Dynamism and context-dependency in the diversification of megadiverse plant groups

Susy Echeverría-Londoño[1,2,3*], Tiina Särkinen[4], Isabel S Fenton[1], Sandra Knapp[1], and Andy Purvis[1,2]

[1]Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK

[2]Department of Life Sciences, Imperial College London, Ascot SL5 7QN, UK

[3]Department of Biology, Kenyon College, Gambier, Ohio 43022, USA

[4]Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK

[*]echeverrialondono1@kenyon.edu

Abstract

Explosive radiations have been considered one of the most important diversification patterns across the Tree of Life, but the subsequent change, movement and extinction of species makes radiations hard to discern and understand as geological time passes. Here we synthesise phylogenetic and distributional data for an ongoing radiation — the mega-diverse plant genus *Solanum* L. — to show how dispersal events and past climatic changes have interacted to shape diversification. We find that despite the vast diversity of *Solanum* lineages in the Neotropics, lineages in the Old World are diversifying more rapidly. This recent explosive diversification coincides with a long-distance dispersal event from the Neotropics, at the time when, and to places where, major climatic changes took place. Two different groups of *Solanum* have migrated and established in Australia, but only the arid-adapted lineages experienced significant increases in their diversification, which is consistent with the continent's long-term climatic trend and the diversification of other arid-adapted groups. Our findings provide a clear example of how successful colonisation of new areas and niches can – but do not always – drive explosive diversifications.

Results and discussion

The uneven distribution of taxonomic diversity among the branches of the tree of life and geographic regions is one of the most intriguing puzzles in biology. The increasing availability of field studies and large phylogenies continues to reveal how some groups are highly diverse compared with their depauperate sister groups. We expect the stochastic nature of the diversification process to produce asymmetric patterns of diversity which do not necessarily

require deterministic explanations (1; 2; 3). Nevertheless, identifying significant patterns of unequal diversity (i.e., those that depart from the null expectation) throughout the tree of life has the potential to reveal mechanistic drivers that have shaped current patterns of biodiversity.

One of the main hypotheses in macroevolution is that unequal diversity can arise from significant differences in diversification rates, both among lineages and through time and space (4; 5; 6; 7). This asymmetry in evolutionary rates is often interpreted as a result of a spectrum of ecological opportunities, such as the opening of niche space by the development of new key traits, extinction of antagonists, and/or colonisation of new open areas (8; 9; 10; 11; 12; 13).

Understanding the evolutionary processes that have promoted the diversification of radiations in older clades is likely to be challenging, due to the difficulty of accounting for extinct species usually not included in molecular phylogenies (14; 15; 16). Furthermore, poor species-level sampling and branch support in the underlying phylogenies used for testing these hypotheses usually limits the power of tests of diversification dynamics. A focus on recently emerged clades that are species-rich, widespread and are underpinned by well-sampled and robust phylogenies will provide us with a more accurate picture of how current patterns of biodiversity have been shaped.

Solanaceae represents one of the most diverse and economically important plant groups with around 3,000 species distributed worldwide. Almost half of the diversity of the family belongs to the mega-diverse genus *Solanum* L. that includes species from all major tropical and subtropical biomes and shows high heterogeneity of richness among its subclades. *Solanum* is traditionally divided into two major groups, the spiny and non-spiny solanums. With robust monophyletic support, the spiny solanums (known as the Leptostemonum clade, or subgenus *Leptostemonum Bitter*) include ca. 420 species and can be recognised by the presence of stellate trichomes, prickles, and long tapering anthers (17). Spiny solanums are most diverse in the Neotropics (ca. 150 species), Australia (ca. 130), and Africa (ca. 79). This high diversity of spiny solanums out of the Americas contrasts with the distribution of the diversity of the non-monophyletic group — the non-spiny solanums — which has more than 90% of its diversity distributed in the Neotropics.

As one of top ten species-rich genera of angiosperms (18), *Solanum* is suggested to have high speciation and diversification rates, especially in the Neotropics which supports most of its diversity and endemism (19; 20; 21). Furthermore, the high heterogeneity of richness observed among the subclades of *Solanum* suggests that clade-specific and tree-wide variation in evolutionary rates could contribute to its diversity dynamics. Where radiating lineages occur and how much they contribute to current diversity and geographic differences in richness among the main two groups of *Solanum* has not yet been explored.

We test whether high heterogeneity of species richness observed among subclades is the result of clade-specific,

tree-wide and/or geographic variation in evolutionary rates using a set of nearly complete phylogenies of extant *Solanum* species (see Methods). The data set was compiled using a combination of molecular data and taxonomic constraints for data-deficient species to account for incomplete taxon sampling under a Bayesian framework (22). Based on a variety of probabilistic methods (likelihood and Bayesian) to infer branch-specific differences in diversification rates along these phylogenies, we infer significant heterogeneity of rates in diversification along the branches of the *Solanum* phylogenies. This variation in evolutionary rates has contributed to the distinguished disparities in species richness among groups, of which the Old World spiny clade and the Petota subclade are supported as the most rapid radiations (Figure 1).

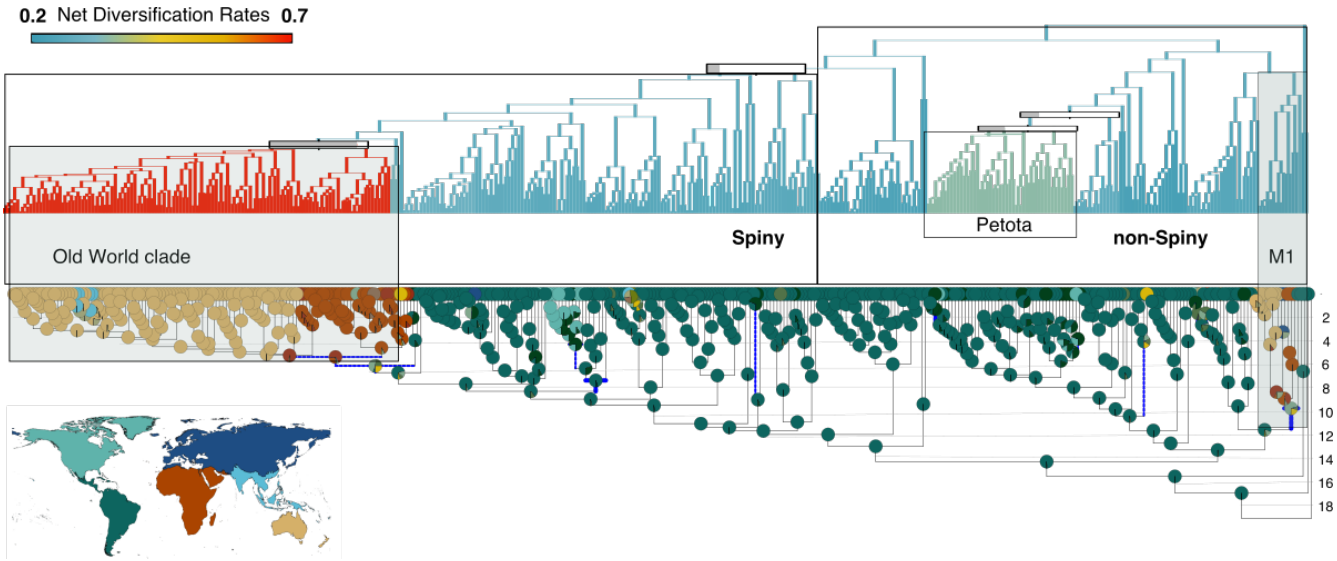

Figure 1. Mean net diversification rates and historical biogeographic reconstruction across the phylogeny of *Solanum*. Thermometers represent the frequency in which nodes are associated with shifts in diversification rates across a distribution of 100 nearly-complete phylogenies of *Solanum* (see Methods for further detail). Pies at each node represent the probability that each region (or the combination of regions) is, according to the model, the ancestral range distribution. The branches highlighted in blue represent the dispersals events inferred in at least the 50% of Biogeographic Stochastic Maps (BSM). The timescale is given in Ma.

The Old World spiny clade is the most diverse monophyletic group within the spiny solanums, with ca. 250 species, and is the most rapid ongoing radiation of extant species of *Solanum*, diversifying nearly as twice as rapidly as any other group, with approximately 0.7 species per million years as compared to the global mean of 0.25 species per million years (Figure 1). The clade is not defined by any single distinctive morphological trait or combination of morphological characters (23) but is usually defined by geography alone (23; 24; 25; 26). Nearly all species of spiny solanums native to the Old World (i.e., Africa, Australia, Asia and the Pacific) belong to this group with unusual rates of diversification, although a few isolated taxa are members of otherwise New World spiny solanum clades (e.g., Lasiocarpa, Acanthophora, and Torva; see 27; 28). Species of these spiny *Solanum* clades are considered

recent introductions via long-distance dispersal with or without human intervention (24; 28; 29; 30).

It has been suggested (31) that a colonisation event from the Neotropics to the Old World accounts for the distribution of the Old World spiny solanums. We confirm this hypothesis by inferring the ancestral ranges of distribution of *Solanum*, under different models of biogeographic history on its phylogeny (32; 33). Here, we find an association between a long-dispersal event from the Neotropics to the Old World ca. 6 Mya (95% HPD 5-7) and the recent rapid radiation of lineages of the Old World spiny clade. This scenario depicts a classic example of "dispersification" sensu (11), where dispersal to a new area promotes a change in diversification either by exploiting newly available resources or through ecological release due to the lack of competitors.

Only some of the clades that historically dispersed from the Neotropics to the Old World resulted in increases in their diversification rates. For example, the first early branching long-dispersal event from the Neotropics to the Old World occurred ca. 10 Mya (95% HPD 7-12) in the crown group of the M1 clade of non-spiny solanums (*Solanum valdiviense* + African-non-spiny + Normania + Archaesolanum), but does not show increased rates (Figure 1). Unlike the spiny solanums, the diversity of non-spiny solanums in the Old World is rather limited — almost ten times less diverse — which translates in significant lower diversification rates than those observed in the Old World spiny solanums —- 0.25 species per million years and 0.7 species per million years, respectively.

Differences in the timing of the dispersal events between non-spiny and spiny solanums and their different diversity patterns suggest that distinct ecological opportunities and evolutionary regimes have shaped the diversity of these groups in the Old World. To understand and identify the drivers that model these different patterns of diversification, we reconstructed the present-day geographic variation in net diversification rates in *Solanum*, estimated as the mean assemblage diversification rates (i.e., the geometric mean of all species' net diversification rates present in a grid cell, Figure 2). Results illustrate a marked heterogeneity of diversification rates across the globe with 'hot' regions of recent rapid diversification of *Solanum* within the Old World, especially in Africa and Australia (Figure 2B and C). Despite most of the richness and endemism of *Solanum* occurring in the Neotropics (Figure 2A), this region is not directly supported by higher diversification rates compared with the Old World (Figure 2B and C).

This mismatch between the high diversity of *Solanum* within the Neotropics and its net diversification rates could be explained by the early and long history of diversification of *Solanum* in this region. Although *Solanum* lineages in the Old World appear to have evolved more rapidly, *Solanum* lineages within the Neotropics could have had more time and opportunity to accumulate before establishing in other regions, creating the current imbalance of diversity (34). The Neotropics is the region where *Solanum* species experienced the greatest expansion in their

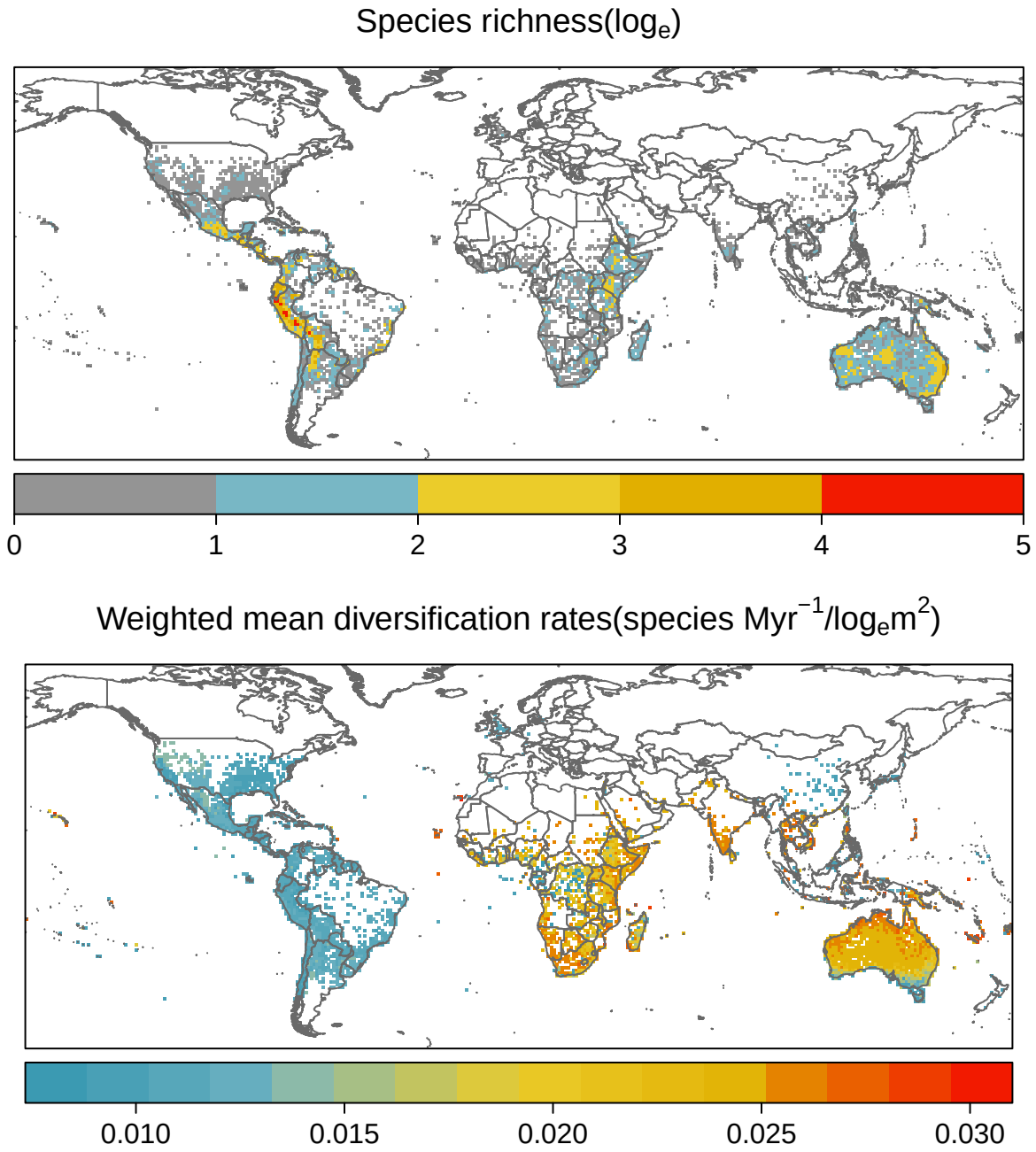

Figure 2. Global distribution of species richness and the net diversification rates of *Solanum* using 1 x 1 degrees grid cells. Weighted mean lineage net diversification rates was calculated as the mean net diversification rate for all the species present in a grid cell assemblage, weighted by the inverse of their range size, $log_e m^2$

ranges, which has also been demonstrated as the highest within the family Solanaceae as a whole (35). The long history of diversification and expansion of *Solanum* within the Neotropics has driven a greater lineage accumulation through time compared with other regions (Figure 3). This region has not only supported the early evolution of *Solanum* but also represents the main source of diversity and origin of biogeographic movements across the globe

(Figure 2). The pattern of highest richness in the Neotropics could therefore be explained by the greater age of *Solanum* species in this region (36).

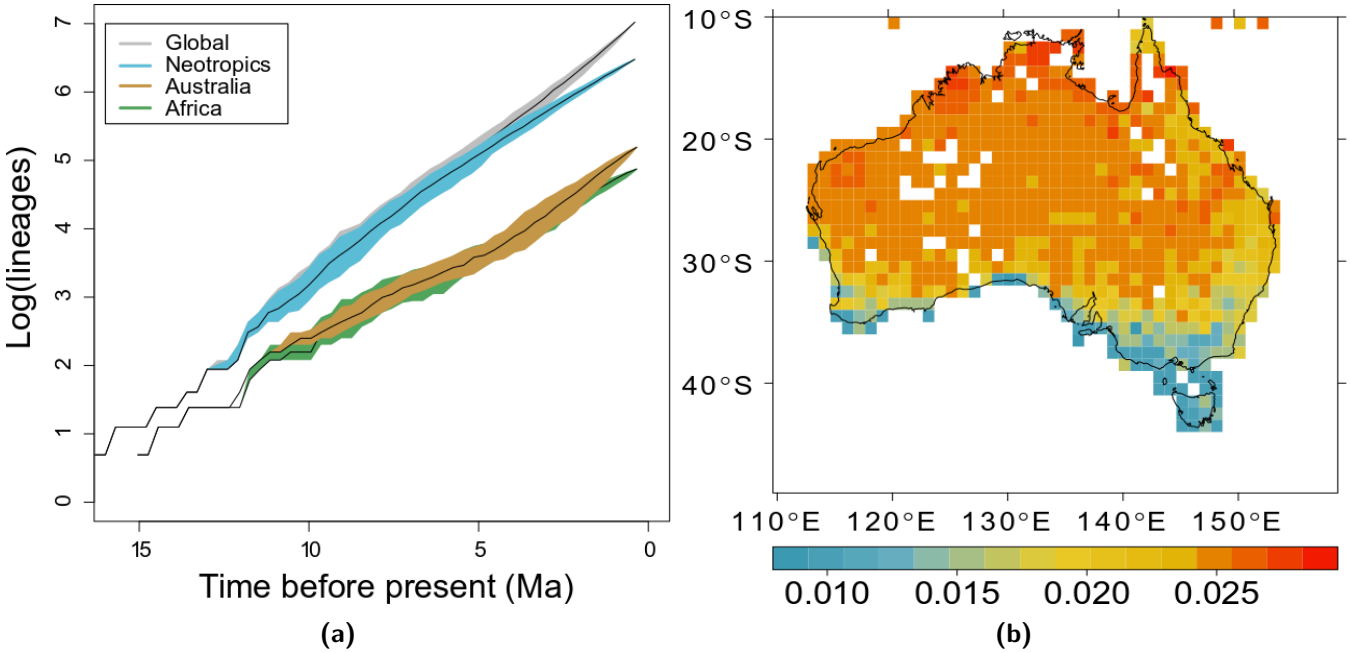

Figure 3. (a) Lineages-through-time plot of *Solanum* species for the global and three regional diversities (Neotropics, Australia, and Africa). (b) Distribution of the net diversification rates of *Solanum* in Australia (same units as in Figure 2). Coloured areas around the median curves shown in black represent two-tailed 95% lower and upper quantiles of the number of lineages across a distribution of 100 nearly-complete phylogenies of *Solanum* (see Methods for further details)

Comparison of temporal patterns of lineage accumulation among major regions shows a distinctive signature of a rapid diversification within Australia ca. 5 Mya (Figure 3a). The timing of diversification and dispersion of spiny solanums in the Old World, as well as their concentration in arid and semi-arid regions of Australia (Figure 3b), suggest a potential link between past environmental changes and radiation of this group. Similar signature has been shown in a wide range of groups from Australia (37; 38; 39; 40; 41; 42; 43; 44; 45; 46; 47). The dramatic climate change in Australia since the breakup of Gondwana is likely to have triggered the radiation of several taxa (46) and the extinction of others (47). The initial stage of aridification in the mid-Miocene (ca. 15 Mya) and the late expansion of desertic regions of Australia in the last 1-4 Mya, has led to the assembly of new biomes, such as the arid, monsoonal and alpine regions, encouraging the establishment and expansion of arid-adapted lineages (e.g., Tremandraceae 41, and see also 37; 38; 45; 46). This pattern and the congruence of the timing of diversification with other arid-adapted Australian groups could indicate that the Old World spiny solanums established and diversified in Australia as the arid environments expanded.

The widespread distribution of Old World spiny solanums in Australia (Figure 3) suggests a positive effect of

distribution area on diversification rates. The large expanse of arid habitat in Australia may have created opportunities for allopatric speciation within the biome itself (45; 48; 49). In contrast, non-spiny solanums found in the temperate and forested areas of the South West Pacific (e.g., Australia, Tasmania, New Zealand and Papua New Guinea) are species-poor with only eight extant species (50). The successful expansion and diversification of spiny solanums in warm and arid zones in Australia, in contrast with the slower diversification of the more mesic non-spiny clades, could indicate the influence of adaptive features to warm and dry environments on species diversification. The sister group of the diverse Old World clade of spiny solanum is the species-poor Elaeagnifolium clade, with five species confined to semi-arid regions of North and South America (21). This suggests some pre-adaptation to dry conditions may have driven diversification in the Old World.

In Africa, species of Old World spiny solanums are usually found in open areas with high temperature and aridity (17; 21), suggesting that the spiny solanums lineages that successfully colonised Australia from Africa were already pre-adapted to arid conditions. Therefore, we propose that the dispersal of spiny solanums to the Old World, the influence of past environmental changes, as well as the pre-adaptive condition of spiny solanums to colonise warm and arid regions, all contributed to the distinctive diversity patterns of *Solanum* that we see currently on the globe.

Implications of the study

The dynamics shaping the diversity of species-rich mega-diverse genera may be more complex than previously thought. A variety of mutually reinforcing drivers appear to have formed current patterns of *Solanum* diversity, which cautions against single-cause hypotheses, such as the influence of "key innovations" often invoked in macroevolutionary studies (51). These findings also raise important questions about the reliability of global studies that assume a single model for the evolution of character change and diversification across large phylogenies (e.g., (52)), or studies that use average values of diversity, trait distribution, or geographical distribution at generic level (e.g., (53)). For instance, (54) and (55) have demonstrated that heterogeneity of diversification rates can have significant consequences for the model selection of trait change that best fits comparative data. In general, therefore, the understanding of the magnitude and location of shifts in diversification rates should be included as prior information in macroevolutionary analyses, especially in analyses of trait evolution (55; 56), or at least models of trait evolution must not implicitly assume that such shifts do not occur.

The results of this study also highlight the importance of context-dependency in the study of diversification dynamics (51; 57; 58). Our study shows that both opportunity and success in geographic movement, as well as correct environmental setting and climatic adaptations, were necessary to spark faster diversification in the spiny clade of *Solanum* in the Old World. Thus, focusing on effects of the combination of events and attributes on changes

in diversification could represent a more productive framework in which to understand diversity processes (51). This combination could also explain that, although many correlates of diversification rates have been reported in the literature, few have demonstrated either generality or high explanatory power. Our results suggest the necessity of considering more complex analyses that involve the integration of phylogenies with other biological and Earth system data sources such as geography, climate, historical biogeography, and physiology.

Conclusions

We demonstrate that *Solanum* is a remarkable study system for understanding the diversification of plants, not only because it contains a recent and ongoing radiation, but also because its diversity is supported by several clades with different and sometimes contrasting evolutionary dynamics. There are still many drivers that need to be explored before we have the whole picture of the diversification of *Solanum*, especially correlations with biogeographic events and morphological traits, but increased taxonomic, phylogenetic, and morphological understanding of *Solanum* now means that we are beginning to uncover many aspects of the macroevolution of this mega-diverse plant genus, and can use it as a model for studying diversity in widespread, species-rich groups in general.

Methods

Phylogenetic relationships and divergence times among the main clades of *Solanum* were obtained from the (20) Solanaceae phylogeny. This tree was built using two nuclear and six plastid loci from 454 *Solanum* species (ca. 34% of the expected number). This phylogeny identifies a total of 37 clades and subclades within the genus, with a vast heterogeneity of species richness within these.

Correcting for non-random taxon sampling

The incomplete nature of the (20) phylogeny could be problematic. Non-random incomplete sampling in phylogenetic studies could significantly affect the estimate of macroevolutionary rates (59). Despite the claims that current speciation-extinction models account for incomplete species sampling (e.g., MEDUSA (6), BAMM (60), TreePar (61), TESS (62), RevBayes (63), etc.), these models often assume a random and even sampling of the species across the phylogeny (i.e., each terminal in the phylogeny is assumed to have the same probability of being sampled). However, this assumption is very often violated in empirical examples due to geographic, temporal and taxonomic sampling biases. To account for non-random incomplete taxon sampling in the diversification analysis, we used the polytomy resolver PASTIS (22), to place missing species in the phylogeny using taxonomic constraints following a birth-death model. This method is a conservative approach for increasing the sampling in phylogenies since it infers

the timing of missing splits under a constant rates birth-death model. Therefore, any inference of rate-heterogeneity in the final analysis will have a strong support since this approach is expected to bias towards the detection of constant-rate models of diversification (22).

Every extant and accepted *Solanum* species was assigned to one of the clades listed in Table 1 in Supplementary information using the (20) phylogeny as backbone. Complete trees were then generated for each clade of *Solanum* using a combination of species with genetic data and taxonomic constrains. For each clade, PASTIS creates an output file in a nexus format, which contains the full set of tree constraints ready to be executed in MrBayes. Therefore, posterior distributions of phylogenies for each clade were then inferred in MrBayes 3.2.3 (64) using a relaxed clock model (independent branch rates – igr prior), with the default (exponential) prior on the distribution of branching rates. A distribution of 100 dated trees was produced from the clade distributions containing nearly all described species of *Solanum*, 1169 species in total. For further information see Supplementary Methods.

Diversification analysis

Several approaches have been developed to detect significant shifts in diversification across the branches of a phylogeny. In this analysis, we mainly used the BAMM approach (65), which aims to provide an improvement on methods that identify heterogeneity in rates only across specific subclades, such as MEDUSA (6). However, other approaches such as TESS (62) and RevBayes (63), were also implemented to assess the robustness of the results obtained in BAMM.

The underlying branching process in BAMM is more complex as the speciation and extinction parameters in the model include the effect of time-dependence in diversification (i.e., the age of a lineage can affect its diversification rate). It also implements a diversity-dependent diversification model where the number of lineages in a clade may affect its diversification rates. This approach uses a reversible-jump Markov Chain Monte Carlo approach to explore a larger space of parameters and candidate models of diversification. Since it follows a Bayesian statistical framework (rather than the maximum likelihood framework like MEDUSA), BAMM implicitly accounts for the uncertainty in parameter estimates by providing a distribution of marginal posterior probabilities instead of point estimates. To implement this approach in the *Solanum* phylogeny, we used the distribution of 100 complete species trees produced in the PASTIS analysis to consider the phylogenetic uncertainty in the estimates of diversification rates. We set the priors of speciation, extinction and the expected number of diversification shifts (γ=1) using the R package BAMMtools v 2.0.5 (66), which identifies the priors of the diversification parameters based on the distribution of divergence times of the phylogeny. For each of the 100 trees, an MCMC analysis was performed with four separate runs of 20 million generations. All the analyses were run using the $C++$ BAMM command line program v 2.5.0 on

the Imperial College London's High-Performance computing cluster (http://www.imperial.ac.uk/admin-services/ict/self-service/research-support/hpc/). We then checked for convergence of the MCMC samples making sure the effect sample size was at least 200 for both the number of evolutionary shifts and likelihood using the CODA R package V 0.16-1 (67). The first 25% of the samples were discarded as burn-in.

For each of the 100 trees run in BAMM, a distribution of 1000 samples of the posterior probabilities of diversification were created. Each sample from the posterior includes either a single event (i.e., the diversification is described by a single time-varying process — no shifts in diversification) or a mixture of two or more shifts and associated parameters. For each set of posterior probabilities, we extracted the list of nodes associated with "core" rate shifts (i.e., rate shift with a marginal probability significantly higher than the probability expected from the prior alone) and calculated the frequency with which these nodes are associated with significant rate shifts across the 100 trees run in BAMM. Finally, we computed the mean of diversification rates for each of the species present in the (20) phylogeny across the pooled distribution of posterior samples produced in the BAMM analysis.

A recent study by (68) raised concerns about several methodological issues in BAMM including a potentially incorrect likelihood function, the strong influence of priors on posterior estimates, and some theoretical errors (e.g., the incorrect use of the Poisson distribution as the error distribution of the prior number of shifts of diversification). To assess the effects of these potential limitations on the results found in this study, we performed several sensitivity analyses focusing on the two most important issues stated by (68) (1) the prior sensitivity of posterior distribution on the number of rate shifts, and (2) the reliability of the diversification rate estimates. None of these limitations influenced the results of our analysis (see Supplementary information for further details)

Geographical patterns of diversification

To determine which regions have accumulated and are accumulating a higher or lower number of *Solanum* lineages across the globe, the average diversification rates per species were displayed to a 1 x 1 degree grid scale map.

84,606 botanical records from 1005 *Solanum* species were extracted from the Solanaceae Source database http://solanaceaesource.org/ on 13 March 2016; and 34,462 records from 215 species were compiled from the Australian National Herbarium, accessed through the Atlas of Living Australia (ALA) website http://www.ala.org.au/, on 16 April 2016. We applied a series of quality control filters to discard the following:

1. Records with latitude and longitude coordinates of 0° , 0° .

2. Records matching coordinates from major herbaria or political centroids at all administrative divisions extracted from (69).

3. Any records with the described country conflicting with the country extracted from their coordinates. This step was performed by overlaying the records with a global administrative polygons extracted from the R package "rworldmap" (70) using the function *"over"* from the R package "sp" v 1.1-0 (71).

4. Records with identical coordinates for a given species

5. Records from non-native, cultivated or naturalized species

Taxonomic names were updated for all records to correct for synonymy using as a reference the accepted name list from http://solanaceaesource.org/. After cleaning, we were left with 64,826 unique records for 1,096 taxa.

Species occurrences were then converted into a presence-absence matrix of a 110 x 110 km equal area grid using the function "lets.presab.points" in the R package letsR v 2.1 (72). Using the mean species-specific rates, we estimated the mean assemblage diversification rates as the geometric mean of all species' rates present in a grid cell. We also computed a weighted version of this, dividing the mean species-specific diversification rates by the inverse of their range size — log of the area (sqm) occupied by each species, to correct for the overestimation of rates in an area as a result of the occurrence of widespread species.

Finally, we reconstructed lineages-through-time plots for the three principal diversity centres for *Solanum* (Neotropics, Australia and Africa) using the R package "paleotree" v 2.3 (73) to visualise the geographical differences in lineage accumulation dynamics among continents.

Ancestral range reconstruction

We investigated the historical biogeography of *Solanum* using the R package "BioGeoBEARS" (32; 33) — Bio-Geography with Bayesian (and likelihood) Evolutionary Analysis in R Scripts. This approach provides a statistical framework to compare traditional models in biogeography such as DIVA (Dispersal-Vicariance Analysis (74)), DEC (Dispersal-Extinction-Cladogenesis (75)), and BayArea (76). Each of these models assumes different processes to reconstruct the ancestral range of lineages. In all of the models, species ranges are allowed to change along the branches by anagenetic evolution through two main events: dispersion (range expansion) and extinction (range contraction). The events allow in cladogenesis vary depending on the fitted model. For example, sympatric speciation subset, or peripatric speciation (i.e., the ancestral range, which is widespread, is completely inherited to one of the daughters whereas the other inherits the range of just a single area) is only allowed in the DEC model. Unlike the DIVA and the BayArea models, the DEC model only allows one daughter to inherent widespread distributions during a vicariant event, since the widespread distribution of both daughters would assume additional events such as post-speciation dispersal (75). In all models, sympatric speciation (i.e., within-area speciation or duplication) is

allowed. In addition to these classic events of biogeography, BioGeoBEARS includes a founder-event speciation event ($+j$) which considers the influence of speciation through long-distance dispersal that are common in Island-like models.

Phylogenetic relationships among species of *Solanum* were taken from the time-calibrated clade credibility (MCC) tree from (20). This phylogeny represents the current state-of-the-art of *Solanum* relationships published so far. Several species were pruned from the tree as follow: (1) species which are considered widely cultivated or with ambiguous native distribution, (2) species considered as synonyms, and therefore duplicated in the tree, and (3) species with a low support of a direct ancestor-descendant relationships leading to negative branches in the MCC. The final pruned phylogeny used for subsequent analyses contained 386 species of *Solanum* with the addition of *Jaltomata andersonii* as the outgroup.

Using the extracted distribution records, we created an occurrence matrix of *Solanum* species into six biogeographic areas — Africa, Australia, Indo-Pacific, Neotropics, Nearctic and Palearctic based on the floral kingdoms defined by (77). The distribution of widespread species was double checked and corrected for potentially recent cultivated and/or naturalised species using the native descriptions available from http: //solanaceaesource.org/ or specific taxonomic monographs. The maximum number of areas was set to three to reduce the complexity and computational time in the analysis.

Three biogeographic models were fitted in the *Solanum* phylogeny and associated geographic distribution using BioGeoBEARS — DIVALIKE, DEC and BAYAREALIKE models. The influence of founder-event speciation event (+J) was also included into each model, resulting in a total of six models. An additional set of six biogeographic models were fitted using a dispersal matrix multiplier to weight the dispersal probability of adjacent areas as 1, 0.5 and 0.001 for easy, medium, and hard dispersal. The model that best describe the empirical data (i.e., optimal fixed model structure) was then chosen using a stepwise selection from the candidate models ranking under the Akaike Information Criterion (AIC (78)).

Once the best-fitting model was identified, we estimated the overall probabilities of the anagenetic and cladogenetic events conditional on the model, the phylogeny, and the geographic distributions from 100 Biogeographic Stochastic Maps (BSM) to account for the uncertainty in the state transitions and ancestral range reconstructions (79). These stochastic maps are similar to simulations of trait change along phylogenies (80) using transition rate models (81). Given the observed range data, the phylogeny and the best fitting model of biogeographic events, the BSM simulates possible histories constraint to the observed ancestral ranges. An average of all the probabilities of the simulated histories will result in the same ancestral state probability obtained under the best fitting model (79).

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Raup DM, Gould SJ, Schopf TJ, Simberloff DS (1973) Stochastic models of phylogeny and the evolution of diversity. *The Journal of Geology* 81(5):525–542.

2. Slowinski JB, Guyer C (1993) Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. *American Naturalist* pp. 1019–1024.

3. Purvis A (1996) *New uses for new phylogenies*, eds. Harvey P, Brown AJL, Smith JM, Nee S. (Oxford University Press, Oxford) Vol. 153.

4. Wiens JJ (2011) The causes of species richness patterns across space, time, and clades and the role of "ecological limits". *The Quarterly Review of Biology* 86(2):75–96.

5. Stanley SM, Signor III PW, Lidgard S, Karr AF (1981) Natural clades differ from "random" clades: simulations and analyses. *Paleobiology* pp. 115–127.

6. Alfaro ME, et al. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* 106(32):13410–13414.

7. Wiens JJ (2015) Faster diversification on land than sea helps explain global biodiversity patterns among habitats and animal phyla. *Ecology letters* 18(11):1234–1241.

8. Simpson GG (1955) *Major features of evolution*. (Columbia University Press: New York).

9. Schluter D (2000) *The ecology of adaptive radiation*. (Oxford University Press).

10. Yoder J, et al. (2010) Ecological opportunity and the origin of adaptive radiations. *Journal of evolutionary biology* 23(8):1581–1596.

11. Moore BR, Donoghue MJ (2007) Correlates of diversification in the plant clade dipsacales: geographic movement and evolutionary innovations. *the american naturalist* 170(S2):S28–S55.

12. Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. *The American Naturalist* 175(6):623–639.

13. Rabosky DL, et al. (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature communications* 4.

14. Liow LH, Quental TB, Marshall CR (2010) When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Systematic Biology* p. syq052.

15. Quental TB, Marshall CR (2010) Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution* 25(8):434–441.

16. Rabosky DL (2010) Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64(6):1816–1824.

17. Vorontsova M, Stern S, Bohs L, Knapp S, , et al. (2013) African spiny solanum (subgenus leptostemonum, solanaceae): a thorny phylogenetic tangle. *Botanical Journal of the Linnean Society* 173(2):176–193.

18. Frodin DG (2004) History and concepts of big plant genera. *Taxon* 53(3):753–776.

19. Olmstead RG, Palmer JD (1997) Implications for the phylogeny, classification, and biogeography of solanum from cpdna restriction site variation. *Systematic Botany* pp. 19–29.

20. Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC evolutionary biology* 13(1):214.

21. Knapp S, Vorontsova MS (2016) A revision of the ”african non-spiny” clade of solanum l.(solanum sections afrosolanum bitter, benderianum bitter, lemurisolanum bitter, lyciosolanum bitter, macronesiotes bitter, and quadrangulare bitter: Solanaceae). *PhytoKeys* 66:1.

22. Thomas GH, et al. (2013) Pastis: an r package to facilitate phylogenetic assembly with soft taxonomic inferences. *Methods in Ecology and Evolution* 4(11):1011–1017.

23. Stern S, Agra MdF, Bohs L (2011) Molecular delimitation of clades within new world species of the “spiny solanums” (solanum subg. leptostemonum). *Taxon* 60(5):1429–1441.

24. Levin RA, Myers NR, Bohs L (2006) Phylogenetic relationships among the “spiny solanums”(solanum subgenus leptostemonum, solanaceae). *American Journal of Botany* 93(1):157–169.

25. Weese TL, Bohs L (2007) A three-gene phylogeny of the genus solanum (solanaceae). *Systematic Botany* 32(2):445–463.

26. Weese TL, Bohs L (2010) Eggplant origins: out of africa, into the orient. *Taxon* 59(1):49–56.

27. Whalen MD, Costich DE, Heiser CB (1981) Taxonomy of solanum section lasiocarpa. *Gentes Herb.(Ithaca)* 12(2):41–129.

28. Aubriot X, Singh P, Knapp S (2016) Tropical asian species show that the old world clade of 'spiny solanums' (solanum subgenus leptostemonum pro parte: Solanaceae) is not monophyletic. *Botanical Journal of the Linnean Society* 181(2):199–223.

29. Levin RA, Watson K, Bohs L (2005) A four-gene study of evolutionary relationships in solanum section acanthophora. *American Journal of Botany* 92(4):603–612.

30. Bohs L (2004) A chloroplast dna phylogeny of solanum section lasiocarpa. *Systematic Botany* 29(1):177–187.

31. Olmstead RG (2013) Phylogeny and biogeography in solanaceae, verbenaceae and bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society* 171(1):80–102.

32. Matzke NJ (2012) Founder-event speciation in biogeobears package dramatically improves likelihoods and alters parameter inference in dispersal–extinction–cladogenesis (dec) analyses. *Front. Biogeogr* 4:210.

33. Matzke NJ (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic biology* p. syu056.

34. McPeek MA, Brown JM (2007) Clade age and not diversification rate explains species richness among animal taxa. *The American Naturalist* 169(4):E97–E106.

35. Dupin J, et al. (2016) Bayesian estimation of the global biogeographical history of the solanaceae. *Journal of Biogeography* 44:887–899.

36. Weir JT, Schluter D (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* 315(5818):1574–1576.

37. Crisp M, Cook L, Steane D (2004) Radiation of the australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities? *Philosophical Transactions of the Royal Society B: Biological Sciences* 359(1450):1551–1571.

38. Crisp MD, Cook LG (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.

39. Crisp MD, Cook LG (2007) A congruent molecular signature of vicariance across multiple plant lineages. *Molecular phylogenetics and evolution* 43(3):1106–1117.

40. Harrington MG, Jackes BR, Barrett MD, Craven LA, Barrett RL (2012) Phylogenetic revision of backhousieae (myrtaceae): Neogene divergence, a revised circumscription of backhousia and two new species. *Australian Systematic Botany* 25(6):404–417.

41. Crayn DM, Rossetto M, Maynard DJ (2006) Molecular phylogeny and dating reveals an oligo-miocene radiation of dry-adapted shrubs (former tremandraceae) from rainforest tree progenitors (elaeocarpaceae) in australia. *American Journal of Botany* 93(9):1328–1342.

42. Blom MP, Horner P, Moritz C (2016) Convergence across a continent: adaptive diversification in a recent radiation of australian lizards in *Proc. R. Soc. B.* (The Royal Society), Vol. 283, p. 20160181.

43. Rabosky DL, Mitchell JS, Chang J (2017) Is bamm flawed? theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic biology*.

44. Harmon LJ, Schulte JA, Larson A, Losos JB (2003) Tempo and mode of evolutionary radiation in iguanian lizards. *Science* 301(5635):961–964.

45. Rabosky DL, Donnellan SC, Talaba AL, Lovette IJ (2007) Exceptional among-lineage variation in diversification rates during the radiation of australia's most diverse vertebrate clade. *Proceedings of the Royal Society of London B: Biological Sciences* 274(1628):2915–2923.

46. Byrne M, et al. (2008) Birth of a biome: insights into the assembly and maintenance of the australian arid zone biota. *Molecular Ecology* 17(20):4398–4417.

47. Byrne M, et al. (2011) Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the australian mesic zone biota. *Journal of Biogeography* 38(9):1635–1656.

48. Losos JB, Schluter D (2000) Analysis of an evolutionary species-area relationship. *Nature* 408(6814):847.

49. Davies TJ, Savolainen V, Chase MW, Goldblatt P, Barraclough TG (2005) Environment, area, and diversification in the species-rich flowering plant family iridaceae. *The American Naturalist* 166(3):418–425.

50. Poczai P, Hyvönen J, Symon DE (2011) Phylogeny of kangaroo apples (solanum subg. archaesolanum, solanaceae). *Molecular biology reports* 38(8):5243–5259.

51. Donoghue MJ, Sanderson MJ (2015) Confluence, synnovation, and depauperons in plant diversification. *New Phytologist*.

52. Zanne AE, et al. (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature* 506(7486):89.

53. Cornwell WK, et al. (2014) Functional distinctiveness of major plant lineages. *Journal of Ecology* 102(2):345–356.

54. Beaulieu JM, O'Meara BC, Donoghue MJ (2013) Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62(5):725–737.

55. Chira AM, Thomas GH (2016) The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. *Journal of evolutionary biology* 29(12):2502–2518.

56. Morlon H (2014) Phylogenetic approaches for studying diversification. *Ecology letters* 17(4):508–525.

57. De Queiroz A (2002) Contingent predictability in evolution: key traits and diversification. *Systematic Biology* 51(6):917–929.

58. Donoghue MJ (2005) Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. *Paleobiology* 31(S2):77–93.

59. Höhna S, Stadler T, Ronquist F, Britton T (2011) Inferring speciation and extinction rates under different sampling schemes. *Molecular biology and evolution* 28(9):2577–2589.

60. Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PloS one* 9(2):e89543.

61. Stadler T (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* 108(15):6187–6192.

62. Höhna S (2015) The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. *Journal of theoretical biology* 380:321–331.

63. Höhna S, et al. (2016) Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic biology* p. syw021.

64. Ronquist F, Huelsenbeck JP (2003) Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.

65. Rabosky DL, Donnellan SC, Grundler M, Lovette IJ (2014) Analysis and visualization of complex macroevolutionary dynamics: an example from australian scincid lizards. *Systematic biology* 63(4):610–627.

66. Rabosky DL, et al. (2014) Bammtools: an r package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*.

67. Plummer M, Best N, Cowles K, Vines K (2006) Coda: Convergence diagnosis and output analysis for mcmc. *R News* 6(1):7–11.

68. Moore BR, Höhna S, May MR, Rannala B, Huelsenbeck JP (2016) Critically evaluating the theory and performance of bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences* 113(34):9569–9574.

69. Edwards EJ, de Vos JM, Donoghue MJ (2015) Doubtful pathways to cold tolerance in plants. *Nature* 521(7552):E5–E6.

70. South A (2011) rworldmap: A new r package for mapping global data. *The R Journal* 3(1):35–43.

71. Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. *R News* 5(2):9–13.

72. Vilela B, Villalobos F (2015) letsr: a new r package for data handling and analysis in macroecology. *Methods in Ecology and Evolution* 6(10):1229–1234.

73. Bapst DW (2012) paleotree: an r package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution* 3(5):803–807.

74. Ronquist F (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46(1):195–203.

75. Ree RH, Moore BR, Webb CO, Donoghue MJ, Crandall K (2005) A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59(11):2299–2311.

76. Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic biology* 62(6):789–804.

77. Cox CB (2001) The biogeographic regions reconsidered. *Journal of Biogeography* 28(4):511–523.

78. Burnham KP, Anderson DR (2002) *Model selection and multi-model inference: a practical information-theoretic approach*. (Springer), 2nd edition, p. 488.

79. Matzke NJ (2016) Stochastic mapping under biogeographical models. PhyloWiki BioGeoBEARS Available at http://phylo.wikidot.com/biogeobears#stochastic_mapping.

80. Huelsenbeck JP, Nielsen R, Bollback JP (2003) Stochastic mapping of morphological characters. *Systematic Biology* 52(2):131–158.

81. Pagel M (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic biology* 48(3):612–622.

Author contributions

S.E-L., designed the study, analysed the data and wrote the manuscript, T.S provided the phylogenetic data (reconstructed the phylogeny), distribution and taxonomic data, I.F, provided advice on the analysis, and contributed to writing the manuscript, particularly the discussion, S.K., provided the taxonomic, distribution and diversity data and supervised the analysis, A.P., designed and supervised the analysis. All authors reviewed the manuscript.