



The role of phylogenetic relatedness on alien plant success depends on the stage of invasion

Ali Omer^{1,2}✉, Trevor Fristoe¹, Qiang Yang¹, Mialy Razanajatovo^{1,3}, Patrick Weigelt^{1,4,5,6}, Holger Kreft^{1,4,6}, Wayne Dawson^{1,7}, Stefan Dullinger^{1,8}, Franz Essl^{1,9}, Jan Pergl^{1,10}, Petr Pyšek^{10,11} and Mark van Kleunen^{1,12}

Darwin's naturalization hypothesis predicts successful alien invaders to be distantly related to native species, whereas his pre-adaptation hypothesis predicts the opposite. It has been suggested that depending on the invasion stage (that is, introduction, naturalization and invasiveness), both hypotheses, now known as Darwin's naturalization conundrum, could hold true. We tested this by analysing whether the likelihood of introduction for cultivation, as well as the subsequent stages of naturalization and spread (that is, becoming invasive) of species alien to Southern Africa are correlated with their phylogenetic distance to the native flora of this region. Although species are more likely to be introduced for cultivation if they are distantly related to the native flora, the probability of subsequent naturalization was higher for species closely related to the native flora. Furthermore, the probability of becoming invasive was higher for naturalized species distantly related to the native flora. These results were consistent across three different metrics of phylogenetic distance. Our study reveals that the relationship between phylogenetic distance to the native flora and the success of an alien species changes from one invasion stage to the other.

The translocation and introduction of species to new regions where they are not native and their successful establishment in the wild (naturalization)¹ is rapidly changing the distribution of biota^{2,3}. A subset of those naturalized alien species—the invasive ones—cause serious environmental and socio-economic problems^{4–7}. Numbers of introduced, naturalized and invasive species are expected to continue to increase^{8,9}, partly boosted by other global changes (for example, climate and land use) and the increasing transportation of people and goods^{10–13}. Thus, understanding the drivers of biological invasions has become a major objective in ecology.

Many studies have tried to determine the features of successful plant invaders by analysing their functional traits (for example, seed mass, height, mode of reproduction), native origins and introduction history across spatial and temporal scales^{14–27}. Another line of research has tried to explain invasion success through taxonomic and phylogenetic comparisons of alien and native floras^{28–30}, because evolutionary relatedness is assumed to reflect the functional similarity of species. This approach has been fostered by the increasing availability of phylogenetic information, which has allowed for more accurate estimates of the evolutionary relatedness of alien to native species^{31–35}.

Interest in using evolutionary relatedness as a potential driver of invasion success started with Darwin's naturalization hypothesis^{32,36,37}. Darwin³⁶ proposed that the more distantly related an alien species is to the native community, the more likely it will establish successfully. The reasoning behind this prediction is that distantly related alien species are functionally more distinct from native species, and it is therefore more likely that they can utilize empty ecological niches in the resident community. However, Darwin³⁶

also proposed that alien species that are closely related to native species should be more successful because their similarity to the native species would indicate that they are pre-adapted to local environmental conditions: the so-called pre-adaptation hypothesis. These seemingly contradictory hypotheses are known as Darwin's naturalization conundrum^{32,38}.

Studies that tested Darwin's naturalization conundrum have found no consistent relationships between invasion success and evolutionary distance to native species^{32,39}. This might reflect that studies considered different temporal and spatial scales; for example, at small scales Darwin's naturalization hypothesis is likely to hold, whereas at large scales the pre-adaptation hypothesis is likely to hold^{32,39,40}. Furthermore, most studies have addressed Darwin's naturalization conundrum without considering possible differences between invasion stages^{28,41–45}, although some have considered both naturalization and invasiveness^{38,46,47}. The process of biological invasion can be subdivided into subsequent stages, and species have to pass different barriers or filters to move from one stage to the next^{1,48}. First, species must overcome biogeographic dispersal barriers via human introduction into a new region. Although Darwin's naturalization conundrum does not make any predictions regarding the introduction stage, accounting for this stage is important because introduced plants, such as the ones introduced intentionally for economic uses (for example, as ornamental plants), are unlikely to be a random phylogenetic subset of the global flora⁴⁹. Ignoring such introduction bias can result in the over- or underestimation of the importance of species' traits and phylogenetic distance for naturalization success²⁵. Once a species has been introduced, phylogenetic relatedness to the native flora may impact

¹Ecology, Department of Biology, University of Konstanz, Konstanz, Germany. ²Department of Forest Management, University of Khartoum, North Khartoum, Sudan. ³Institute of Landscape and Plant Ecology (320a), University of Hohenheim, Stuttgart, Germany. ⁴Biodiversity, Macroecology & Biogeography, University of Goettingen, Göttingen, Germany. ⁵Campus-Institut Data Science, Göttingen, Germany. ⁶Centre of Biodiversity and Sustainable Land Use (CBL), University of Goettingen, Göttingen, Germany. ⁷Department of Biosciences, Durham University, Durham, UK. ⁸Division of Biodiversity Dynamics and Conservation, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria. ⁹BioInvasions, Global Change, Macroecology-Group, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria. ¹⁰Department of Invasion Ecology, Czech Academy of Sciences, Institute of Botany, Průhonice, Czech Republic. ¹¹Department of Ecology, Charles University, Prague, Czech Republic. ¹²Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China. ✉e-mail: ali.omer@uni-konstanz.de

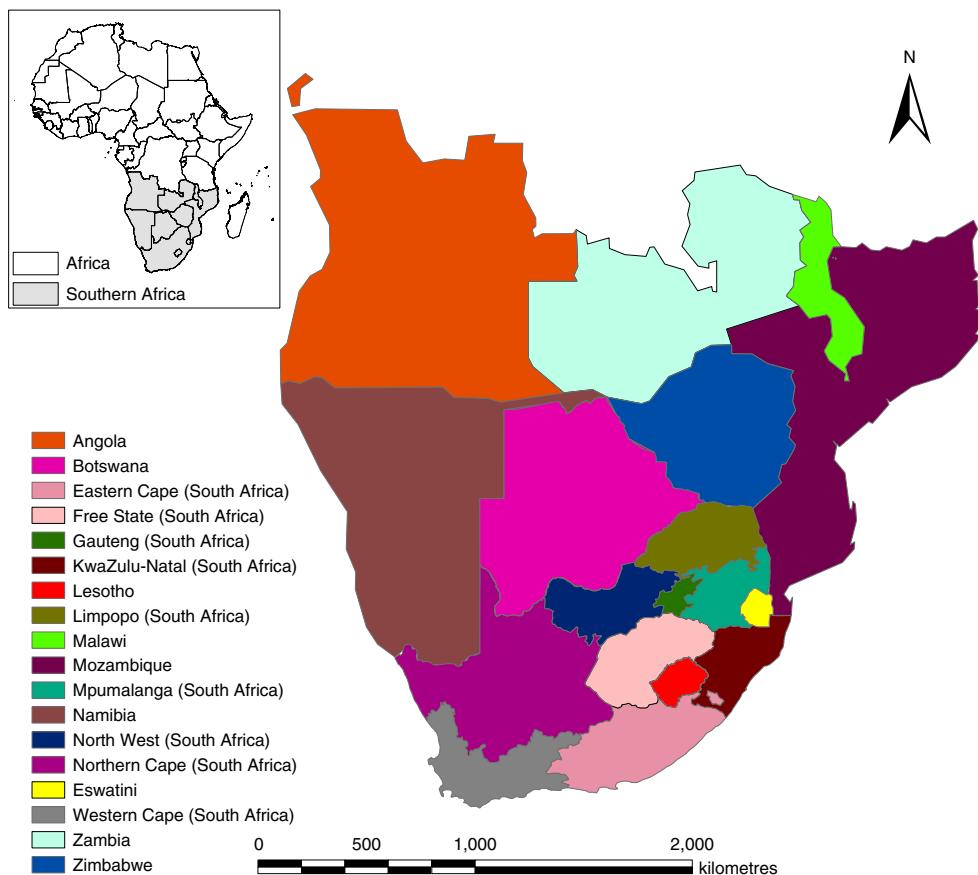


Fig. 1 | Map of Africa and the study regions showing the ten countries of Southern Africa. Map showing Angola, Botswana, Eswatini, Lesotho, Malawi, Mozambique, Namibia, South Africa, Zambia and Zimbabwe. Because we had separate native species occurrence data for the nine provinces of South Africa, the total number of Southern African regions was 18. The map was produced using ArcMap Software v10.8 (<https://desktop.arcgis.com/en/arcmap/>).

its success in subsequent stages in different ways. It is plausible that to establish self-sustaining populations (that is, to become naturalized) in a region, it is important that a species is pre-adapted to the new environment⁵⁰, which is likely the case when there are closely related natives^{28,39,51}. However, to reach the next stage of invasion (that is, become dominant and widespread—being invasive), it may be beneficial to be different from the native species and capitalize on unoccupied niches, which is likely when there are no closely related natives. We hence hypothesize that distinguishing between the three different stages of invasion^{39,52,53} might provide additional insights into Darwin’s naturalization conundrum.

The evolutionary relatedness of naturalized species to native floras has typically been compared with the relatedness expected if the naturalized species had been drawn randomly from the global flora^{29,54,55}. Such a test relies on the unrealistic assumption that there has been no bias with regard to the phylogenetic affinity and characteristics of the introduced species and that all species have had the same opportunity to naturalize. However, the pool of introduced species from which the naturalized ones have emerged is often phylogenetically biased⁴⁹. Information on the pool of alien species introduced into a region for cultivation, in addition to information on naturalized and native species, would be needed to account for biases that might be associated with the selection of species for introduction^{28,42}. Furthermore, when assessing drivers of plant invasions, one should consider that relationships may be nonlinear (hump or U-shaped)¹⁸. To improve our understanding of the drivers of plant invasion success, we hence need to quantify and account for biases associated with the introduction stage^{25,56} and to test for nonlinear effects⁴⁰.

Here, we study 5,091 alien angiosperm species that have been introduced into Southern Africa for cultivation⁵⁷. Combined with lists of the global flora, the naturalized and native species within the region of Southern Africa and the species that have become invasive within the country of South Africa (Fig. 1), we test Darwin’s naturalization conundrum at the three different stages of the invasion process. Specifically, we asked how phylogenetic relatedness to native species is correlated with the subsequent stages of introduction for cultivation, naturalization and invasiveness of alien species.

Results

The introduction and naturalization transitions. We found that of the 299,834 angiosperm species that are not native to Southern Africa, 974 (0.32%) have become naturalized in at least one of Southern Africa’s subregions. Among the phylogenetic distance models explaining the likelihood of being naturalized, phylogenetic distance to the most closely related native species (PD_{Min}) provided the best model fit (had the lowest Akaike’s information criterion), followed by weighted mean phylogenetic distance to the native species (PD_{wMean}) and mean phylogenetic distance to the native species (PD_{Mean}) (Fig. 2 and Supplementary Table 1). The likelihood of naturalization of a species from the global species pool decreased significantly for species more distantly related to the native flora according to all three phylogenetic indices (Fig. 3a–c, Supplementary Fig. 2a and Supplementary Table 1).

When accounting for cultivation bias by considering the two transitions—to the introduction for cultivation and from the introduction for cultivation to naturalization—separately, we found that of the 299,834 angiosperm species that are not native to Southern

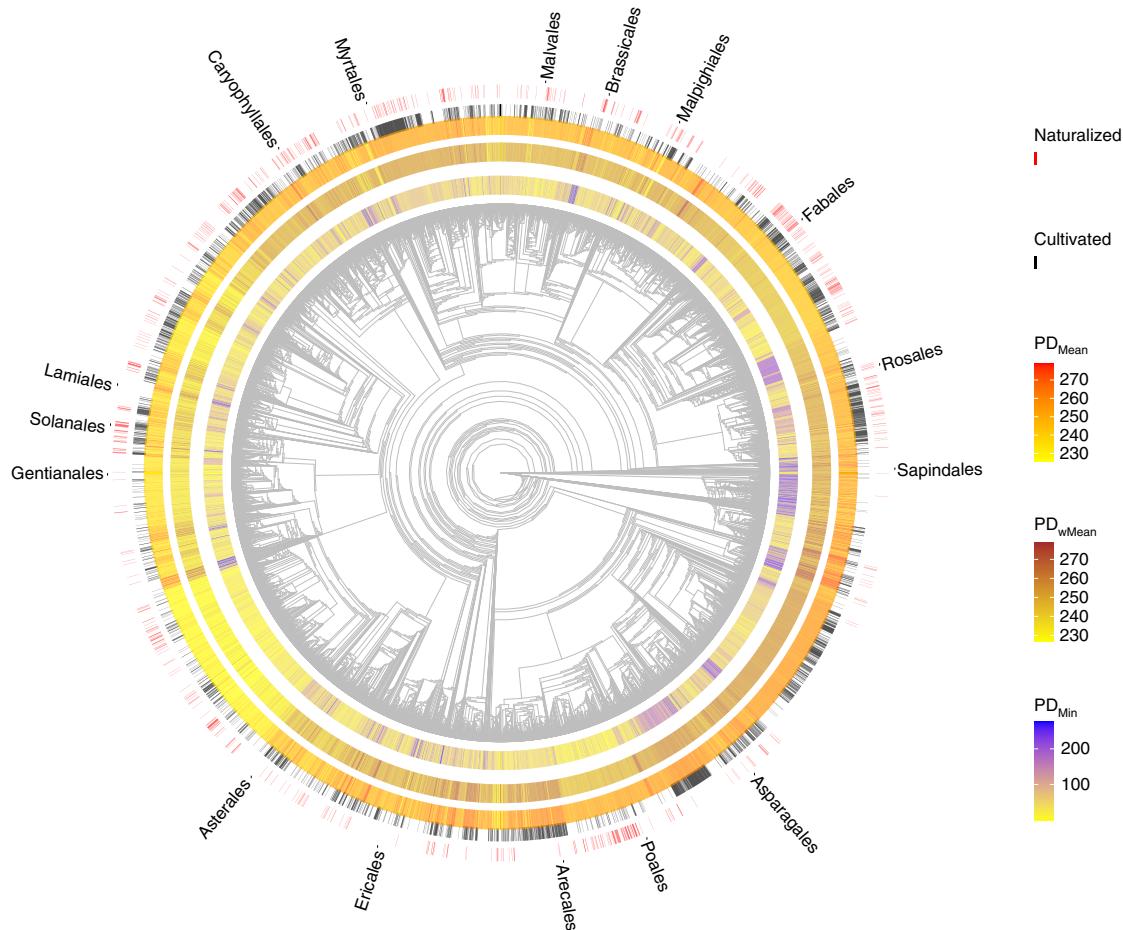


Fig. 2 | Phylogenetic tree of the global species pool showing all cultivated and naturalized alien species in Southern Africa and the three indices of phylogenetic distance. The tree includes all non-native angiosperms with accepted names in The Plant List ($n=299,834$ species). The three inner rings indicate in heat colours the three indices of phylogenetic distance to the native flora of Southern Africa: PD_{Mean}, PD_{wMean} and PD_{Min}. The outer black and red stripes indicate cultivated species ($n=5,091$) and naturalized species ($n=974$), respectively. The names of major clades (that is, plant orders) are provided for orientation.

Africa, 5,091 (1.69%) have been introduced for cultivation, of which 592 (11.6%) have become naturalized in at least one subregion of Southern Africa. For the first transition, mean phylogenetic distance weighted by range size of the natives (PD_{wMean}) provided the best model fit, whereas unweighted mean phylogenetic distance provided the best model fit for the second transition (PD_{Mean}; Supplementary Table 1). Regardless of which distance index was chosen, the probability of a species being introduced from the global species pool into Southern Africa increased with phylogenetic distance to the native flora (Fig. 3d–f, Supplementary Fig. 1a and Supplementary Table 1), whereas the opposite was true for the probability of subsequent naturalization (Fig. 3g–i and Supplementary Table 1). However, the amount of variation explained by the models was relatively small (all Nagelkerke $R^2 < 0.1$; Supplementary Table 1).

The naturalization-invasiveness transition. Because information on the invasiveness of naturalized aliens is not available for the entire region of Southern Africa, we tested for a phylogenetic signal in the naturalization-invasiveness transition within the naturalized cultivated flora of the country of South Africa. Among the 524 introduced cultivated species that have naturalized in the country of South Africa, 310 (59.1%) are considered invasive. In phylogenetic distance models explaining the likelihood of being invasive in the country of South Africa, PD_{Min} provided the best model fit, followed

by PD_{wMean} and PD_{Mean} (Supplementary Table 2). According to all three models, naturalized species were more likely to become invasive if the phylogenetic distance to the most closely related native species is large (Supplementary Fig. 1c, Fig. 4 and Supplementary Table 2).

Discussion

Because many studies on biological invasions lack crucial data on which species have been introduced but failed to establish, we took advantage of a comprehensive list of alien plant species that have been introduced into Southern Africa for cultivation. This list allowed us to account for cultivation biases in our tests of Darwin's naturalization conundrum along the different stages of the invasion process. Our results show that alien species distantly related to native species were more likely to be introduced for cultivation. Once introduced for cultivation, however, those that are more closely related to native species were more likely to naturalize. By contrast, among the naturalized species in the country of South Africa, the ones most distantly related to the natives were more likely to become invasive. Our results thus show that phylogenetic distance to the native flora has opposing effects on the subsequent transitions during the invasion process.

Humans have introduced thousands of plant species from their native regions into foreign lands. Although some of those

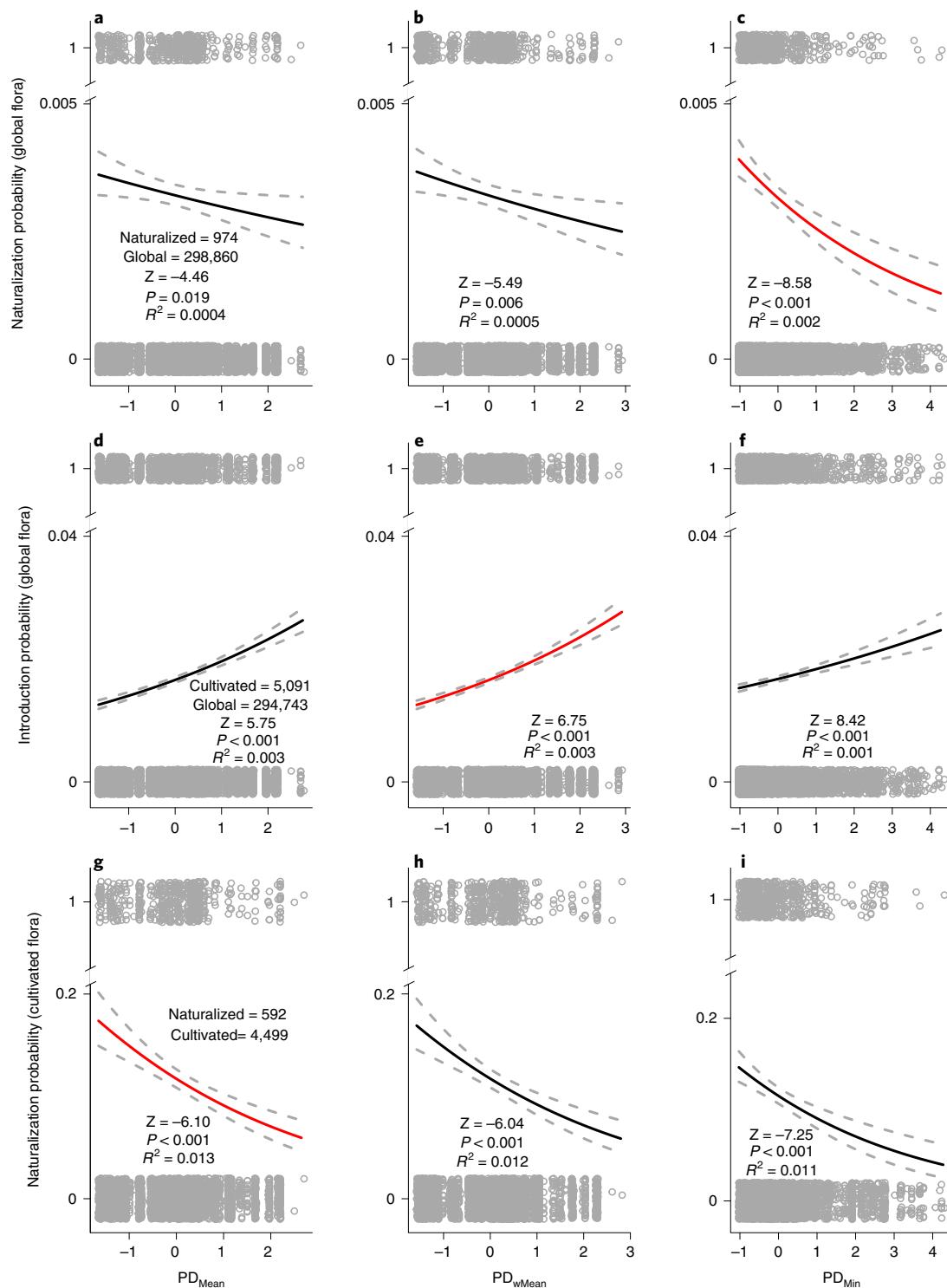


Fig. 3 | The probabilities of introduction for cultivation and naturalization of alien species in Southern Africa. The probabilities of naturalization of species from the global flora outside Southern Africa (**a–c**), of introduction for cultivation from the global flora outside Southern Africa (**d–f**) and of naturalization of introduced cultivated plant species in Southern Africa (**g–i**) in relation to the different phylogenetic distance indices: PD_{Mean} (**a,d,g**), PD_{wMean} (**b,e,h**), PD_{Min} (**c,f,i**). The solid black and red lines are the predicted relationships from binomial GLMs (red lines indicate the best fitting model, Supplementary Table 1), and the dashed grey lines indicate the 95% confidence intervals. The values of the phylogenetic distance indices of the successful and unsuccessful species are shown as open dots at one and zero, respectively (jittered to increase visibility). The values of the indices were standardized to a mean of zero and a standard deviation of one. Note the y axis breaks. Statistical significance according to GLM models (two-sided tests) is indicated in the figure as exact P values.

introductions were accidental, most were intentional for cultivation purposes^{40,58–61}. Moreover, it is likely that the species that have been introduced are not a random selection from the global flora but

possess certain characteristics that make them of interest for cultivation^{25,56}. This is also reflected in their phylogenetic distance to the native flora because we found that introduction of alien species for

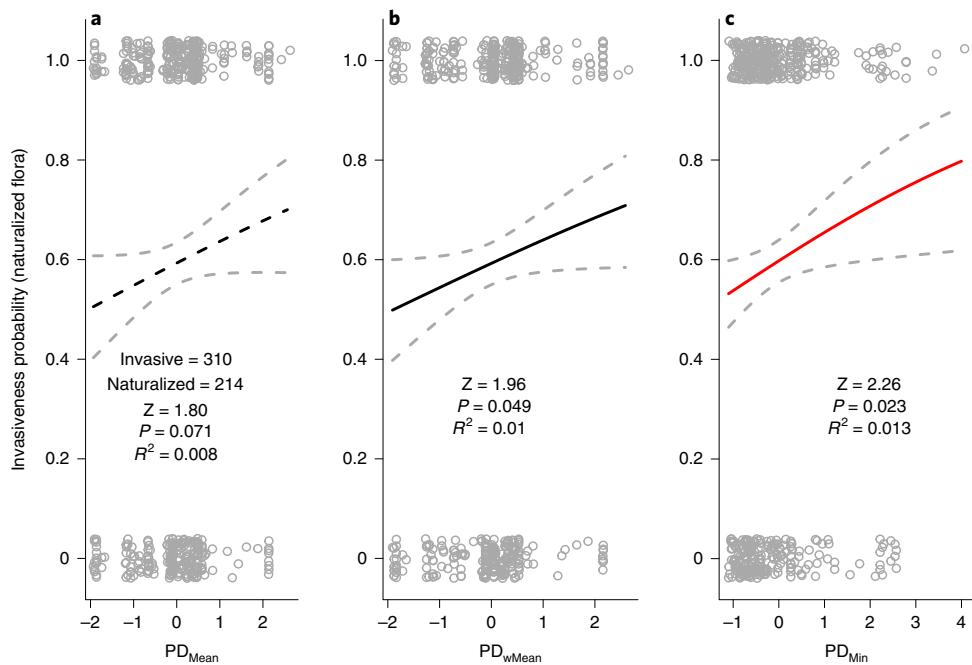


Fig. 4 | The probability of becoming invasive for naturalized alien plants that have been introduced for cultivation ($n=524$) in the country of South Africa in relation to the different phylogenetic distance indices. The phylogenetic distance indices are PD_{Mean} (a), PD_{wMean} (b) and PD_{Min} (c). The solid or dashed black and red lines are the predicted relationships from binomial GLMs (the red line indicates the best fitting model, and the dashed line indicates a non-significant relationship, Supplementary Table 2), and the dashed grey lines indicate the 95% confidence intervals. The values of the phylogenetic distance indices of the successful and unsuccessful species are shown as open dots at one and zero, respectively (jittered to increase visibility). The values of the indices were standardized to a mean of zero and a standard deviation of one. Statistical significance according to GLM models (two-sided tests) is indicated in the figure as exact P values.

cultivation in Southern Africa was positively associated with phylogenetic distance to the native flora (Figs. 2 and 3d–f and Supplementary Table 1). This most likely indicates that non-native species with characteristics that are missing in the native flora were more likely to have been prospected for cultivation in Southern Africa. For example, the fact that Australian *Eucalyptus* species—which have no close relatives in Southern Africa—grow faster and produce better wood than most native Southern African trees, made them very attractive for cultivation in forestry plantations⁶². The same is true for the introduction of other woody species into the country of South Africa that were planted for desertification control and to reduce firewood shortages⁶³. In other words, our results are in line with the idea that humans might preferentially introduce plant species with characteristics that could provide economic and social benefits or ecosystem services not provided by species of the native flora.

Once introduced for cultivation, not all alien species manage to grow and reproduce outside cultivation. We found that the naturalization success of species introduced for cultivation was negatively associated with phylogenetic distance to the native flora (Figs. 2 and 3g–i and Supplementary Table 1). In other words, and in line with the pre-adaptation hypothesis, introduced species that have closely related, and likely ecologically similar, species in the native Southern African flora were more likely to naturalize. This pattern was even visible when we considered the naturalization of non-cultivated species only (Supplementary Fig. 2) and the naturalization of all species in the global flora that are not native to Southern Africa (Fig. 3a–c and Supplementary Table 1), despite the biased introduction of cultivated species that are distantly related to the native flora. This suggests that even if there is no information on which alien species have been introduced and failed to establish, comparing naturalized species with the global flora, as done in previous studies^{29,54,55}, provides some indication of the importance of phylogenetic distance.

Naturalization in a new range depends on several abiotic and biotic factors¹. These factors can act as filters that determine which species can, in principle, grow in the region, but biotic interactions such as competition can also provide resistance against the ultimate establishment of those species. It has been shown that, at least at the local scale, environmental pre-adaptation and biotic resistance are both important for alien plant naturalization^{38,43,64}. However, in line with our results, at large spatial scales, environmental filtering is usually the most decisive factor for naturalization success^{28,39,51}. One potential explanation, the pre-adaptation hypothesis proposed by Darwin³⁶, is that introduced species closely related to the native flora share features with those native species that allow for survival and reproduction in the new range. Indeed, a study that used functional traits of alien and native plants found that the similarity of both groups facilitates naturalization²¹. These shared characteristics make the alien species pre-adapted to the new environmental conditions. Although it is likely that closely related species have stronger competitive impacts on each other^{65–68} (but see Dostál⁶⁹), at large scales, the pre-adaptation effect may overrule the drawback of stronger competition from close relatives^{32,41,70}. Overall, our results at the naturalization stage support the results of previous studies that examined the same mechanism³⁹.

Although the naturalization of introduced alien plants for cultivation in Southern Africa might have occurred at sites where closely related native species happened to be present, this is less likely the case for naturalized species that have become widespread and locally dominant (invasive). For a naturalized species to spread and become dominant, it has to interact with increasing numbers of native species. Those interactions are expected to be less detrimental for the alien species if it is more dissimilar to the natives because this allows it to avoid strong competition and common enemies by occupying vacant niches^{70,71}. In line with this idea, we found that

all naturalized species (Supplementary Fig. 1c) and the subset of cultivated naturalized species in the country of South Africa (Fig. 4) were more likely to be invasive if they have no close relatives in the native flora. The same non-significant trend was also found for the subset of non-cultivated species in the country of South Africa for the phylogenetic distance metric that provided the best model fit (PD_{Min} ; Supplementary Fig. 3). A positive relationship between phylogenetic distance and invasiveness has also been shown for alien trees and shrubs in Southern Africa⁷². Many other studies addressing Darwin's naturalization conundrum in plants also found support for Darwin's naturalization hypothesis at the invasiveness stage^{21,28,39,45,51,73}. However, a recent study on the abundance of alien birds⁷⁴ and a study on alien plant dominance in vegetation plots⁴⁶ found support for Darwin's pre-adaptation hypothesis. This suggests that the role of phylogenetic relatedness in driving invasiveness may differ between groups of organisms and depends on the spatial scale considered.

We used three different phylogenetic distance indices, but because they are correlated, they revealed similar patterns. The time-calibrated phylogeny of Smith and Brown⁷⁵, which we used to construct the global phylogeny, includes 86% of all genera. Consequently, the phylogenetic resolution of our tree is relatively good within families but is not within most genera. This might particularly affect estimates of PD_{Min} . Nonetheless, because phylogenetic distances between species of the same genus are usually shorter than for species belonging to different genera or families, PD_{Min} should still provide a good proxy for the presence of closely related species, and therefore is a useful phylogenetic distance measure for the question at hand. Although all phylogenetic distance measures revealed the same patterns, the ones that provided the best model fit differed among invasion stages (Fig. 3 and Supplementary Table 1). For the probability of introduction for cultivation, PD_{wMean} had the best model fit, suggesting that particularly alien species that have no widely distributed native relatives are likely to be introduced for cultivation. For subsequent naturalization of the cultivated species, unweighted PD_{Mean} had the best model fit, suggesting that for naturalizing, each native species is equally important, irrespective of its range size. For the probability of naturalized species becoming invasive, PD_{Min} had the best model fit, suggesting that distance to the closest relatives is most important for becoming widespread and dominant.

In conclusion, our results show that the direction of the effect of phylogenetic distance of alien plants to the native flora alternated from one invasion stage to another in Southern Africa. Although introduction for cultivation was positively associated with phylogenetic distance to the native flora, the opposite was true for subsequent naturalization success, but invasiveness was again positively associated with phylogenetic distance. Although the three different phylogenetic distance indices (PD_{Mean} , PD_{wMean} and PD_{Min}) showed the same patterns along the cultivation–naturalization–invasiveness transitions, different phylogenetic indices provided the best fit model in each stage. Thus, accounting for the different invasion stages and considering multiple phylogenetic distance metrics provide additional insights into Darwin's naturalization conundrum. Finally, yet importantly, the two seemingly opposing hypotheses of Darwin are not in conflict. Rather, the mechanisms underlying them act concurrently, and the one that dominates depends on the alien species' stage along the invasion process.

Methods

Study area. Our study focuses on the region of Southern Africa, which includes ten countries: Angola, Botswana, Eswatini, Lesotho, Malawi, Mozambique, Namibia, South Africa, Zambia and Zimbabwe, with a land area of approximately 4,000,000 km² (ref. ⁷²). Because we had separate native species occurrence data for the nine provinces of the country of South Africa, the total number of Southern African subregions was 18 (Fig. 1). Southern Africa has had a long history of plant introductions, which started in the late 18th century with the arrival of the European settlers⁷⁶.

The cultivated, naturalized and invasive alien species lists. The basis for our study is a list of over 8,000 taxa (including species and infraspecific taxa; here jointly referred to as 'species') reported as being cultivated in Southern Africa in the book *Cultivated Plants of Southern Africa*⁷⁷. To allow the alignment of the list of species with other datasets used in this study (see below), we validated and synchronized the taxonomic names according to The Plant List (v.1.1; <http://www.theplantlist.org/>) using the R package 'Taxonstand'⁷⁷. We then removed the species that, according to The Global Inventory of Floras and Traits database (GIFT)⁷⁸, are native to Southern Africa. GIFT data had been extracted using the R package 'RMySQL' v.0.10.20. Moreover, we removed ferns and gymnosperms because they are phylogenetically very distinct from angiosperms. The final list of introduced species comprised 5,091 cultivated alien angiosperm species.

To identify which of the cultivated alien species have become naturalized in Southern Africa, we used the Global Naturalized Alien Flora database (GloNAF)⁷⁹. GloNAF includes lists of naturalized vascular plant taxa for over 1,000 regions (usually administrative regions, such as countries, states and provinces) around the globe⁷⁹. Of the 5,091 cultivated species, 592 (11.6%) have managed to naturalize in at least one of the GloNAF regions⁷⁹ that are part of Southern Africa.

Because the invasiveness of naturalized species might vary among the different subregions of Southern Africa, and because data on invasiveness were not available for all those subregions, we restricted the analysis of how invasiveness relates to phylogenetic distance between the alien and native species to the subset of alien species that are naturalized in the country of South Africa ($n=524$). For this, we used the most recent list of invasive plant species for the country of South Africa by Zengeya and Wilson⁸⁰, in which invasive species are defined as widely spread alien species, in accordance with the unified invasion framework of Blackburn et al.⁴⁸.

Global angiosperm species pool. To be able to test whether the cultivated introduced plant species as well as the naturalized species are more or less phylogenetically related to the native flora of Southern Africa than expected by chance, we needed a list of the global species pool outside Southern Africa. To obtain such a list, we extracted all 324,808 accepted names of angiosperm taxa from The Plant List (<http://www.theplantlist.org/>, last accessed May 2019). From this list, we excluded the species that, according to GIFT, are native to Southern Africa ($n=24,974$), resulting in a global pool of 299,834 species that are not native to Southern Africa.

Phylogenetic tree of the global flora. To compute phylogenetic distance indices between the species from the global species pool outside Southern Africa and the native species of Southern Africa, we constructed a global phylogenetic tree of all plant species with accepted names according to The Plant List ($n=326,101$; for more details see van Kleunen et al.⁴⁹). In brief, as a basis for this phylogeny, we used the time-calibrated phylogeny of Smith and Brown⁷⁵, which is currently the most comprehensive phylogeny for seed plants. Of the global flora, 2,056 genera (14.0%) and 27,064 species (8.3%) were missing from the base tree. We manually added these missing species to their genus- or family-level root nodes in the base tree using the R package 'phytools'⁸¹. Because we were only interested in seed plants, we removed all non-seed plants from the phylogeny. This resulted in a global angiosperm phylogeny of 324,808 species (Fig. 2), of which ($n=24,974$) are native to Southern Africa.

Phylogenetic distance indices. Several metrics can be used to describe phylogenetic relatedness between plant species³². Because phylogenetic grain (the depth within a phylogeny) might affect the results of phylogenetic analyses⁷³, we selected the three most commonly used phylogenetic distance indices that describe phylogenetic composition at the deeper branches and at the tips of the tree³². First, to capture phylogenetic composition at the deeper branches of the tree, we calculated PD_{Mean} . This measure assumes that the whole native flora (irrespective of the abundances of the species) contributes equally to the success or failure of the alien species. Second, we calculated PD_{wMean} . This measure weights the phylogenetic distance by the occurrence frequency of the native species, assuming that the alien species will be more likely to interact with the widespread native species. As a measure of how widespread a native species is, we used the number of Southern African subregions (total $n=18$; Fig. 1) in which it occurs according to GIFT. Lastly, to capture phylogenetic composition at the tips of the tree, we calculated PD_{Min} . This measure assumes that the success of an alien species is driven by its distance to the closest native relative because they most likely use similar resources and share similar enemies and mutualists.

Using the global angiosperm phylogeny, we calculated for each species introduced for cultivation in Southern Africa its PD_{Mean} , PD_{wMean} and PD_{Min} to the native flora of Southern Africa. In addition, to test for the effect of phylogenetic relatedness on the probability that a naturalized species becomes invasive in the country of South Africa, we calculated for each cultivated naturalized species in the country of South Africa its PD_{Mean} , PD_{wMean} and PD_{Min} to the native flora of the country of South Africa ($n=15,382$, according to GIFT).

Statistical analysis. All statistical analyses were performed in R v.3.6.1 (ref. ⁸²). To test how introduction for cultivation, naturalization and invasiveness are associated with the phylogenetic distance indices, PD_{Mean} , PD_{wMean} and PD_{Min} , we used a series of complementary generalized linear models (GLMs) with binomial

error distribution with either a clog-log or a logit link function (see below). To see how the naturalization success of all non-native species relates to phylogenetic distance if not accounting for a potential cultivation bias, we first tested among the entire non-native global flora how the likelihood of naturalization in Southern Africa relates to the phylogenetic distance indices. Because in these analyses, the binomial response variable had many more zeros than ones, we used the clog-log link function⁴³. Then to account for cultivation bias, we broke this down into two further analyses. First, we tested using the entire global flora not native to Southern Africa how the likelihood of being introduced into Southern Africa for cultivation relates to the phylogenetic distance indices. We again used the clog-log link function in this analysis. Second, we tested among the species introduced for cultivation in Southern Africa how the likelihood of naturalization relates to the phylogenetic distance indices. Because for this analysis, the numbers of zeros and ones were more similar, we used the logit link function.

Finally, we tested for the subset of cultivated species that have become naturalized in the country of South Africa ($n=524$) how the likelihood of invasiveness relates to the phylogenetic distance indices. We also used the logit link function for this analysis. For the invasiveness analysis, we also assessed the robustness of the results when considering all naturalized species in the country of South Africa ($n=874$ species), irrespective of whether they were introduced for cultivation ($n=524$) or not ($n=350$). Because the results of models considering all naturalized species and the results of models considering only the naturalized cultivated species were comparable, we present the results from the analyses using the cultivated species in the main article and the results from the analyses using all species in the Supplementary Table 2. Because the analysis of invasiveness was only possible for the country of South Africa, whereas the other analyses were done for entire region of Southern Africa, there was a scale mismatch. To assess how sensitive our results are to this scale mismatch, we also performed the analysis of naturalization success among cultivated species for the country of South Africa only. The results of this analysis are comparable with those of the analysis for all of Southern Africa (Supplementary Table 5).

Because pairwise Pearson correlations between some of the phylogenetic indices were strong to very strong (PD_{Mean} and PD_{wMean} ; $r=0.99$; PD_{Mean} and PD_{Min} ; $r=0.39$; PD_{Min} and PD_{wMean} ; $r=0.42$), we ran separate GLMs for each. We then used Akaike's information criterion to identify which of the three phylogenetic distance indices resulted in the best model fit as in Malecore et al.⁴³. To test for potential nonlinear hump- or U-shaped relationships of introduction, naturalization and invasiveness probabilities with the phylogenetic distance indices, we also included a quadratic term for each distance index after centring the indices to means of zero. Moreover, to facilitate comparisons of the estimates within and between the models, we also scaled each explanatory variable to a standard deviation of one⁴⁴. When the quadratic term was not significant ($P>0.05$), we removed it from the model. To calculate the explained deviance (R^2) by the explanatory variables in our models, we calculated the Nagelkerke pseudo R^2 for each GLM⁴⁵ using the 'rcompanion' R package⁴⁶. To account for phylogenetic non-independence of the species, we also ran phylogenetically corrected GLMs using the 'phyloglm' function of the R package 'phyloglm', v.2.6.2 (ref. ⁴⁷). However, because the results were very similar to those of the standard GLMs, we present the results of the phylogenetically corrected GLMs in Supplementary Tables 3 and 4.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The core dataset of this study is available at (<https://doi.org/10.6084/m9.figshare.19597093.v1>). We also used the GloNAF dataset available at (<https://idata.idiv.de/DDM/Data>ShowData/257>), The Global Inventory of Floras and Traits database available up on request at (<https://gift.uni-goettingen.de/home>), The Plant List available at (<http://www.theplantlist.org/>), and the phylogeny of Smith and Brown (2018) available at (https://github.com/FePhyFoFum/big_seed_plant_trees).

Received: 25 November 2021; Accepted: 7 July 2022;

Published online: 11 August 2022

References

- Richardson, D. M. et al. Naturalization and invasion of alien plants: concepts and definitions. *Divers. Distrib.* **6**, 93–107 (2000).
- van Kleunen, M. et al. Global exchange and accumulation of non-native plants. *Nature* **525**, 100–103 (2015).
- Capinha, C., Essl, F., Seebens, H., Moser, D. & Pereira, H. M. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* **348**, 1248–1251 (2015).
- Vila, M. & Hulme, P. E. in *Impact of Biological Invasions on Ecosystem Services Vol. 12 Invading Nature – Springer Series in Invasion Ecology* (eds Vila, M. & Hulme, P. E.) 1–14 (Springer, 2017).
- Pyšek, P. et al. A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. *Glob. Chang. Biol.* **18**, 1725–1737 (2012).
- Pyšek, P. et al. Scientists' warning on invasive alien species. *Biol. Rev.* **95**, 1511–1534 (2020).
- Bacher, S. et al. Socio-economic impact classification of alien taxa (SEICAT). *Methods Ecol. Evol.* **9**, 159–168 (2018).
- Seebens, H. et al. No saturation in the accumulation of alien species worldwide. *Nat. Commun.* **8**, 14435 (2017).
- Seebens, H. et al. Projecting the continental accumulation of alien species through to 2050. *Glob. Chang. Biol.* **27**, 970–982 (2021).
- Kriticos, D. J., Sutherst, R. W., Brown, J. R., Adkins, S. W. & Maywald, G. F. Climate change and the potential distribution of an invasive alien plant: *Acacia nilotica* ssp. *indica* in Australia. *J. Appl. Ecol.* **40**, 111–124 (2003).
- Thuiller, W., Richardson, D. M. & Midgley, G. F. in *Biological Invasions* (ed. Nentwig, W.) 197–211 (Springer, 2007).
- Hobbs, R. J. in *Invasive Species in a Changing World* (eds Mooney, H. A. & Hobbs, R. J.) 55–64 (Island Press, 2000).
- Seebens, H. et al. Global trade will accelerate plant invasions in emerging economies under climate change. *Glob. Chang. Biol.* **21**, 4128–4140 (2015).
- Razanajatovo, M. et al. Plants capable of selfing are more likely to become naturalized. *Nat. Commun.* **7**, 13313 (2016).
- Bucharova, A. & van Kleunen, M. Introduction history and species characteristics partly explain naturalization success of North American woody species in Europe. *J. Ecol.* **97**, 230–238 (2009).
- Ordonez, A., Wright, I. J. & Olf, H. Functional differences between native and alien species: a global-scale comparison. *Funct. Ecol.* **24**, 1353–1361 (2010).
- van Kleunen, M., Weber, E. & Fischer, M. A meta-analysis of trait differences between invasive and non-invasive plant species. *Ecol. Lett.* **13**, 235–245 (2010).
- van Kleunen, M., Dawson, W. & Maurel, N. Characteristics of successful alien plants. *Mol. Ecol.* **24**, 1954–1968 (2015).
- Essl, F. et al. Drivers of the relative richness of naturalized and invasive plant species on Earth. *AoB Plants* **11**, plz051 (2019).
- Winkler, D. E., Gremer, J. R., Chapin, K. J., Kao, M. & Huxman, T. E. Rapid alignment of functional trait variation with locality across the invaded range of Sahara mustard (*Brassica tournefortii*). *Am. J. Bot.* **105**, 1188–1197 (2018).
- Divíšek, J. et al. Similarity of introduced plant species to native ones facilitates naturalization, but differences enhance invasion success. *Nat. Commun.* **9**, 4631 (2018).
- Banerjee, A. K., Prajapati, J., Bhowmick, A. R., Huang, Y. & Mukherjee, A. Different factors influence naturalization and invasion processes – a case study of Indian alien flora provides management insights. *J. Environ. Manag.* **294**, 113054 (2021).
- Ni, M. et al. Invasion success and impacts depend on different characteristics in non-native plants. *Divers. Distrib.* **27**, 1194–1207 (2021).
- Fristoe, T. S. et al. Dimensions of invasiveness: links between local abundance, geographic range size, and habitat breadth in Europe's alien and native floras. *Proc. Natl Acad. Sci. USA* **118**, e2021173118 (2021).
- Omer, A. et al. Characteristics of the naturalized flora of Southern Africa largely reflect the non-random introduction of alien species for cultivation. *Ecography* **44**, 1812–1825 (2021).
- Pyšek, P. et al. Naturalization of central European plants in North America: species traits, habitats, propagule pressure, residence time. *Ecology* **96**, 762–774 (2015).
- Omer, A., Kordofani, M., Gibreel, H. H., Pyšek, P. & van Kleunen, M. The alien flora of Sudan and South Sudan: taxonomic and biogeographical composition. *Biol. Invasions* **23**, 2033–2045 (2021).
- Duncan, R. P. & Williams, P. A. Darwin's naturalization hypothesis challenged. *Nature* **417**, 608–609 (2002).
- Daehler, C. C. Darwin's naturalization hypothesis revisited. *Am. Nat.* **158**, 324–330 (2001).
- Pyšek, P. Is there a taxonomic pattern to plant invasions? *Oikos* **82**, 282–294 (1998).
- Tan, J., Pu, Z., Ryberg, W. A. & Jiang, L. Resident–invader phylogenetic relatedness, not resident phylogenetic diversity, controls community invasibility. *Am. Nat.* **186**, 59–71 (2015).
- Thuiller, W. et al. Resolving Darwin's naturalization conundrum: a quest for evidence. *Divers. Distrib.* **16**, 461–475 (2010).
- Loiola, P. P. et al. Invaders among locals: alien species decrease phylogenetic and functional diversity while increasing dissimilarity among native community members. *J. Ecol.* **106**, 2230–2241 (2018).
- Lososová, Z. et al. Alien plants invade more phylogenetically clustered community types and cause even stronger clustering. *Glob. Ecol. Biogeogr.* **24**, 786–794 (2015).
- Marx, H. E., Giblin, D. E., Dunwiddie, P. W. & Tank, D. C. Deconstructing Darwin's naturalization conundrum in the San Juan Islands using community phylogenetics and functional traits. *Divers. Distrib.* **22**, 318–331 (2016).
- Darwin, C. *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).

37. Proches, Š., Wilson, J. R. U., Richardson, D. M. & Rejmánek, M. Searching for phylogenetic pattern in biological invasions. *Glob. Ecol. Biogeogr.* **17**, 5–10 (2008).
38. Diez, J. M., Sullivan, J. J., Hulme, P. E., Edwards, G. & Duncan, R. P. Darwin's naturalization conundrum: dissecting taxonomic patterns of species invasions. *Ecol. Lett.* **11**, 674–681 (2008).
39. Cadotte, M. W., Campbell, S. E., Li, S. P., Sodhi, D. S. & Mandrak, N. E. Preadaptation and naturalization of nonnative species: Darwin's two fundamental insights into species invasion. *Annu. Rev. Plant Biol.* **69**, 661–684 (2018).
40. van Kleunen, M., Bossdorf, O. & Dawson, W. The ecology and evolution of alien plants. *Annu. Rev. Ecol. Evol. Syst.* **49**, 25–47 (2018).
41. Park, D. S., Feng, X., Maitner, B. S., Ernst, K. C. & Enquist, B. J. Darwin's naturalization conundrum can be explained by spatial scale. *Proc. Natl. Acad. Sci. USA* **117**, 10904–10910 (2020).
42. Diez, J. M. et al. Learning from failures: testing broad taxonomic hypotheses about plant naturalization. *Ecol. Lett.* **12**, 1174–1183 (2009).
43. Malecore, E. M., Dawson, W., Kempel, A., Müller, G. & van Kleunen, M. Nonlinear effects of phylogenetic distance on early-stage establishment of experimentally introduced plants in grassland communities. *J. Ecol.* **107**, 781–793 (2019).
44. Schaefer, H., Hardy, O. J., Silva, L., Barraclough, T. G. & Savolainen, V. Testing Darwin's naturalization hypothesis in the Azores. *Ecol. Lett.* **14**, 389–396 (2011).
45. Strauss, S. Y., Webb, C. O. & Salamin, N. Exotic taxa less related to native species are more invasive. *Proc. Natl. Acad. Sci. USA* **103**, 5841–5845 (2006).
46. Li, S.-p. et al. The effects of phylogenetic relatedness on invasion success and impact: deconstructing Darwin's naturalisation conundrum. *Ecol. Lett.* **18**, 1285–1292 (2015).
47. Pellock, S., Thompson, A., He, K., Mecklin, C. & Yang, J. Validity of Darwin's naturalization hypothesis relates to the stages of invasion. *Community Ecol.* **14**, 172–179 (2013).
48. Blackburn, T. M. et al. A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 (2011).
49. van Kleunen, M. et al. Economic use of plants is key to their naturalization success. *Nat. Commun.* **11**, 3201 (2020).
50. Broennimann, O. et al. Distance to native climatic niche margins explains establishment success of alien mammals. *Nat. Commun.* **12**, 2353 (2021).
51. Carboni, M. et al. What it takes to invade grassland ecosystems: traits, introduction history and filtering processes. *Ecol. Lett.* **19**, 219–229 (2016).
52. Milbau, A. & Stout, J. C. Factors associated with alien plants transitioning from casual, to naturalized, to invasive. *Conserv. Biol.* **22**, 308–317 (2008).
53. Dawson, W., Burslem, D. F. R. P. & Hulme, P. E. Factors explaining alien plant invasion success in a tropical ecosystem differ at each stage of invasion. *J. Ecol.* **97**, 657–665 (2009).
54. Rejmánek, M. in *Invasive Species and Biodiversity Management* (eds Schei, P. J. & Vilken, A.) 79–102 (Kluwer Academic, 1998).
55. Rejmánek, M. A theory of seed plant invasiveness: the first sketch. *Biol. Conserv.* **78**, 171–181 (1996).
56. Maurel, N., Hanspach, J., Kuhn, I., Pysek, P. & van Kleunen, M. Introduction bias affects relationships between the characteristics of ornamental alien plants and their naturalization success. *Glob. Ecol. Biogeogr.* **25**, 1500–1509 (2016).
57. Glen, H. F. *Cultivated Plants of Southern Africa: Botanical Names, Common Names, Origins, Literature* (National Botanical Institute, 2002).
58. Reichard, S. H. & White, P. Horticulture as a pathway of invasive plant introductions in the United States. *Bioscience* **51**, 103–113 (2001).
59. Faulkner, K. T., Robertson, M. P., Rouget, M. & Wilson, J. R. U. Understanding and managing the introduction pathways of alien taxa: South Africa as a case study. *Biol. Invasions* **18**, 73–87 (2016).
60. Dodd, A. J., Burgman, M. A., McCarthy, M. A. & Ainsworth, N. The changing patterns of plant naturalization in Australia. *Divers. Distrib.* **21**, 1038–1050 (2015).
61. Lambdon, P.-W. et al. Alien flora of Europe: species diversity, temporal trends, geographical patterns and research needs. *Preslia* **80**, 101–149 (2008).
62. Bennett, B. M. Naturalising Australian trees in South Africa: climate, exotics and experimentation. *J. South Afr. Stud.* **37**, 265–280 (2011).
63. Richardson, D. M. et al. in *Biological Invasions in South Africa* (eds van Wilgen, B. W. et al.) 67–96 (Springer, 2020).
64. Li, S.-p. et al. Contrasting effects of phylogenetic relatedness on plant invader success in experimental grassland communities. *J. Appl. Ecol.* **52**, 89–99 (2015).
65. Duarte, M., Verdú, M., Cavieres, L. A. & Bustamante, R. O. Plant–plant facilitation increases with reduced phylogenetic relatedness along an elevation gradient. *Oikos* **130**, 248–259 (2021).
66. Verdú, M., Rey, P. J., Alcántara, J. M., Siles, G. & Valiente-Banuet, A. Phylogenetic signatures of facilitation and competition in successional communities. *J. Ecol.* **97**, 1171–1180 (2009).
67. Valiente-Banuet, A. & Verdú, M. Plant facilitation and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* **44**, 347–366 (2013).
68. Anacker, B. L. & Strauss, S. Y. Ecological similarity is related to phylogenetic distance between species in a cross-niche field transplant experiment. *Ecology* **97**, 1807–1818 (2016).
69. Dostál, P. Plant competitive interactions and invasiveness: searching for the effects of phylogenetic relatedness and origin on competition intensity. *Am. Nat.* **177**, 655–667 (2011).
70. Levin, S. C., Crandall, R. M., Pokoski, T., Stein, C. & Knight, T. M. Phylogenetic and functional distinctiveness explain alien plant population responses to competition. *Proc. R. Soc. B* **287**, 20201070 (2020).
71. Williams, E. W., Zeldin, J., Semski, W. R., Hipp, A. L. & Larkin, D. J. Phylogenetic distance and resource availability mediate direction and strength of plant interactions in a competition experiment. *Oecologia* **197**, 459–469 (2021).
72. Bezing, S. B., Davies, J. T., Yessoufou, K., Maurin, O. & Van der Bank, M. Revisiting Darwin's naturalization conundrum: explaining invasion success of non-native trees and shrubs in Southern Africa. *J. Ecol.* **103**, 871–879 (2015).
73. Trotta, L. B., Siders, Z. A., Sessa, E. B. & Baiser, B. The role of phylogenetic scale in Darwin's naturalization conundrum in the critically imperilled pine rockland ecosystem. *Divers. Distrib.* **27**, 618–631 (2021).
74. Sol, D. et al. A test of Darwin's naturalization conundrum in birds reveals enhanced invasion success in the presence of close relatives. *Ecol. Lett.* **25**, 661–672 (2022).
75. Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
76. Henderson, L. Comparisons of invasive plants in Southern Africa originating from southern temperate, northern temperate and tropical regions. *Bothalia* **36**, 201–222 (2006).
77. Cayuela, L., Stein, A. & Oksanen, J. Taxonstand: Taxonomic Standardization of Plant Species Names. R package version 2.2. <https://CRAN.R-project.org/package=Taxonstand> (R Foundation for Statistical Computing, Vienna, 2019).
78. Weigelt, P., König, C. & Kreft, H. GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography. *J. Biogeogr.* **47**, 16–43 (2020).
79. van Kleunen, M. et al. The Global Naturalized Alien Flora (GloNAF) database. *Ecology* **100**, e02542 (2019).
80. Zengeya, T. A. & Wilson, J. R. (eds) *The Status of Biological Invasions and Their Management in South Africa in 2019* (South African National Biodiversity Institute and DSI-NRF Centre of Excellence for Invasion Biology, 2021).
81. Revell, L. J. phytobase: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
82. R: A Language and Environment for Statistical Computing v.3.6.1 (R Foundation for Statistical Computing, 2019).
83. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* Vol. 574 (Springer, 2009).
84. Schielzeth, H. Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**, 103–113 (2010).
85. Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).
86. rcompanion: Functions to support extension education program evaluation v. 2.4.1 (R Foundation for Statistical Computing, 2021).
87. Tung Ho, L. S. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).

Acknowledgements

A.O. thanks the Ministry of Higher Education of Sudan and the University of Konstanz for funding, and the International Max Planck Research School for Quantitative Behaviour, Ecology and Evolution for support. M.v.K. thanks the German Research Foundation DFG for funding (grants 264740629 and 432253815). M.R. thanks the DFG for funding (grant RA 3009/1-1). P.P. and J.P. were supported by EXPRO grant no. 19-28807X (Czech Science Foundation) and long-term research development project RVO 67985939 (Czech Academy of Sciences). F.E. appreciates funding from the Austrian Science Foundation FWF (grant I 3757-B29). We thank C. Gommel, K. Mamona, V. Pasqualeto and B. Rüter for help with data extraction, and L. Henderson for providing lists of naturalized species for South Africa.

Author contributions

A.O. and M.v.K. conceived and designed the study. A.O., Q.Y. and M.v.K. compiled data used in this study. A.O. led the analysing with major inputs from M.v.K. and further inputs from P.P., S.D., T.F. and Q.Y. A.O. led the writing with major inputs from M.v.K. and further inputs from T.F., Q.Y., M.R., P.W., H.K., W.D., S.D., F.E., J.P. and P.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-022-01216-9>.

Correspondence and requests for materials should be addressed to Ali Omer.

Peer review information *Nature Plants* thanks Richard Duncan and Shaopeng Li for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection We obtained GloNAF and GIFT datasets using the R package "RMySQL" (version 0.10.20).

Data analysis The main data process and analysis were done using R version 4.1.0. We used the R "Taxonsatad" (version 2.2) to validate and synchronized the taxonomic names according to The Plant List. We constructed a global phylogenetic tree of all plant species with accepted names according to The Plant List using the R package "phytools" (version 1.0-1). We did the phylogenetically corrected GLMs using the 'phyloglm' function of the R package 'phylolm' (version 2.6.2). We calculated the Nagelkerke Pseudo R² for each GLM using the "rcompanion" R package (version 2.4.1). The map in Figure 1 was produced using ArcMap Software version 10.8 (Website: <https://desktop.arcgis.com/en/arcmap/>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The core dataset of this study is available at (<https://doi.org/10.6084/m9.figshare.19597093.v1>). We also used the GloNAF dataset available at (<https://idata.idiv.de/DDM/Data>ShowData/257>), The Global Inventory of Floras and Traits database available up on request at (<https://gift.uni-goettingen.de/home>), The Plant List available at (<http://www.theplantlist.org/>), and the phylogeny of Smith and Brown (2018) available at (https://github.com/FePhyFoFum/big_seed_plant_trees).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We collected data on 5,091 alien angiosperm species that have been introduced to Southern Africa for cultivation. Then combined with lists of the global flora non-native to Southern Africa ($n= 299,834$ species), the naturalized and native species within the region of Southern Africa, and the species that have become invasive within the country of South Africa. This novel combination of datasets was used to test the Darwin's naturalization conundrum at the different stages of the invasion process. Specifically, we showed how phylogenetic relatedness to native species is correlated with the subsequent stages of introduction, naturalization and invasiveness of alien species.

Research sample

Our study based on several existing datasets. The list of cultivated flora of Southern Africa was collected from the book 'Cultivated Plants of Southern Africa: Botanical Names, Common Names, Origins, Literature' (Glen 2000, Jacana: National Botanical Institute), the list of naturalized flora of Southern Africa was extracted from Global Naturalized Alien Flora (GloNAF, van Kleunen et al. 2019: Ecology: DOI: 10.1002/ecy.2542), the list of invasive flora of South Africa from (Zengetya and Wilson 2021: South African National Biodiversity Institute: DIO: <http://dx.doi.org/10.5281/zenodo.3947613>), and the list of native flora of Southern Africa from Global Inventory of Floras and Traits (GIFT; Weigelt et al. 2019; Journal of Biogeography; DIO: 10.1111/jbi.13623). Finally, the global angiosperm phylogeny was based on the phylogeny of Smith and Brown (2018; American journal of Botany; DIO: 10.1002/ajb2.1019). We used these datasets because, to our knowledge, they represent the most comprehensive available data on the cultivated flora of Southern Africa.

Sampling strategy

We extracted the list of over 8, 000 species that are known to be cultivated in Southern Africa from the book 'Cultivated Plants of Southern Africa' (Glen 2000, Jacana: National Botanical Institute). Then, we validated and synchronized the taxonomic names according to The Plant List (version 1.1; <http://www.theplantlist.org/>) using the R package 'Taxonstand'. Moreover, we removed the species that, according to The Global Inventory of Floras and Traits database (GIFT), are native to Southern Africa. Finally, we removed ferns and gymnosperms as they are phylogenetically very distinct from the angiosperms. The final list of introduced species comprised 5,091 cultivated alien angiosperm species.

Data collection

The list of cultivated flora of Southern Africa was manually digitized to an Excel sheet by C. Gommel, K. Mamonova, V. Pasqualetto and B. Rüter.

Timing and spatial scale

GloNAF and GIFT datasets were accessed on 23 September 2020. The Smith and Brown Phylogeny was accessed on 18 January 2019. The invasive Flora of South Africa were downloaded on 1 April 2021. GloNAF, GIFT, and Smith and Brown Phylogeny datasets have a global spatial scale.

Data exclusions

To allow the alignment of the list of species with other datasets used in this study, we validated and synchronized the taxonomic names according to The Plant List (version 1.1; <http://www.theplantlist.org/>), we then removed species with unaccepted names. Moreover, we removed ferns and gymnosperms as they are phylogenetically very distinct from the angiosperms.

Reproducibility

As our study mainly analyzed existing datasets, we did not test for reproducibility. However, we will provide a link to all the processed data.

Randomization

We analyzed existing datasets, so we did not need to allocate samples to groups.

Blinding

Blinding was not relevant to this study. Because the study has no experimental groups (it analysed existing datasets). All datasets included in this study were extracted from authoritative sources and from already published literature as mentioned above.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging