, J.J. (2008) Phylogeny and Historical Biogeography of Geraniaceae in Relation to Climate Changes and Pollination Ecology. *Systematic Botany*, **33**, 326-342.

Galeotti, S., DeConto, R., Naish, T., Stocchi, P., Florindo, F., Pagani, M., Barrett, P., Bohaty, S.M., Lanci, L., Pollard, D., Sandroni, S., Talarico, F.M. & Zachos, J.C. (2016) Antarctic Ice Sheet variability across the Eocene-Oligocene boundary climate transition. *Science*, **352**, 76.

Galley, C., Bytebier, B., Bellstedt, D.U. & Linder, H.P. (2007) The Cape Element in the Afrotemperate Flora: From Cape to Cairo? *Proceedings: Biological Sciences*, **274**, 535-543.

Gamerro, J.C. & Barreda, V. (2008) New fossil record of Lactoridaceae in southern South America: a palaeobiogeographical approach. *Botanical Journal of the Linnean Society*, **158**, 41-50.

GBIF.org (2019a) (21 May 2019).

GBIF.org (2019b) (27 February 2019)b.

GBIF.org (2019c) (27 February 2019).

GBIF.org (2019d) (26 February 2019.

Gittenberger, E. (1991) What about non-adaptive radiation? *Biological Journal of the Linnean Society*, **43**, 263-272.

Goldblatt, P. & Manning, J.C. (2000) The Long-Proboscid Fly Pollination System in Southern Africa. *Annals of the Missouri Botanical Garden*, **87**, 146-170.

Harbaugh, D.T., Wagner, W.L., Allan, G.J., Zimmer, E.A. & Bellwood, D. (2009) The Hawaiian Archipelago Is a Stepping Stone for Dispersal in the Pacific: An Example from the Plant Genus Melicope (Rutaceae). *Journal of Biogeography*, **36**, 230-241.

Heleno, R. & Vargas, P. (2015) How do islands become green? *Global Ecology and Biogeography*, **24**, 518-526.

Hennig, W. (1965) Phylogenetic Systematics. *Annual Review of Entomology*, **10**, 97-116.

Higgins, S.I., Nathan, R. & Cain, M.L. (2003) ARE LONG-DISTANCE DISPERSAL EVENTS IN PLANTS USUALLY CAUSED BY NONSTANDARD MEANS OF DISPERSAL? *Ecology*, **84**, 1945-1956.

Hijmans, R.J. (2019) raster: Geographic Data Analysis and Modeling. R package version 2.8-19. .

Hovenkamp, P. (1997) Vicariance Events, not Areas, Should be Used in Biogeographical Analysis. *Cladistics*, **13**, 67-79.

Humphries, C.J. (2000) Form, space and time; which comes first? *Journal of Biogeography*, **27**, 11-15.

Iglesias, A. (2016) New Upper Cretaceous (Campanian) Flora from James Ross Island, Antarctica. *Ameghiniana*, **53**, 358-374, 17.

Janssens, S.B., Groeninckx, I., De Block, P.J., Verstraete, B., Smets, E.F. & Dessein, S. (2016) Dispersing towards Madagascar: Biogeography and evolution of the Madagascan endemics of the Spermacoceae tribe (Rubiaceae). *Molecular Phylogenetics and Evolution*, **95**, 58-66.

Jürgens, N. (1997) Floristic biodiversity and history of African arid regions. *Biodiversity & Conservation*, **6**, 495-514.

Katinas, L., Crisci, J.V., Hoch, P., Tellería, M.C. & Apodaca, M.J. (2013) Trans-oceanic dispersal and evolution of early composites (Asteraceae). *Perspectives in Plant Ecology, Evolution and Systematics*, **15**, 269-280.

Lieberman, B.S. (2003) *Paleobiogeography: The Relevance of Fossils to Biogeography*.

Linder, H., Lovett, J., Mutke, J., Barthlott, W., Jürgens, N., Rebelo, T. & Küper, W. (2005) *A numerical re-evaluation of the sub-Saharan Phytochoria of mainland Africa*.

Linder, H.P. (2003) The radiation of the Cape flora, southern Africa. *Biological Reviews*, **78**, 597-638.

Linder, H.P. (2017) East African Cenozoic vegetation history. *Evolutionary Anthropology: Issues, News, and Reviews*, **26**, 300-312.

Livermore, R., Nankivell, A., Eagles, G. & Morris, P. (2005) Paleogene opening of Drake Passage. *Earth and Planetary Science Letters*, **236**, 459-470.

Maddison, W.P. & Maddison, D.R. (2018) Mesquite: a modular system for evolutionary analysis. Version 3.6

Marais, E.M. (1994) Taxonomic studies in Pelargonium section Hoarea (Geraniaceae) (PhD thesis). *University of Stellenbosch, Stellenbosch, South Africa.*,

Martínez-Cabrera, H.I. & Peres-Neto, P.R. (2013) Shifts in Climate Foster Exceptional Opportunities for Species Radiation: The Case of South African Geraniums. *PLOS ONE*, **8**, e83087.

Matzke, N. (2013a) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts.

Matzke, N. (2013b) *Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing*.

Matzke, N.J. (2014) Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, **63**, 951-970.

Mucina, L., Devos, N., Barker, N. & Nordenstam, B. (2010) *A Multi-Locus Phylogeny of Euryops (Asteraceae, Senecioneae) Augments Support for the "Cape to Cairo" Hypothesis of Floral Migrations in Africa*.

Myers, J.A., Vellend, M., Gardescu, S. & Marks, P.L. (2004) Seed dispersal by white-tailed deer: implications for long-distance dispersal, invasion, and migration of plants in eastern North America. *Oecologia*, **139**, 35-44.

Nathan, R. (2006) Long-Distance Dispersal of Plants. *Science*, **313**, 786.

Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution*, **23**, 638-647.

Newman, E., Manning, J. & Anderson, B. (2014) Matching floral and pollinator traits through guild convergence and pollinator ecotype formation. *Annals of botany*, **113**, 373-384.

Nicotra, A.B., Chong, C., Bragg, J.G., Ong, C.R., Aitken, N.C., Chuah, A., Lepschi, B. & Borevitz, J.O. (2016) Population and phylogenomic decomposition via genotyping-by-sequencing in Australian Pelargonium. *Molecular Ecology*, **25**, 2000-2014.

Palazzesi, L., Gottschling, M., Barreda, V. & Weigend, M. (2012) First Miocene fossils of Vivianiaceae shed new light on phylogeny, divergence times, and historical biogeography of Geraniales. *Biological Journal of the Linnean Society*, **107**, 67-85.

Pole, M. (2001) *Can long-distance dispersal be inferred from the New Zealand fossil record*.

R-Core-Team (2019) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.*,

Ree, R.H. & Sanmartín, I. (2009) Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, **36**, 1211-1220.

Ree, R.H., Moore, B.R., Webb, C.O. & Donoghue, M.J. (2005) A LIKELIHOOD FRAMEWORK FOR INFERRING THE EVOLUTION OF GEOGRAPHIC RANGE ON PHYLOGENETIC TREES. *Evolution*, **59**, 2299-2311.

Reeves, C. & De Wit, M. (2000) Making ends meet in Gondwana: retracing the transforms of the Indian Ocean and reconnecting continental shear zones. *Terra Nova*, **12**, 272-280.

Renner, S. (2004) Plant Dispersal across the Tropical Atlantic by Wind and Sea Currents. *International Journal of Plant Sciences*, **165**, S23-S33.

Renner, S.S., Strijk, J.S., Strasberg, D. & Thébaud, C. (2010) Biogeography of the Monimiaceae (Laurales): a role for East Gondwana and long-distance dispersal, but not West Gondwana. *Journal of Biogeography*, **37**, 1227-1238.

Revell, L.J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol*, 217-223.

Ronquist, F. (2002) *Parsimony analysis of coevolving species associations*.

Ronquist, F.C.F.p.d.J. (1994) Ancestral Areas and Parsimony. *Systematic Biology*, **43**, 267-274.

Röschenbleck, J., Albers, F., Müller, K., Weinl, S. & Kudla, J. (2014) Phylogenetics, character evolution and a subgeneric revision of the genus Pelargonium (Geraniaceae). *Phytotaxa*, **159**

Rosen, D.E. (1978) Vicariant Patterns and Historical Explanation in Biogeography. *Systematic Zoology*, **27**, 159-188.

Sanmartin, I. (2007) *Event-based biogeography: Integrating patterns, processes, and time*.

Sanmartin, I. (2016) Breaking the chains of parsimony: the development of parametric methods in historical biogeography. pp. 239-243.

Sanmartín, I. (2012) Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach*, **5**, 555-568.

Sanmartín, I. & Ronquist, F. (2004) Southern Hemisphere Biogeography Inferred by Event-Based Models: Plant versus Animal Patterns. *Systematic Biology*, **53**, 278-298.

Sanmartín, I., Anderson Cajsa, L., Alarcon, M., Ronquist, F. & Aldasoro Juan, J. (2010) Bayesian island biogeography in a continental setting: the Rand Flora case. *Biology Letters*, **6**, 703-707.

Scher, H.D. & Martin, E.E. (2006) Timing and Climatic Consequences of the Opening of Drake Passage. *Science*, **312**, 428.

Ségalen, L., Lee-Thorp, J.A. & Cerling, T. (2007) Timing of C4 grass expansion across sub-Saharan Africa. *Journal of Human Evolution*, **53**, 549-559.

Sommerfeld, A., Prömmel, K. & Cubasch, U. (2016) The East African Rift System and the impact of orographic changes on regional climate and the resulting aridification. *International Journal of Earth Sciences*, **105**, 1779-1794.

South African National Biodiversity Institute, S. (2018) Dataset.

Struck, M. (1997) Floral divergence and convergence in the genusPelargonium (Geraniaceae) in southern Africa: Ecological and evolutionary considerations. *Plant Systematics and Evolution*, **208**, 71-97.

Sytsma, K.J., Spalink, D. & Berger, B. (2014) Calibrated chronograms, fossils, outgroup relationships, and root priors: re-examining the historical biogeography of Geraniales. *Biological Journal of the Linnean Society*, **113**, 29-49.

Taylor, D.W. (1991) Paleobiogeographic relationships of Andean angiosperms of Cretaceous to Pliocene age. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **88**, 69-84.

Thiv, M. & Meve, U. (2007) *A phylogenetic study of Echidnopsis Hook. f. (Apocynaceae-Asclepiadoideae) - Taxonomic implications and the colonization of the Socotran archipelago*.

Thiv, M., van der Niet, T., Rutschmann, F., Thulin, M., Brune, T. & Linder, H.P. (2011) Old—New World and trans-African disjunctions of Thamnosma (Rutaceae): Intercontinental long-distance dispersal and local differentiation in the succulent biome. *American Journal of Botany*, **98**, 76-87.

Upchurch, P. & Hunn, C.A. (2002) "Time": the neglected dimension in cladistic biogeography? *Geobios*, **35**, 277-286.

van de Kerke, S.J., Shrestha, B., Ruhlman, T.A., Weng, M.-L., Jansen, R.K., Jones, C.S., Schlichting, C.D., Hosseini, S., Mohammadin, S., Schranz, M.E. & Bakker, F.T. (2019) Plastome based phylogenetics and younger crown node age in Pelargonium. *Molecular Phylogenetics and Evolution*, **137**, 33-43.

Van der Walt, J.J.A., Vorster, P.J. & Ward-Hilhorst, E. (1977) *Pelargoniums of Southern Africa*. Purnell.

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

Waters, J.M. & Roy, M.S. (2004) Out of Africa: The Slow Train to Australasia. *Systematic Biology*, **53**, 18-24.

Weng, M.-L., Ruhlman, T.A., Gibby, M. & Jansen, R.K. (2012) Phylogeny, rate variation, and genome size evolution of Pelargonium (Geraniaceae). *Molecular Phylogenetics and Evolution*, **64**, 654-670.

Wichura, H., Bousquet, R., Oberhänsli, R., Strecker, M.R. & Trauth, M.H. (2010) Evidence for middle Miocene uplift of the East African Plateau. *Geology*, **38**, 543-546.

Wichura, H., Jacobs, L.L., Lin, A., Polcyn, M.J., Manthi, F.K., Winkler, D.A., Strecker, M.R. & Clemens, M. (2015) A 17-My-old whale constrains onset of uplift and climate change in east Africa. *Proceedings of the National Academy of Sciences*, **112**, 3910.

Wikström, N., Savolainen, V. & Chase Mark, W. (2001) Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **268**, 2211-2220.

Wu, Z.-Y., Liu, J., Provan, J., Wang, H., Chen, C.-J., Cadotte, M.W., Luo, Y.-H., Amorim, B.S., Li, D.-Z. & Milne, R.I. (2018) Testing Darwin's transoceanic dispersal hypothesis for the inland nettle family (Urticaceae). *Ecology Letters*, **21**, 1515-1529.

Yoder, A.D. & Nowak, M.D. (2006) Has Vicariance or Dispersal Been the Predominant Biogeographic Force in Madagascar? Only Time Will Tell. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 405-431.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001) Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science*, **292**, 686.

Zeng, X., Zheng, Y., Peng, R. & Kuro-o, M. (2011) Exploring Patterns and Extent of Bias in Estimating Divergence Time from Mitochondrial DNA Sequence Data in a Particular Lineage: A Case Study of Salamanders (Order Caudata). *Molecular Biology and Evolution*, **28**, 2521-2535.