

# Bayesian estimation of the global biogeographical history of the Solanaceae

Julia Dupin<sup>1\*</sup>, Nicholas J. Matzke<sup>2</sup>, Tiina Särkinen<sup>3</sup>, Sandra Knapp<sup>4</sup>, Richard G. Olmstead<sup>5</sup>, Lynn Bohs<sup>6</sup> and Stacey D. Smith<sup>1</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO 80309, USA, <sup>2</sup>Division of Ecology, Evolution and Genetics, Research School of Biology, The Australian National University, ACT 2601 Canberra, Australia, <sup>3</sup>Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK, <sup>4</sup>Department of Life Sciences, Natural History Museum, London SW7 5BD, UK, <sup>5</sup>Department of Biology and Burke Museum, University of Washington, Seattle, Washington 98195, USA, <sup>6</sup>Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

## ABSTRACT

**Aim** The tomato family Solanaceae is distributed on all major continents except Antarctica and has its centre of diversity in South America. Its worldwide distribution suggests multiple long-distance dispersals within and between the New and Old Worlds. Here, we apply maximum likelihood (ML) methods and newly developed biogeographical stochastic mapping (BSM) to infer the ancestral range of the family and to estimate the frequency of dispersal and vicariance events resulting in its present-day distribution.

**Location** Worldwide.

**Methods** Building on a recently inferred megaphylogeny of Solanaceae, we conducted ML model fitting of a range of biogeographical models with the program 'BioGeoBEARS'. We used the parameters from the best fitting model to estimate ancestral range probabilities and conduct stochastic mapping, from which we estimated the number and type of biogeographical events.

**Results** Our best model supported South America as the ancestral area for the Solanaceae and its major clades. The BSM analyses showed that dispersal events, particularly range expansions, are the principal mode by which members of the family have spread beyond South America.

**Main conclusions** For Solanaceae, South America is not only the family's current centre of diversity but also its ancestral range, and dispersal was the principal driver of range evolution. The most common dispersal patterns involved range expansions from South America into North and Central America, while dispersal in the reverse direction was less common. This directionality may be due to the early build-up of species richness in South America, resulting in large pool of potential migrants. These results demonstrate the utility of BSM not only for estimating ancestral ranges but also in inferring the frequency, direction and timing of biogeographical events in a statistically rigorous framework.

## Keywords

BioGeoBEARS, biogeographical stochastic mapping, directionality, dispersal, historical biogeography, Solanaceae

\*Correspondence: Julia Dupin, Department of Ecology and Evolutionary Biology, University of Colorado Boulder, 1800 Colorado Ave, Boulder, CO 80309, USA.  
E-mail: julia.dupin@colorado.edu

## INTRODUCTION

The growing availability of large phylogenies together with developments in statistical methods provides researchers with new opportunities to explore the complex biogeographical history of large cosmopolitan clades (Matzke, 2014; Yu *et al.*, 2015). Worldwide distributions can be achieved through dispersal events, vicariance, or a combination of the two (Yoder

& Nowak, 2006; Nauheimer *et al.*, 2012). For example, Gondwanan clades that arose and diversified before continental separation may become cosmopolitan largely through vicariance (Gamble *et al.*, 2008; Rasmussen & Cameron, 2010). By contrast, groups that have evolved subsequent to continental separation can only achieve wide distributions through long-distance dispersal events (Givnish *et al.*, 2004; Szövényi *et al.*, 2008). The relative importance of these

processes is expected to vary across groups of organisms given intrinsic differences in dispersal ability and level of ecological specialization (Gillespie *et al.*, 2012; Edwards & Donoghue, 2013).

The interplay of vicariance and dispersal in shaping cosmopolitan distributions has been relatively well studied in plants. For many widely distributed plant families, dispersal appears to be the principal driver of range evolution (Givnish & Renner, 2004; Christenhusz & Chase, 2013), whether by water (Gallaher *et al.*, 2015), wind (Muñoz *et al.*, 2004) or animals (Nogales *et al.*, 2012). By contrast, relatively few widespread families show strong signatures of vicariance in shaping their present-day distributions (Barker *et al.*, 2007; Mao *et al.*, 2012). This pattern may be attributable to the fact that long-distance dispersal of seeds or other germplasm, while limited on ecological time-scales (Cain *et al.*, 2000), appears to be relatively frequent on macroevolutionary time-scales (Renner, 2004; Nathan, 2006). Moreover, the establishment of distantly dispersed lineages may be favoured by the absence of their native competitors, pathogens and predators (Janzen, 1970; Howe & Miriti, 2004). One caveat, however, with respect to the apparent predominance of dispersal is that many of the studied lineages are relatively young (but see Beaulieu *et al.*, 2013). Also, vicariance becomes harder to identify in older clades as subsequent dispersal and population movement, along with local extinction, can obscure the original geographical signature (e.g. Clayton *et al.*, 2009).

Here, we examine the biogeographical history of a relatively young plant family, the Solanaceae. This clade of about 2,800 species began to diversify roughly 50–65 million years ago (Ma) (Särkinen *et al.*, 2013; Magallón *et al.*, 2015) and is presently distributed on all continents except Antarctica. The Solanaceae contains a large number of important crops, such as potato (*Solanum tuberosum* L.), tomato (*Solanum lycopersicum* L.) and tobacco (*Nicotiana tabacum* L.), and has long been the focus of genetic, biochemical and morphological studies (e.g. Pabón-Mora & Litt, 2011; Sato *et al.*, 2012; Itkin *et al.*, 2013). Solanaceae is one of the groups in the Asterids I (Lamiids) that is especially species rich in the Neotropics, along with the Verbenaceae and Bignoniaceae (Olmstead, 2013), and the Acanthaceae (Tripp & McDade, 2014). Its sister group is the large morning glory family, Convolvulaceae, which also has a worldwide distribution. Despite its economic importance, the biogeographical history of Solanaceae has received relatively little attention beyond taxonomic or floristic surveys (e.g. Gentry & D'Arcy, 1986; Hepper & Jaeger, 1986). In a recent review of Solanaceae biogeography in the context of its phylogenetic history, Olmstead (2013) suggested that the family's cosmopolitan distribution is due to repeated dispersals from South America to both nearby and distant continents. However, identifying the timing, number and direction of these events was limited by the level of taxon sampling (c. 5% of species of the family) and the lack of a time-calibrated phylogeny.

In this study, we combine a recent dated phylogeny that includes nearly 40% of the species in the family (Särkinen

*et al.*, 2013) with newly developed biogeographical stochastic mapping (BSM; Matzke, 2015) to estimate the biogeographical events that account for the global distribution of the Solanaceae. Stochastic mapping (Nielsen, 2002; Huelsenbeck *et al.*, 2003) is a simulation approach that builds on likelihood models of trait evolution and, in addition to estimating ancestral states at nodes, provides possible histories of changes along branches. By summarizing across many of these possible histories, we can obtain estimates of the number and phylogenetic location of various types of events (e.g. vicariance, dispersal between areas) along with measures of uncertainty. This approach was originally developed for mapping mutations onto phylogenies (Nielsen, 2002) and was later expanded to accommodate morphological characters (Huelsenbeck *et al.*, 2003). Just as the morphological stochastic mapping was built on Pagel's (1999) likelihood models for trait evolution, the development of stochastic mapping for biogeographical patterns relies on existing likelihood models describing how geographical ranges evolve (Ree & Smith, 2008; Matzke, 2014). In the context of Solanaceae biogeography, we apply likelihood methods and BSM to (1) infer the most likely ancestral range of the family and major groups within it, (2) assess the relative contribution of vicariance and dispersal events to the distribution of extant taxa and (3) detect directionality in dispersal between areas. In addition to elucidating the biogeographical history of this economically and floristically important family, these results provide new insights into the relative importance of alternate dispersal routes in the spread of plant clades to new areas.

## MATERIALS AND METHODS

### Solanaceae phylogeny and species distribution

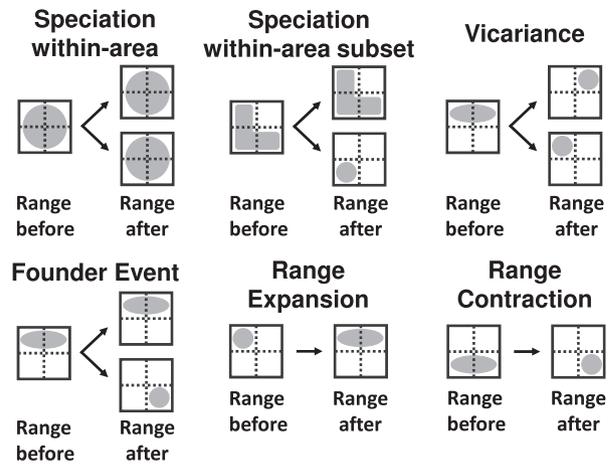
We used the time-calibrated maximum clade credibility (MCC) tree from Särkinen *et al.* (2013), which was estimated using two nuclear and six plastid loci from 1075 species. This taxon sampling includes all but three of the 98 Solanaceae genera [*Darcyanthus* Hunz., *Capsicophysalis* Averett & M. Martínez and *Tubocapsicum* (Wettst.) Makino] and nearly 40% of all species. For this study, we pruned: (1) taxa that are widely cultivated and whose native distributions have been obscured by extensive human transport, and (2) taxa that were duplicated in the phylogeny (see Table S1.1 in Appendix S1 in Supporting Information). The phylogeny was pruned using 'ape' package (Paradis *et al.*, 2004) in R (R Core Team, 2015). We also updated species names according to the most recent literature (Table S1.1 in Appendix S1). The final pruned phylogeny used for downstream analyses contained 1044 species.

The current distribution for each species in the phylogeny was determined using numerous literature sources (e.g. Benítez de Rojas & D'Arcy, 1997; Garcia & Olmstead, 2003; Dillon *et al.*, 2009), online databases (Solanaceae Source, <http://solanaceae-source.org/>, last accessed on November 2015; TROPICOS, Missouri Botanical Garden, <http://www.tropicos.org>, last

accessed on November 2015; Global Biodiversity Information Facility, <http://www.gbif.org>, last accessed on November 2015), and experts' input (Table S1.2 in Appendix S1). Considering current distribution patterns within the family, we chose to recognize seven major areas: South America (SAM), Central America (CAm), Caribbean (Car), North America (NAm), Eurasia (EU), Africa (AF) and Australia (OZ; includes Australia, other islands of Oceania and the Hawaiian islands). The decision to focus on these seven areas reflects the need to balance model complexity (i.e. the number of dispersal rates, within-area extinction rates) with the ability to detect major biogeographical shifts within this widely distributed family. The Caribbean and Central America were maintained as separate areas to examine the origins of the many solanaceous lineages endemic to those regions. We grouped the six species native to Hawaii (*Solanum incompletum* Dunal, *S. sandwicense* Hook. & Arn., *S. viridifolium* Dunal, *Nothocestrum latifolium* A. Gray, *N. longifolium* A. Gray, and *Lycium sandwicense* A. Gray) with the Australian taxa (Table S1.2 in Appendix S1). While dispersal to Hawaii and remote Pacific islands represents a long-distance dispersal from any potential source, three of them are nested within clades found in Australia and New Zealand (*S. incompletum*, *S. sandwicense* and *S. viridifolium*), whereas the others represent putative Eurasian or New World ancestry (Levin & Miller, 2005; Levin *et al.*, 2006; Olmstead *et al.*, 2008; Vorontsova *et al.*, 2013). Our divisions are similar to those used in other studies of widespread clades (e.g. Buerki *et al.*, 2011; Sessa *et al.*, 2012) and will allow us to detect intercontinental movements as well as shorter range dispersal events (e.g. South America to the Caribbean).

### Ancestral range estimation

We used the R package 'BioGeoBEARS' (Matzke, 2013) to compare biogeographical models and estimate ancestral ranges in the Solanaceae. 'BioGeoBEARS' implements maximum likelihood (ML) methods that replicate the key assumptions of three most commonly used methods in historical biogeography, namely DEC (dispersal–extinction–cladogenesis; Ree & Smith, 2008), DIVA (dispersal–vicariance analysis; Ronquist, 1997) and BayArea (Bayesian Inference of Historical Biogeography for Discrete Areas; Landis *et al.*, 2013). These three methods were originally developed in different frameworks (likelihood for DEC, parsimony for DIVA, and Bayesian for BayArea), but are all represented as likelihood models in 'BioGeoBEARS' to allow for direct comparison. The latter two models are, thus, not identical to their original formulation and are referred to as DIVALIKE and BAYAREALIKE within 'BioGeoBEARS' (Matzke, 2013). Collectively, these models allow for a wide range of processes, including within-area speciation, vicariance, range expansion (dispersal to a new area) and range contraction (extinction in an area) (Fig. 1). We also tested models with and without founder-event speciation, which is incorporated with the  $j$  parameter. In such an event, range switching (e.g. South America to North America) occurs at a lineage-splitting



**Figure 1** Diagrams of different types of biogeographical events allowed in the models tested in this study (adapted from Matzke, 2015). Cladogenetic events include within-area speciation, vicariance and founder events. Anagenetic events include range expansion and range contraction.

event (a node in the phylogeny), leaving one daughter lineage in a new range and the other daughter lineage retaining the ancestral range. Such range switching events are restricted to nodes (instead of occurring along branches) as it is considered unlikely that an entire lineage would simultaneously disperse to a new area and go extinct in its ancestral area (Matzke, 2014).

We also incorporated time-stratified dispersal multiplier matrices in the model fitting to account for the changing distances between the regions over geological time. We divided the history of Solanaceae into three strata: 50–24, 24–10, and 10 Ma to present. We began the strata at 50 Ma given the estimated depth of the Solanaceae phylogeny (*c.* 49 Ma for the stem group age; Särkinen *et al.*, 2013). The breaks at 24 and 10 Ma reflect recent studies showing significant shifts in plant dispersal between South and North America at these time points (Bacon *et al.*, 2015). Also, the 10-Ma threshold approximates a new estimated age of the closure of the Panama isthmus during the mid-Miocene (Montes *et al.*, 2012; Bacon *et al.*, 2013; but see O'Dea *et al.*, 2016). The dispersal multiplier matrices for each of these strata give the relative probability of dispersal between areas and are roughly scaled to represent the relative distance between the areas during each time slice (Appendix S1). We examined models using these multiplier matrices directly and also with the matrices modified by the  $w$  parameter, also estimated with ML (Appendix S1).

In total, we tested a set of 18 models (Table S1.3 in Appendix S1) that varied in the number and types of free parameters included and in the type of dispersal multiplier matrices used. The free parameters were  $w$ ,  $d$  (the base rate of range expansion),  $j$  (the per-event weight of founder-event speciation at cladogenesis; Matzke, 2014) and  $e$  (the rate of range contraction). For models without the dispersal multiplier matrices, the probabilities of all dispersal events

are equal (set to 1). Because not all of the models are nested, we used the Akaike information criterion (AIC) (Burnham & Anderson, 2002) to select among the 18 models; the best fit model with the lowest AIC score was used to infer the relative probabilities of ancestral ranges within the phylogeny.

### Estimation of number and type of biogeographical events

We estimated the number and type of biogeographical events using BSM implemented in 'BioGeoBEARS' (Matzke, 2015). Previous implementations of stochastic mapping used transition rate models (Pagel, 1999) to simulate the histories of mutations or trait changes (Huelsenbeck *et al.*, 2003). Whereas the transition rate models involve only trait gain and loss, the biogeographical models used in BSM will include a range of anagenetic and cladogenetic events (Fig. 1). After providing a biogeographical model with specified parameters, BSM generates simulated histories ('realizations'), including the times and locations of all events along the branches in that simulation. These realizations of possible histories are constrained to produce the observed data given the phylogeny, and averaging over many realizations will result in the same ancestral state probabilities as those calculated analytically under the ML model (see Appendix S1 for description and BSM algorithm validation). Biogeographical events possible under the models include within-area speciation, vicariance and dispersal events (range expansions and founder events; Fig. 1). For cases where a dispersal event occurs from a widespread ancestor occupying two or more areas, the exact source area was simulated using the dispersal probability multipliers matrix as modified by  $w$ . We conducted BSM on the Särkinen *et al.* (2013) MCC tree using a time-stratified four-parameter model (DEC+ $j+w$ ) that produced a significantly better fit to the data compared with other tested models (see Results). Event frequencies were estimated by taking the mean and standard deviation of event counts from 100 BSMs. All stochastic maps and derived statistical estimates are conditioned on not just the phylogeny, the observed range data, and the model, but the inferred model parameters and the implicit Yule process (no lineage extinction) assumption shared by DEC and all other biogeographical models considered here.

## RESULTS

### Ancestral range estimation and time-stratified analyses

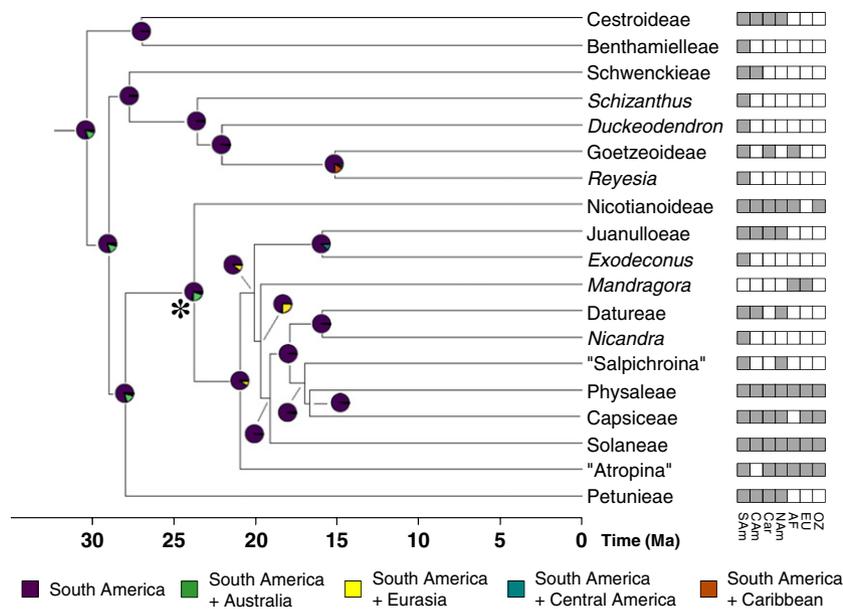
The AIC model selection strongly supported the DEC+ $j+w$  model including the dispersal multiplier matrices as the best fit. This is among the more complex models fitted (with four free parameters), and it was 47 AIC units lower than the second best model (DEC+ $j$ ) (Table S1.3 in Appendix S1). Across all the models, the inclusion of the  $j$  parameter (or

the possibility of a founder-event speciation where one of the lineages occupies an area not present in the ancestral range) consistently improved model fit, suggesting that range expansions alone are not sufficient to account for movements to new areas. Such an increase in the likelihood in models with  $j$  has been observed in other studies where, as in our case, the areas under consideration are continents or other large regions and thus, many lineages are single area endemics (Litsios *et al.*, 2014; Voelker *et al.*, 2014; Thacker, 2015). It is also notable that the best model employed the user-specified dispersal matrices with the  $w$  modifier parameter. This result indicates that scaling the dispersal probabilities to the relative distance of the areas results in higher model likelihood. The  $w$  parameter, acting as an exponent on the dispersal matrices, improves fit to our data (Table S1.3 in Appendix S1, + $j$  models vs. + $j+w$  models).

Ancestral range estimations under this best fitting model (DEC+ $j+w$ ) showed that the most probable ancestral area for extant species of Solanaceae is South America ( $P = 0.8$ , with 0.14 for South America + Australia, and 0.06 for other state combinations) (Fig. 2 & Appendix S2). Other deep nodes in the family also have South America as the estimated range, supporting the idea that this area is the centre of origin of the family and major clades within it (Fig. 2). The first two clades in the family confidently inferred as having an ancestral range outside of the New World are the Hyoscyameae (*c.* 12 Ma), nested within Atropina, which has Eurasia as the most probable ancestral state (0.95), and the clade Anthocercidae (*c.* 9 Ma), nested within Nicotianoideae, which has Australia as the most probable ancestral range (0.98) (Appendix S2).

### Estimation of number, type, and directionality of biogeographical events

A summary of our BSMs revealed that most biogeographical events comprise within-area speciation (76%) and dispersals (20%), with only a few vicariant events (3%) (Table 1; see Appendix S3 for an example of a BSM). The high number of within-area speciation events was expected given the large size of our regions (e.g. North America, Eurasia). The number of these events within each area was closely related to species richness, where, for example, 60% of the within-area speciation events occurred within South America, which is home to about 64% of the taxa (Table S1.4 in Appendix S1). Among the dispersal events, range expansions were much more common than founder events (mean of 218 vs. 38; Table 1) and were clustered towards the present (Fig. 3). The relative rarity of vicariance (mean = 42 events; Table 1) may reflect the fact that the diversification of Solanaceae postdates the Gondwanan break-up (180 Ma, early Jurassic – 94 Ma, mid-Cretaceous; McLoughlin, 2001). Most of the estimated vicariance events involved adjacent areas (e.g. South America and North America) (Table S1.5 in Appendix S1) and occurred following a range expansion. For example, the ancestor of the genus *Jaltomata* was estimated



**Figure 2** Maximum likelihood ancestral range estimation in Solanaceae, using the best model DEC+*j*+*w* (model 16 in Table S1.3 in Appendix S1). This tree is a simplified version of our 1044-tip tree (Fig. 3, Appendix S2 & Appendix S3), and it includes 19 monophyletic groups within Solanaceae, representing either genera (italicized) or higher order taxa (not italicized). The pie diagrams at nodes show the relative probability of the possible states (areas or combinations of areas). The boxes on the right show the native ranges of taxa within these clades, including South America (SAm), Central America (CAm), the Caribbean (Car), North America (NAm), Africa (AF), Eurasia (EU) and Australia (OZ). The asterisk marks the ancestor for the '*x* = 12' clade, a group that shares the base number of 12 chromosomes and comprises roughly 85% of species in the family (Olmstead & Sweere, 1994). Outgroups are not shown.

**Table 1** Summary of biogeographical stochastic mapping counts for the Solanaceae using the DEC+*j*+*w* model. Mean numbers of the different types of events estimated are shown here along with standard deviations. No range contractions were estimated because the relevant model parameter (*e*) was not required in the best fitting model (Table S1.3 in Appendix S1).

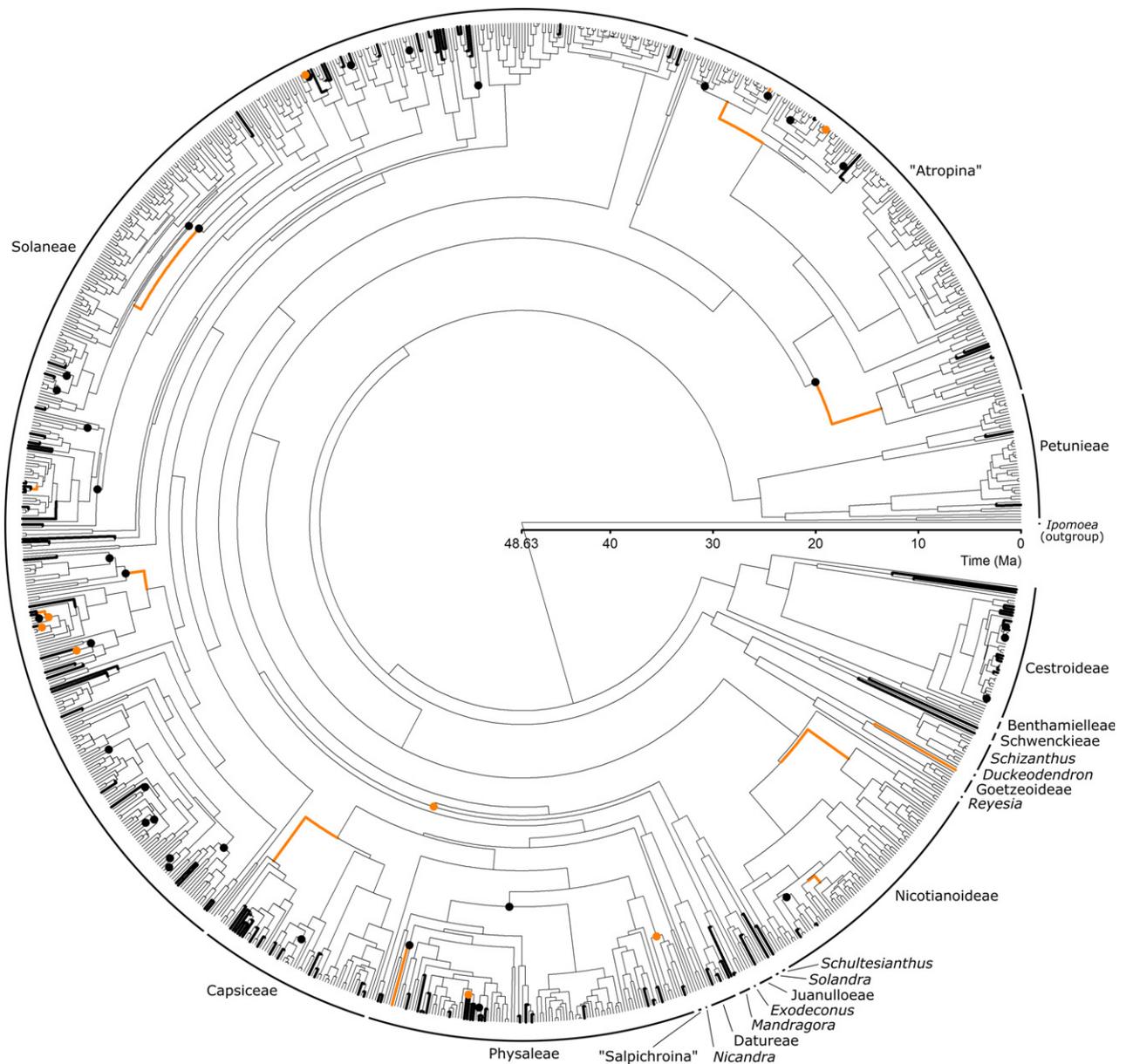
Mode	Type	Mean (SD)	%
Within-area speciation	Speciation	883.38 (6.99)	70.11
	Speciation – subset	79.28 (7.77)	6.29
Dispersal	Founder events	38.26 (4.25)	3.03
	Range expansions	217.55 (4.23)	17.26
	Range contractions	0	0
Vicariance	Vicariance	41.53 (4.18)	3.29
Total		1,260.00 (4.23)	100.00

to have expanded its range from South America to South and North America, and this expansion was followed by a division into subclades most diverse in South and North America (Appendix S2; see also Mione *et al.*, 1994).

Focusing on dispersal events, we found that movement patterns varied tremendously across areas. The highest number of dispersals involved movements from South America to Central America (*c.* 49 of 256 total estimated events), closely followed by movements from South America to North America (*c.* 42 of 256) (Fig. 4a). Overall, South America was

the source for 47% of the estimated dispersal events. North America was the next most common source (20%), and most of these dispersals were towards Central America (Figs 4a & 5). In total, dispersals among the four New World areas accounted for 81% of the events, while dispersal among the three Old World regions or between the Old World and New World were less common (10% and 9%, respectively). Among the Old World areas, Australia and the nearby islands of Oceania were the least common source and sink for dispersal events, and three of the estimated 11 dispersals into the area comprise the Hawaiian taxa previously shown to have originated from Eurasia and the New World (Levin & Miller, 2005; Levin *et al.*, 2006; Olmstead *et al.*, 2008; Vortontsova *et al.*, 2013).

The relative importance of different regions as sources and sinks for dispersal differed to some degree depending on the type of dispersal event (range expansions vs. founder events). Since the vast majority of dispersal events were range expansions, the pattern of those movements (Fig. 4b) largely mirrors the overall pattern (Fig. 4a), where South America is the most common source and Central America is the most common sink. By contrast, the most frequent founder-event type involved dispersal from South America to North America (Fig. 4c), with the other transitions (e.g. to Central America, to Africa) occurring with roughly equal frequency. Thus, whereas the range expansions commonly occur between adjacent areas, the founder events more often involve distant



**Figure 3** Representation of timing of dispersal events in the Solanaceae chronogram. Range expansions = thick branches, founder events = full circles. Transoceanic dispersals are represented in orange and non-transoceanic dispersals in black. Thicker branches and nodes with circles represented here were found in at least 95% of BSM realizations. Clade names follow Fig. 2.

related regions. Indeed, movements between the Old World and New World accounted for just 6% of estimated range expansions, but 24% of the estimated founder events.

Regardless of the type of dispersal event, we inferred strong asymmetry in the movements between areas. For example, dispersal events from Central America to the Caribbean were more than twice as common as those in the opposite direction ( $10.12 \pm 1.41$  vs.  $4.20 \pm 1.14$ ; Fig. 4a). Such directionality was observed for nearly all pairs of areas (compare upper diagonals and lower diagonals in Fig. 4) and was most marked for events involving South America (see also Fig. 5). The most prominent exceptions to this overall trend are dispersals between Africa and Eurasia, which have

occurred in approximately equal numbers in both directions (Figs 4 & 5). The general asymmetry of transitions was consistent across all of the individual BSM realizations (chi-square contingency analysis,  $P < 1 \times 10^{-45}$  for each of the 100 realizations).

## DISCUSSION

Our analyses of the historical biogeography of Solanaceae confirm that the early evolution of the family took place in South America. Moreover, major clades in the family, such as the large 'x = 12' clade (genera with a base chromosome number of 12) and the Solaneae (*Solanum*+*Jaltomata*), are

**(a) Summary of dispersal events counts (and standard deviations)**

	SAm	CAm	Car	NAm	AF	EU	OZ	
SAm	-	48.6 (2.63)	15.13 (1.62)	41.65 (2.83)	5.67 (0.89)	5.03 (1.14)	3.58 (0.77)	119.66 47%
CAm	11.35 (2.32)	-	10.12 (1.41)	23.26 (4.21)	0.22 (0.19)	0.35 (0.34)	0.32 (0.31)	45.62 18%
Car	1.21 (0.94)	4.20 (1.14)	-	3.09 (1.02)	0.22 (0.27)	0.11 (0.17)	0.14 (0.08)	8.97 4%
NAm	9.12 (1.51)	32.44 (4.01)	7.39 (1.20)	-	0.59 (0.29)	2.33 (0.78)	1.34 (0.52)	53.21 20%
AF	0.03 (0.21)	0.04 (0.22)	0.07 (0.33)	0.13 (0.31)	-	10.15 (1.39)	3.89 (0.83)	14.31 6%
EU	0.09 (0.22)	0.08 (0.32)	0.11 (0.21)	1.17 (0.32)	8.28 (1.32)	-	1.78 (0.37)	11.51 4%
OZ	0.06 (0.12)	0.04 (0.21)	0.04 (0.11)	0.12 (0.34)	1.39 (0.71)	0.88 (0.27)	-	2.53 1%
	21.86 9%	85.4 33%	32.86 13%	69.42 27%	16.37 6%	18.85 8%	11.05 4%	255.81 100%

**(b) Range expansion event counts (and standard deviations)**

	SAm	CAm	Car	NAm	AF	EU	OZ	
SAm	-	45.79 (3.51)	12.52 (1.9)	30.01 (3.24)	3.16 (1.42)	2.2 (1.29)	2.61 (1.03)	96.29 44%
CAm	10.7 (2.58)	-	9.11 (2.09)	22.23 (4.41)	0.17 (0.44)	0.28 (0.53)	0.3 (0.48)	42.79 20%
Car	1.15 (1.03)	3.78 (1.65)	-	2.94 (1.51)	0.22 (0.46)	0.09 (0.32)	0.14 (0.4)	8.32 4%
NAm	7.88 (2.01)	31.81 (4.25)	6.53 (1.7)	-	0.32 (0.55)	1.16 (0.91)	0.71 (0.77)	48.41 22%
AF	0.02 (0.14)	0.04 (0.2)	0.06 (0.24)	0.12 (0.33)	-	8.24 (1.88)	1.31 (1.13)	9.79 4%
EU	0.07 (0.26)	0.07 (0.26)	0.11 (0.31)	0.74 (0.71)	8.02 (1.52)	-	1.2 (0.88)	10.21 5%
OZ	0.05 (0.22)	0.04 (0.2)	0.04 (0.2)	0.12 (0.33)	0.93 (0.88)	0.56 (0.72)	-	1.74 1%
	19.87 9%	81.53 37%	28.37 13%	56.16 26%	12.82 6%	12.53 6%	6.27 3%	217.55 100%

**(c) Founder event counts (and standard deviations)**

	SAm	CAm	Car	NAm	AF	EU	OZ	
SAm	-	2.81 (1.71)	2.61 (1.26)	11.64 (2.51)	2.51 (1.25)	2.83 (1.12)	0.97 (0.74)	23.37 61%
CAm	0.65 (0.78)	-	1.01 (1.24)	1.03 (0.92)	0.05 (0.22)	0.07 (0.26)	0.02 (0.14)	2.83 7%
Car	0.06 (0.24)	0.42 (0.59)	-	0.15 (0.36)	0 (0)	0.02 (0.14)	0 (0)	0.65 2%
NAm	1.24 (1.07)	0.63 (0.84)	0.86 (0.51)	-	0.27 (0.47)	1.17 (0.79)	0.63 (0.54)	4.8 13%
AF	0.01 (0.12)	0 (0)	0.01 (0.13)	0.01 (0.11)	-	1.91 (0.89)	2.58 (0.91)	4.52 12%
EU	0.02 (0.14)	0.01 (0.12)	0 (0)	0.43 (0.54)	0.26 (0.65)	-	0.58 (0.54)	1.3 3%
OZ	0.01 (0.12)	0 (0)	0 (0)	0 (0)	0.46 (0.56)	0.32 (0.47)	-	0.79 2%
	1.99 5%	3.87 10%	4.49 12%	13.26 35%	3.55 9%	6.32 17%	4.78 12%	35.26 100%

**Figure 4** Number of dispersal events estimated in the history of Solanaceae with biogeographical stochastic mapping. Counts of dispersal events were averaged across the 100 BSMs and are presented here with standard deviations in parentheses. Total event counts are given in (a) and divided among the two types of dispersals in (b) and (c) (see Fig. 1 for depictions of range expansions and founder events). Colour temperature indicates the frequency of events; the warmer the colour, the more common the event. Note that given the standard deviations, values in green are often not different from zero. The ancestral states (where the lineage dispersed from) are given in the row, and the descendant states (where the lineage dispersed to) are given in the column. The sum and correspondent percentages of events involving each area, either as a source for dispersal (the rows) or as a sink (the columns), are given on the margins. Area names in rows and columns are South America (SAm), Central America (CAm), the Caribbean (Car), North America (NAm), Africa (AF), Eurasia (EU) and Australia (OZ).

also estimated to have originated in South America (Fig. 2), despite the fact that these groups today are very diverse in the Old World. Our best model, which included time-stratified matrices representing the continental area distances over time, presented a significantly better fit than those that did not incorporate palaeogeographical information. This

suggests that distance is an important factor when estimating dispersal events between areas in Solanaceae. Finally, Solanaceae presents a marked directionality of dispersals over its history, with range expansions and founder events being significantly more common from South America to other areas than in the opposite direction.



**Figure 5** Summary of dispersal events estimated with biogeographical stochastic mapping in the history of Solanaceae. Each of the seven areas in the analysis is shaded by its species richness (Table S1.4 in Appendix S1), with darker areas being more species rich. Some species are native to more than one area and thus contribute to species richness for more than one area. The arrows between areas represent direction and frequency of dispersal events. Only event counts that presented a mean of 0.95 or higher (Fig. 4a) are depicted as arrows here; arrow line thickness corresponds to natural log of the events counts.

### The ancestral range of Solanaceae

Our estimation of biogeographical history shows that, for Solanaceae, South America is not only the family's distributional centre but also its ancestral range (Fig. 2). The radiation of the family appears to have continued in South America up to *c.* 15 Ma, before the first lineages established on other continents (Fig. 2, Appendix S2). Such a clustering of long-distance dispersal events in the last 20 million years (Myr) was also observed in the Acanthaceae, a similarly species rich and largely Neotropical family (Tripp & McDade, 2014). Across the history of the Solanaceae, we estimate about 120 dispersals from South America to new areas, mostly to adjacent regions within the New World (Figs 4 & 5). This tendency for geographical movements to involve proximate areas is reflected in the fact that models that incorporated distance between areas through dispersal matrices were consistently a better fit for the data (Table S1.3 in Appendix S1), as has been observed in studies of other taxa (e.g. Matos-Maraví *et al.*, 2014).

Although short range movements account for most of the spread of Solanaceae from South America, our results provide evidence of multiple long-distance dispersals to Africa, Australia and Eurasia. These dispersals to the Old World occur long after the separation of Gondwana (*c.* 94 Ma) and therefore were likely to involve transoceanic movements. The Solanaceae have fruit types ranging from dry to fleshy and a variety of dispersal agents, including wind, water, and many animals such as birds, bats, small rodents and ants (Knapp, 2002). Thus, fruits could be blown in wind currents, float across the ocean or be carried to new regions by migrating or rafting animals. Our results show that dry fruited and, more commonly, fleshy fruited lineages have experienced range expansions, including transoceanic movements (Table S1.7 in Appendix S1), consistent with previous studies in the family (Olmstead, 2013). Transoceanic dispersal has

been hypothesized to explain the spread of other New World families to the Old World (Perret *et al.*, 2013; Tripp & McDade, 2014). Still, we cannot exclude the possibility that these South American taxa reached the Old World through a series of shorter dispersal events (e.g. via a northern route through Beringia) followed by subsequent extinction in the intervening areas (Davis *et al.*, 2002).

Our estimate of the number of dispersal events from the New World to the Old World is strikingly similar to previous studies, despite the use of different methods. In surveying the biogeography of Solanaceae, Olmstead (2013) suggested that a total of 15–17 long-distance dispersal events from South America would be needed to account for the present distribution, assuming a most parsimonious reconstruction. Our model-based stochastic mapping estimates  $20 \pm 2.04$  such events (summing all New World to Old World transitions, Fig. 3, Table S1.6 in Appendix S1). Both of these totals are minimum estimates of the actual number of long-distance dispersals in the history of extant Solanaceae because only *c.* 40% of all species were sampled. For example, the unsampled taxa include the monotypic genus *Tubocapsicum*, which is endemic to eastern temperate Asia and could represent an independent colonization from the New World. However, previous studies and overall morphology suggest that *Tubocapsicum* is closely related to the Hawaiian *Nothoestrum* and the African *Discopodium*, a placement that would favour an Old World origin as opposed to long-distance dispersal from the New World (Olmstead *et al.*, 2008). By assigning Hawaiian native species to Australia in our analysis, we were able to include the one inferred dispersal to Hawaii from the New World (*Lycium*) in our estimates. We recognize that all dispersals to Hawaii, however, represent long-distance dispersals regardless of source area. Although a complete family-level phylogeny would be needed to arrive at a final estimate of the total number of New World to Old World dispersals,

we expect this study captures the vast majority of these events.

### Patterns of speciation, vicariance and dispersal in Solanaceae history

Our simulations of biogeographical history using BSM identified within-area speciation as the most frequent type of event across the phylogeny (Table 1). This result is likely to reflect the scale of this analysis (global) and the size of the regions under consideration. The preponderance of within-area speciation events is consistent with the large clades of Solanaceae that are endemic to single areas as defined by our study. For example, the 31 species of Anthocercidae are restricted to Australia and New Caledonia and are inferred to have diversified entirely within Australia (Appendix S2). Our study shows the importance of events at a global scale, but we acknowledge the need for future studies with a finer division of the regions, especially of South America, given its importance in the Solanaceae history. Such a division would likely reveal that many of these speciation events are actually associated with shifts in geographical range not revealed in our continental-scale analysis (see also Sanmartín & Ronquist, 2004).

Among the remaining types of events, we found that dispersal was the principal driver of range evolution, occurring about six times more often than vicariance. While vicariant events appear to have been important for range evolution in many animal groups (e.g. Giribet *et al.*, 2012), dispersal seems to be the most common factor shaping the distribution of plant clades, even those whose origins date to Pangea or Gondwana (Sanmartín & Ronquist, 2004). In the case of the Solanaceae, the crown age for the family is roughly 60 Myr after the separation of South America and Africa, making Gondwanan vicariance a less likely explanation for the family's pantropical distribution. Indeed, we inferred multiple dispersal events between South America and the Old World (Figs 4 & 5), and these are confined to the last 15 Myr of the Solanaceae evolution (Fig. 3).

Dispersal events between both proximate and distant regions appear to have been frequent in the history of the Solanaceae, and they may have been facilitated by colonization of similar niches. For example, the genus *Lycium*, distributed across all of the recognized regions except for Central America (Appendix S2, Table S1.2 in Appendix S1), is restricted to dry habitats (Levin & Miller, 2005). Moreover, all of the species inferred to show recent range expansions along their terminal branch (Fig. 3) are found in a single type of terrestrial ecoregion (e.g. dry, tropical or temperate), with the vast majority (80%) from wet tropical areas (Table S1.7 in Appendix S1). This pattern of niche conservatism has been documented in many plant groups (e.g. Francisco-Ortega *et al.*, 2001; Martínez-Meyer & Peterson, 2006), leading to the idea that, at least in some clades, it is easier to move than to evolve (Edwards & Donoghue, 2013). In Solanaceae, additional studies will be needed to assess the

extent of niche conservatism and to test the specific factors, biotic or abiotic, that affect colonization of new areas. Certainly, despite the tendency for clades of Solanaceae to be restricted to particular environments (Olmstead, 2013), the family as a whole is found in a wide range of habitats, from the driest deserts to wet forests in both tropical and temperate zones, indicating that major niche shifts must have occurred during its history.

### Directionality of dispersal events

One of the most striking patterns to emerge from our statistically robust estimation of biogeographical history was the strong asymmetry of dispersals. For every pair of areas considered, transitions were consistently higher in one of the two directions. The directionality was most notable for range expansion events, where for example, dispersals from South America to North America were over four times more common than dispersals in the opposite direction. Similar patterns of dispersal asymmetry have been observed in other studies at deep and recent time-scales (Sanmartín *et al.*, 2001, 2007). For example, Bacon *et al.* (2015) found the overall migration rates for animals and plants from South to North America during the last 6–7 Myr was *c.* 30% higher than in the reverse direction.

In the case of Solanaceae, the strong directionality in transitions from South America to other regions may be attributable to the age of the South American lineages and their species richness. As our ancestral range estimations show, the origin and early diversification of the family took place in South America, making this region the principal source for migrants for most of the past 50 Myr (Fig. 2, Appendix S2). In our dataset, South American species comprise roughly half of the taxa (Table S1.4) and also account for roughly half of the dispersal events (Fig. 4a). We observed a similarly close relationship between the species richness of an area and the number of dispersals out of that area for the all of remaining regions, with the exception of Australia (home to 13% of taxa but only the source for 1% of dispersals, Table S1.4). Although more extensive analyses would be required to formally test the relationship between species richness and dispersal, this congruence suggests that extrinsic explanations, such as directional wind or water currents or migratory paths of dispersers (Renner, 2004; Sanmartín *et al.*, 2007; Gillespie *et al.*, 2012), may not be needed to explain the apparent biases in inter-area movements.

### BSM as an approach to estimate biogeographical history of clades

As demonstrated by this study, the application of stochastic mapping to the estimation of biogeographical history gives researchers new power to address questions about patterns of range evolution. Whereas parsimony and likelihood reconstruction provide estimates of ancestral ranges at nodes, stochastic mapping also gives possible histories along

branches and at nodes, including any of the events incorporated in the model. By summarizing over many histories, we can extract distributions for the numbers of the events and determine the relative importance of each type of event (e.g. range expansions, vicariance) in shaping present-day distributions. Inferring the numbers and types of biogeographical events can also be achieved in a parsimony framework (e.g. Sanmartín & Ronquist, 2004); however, this approach requires user-defined costs for different events and an inferred area cladogram. Recently developed Bayesian MCMC approaches (Höhna *et al.*, 2014, 2016) could also be used for estimating distributions of event counts and timing of events and have the benefit of directly incorporating uncertainty in the rate parameters. While implemented models are currently limited to DEC, additional models and tools are likely to be added in the near future (Matzke, pers. comm.).

Although our study used BSM primarily to examine the frequency and directionality of events, this approach could be used to explore other questions, such as how the prevalence of different events (e.g. range expansions or vicariance) varies across clades. Moreover, subsequent studies targeting particular clades could employ a more fine-scale division of geographical areas, revealing dynamics that may not be apparent in this broad family-level analysis. We also envision that future work in biogeography will lead to the extension of the models with additional types of events, and as long as the models are developed in a probabilistic framework, the frequency of these events can be estimated with BSM. Applying these new methods to other plant families of New World origin would provide a powerful test of the generality of the patterns of speciation, dispersal and vicariance that we have inferred for the Solanaceae.

## ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation grants to S.D.S. (NSF DEB-1413855), S.K. and L.B. (NSF DEB-0316614), and R.G.O. (NSF DEB-1020369). The authors acknowledge the Research Computing high-performance facilities at the University of Colorado Boulder for providing computational assistance. The Solanaceae phylogeny we used was done with support of NSF-DEB-0316614 to L.B. and S.K. N.J.M. and 'BioGeoBEARS' were supported by NSF DEB-0919124, the National Institute for Mathematical and Biological Synthesis sponsored by NSF Award EFJ-0832858, with additional support from the University of Tennessee, Knoxville. N.J.M. was also funded by the Discovery Early Career Researcher Award DE150101773, the Australian Research Council and the Australian National University.

## REFERENCES

- Bacon, C.D., Mora, A., Wagner, W.L. & Jaramillo, C.A. (2013) Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical Journal of the Linnean Society*, **171**, 287–300.
- Bacon, C.D., Silvestro, D., Jaramillo, C., Smith, B.T., Chakrabarty, P. & Antonelli, A. (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences USA*, **112**, 6110–6115.
- Barker, N.P., Weston, P.H., Rutschmann, F. & Sauquet, H. (2007) Molecular dating of the “Gondwanan” plant family Proteaceae is only partially congruent with the timing of the break-up of Gondwana. *Journal of Biogeography*, **34**, 2012–2027.
- Beaulieu, J.M., Tank, D.C. & Donoghue, M.J. (2013) A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evolutionary Biology*, **13**, 80.
- Benítez de Rojas, C. & D’Arcy, W.G. (1997) The Genus *Lycianthes* (Solanaceae) in Venezuela. *Annals of the Missouri Botanical Garden*, **84**, 167–200.
- Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N. & Sanmartín, I. (2011) An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography*, **38**, 531–550.
- Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, USA.
- Cain, M.L., Milligan, B.G. & Strand, A.E. (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.
- Christenhusz, M.J. & Chase, M.W. (2013) Biogeographical patterns of plants in the Neotropics—dispersal rather than plate tectonics is most explanatory. *Botanical Journal of the Linnean Society*, **171**, 277–286.
- Clayton, J.W., Soltis, P.S. & Soltis, D.E. (2009) Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical Angiosperm family (Simaroubaceae, Sapindales). *Systematic Biology*, **58**, 395–410.
- Davis, C.C., Bell, C.D., Mathews, S. & Donoghue, M.J. (2002) Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences USA*, **99**, 6833–6837.
- Dillon, M.O., Tu, T., Xie, L., Quipuscoa Silvestre, V. & Wen, J. (2009) Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian Deserts along the western coast of South America. *Journal of Systematics and Evolution*, **47**, 457–476.
- Edwards, E.J. & Donoghue, M.J. (2013) Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany*, **64**, 4047–4052.
- Francisco-Ortega, J., Barber, J.C., Santos-Guerra, A., Febles-Hernandez, R. & Jansen, R.K. (2001) Origin and evolution of the endemic genera of Gonosperminae (Asteraceae: Anthemideae) from the Canary Islands: evidence from nucleotide sequences of the internal transcribed spacers of

- the nuclear ribosomal DNA. *American Journal of Botany*, **88**, 161–169.
- Gallaher, T., Callmander, M.W., Buerki, S. & Keeley, S.C. (2015) A long distance dispersal hypothesis for the Pandanaceae and the origins of the *Pandanus tectorius* complex. *Molecular Phylogenetics and Evolution*, **83**, 20–32.
- Gamble, T., Bauer, A.M., Greenbaum, E. & Jackman, T.R. (2008) Out of the blue: a novel, trans-Atlantic clade of geckos (Gekkota, Squamata). *Zoologica Scripta*, **37**, 355–366.
- Garcia, V.F. & Olmstead, R.G. (2003) Phylogenetics of tribe Anthocercideae (Solanaceae) based on *ndhF* and *trnL/F* sequence data. *Systematic Botany*, **28**, 609–615.
- Gentry, J.L. & D'Arcy, W.G. (1986) *Solanaceae of Mesoamerica*, pp. 15–26. *Solanaceae Biology and Systematics*. Columbia University Press. New York, USA.
- Gillespie, R.G., Baldwin, B.G., Waters, J.M., Fraser, C.I., Nikula, R. & Roderick, G.K. (2012) Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology and Evolution*, **27**, 47–56.
- Giribet, G., Sharma, P.P., Benavides, L.R., Boyer, S.L., Clouse, R.M., De Bivort, B.L., Dimitrov, D., Kawachi, G.Y., Muriene, J. & Schwendinger, P.J. (2012) Evolutionary and biogeographical history of an ancient and global group of arachnids (Arachnida: Opiliones: Cyphophthalmi) with a new taxonomic arrangement. *Biological Journal of the Linnean Society*, **105**, 92–130.
- Givnish, T.J. & Renner, S.S. (2004) Tropical intercontinental disjunctions: Gondwana breakup, immigration from the Boreotropics, and transoceanic dispersal. *International Journal of Plant Sciences*, **165**, S1–S6.
- Givnish, T.J., Millam, K.C., Evans, T.M., Hall, J.C., Pires, J.C., Berry, P.E. & Sytsma, K.J. (2004) Ancient vicariance or recent long-distance dispersal? Inferences about phylogeny and South American-African disjunctions in Rapateaceae and Bromeliaceae based on *ndhF* sequence data. *International Journal of Plant Sciences*, **165**, S35–S54.
- Hepper, F. & Jaeger, P. (1986) *A review of the genus Solanum in Africa*, pp. 41–55. *Solanaceae Biology and Systematics*. Columbia University Press. New York, USA.
- Höhna, S., Heath, T.A., Boussau, B., Landis, M.J., Ronquist, F. & Huelsenbeck, J.P. (2014) Probabilistic graphical model representation in phylogenetics. *Systematic Biology*, **63**, 753–771.
- Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P. & Ronquist, F. (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, **65**, 726–736.
- Howe, H.F. & Miriti, M.N. (2004) When seed dispersal matters. *BioScience*, **54**, 651–660.
- Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. (2003) Stochastic mapping of morphological characters. *Systematic Biology*, **52**, 131–158.
- Itkin, M., Heinig, U., Tzfadia, O. *et al.* (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, **341**, 175–179.
- Janzen, D.H. (1970) Herbivores and the number of tree species in tropical forests. *The American Naturalist*, **104**, 501–528.
- Knapp, S. (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *Journal of Experimental Botany*, **53**, 2001–2022.
- Landis, M.J., Matzke, N.J., Moore, B.R. & Huelsenbeck, J.P. (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, **62**, 789–804.
- Levin, R.A. & Miller, J.S. (2005) Relationships within tribe Lycieae (Solanaceae): paraphyly of *Lycium* and multiple origins of gender dimorphism. *American Journal of Botany*, **92**, 2044–2053.
- Levin, R.A., Myers, N.R. & Bohs, L. (2006) Phylogenetic relationships among the “spiny solanums” (*Solanum* subgenus *Leptostemonum*, Solanaceae). *American Journal of Botany*, **93**, 157–169.
- Litsios, G., Pearman, P.B., Lanterbecq, D., Tolou, N. & Salamin, N. (2014) The radiation of the clownfishes has two geographical replicates. *Journal of Biogeography*, **41**, 2140–2149.
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L. & Hernández-Hernández, T. (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, **207**, 437–453.
- Mao, K., Milne, R.I., Zhang, L., Peng, Y., Liu, J., Thomas, P., Mill, R.R. & Renner, S. (2012) Distribution of living Cupressaceae reflects the breakup of Pangea. *Proceedings of the National Academy of Sciences USA*, **109**, 7793–7798.
- Martínez-Meyer, E. & Peterson, A.T. (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography*, **33**, 1779–1789.
- Matos-Maraví, P., Águila, R.N., Peña, C., Miller, J.Y., Sourakov, A. & Wahlberg, N. (2014) Causes of endemic radiation in the Caribbean: evidence from the historical biogeography and diversification of the butterfly genus *Calisto* (Nymphalidae: Satyrinae: Satyrini). *BMC Evolutionary Biology*, **14**, 199.
- Matzke, N.J. (2013) BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2.1. Available at: <http://CRAN.R-project.org/package=BioGeoBEARS> (accessed 27 July 2013).
- 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.
- Matzke, N.J. (2015) Stochastic mapping under biogeographical models. Available at: [http://phylo.wikidot.com/biogeobears#stochastic\\_mapping](http://phylo.wikidot.com/biogeobears#stochastic_mapping) (accessed 1 June 2015).
- McLoughlin, S. (2001) The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, **49**, 271–300.

- Mione, T., Olmstead, R.C., Jansen, R.K. & Anderson, G.J. (1994) Systematic implications of chloroplast DNA variation in *Jaltomata* and selected Physaloid Genera (Solanaceae). *American Journal of Botany*, **81**, 912.
- Montes, C., Cardona, A., McFadden, R., Morón, S., Silva, C., Restrepo-Moreno, S., Ramírez, D., Hoyos, N., Wilson, J., Farris, D., Bayona, G.A., Jaramillo, C.A., Valencia, V., Bryan, J. & Flores, J.A. (2012) Evidence for middle Eocene and younger land emergence in central Panama: implications for Isthmus closure. *Geological Society of America Bulletin*, **124**, 780–799.
- Muñoz, J., Felicísimo, Á.M., Cabezas, F., Burgaz, A.R. & Martínez, I. (2004) Wind as a long-distance dispersal vehicle in the Southern Hemisphere. *Science*, **304**, 1144–1147.
- Nathan, R. (2006) Long-distance dispersal of plants. *Science*, **313**, 786–788.
- Naueheimer, L., Metzler, D. & Renner, S.S. (2012) Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytologist*, **195**, 938–950.
- Nielsen, R. (2002) Mapping mutations on phylogenies. *Systematic Biology*, **51**, 729–739.
- Nogales, M., Heleno, R., Traveset, A. & Vargas, P. (2012) Evidence for overlooked mechanisms of long-distance seed dispersal to and between oceanic islands. *New Phytologist*, **194**, 313–317.
- O’Dea, A., Lessios, H.A., Coates, A.G. *et al.* (2016) Formation of the Isthmus of Panama. *Science Advances*, **2**, e1600883.
- Olmstead, R.G. (2013) Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society*, **171**, 80–102.
- Olmstead, R.G. & Sweere, J.A. (1994) Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology*, **43**, 467–481.
- Olmstead, R.G., Bohs, L., Migid, H.A., Santiago-Valentin, E., Garcia, V.F. & Collier, S.M. (2008) A molecular phylogeny of the Solanaceae. *Taxon*, **57**, 1159–1181.
- Pabón-Mora, N. & Litt, A. (2011) Comparative anatomical and developmental analysis of dry and fleshy fruits of Solanaceae. *American Journal of Botany*, **98**, 1415–1436.
- Pagel, M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology*, **48**, 612–622.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Perret, M., Chautems, A., De Araujo, A.O. & Salamin, N. (2013) Temporal and spatial origin of Gesneriaceae in the New World inferred from plastid DNA sequences. *Botanical Journal of the Linnean Society*, **171**, 61–79.
- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. Available at: <http://www.R-project.org/>. Last accessed on 2 November 2015
- Rasmussen, C. & Cameron, S.A. (2010) Global stingless bee phylogeny supports ancient divergence, vicariance, and long distance dispersal. *Biological Journal of the Linnean Society*, **99**, 206–232.
- Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.
- Renner, S. (2004) Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences*, **165**, S23–S33.
- Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195–203.
- Sanmartín, I. & Ronquist, F. (2004) Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*, **53**, 216–243.
- Sanmartín, I., Enghoff, H. & Ronquist, F. (2001) Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society*, **73**, 345–390.
- Sanmartín, I., Wanntorp, L. & Winkworth, R.C. (2007) West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. *Journal of Biogeography*, **34**, 398–416.
- Särkinen, T., Bohs, L., Olmstead, R.G. & Knapp, S. (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology*, **13**, 214.
- Sato, S., Tabata, S., Hirakawa, H. *et al.* (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.
- Sessa, E.B., Zimmer, E.A. & Givnish, T.J. (2012) Phylogeny, divergence times, and historical biogeography of New World *Dryopteris* (Dryopteridaceae). *American Journal of Botany*, **99**, 730–750.
- Szövényi, P., Terracciano, S., Ricca, M., Giordano, S. & Shaw, A.J. (2008) Recent divergence, intercontinental dispersal and shared polymorphism are shaping the genetic structure of amphi-Atlantic peatmoss populations. *Molecular Ecology*, **17**, 5364–5377.
- Thacker, C.E. (2015) Biogeography of goby lineages (Gobiiformes: Gobioidi): origin, invasions and extinction throughout the Cenozoic. *Journal of Biogeography*, **42**, 1615–1625.
- Tripp, E.A. & McDade, L.A. (2014) A rich fossil record yields calibrated phylogeny for Acanthaceae (Lamiales) and evidence for marked biases in timing and directionality of intercontinental disjunctions. *Systematic Biology*, **63**, 660–684.
- Voelker, G., Peñalba, J.V., Huntley, J.W. & Bowie, R.C.K. (2014) Diversification in an Afro-Asian songbird clade (*Erythropgia–Copsychus*) reveals founder-event speciation

via trans-oceanic dispersals and a southern to northern colonization pattern in Africa. *Molecular Phylogenetics and Evolution*, **73**, 97–105.

- Vorontsova, M., Stern, S., Bohs, L. & Knapp, S. (2013) African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. *Botanical Journal of the Linnean Society*, **173**, 176–193.
- 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.
- Yu, Y., Harris, A.J., Blair, C. & He, X. (2015) RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**, 46–49.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Supplementary tables and algorithm validation.

**Appendix S2** Maximum likelihood ancestral range estimates.

**Appendix S3** Example of a BSM realization.

### DATA ACCESSIBILITY

The phylogeny used for these analyses is available through Dryad Digital Repository <http://dx.doi.org/10.5061/dryad.6gd57>, and the Natural History Museum Data Portal, <http://data.nhm.ac.uk/>.

### BIOSKETCH

**Julia Dupin** is a PhD student at the Department of Ecology and Evolutionary Biology at University of Colorado Boulder. She is interested in understanding patterns in plant biogeography and evolution, and in particular, how the interplay between historical biogeographical events and environmental factors shape species distributions and traits in plants.

Author contributions: J.D. and S.D.S conceived the study; J.D., T.S., S.K., R.G.O., L.B. gathered the data; N.J.M. developed the method; J.D. and S.D.S led the analyses with fundamental contributions from N.J.M.; J.D. and S.D.S led the writing with revisions from N.J.M., T.S., S.K., R.G.O. and L.B. All co-authors read and approved the article.

---

Editor: Isabel Sanmartín