Evolution of the latitudinal diversity gradient in the hyperdiverse ant genus *Pheidole*

Evan P. Economo[1,2*], Jen-Pan Huang[2], Georg Fischer[1], Eli M. Sarnat[1], Nitish Narula[1], Milan Janda[3,4], Benoit Guénard[5], John T. Longino[6], L. Lacey Knowles[2]

*correspondence: Evan P. Economo, evaneconomo@gmail.com

[1]*Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan, 904-0495*

[2]*Department of Ecology & Evolutionary Biology, Museum of Zoology, University of Michigan, USA*

[3]*National Laboratory for Ecological Analysis and Synthesis (LANASE), ENES, UNAM, Morelia, Mexico*

[4]*Biology Centre of Czech Academy of Sciences, Ceske Budejovice, Czech Republic*

[5]*The University of Hong Kong, School of Biological Sciences, Hong Kong, SAR, China*

[6]*Department of Biology, University of Utah, USA*

Abstract

Aim

The latitudinal diversity gradient is the dominant pattern of life on Earth, but a consensus understanding of its origins has remained elusive. The analysis of recently diverged, hyper-rich invertebrate groups provides an opportunity to analyze latitudinal patterns with the statistical power of large trees while minimizing potentially confounding variation in ecology and history. Here, we synthesize global phylogenetic and macroecological data on a hyperdiverse (>1100 species) ant radiation, *Pheidole*, and evaluate the roles of three general explanations for the latitudinal gradient: variation in diversification rate, tropical conservatism, and ecological regulation.

Location

Global.

Time Period

The past 35 million years.

Major taxa studied

The ant genus *Pheidole*.

Methods

We assembled geographic data for 1499 species and morphospecies, and inferred a dated phylogeny of *Pheidole* of 449 species, including 150 species newly sequenced for this study. We inferred macroevolutionary rates with BAMM and GeoSSE, tested for correlations between diversification rate and latitude, and examined patterns of diversification as *Pheidole* spread around the globe.

Results

We found that *Pheidole* diversification occurred in series of bursts when new continents were colonized, followed by a slowdown in each region. There was no evidence of systematic variation of net diversification rates with latitude across any of the methods. Additionally, we found latitudinal affinity is moderately conserved with a Neotropical ancestor and phylogenetic inertia alone is sufficient to produce the gradient pattern.

Main Conclusions

Overall our results are consistent with tropical conservatism explaining the diversity gradient, while providing no evidence that diversification rate varies systematically with latitude. There is evidence of ecological regulation on continental scales through the pattern of diversification after colonization. These results shed light on the mechanisms underlying the diversity gradient, while contributing toward a much-needed invertebrate perspective on global biodiversity dynamics.

INTRODUCTION

Understanding how ecological and evolutionary processes interact with historical factors to shape global biodiversity patterns remains a major goal of biology. The latitudinal diversity gradient (LDG) is the most general biogeographic pattern, yet we still lack a consensus understanding of its mechanisms[1-4]. This is likely because many biological, physical, and historical factors that could plausibly affect diversity vary systematically with latitude, and thus a large number of hypotheses have been developed to explain the pattern. However, testing the predictions of different hypotheses empirically and evaluating their relative merits has proven to be a challenge.

Recently, the synthesis of large-scale geographic datasets along with large-scale phylogenetic data has provided new opportunities for empirical evaluation of hypotheses for the mechanisms underlying the LDG. These tests have mainly focused on vertebrates[5-11] and woody plants[12], since those are the taxa with large-scale comprehensive data available. Several pioneering studies have examined insect diversification with latitude[13-16], but these have been focused on relatively small groups and/or lacked comprehensive distributional data.

Among invertebrates, ants are emerging as an exemplar taxon for global biodiversity studies. Ants are ecologically dominant in most terrestrial ecosystems and are, for an insect group, relatively well documented scientifically. Moreover, their diversity is high, but not intractably so, with richness on the same order as major vertebrate groups (~15,000 described ant species). Recently, a new comprehensive dataset has been compiled which gives the known geographic distribution of all described ant species across >400 geographic regions around the globe[17]. These data, combined with progress toward reconstructing the ant tree of life[15,18-20],

allow for inferences of the evolutionary underpinnings of large-scale diversity patterns in ants.

Here, we use globally distributed, hyperdiverse (>1100 described species) ant genus *Pheidole* as a model taxon to test hypotheses for the latitudinal diversity gradient. While over a hundred hypotheses have been proposed to explain the gradient[1-3], these can roughly be sorted into three umbrella hypotheses which we use to frame our study: i) the *Diversification Rate hypothesis* (DRH), ii) the *Tropical Conservatism Hypothesis* (TCH), and iii) the *Ecological Regulation Hypothesis* (ERH).

First, the *Diversification Rate Hypothesis* posits that there is some causal factor that affects speciation and/or extinction rates and varies with latitude. This leads to a latitudinal disparity in species accumulation rate that underlies the gradient, rather than any top down regulation. Many such factors have been proposed. For example, temperature may affect mutation rates, which in turn could affect the rates of evolution of reproductive incompatibilities[21]. Or, extinction rates could be higher in the temperate zone to due greater climatic variability. The prediction of the DRH is straightforward: net diversification rate inferred from a phylogeny should be higher in tropical lineages compared with extratropical lineages.

Second, the *Tropical Conservatism Hypothesis* (TCH)[4,22] posits that the relative youth of colder temperate biomes combined with the inertia of phylogenetic niche conservatism [23] has limited the accumulation of diversity in the temperate zone. In this scenario, net diversification rates do not necessarily vary with latitude, and main difference in richness is due to time for diversification. This idea is based on the fact that historically the Earth has been much warmer than it is now, and much of what is now the temperate zone was covered by “megathermal” biomes. This hypothesis is supported by the fossil record: for example many lineages that used to

occur in the Palearctic are now limited to tropical latitudes. This is true for ants as well; the Baltic amber ant fauna from the late Eocene has been shown to have more affinity to modern Indo-Australian faunas than modern Palaearctic faunas[24]. The main prediction of this hypothesis is that the ancestral region of most groups is the tropics, transitions out of the tropical zone are rare, and thus the temperate clades are younger and nested within tropical clades. The fact that transition from the tropical to the temperate zones should be difficult is consistent with the many nontrivial adaptations that ectothermal organisms such as ants need in order to solve the problems associated with overwintering in the higher latitudes. One additional key prediction of the TCH is that temperate diversification matches the timing of global cooling: specifically, that diversification of cold-adapted lineages accelerated after the Oligocene cooling 34mya.

Finally, the *Ecological Regulation Hypothesis* (ERH) posits that due to some causal factor that varies with latitude, for example available energy, more species can coexist locally and regionally in tropical ecosystems than in temperate ecosystems. In this case, diversity is saturated at or near some ecological limits, and this "carrying capacity" of species varies with latitude regulating diversity from the top-down[25]. This is perhaps due to limitations on species coexistence that are driven by productivity or other factors[4,26]. Speciation and extinction rates may vary over time to regulate richness at the requisite quota for a geographic region, but are not causally responsible for the disparity in diversity. Likewise, latitudinal affinity may be highly conserved or evolve quickly, but it would be immaterial to the origins of the gradient if diversity is regulated at levels that vary with latitude.

In a parallel study [27], Economo et al. examined latitudinal patterns across 262 ant clades and tested hypotheses for the latitudinal gradient. That taxon-wide analysis focused on deeper timescales, and lacks phylogenetic resolution within recent radiations. They found that tropical

lineages are more ancient than extratropical lineages, which mainly arose since the Oligocene cooling (past 34my), consistent with the TCH. Further, they found that diversification rate is highly heterogeneous but uncorrelated with latitude among ant clades, inconsistent with the DRH. Due to the limitations of phylogenetic data at such broad scales, they could not explicitly test for ecological regulation (ERH).

As with other studies on broad taxonomic scales (e.g. birds[6], mammals[8,28]), the analyses across all ants provide the advantages of the statistical power of large numbers and a deep-time perspective. However, as many ecological, functional trait, and historical factors could contribute to variation in macroevolutionary rates at these deeper phylogenetic scales, this variation among clades may obscure underlying latitudinal effects that could be detectable among more similar, closely related lineages. For example, ant diversification has been shown to be related to functional traits varying among genera[29]. Moreover, latitudinal gradients are often present within individual clades that evolved recently, and macrophylogenetic studies may miss the relevant scale of variation. Thus, the analysis of closely related lineages within younger, hyper-rich radiations provides a necessary complement to large-scale taxon-wide studies. In highly diverse groups, these radiations can provide both the statistical power of large numbers, while controlling to some degree for differences in ecology, functional traits, and historical factors.

The global radiation of *Pheidole* arose entirely since the Oligocene cooling (last 34my), during which time it has evolved a latitudinal gradient echoing the pattern for all ants. Thus, *Pheidole* presents an opportunity to examine diversification dynamics in this most recent period since the Oligocene, a period where many ant lineages transitioned out of the tropics, complementing the deeper timescales of the ant-wide study. While the low number of older extratropical ant lineages is consistent with the TCH, there is still an open question of whether

niche conservatism or diversification rate differences explain the emergence of the gradient since the Oligocene. According to the TCH, the tropical ancestry of *Pheidole* combined with phylogenetic niche conservatism is sufficient to explain why there are more species in the tropics. The DRH predicts that *Pheidole* diversified more rapidly in the tropics, and this explains the disparity in diversity. Finally, we examine whether there is evidence for ecological limits to *Pheidole* diversification (ERH) by examining whether the radiation is undergoing pulse-like bursts of diversification. If the latter, we would expect a series of pulses as the genus colonized different regions around the globe.

Here, we reconstruct a new global *Pheidole* phylogeny, the most comprehensive to date, increasing substantially the taxonomic and geographic coverage from previous studies of the genus[30-33]. We use the new phylogeny and geographic data from the GABI database to test predictions of the three umbrella hypotheses for the latitudinal gradient. There is no biological reason why two or more of these mechanisms cannot be simultaneously operating (e.g. diversity can be regulated and speciation rate can vary systematically with latitude, or niche conservatism can occur along with diversity regulation). Thus, our goal is to rule out hypotheses rather than isolate a single exclusive answer. The analysis of this famously hyperdiverse radiation will advance our general understanding of the latitudinal gradient, the most pervasive pattern of life on Earth.

METHODS:

Geographic Data

Our geographic data are based primarily on the Global Ant Biodiversity Informatics Project (GABI) database[17] which can be viewed through the website antmaps.org[37], and secondarily on the personal collection records of the authors. The former focuses on described species, while the second was used to supplement data on morphospecies for taxa included on the tree.

Phylogeny reconstruction

Taxon Selection: Compared with many other large ant radiations, the effort to reconstruct the phylogenetic history of *Pheidole* is relatively far along. A series of studies, beginning with Moreau[30] and followed by other studies[31-33] has produced a broad picture of diversification in the genus. However, for the purposes of understanding geographic patterns of diversification, having a larger, and more proportionally sampled phylogeny will provide additional statistical power and more robust results. Thus, we continued sampling *Pheidole* taxa for sequencing, focusing on sampling more taxa from the Neotropics, Madagascar, and SE Asia. In all, we increased the species number from 295 taxa in the most recent global *Pheidole* phylogeny[31] to 449 taxa in the current contribution (Table S2).

Estimation of Sampling Completeness: One source of uncertainty in large-scale analyses of diversity is bias in taxonomic completeness overall and among different areas, particularly in relatively poorly known groups such as insects. While there is still a pronounced latitudinal

gradient in *Pheidole* even among described species, there are undoubtedly many undescribed species in the genus, and it is probable they are disproportionately found in the tropics. While accounting for unobserved species is a challenge in any analysis, we devised a back-of-the-envelope method to calculate sampling completeness across areas given the information in hand, and use these estimates in our analysis of diversification rate. The details of our calculation are in the Supplementary Information.

DNA Sequencing: We added 154 *Pheidole* taxa to existing molecular datasets by sequencing three loci, COI, LR, and H3L. Ant samples from field collections fixed in 95% EtOH were extracted for DNAs using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). The whole ant body was incubated in the extraction buffer without grinding during the first step, and then the complete ant specimen was removed before filtering and cleaning the extracts via a provided column. Extracted DNAs were subsequently used for PCR reactions for one mitochondrial (CO1)[38] and two nuclear (His3.3B and Lop1) regions. Each reaction contained 0.5 ul of extracted DNA, 1ul of 10 × buffer, 0.75 ul of MgCl2, 0.5 ul of 10mN dNTPs, 0.2 ul of 1% BSA, 0.4ul of of each primer, 0.04ul of Tag DNA polymerase (Invitrogen, USA), and ddH_2O to make a total of 10 ul reaction. Standard PCR procedures were employed with annealing temperatures of 52, 60, and 60 C for CO1, His3.3B, and Lop1 regions, respectively. The amplicons were sequenced via a ABI^{3700} machine by the Sequencing Core at the University of Michigan. Sequences were checked using SeqMan (DNAStar Inc., USA).

Phylogenetic inference: We used Bayesian methods to infer a dated *Pheidole* phylogeny including 449 ingroup taxa (Table S2), using eight nuclear loci [His3.3B (histone H3.3B F1

copy), Lop1 (long wavelength sensitive opsin 1), GRIK2 (glutamate receptor ionotropic, kainate 2-like), unc_4 (unc-4 homeodomain gene), LOC15 (uncharacterized locus LOC15), CAD (carbomoylphosphate synthase), EF-1α F2 (elongation factor 1-alpha F2), Top1 (DNA topoisomerase 1)], and one mitochondrial locus [CO1 (cytochrome oxidase 1)]. To generate codon-aware alignments for these loci, we first searched NCBI's non-redundant CDS database[39] for reliable amino acid sequences for all loci and retrieved such sequences for seven of the nine loci with the following accession numbers: AIM2284.1 (CAD), ABW70333.1(CO1), EZA53539.1 (EF-1α F2), EGI60526.1 (His3.3B), ABW36758.1 (Lop1), EGI59282.1 (unc-4), and AIM43286.1 (Top1). These sequences were used as references for generating codon-aware alignments. The CAD, unc-4, and Top1 alignments generated using MAFFT v7.205 [40] (--retree 4; --maxiterate 1000) showed no frameshift mutations and/or insertions and deletions. However, the CO1, EF-1α F2, His3.3B, and Lop1 alignments did not match the reference sequences, showing disruptions in the translated amino acid alignments (such as the presence of numerous stop codons). For these loci, we used a codon-aware alignment software, MACSE v1.01b [41], to generate the alignments. Reverse translations of the reliable amino acid reference sequences, accounting for all possibilities at each codon position, were passed as reliable input sequences to the software, we were able to assign codon positions within the exons in these seven loci. The resulting alignments were manually inspected and cleaned using Geneious R8 software. Furthermore, we identified, extracted, and separately aligned intronic regions wherever necessary. The remaining two loci, LOC15 and GRIK-2, were aligned using MAFFT. We concatenated all nine alignments and once again manually cleaned the master alignment, resulting in an alignment containing 8839 sites.

We used PartitionFinder v1.1.1[42] to determine the partitioning scheme and corresponding models of molecular evolution. The model scope included HKY, HKY+Γ, SYM, SYM+Γ, GTR, GTR+Γ, TrN, TrN+Γ, K80, K80+Γ, TrNef, TrNef+Γ, JC, and JC+Γ, branch lengths were set to 'linked', and model selection and comparison was set to Bayesian Information Criterion (BIC). PartitionFinder identified an optimal scheme containing 16 partitions (Table S3). A ClockstaR analysis indicated that a single linked clock across all partitions.

Our primary phylogenetic inference was conducted in BEAST2 v2.1.3[43], but we first performed maximum likelihood (ML) reconstruction in RAxML v8.0.25[44]. Using the partitioning scheme described above and the GTR+Γ model, we ran 75 ML inferences with 1000 bootstraps to find the ML tree. Using the *chronos* function in the *ape* package in R[45], we scaled the tree by calibrating the root node to a range of 50-60my. This tree was used as the starting tree for the BEAST2 analyses, but the topology was not fixed. The root node (mrca of *Pheidole* and *Cephalotes*) calibration was set to normal distribution (mean: 58.0 mya, sigma, 4.8my), following results for that node from an extensive dating analysis on the subfamily Myrmicinae[19]. We used a relaxed lognormal clock model linked across partitions (according to the ClockstaR analysis), and used the partitioning scheme and models identified with PartitionFinder. Six independent analyses were run and chains were stopped between 45 and 80 million generations, after we observed convergence using Tracer software v1.6.0 (Rambaut 2014). We discarded the leading 33% of saved states as burnin, combined the remaining trees from all runs to create the posterior set, and generated the Maximum Clade Credibility tree and nodes set to median height. This tree was used for all subsequent analyses.

Macroevolutionary rate inference

We took two complimentary approaches to estimating macroevolutionary rates and potential dependencies on latitude, based on the programs BAMM[46] and GeoSSE[47]. We sought conclusions about our data that are robust to methodological assumptions and implementation. In general, our focus was on evaluating evidence for latitudinal differences in net diversification rate and not speciation and extinction individually, because the latter are more difficult to assess.

BAMM and GeoSSE each have strengths and weaknesses. One advantage of BAMM is that complex mixture models can be assessed with rate shifts across the tree, including accelerating and decelerating diversification rates. While trait-dependent diversification models are not fit directly, trait-diversification correlations can be assessed *post hoc* using structured rate permutations that account for phylogenetic dependency[48]. We use BAMM to test for correlations between latitude and net diversification rate, and evaluate evidence of decelerating diversification (ecological regulation of diversity) overall and in relation to the colonization of continents. GeoSSE, on the other hand, fits combined geographic speciation-extinction-dispersal models that can be formally tested against geography-independent models. However, GeoSSE does not allow for rate-shifts that are unrelated to the focal trait and considers time-invariant models only (other than related to trait changes). These restrictions can lead to model misspecification and poor performance under some conditions, particularly type-I errors[49]. We use GeoSSE to evaluate trait (i.e. latitude) dependent macroevolution and rates of dispersal between the regions.

BAMM approach: Net-diversification, speciation and extinction rates through time based on the *Pheidole* species tree were estimated using the program BAMM V2.5 The initial values for speciation rate, rate shift, and extinction rate were estimated using the setBAMMpriors function

from the R package BAMMtools[50] and specified in the BAMM control file. Specifically, a total of 2×10^8 generations of rjMCMC searches with samples stored every 8000 generations were performed using speciation-extinction. A total of 1000 post burnin samples (50%) were retained. We performed two BAMM runs for each of three assumptions about sampling completeness (L, M, H) accounted for by changing the GlobalSamplingFraction parameter (0.3, 0.22, 0.18) respectfully (see Supplemental Information). To account for potential oversampling of Nearctic species, we performed a series of runs where we lowered the number of Nearctic species by randomly pruning 21 (of total 48) Nearctic tips from the tree ten times and performed a BAMM run on each replicate, using the M assumption for the GlobalSamplingFraction parameter.

Using the posteriors generated from these MCMC runs, we sought to 1) explore the overall pattern of *Pheidole* diversification, 2) assess whether there is evidence of diversity regulation, particularly decelerating diversification over time and after colonization of new areas, and 3) test for latitudinal dependency in diversification rate while accounting for phylogenetic non-independence. We visualized the linage specific diversification with the plot.bammdata function from BAMMtools, and the time plot of clade-specific diversification rate was plotted with the plotRateThroughTime function. We used STRAPP (e.g. the traitDependentBAMM function in BAMMtools) to test for significance of any latitude-diversification correlations. We tested for diversification rate vs. either tropicality index or absolute midpoint latitude (one-tailed, 10000 iterations, Spearman's rho as test statistic). We also checked whether our results were robust to using Pearson correlation as test statistic or coding latitude as a binary variable and using Mann-Whitney test (tropicality>0 or tropicality<0).

GeoSSE approach: The GeoSSE model[47] implements trait-dependent diversification in a

geographic context, where lineages can occur in one of two areas (A or B) or both (AB). Extinction and speciation may be different for the different states (controlled by rate parameters sA, sB, xA, xB), and dispersal (transition from A->AB and B->AB, with rates dA and dB respectively). An additional speciation process (with rate, sAB) splits AB into two species localized to the two different areas. Geography-dependent diversification (such as with latitude) can be inferred by comparing models in which the parameters are free to differ among the two areas or constrained to be equal.

The hypothesis of different diversification rates between tropical and extratropical lineages does not lend itself to a single model assumption about rate constraints that can be tested against others. Since net diversification rate is the difference between speciation and extinction rates, either or both could vary and lead to a difference in net diversification (or both could vary between regions, but in such a way that there was still no difference in net diversification). Our approach was thus to create a set of candidate models with different constraints, fit them with maximum likelihood, and compare the models using Akaike's Information Criterion (AIC) to get a set of models that fit the data reasonably well (i.e. $\Delta AIC<4$). We then compared the differences in net diversification and dispersal between the regions across models in the top model set. If the top models disagree on differences in net diversification or dispersal rates, we consider the results to be inconclusive on rate differences.

The candidate model set included a full model where all three rates are different across the two areas, or that one or more rates are constrained to be equal (e.g., one or more of, $sA=sB$, $xA=xB$, $dA=dB$), which makes a total of ($2^3=8$ models). In initial testing, we found that the between-area speciation rate (sAB) generally converged toward zero and never improved the model fit, so to reduce model space we set it to zero ($sAB=0$) for subsequent analyses. To ensure

we converge on the global model, we first fitted the full model using 20 total parameter values including the values suggested by starting.point.geosse(), and 19 parameter sets consisting of uniform random numbers between 0 and 1. For the rest of the models, we used the suggested starting point, the full model ML parameters, and 18 random parameter sets as starting points. To estimate and visualize uncertainty in model parameters for representative results, we ran mcmc Bayesian inference on the full model using a chain of 10000 generations, using the ML parameter values as the starting point.

As with the BAMM analyses, we repeated the inference with different values of sampling fraction (L, M, H). Since the new world is less broken up by biogeographic barriers, we also performed the same analyses on the New World *Pheidole* clade with the Old World lineage removed. Since we detected possible oversampling of Nearctic species we performed an additional analysis (M*) with state-specific sampling fraction that accounts for this difference. Finally, we performed the analyses with the M global sampling fraction on the 10 trees where Nearctic species have been culled, using either all *Pheidole* or only New World taxa. Again, if our conclusions about net diversification differences or dispersal rate differences are sensitive to these uncertainties in sampling fraction, we will consider them not to be robust. The above analyses were performed with the *diversitree* package[51] in R.

Phylogenetic niche conservatism: We performed additional analyses to evaluate the degree to which latitudinal affinity is conserved in *Pheidole*. For this, we first calculated two measures of phylogenetic signal—Blomberg's K[52] and Pagel's lambda[53]—treating absolute latitudinal midpoint as a continuous trait, using the phylosig() function in the R package *phytools*[54]. Second, to estimate the overall evolutionary rates, we fit a discrete models of character evolution (treating

latitudinal affinity as a binary variable) using the fitDiscrete() function in the R package *geiger*[55]. To visualize the evolution of latitudinal affinity, we performed 100 stochastic character maps on the empirical tree using the make.simmap() function, and plotted a summary of state probabilities with the function densityMap(), both from the *phytools* package. Finally, to estimate whether the inferred rate of evolution combined with tropical ancestral state is consistent with the observed richness difference even in the absence of diversity regulation and diversification rate differences, we simulated niche evolution on the empirical tree and maximum likelihood model with the sim.history() function from *phytools*. While tree shape and trait state are not necessarily independent (i.e. the dependent model is implemented in the GeoSSE analysis), this analysis asks whether we would be likely to observe a gradient even if they were independent, given that *Pheidole* likely has a tropical ancestor and given the rate that latitudinal affinity evolves. *Pheidole* likely has a tropical ancestor as the most basal *Pheidole* species, *P. fimbriata*, and the sister lineage of *Pheidole*, *Cephalotes* + *Procryptocerus*, are tropical[19,30].

RESULTS

Pheidole exhibits a latitudinal diversity gradient that is overall similar to ants as a whole (Fig. 1). The BEAST analysis inferred a well-supported phylogeny whose major features are consistent with previous studies (Figs. 2, S1). The crown age of the group (i.e. the mrca of *Pheidole fimbriata* with the rest of *Pheidole*) is inferred here to be slightly younger than in a previous study (~29mya vs. ~37mya in ref[31]), although closer to the crown age inferred in other recent broader scale phylogenies[19].

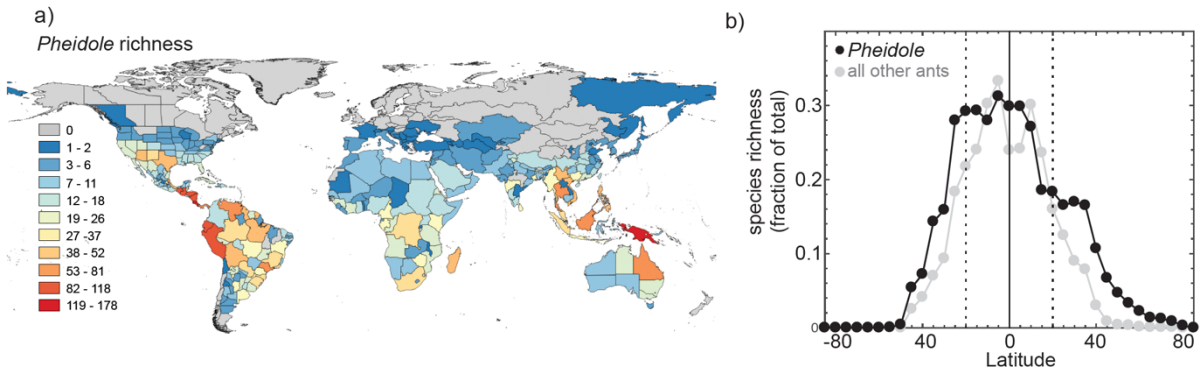

Figure 1 | Global patterns of *Pheidole* species richness. a) Richness plotted by geographic region and b) 5-degree latitudinal band for 1138 described species/subspecies and 361 morphospecies. For comparison, latitudinal distribution of 13771 ant species excluding *Pheidole* are also depicted. Latitudinal richness is expressed as fraction of total richness (either 1499 or 13,771 for *Pheidole* or all other ants, respectively).

According to the BAMM analysis, the hyperdiversification of *Pheidole* began after an acceleration approximately 15-16 mya, and all species except for two basal species (*P. fimbriata* and *P. rhea*) are descended from this lineage. Diversification initially occurred in the New World, exhibiting a decelerating trend over time. Around 13mya, a single lineage colonized the Old World and this was associated with another burst of diversification followed by a slowdown in a clade encompassing Asia and Africa. Madagascar and Australia-NG were later colonized, followed by accelerations and subsequent decelerations in each clade (Figs. 2, S1, S2). There were several other accelerations that were not obviously associated with geographic transitions, including one clade in the new world and the *megacephala* group in Afrotropics. This general pattern of sequential colonization-acceleration-deceleration pattern is robust to changing the sampling fraction parameter, although as one would expect the inferred degree of deceleration becomes less pronounced if one assumes that more species are left to be sampled.

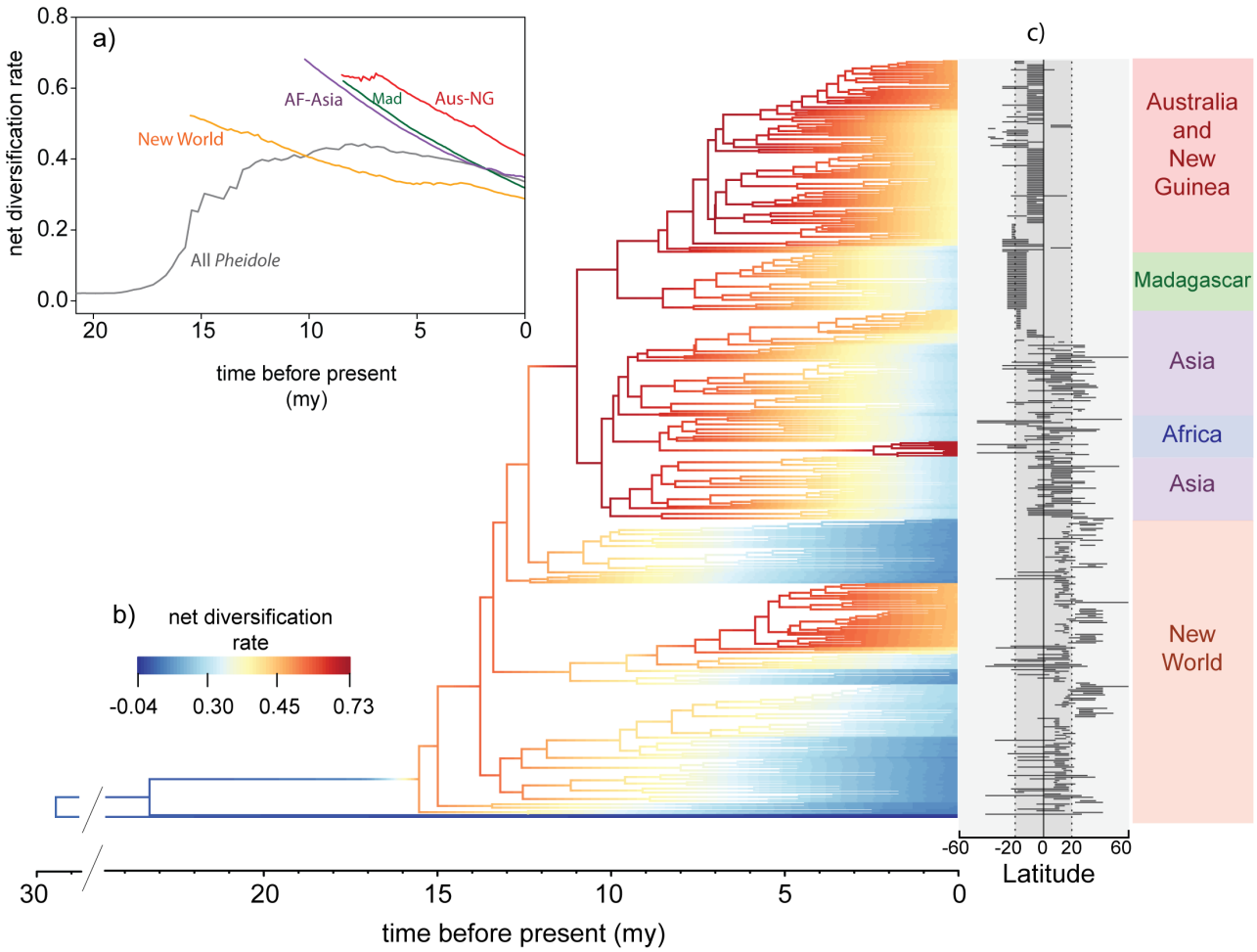

Figure 2 | Diversification rate dynamics inferred with BAMM from a phylogeny of 449 *Pheidole* species. a) Median diversification rates through time of the major *Pheidole* clades. The New World median excludes the two basal species (*P. rhea* and *P. fimbriata*) that fall outside the initial acceleration of *Pheidole* diversification. b) The mcc phylogeny colored with inferred net diversification rate. c) Latitudinal extent of all 449 taxa included in the tree. A high-resolution version with taxon names visible is presented in Figure S1.

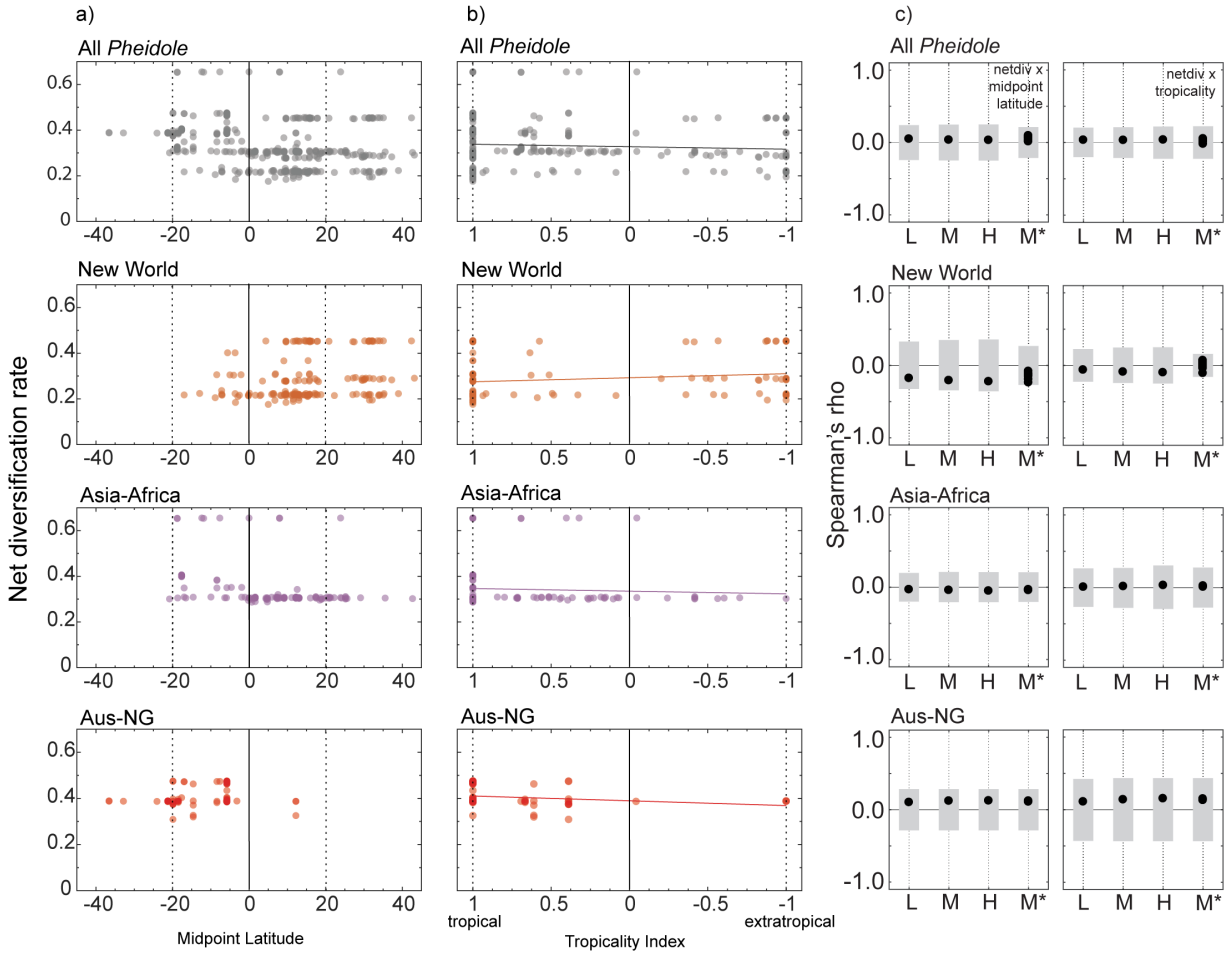

Figure 3 | Net diversification rate inferred with BAMM as a function of latitude. Diversification rate of each *Pheidole* species (present day) inferred with BAMM using the "M" assumption of sampling completeness per species a) as a function of latitudinal midpoint and b) tropicality index, which varies from -1 for a species with a range located competely outside the tropics to 1 for a species confined to the tropics. c) Spearman correlations (black dots) for net diversification and either absolute midpoint latitude (left) or tropicality (right), where the grey boxes reflect 95% null distribution generated with STRAPP. L, M, H, reflect different assumptions about unsampled species (low, medium, high estimates of total numbers of Pheidole), while M* are 10 trees where temperate species have been culled to account for possible sampling bias (see methods).

The extratropical lineages generally belong to young clades nested within larger tropical clades (Figs. 2, S1). While diversification rate varies across the genus to a degree, we could not detect a significant correlation (assessed with STRAPP) between BAMM-inferred net diversification rate and either absolute midpoint latitude or tropicality index across any of the analyses we performed (Fig. 3). These results were similar across variation in the assumed global sampling fractions, whether we calculated correlations for individual clades or the whole tree,

and including trees where Nearctic species were culled to account for possible uneven sampling. Although significance tests were one-tailed for higher diversification in the tropics, we also note that none of the observed correlation coefficients were outside the null range in either direction.

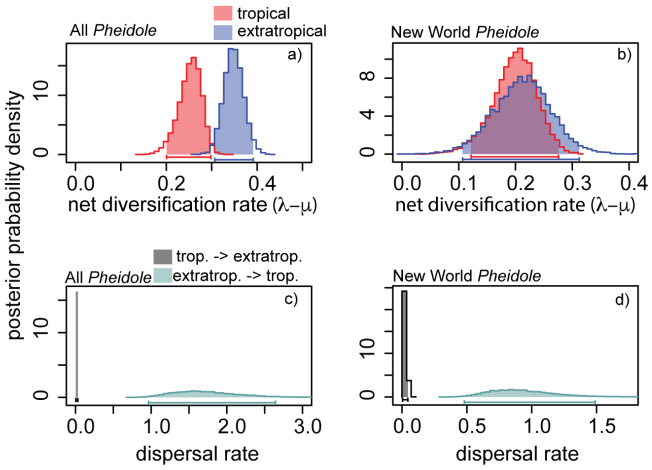

Figure 4 | Posterior probability distributions of *Pheidole* macroevolutionary rate parameters inferred with the GeoSSE model. Net diversification rate (a, b) and dispersal rate (c, d) inferred from the (a,c) global *Pheidole* phylogeny and (b,d) new world *Pheidole* only. The global analysis presented assumed used the M sampling fraction while the New World analysis assumed M*, which corrects for potential latitudinal sampling bias (see text for results using other assumptions).

For the global *Pheidole*, GeoSSE inferred a slightly elevated net diversification rate for *extratropical* lineages (1.2x-1.6x of net diversification rate in the top models, Fig. 4, Table S4). This result was robust to different global sampling fractions. When the New World *Pheidole* are considered alone, net diversification rate is also slightly elevated by a similar degree. However, when uneven latitudinal sampling fractions in the New World are accounted for—either by altering the sampling fraction parameter in GeoSSE or by culling species from the tree—the difference in net diversification rate disappears and statistically a single rate model is preferred (Fig. 4, Table S4). Broad differences in dispersal rate were statistically supported in all of the GeoSSE analyses, with extratropical lineages transitioning to the tropics more easily than the reverse.

The extratropical lineages are clustered with each other on the tree, although it is clear there were numerous transitions out of the tropics (Fig. 5). The tests for phylogenetic signal in latitudinal affinity for Blomberg's K ($K=0.34$, $p<0.002$) and Pagel's lambda ($lambda=0.95$, $p<10^{-57}$) were both highly significant. Symmetric and asymmetric models of discrete character

evolution both fit the data comparably well (symmetric model $q_{trop->etrop}=q_{etrop->trop}=0.015$, AICc=235.5, asymmetric model $q_{trop->etrop}= 0.013$, $q_{etrop->trop} =0.060$, AICc=234.9). Simulations of character evolution on the empirical phylogeny show that a latitudinal gradient is the most common outcome if one assumes a tropical ancestor and either model for the inferred rate of evolution of latitudinal affinity (Fig. 5).

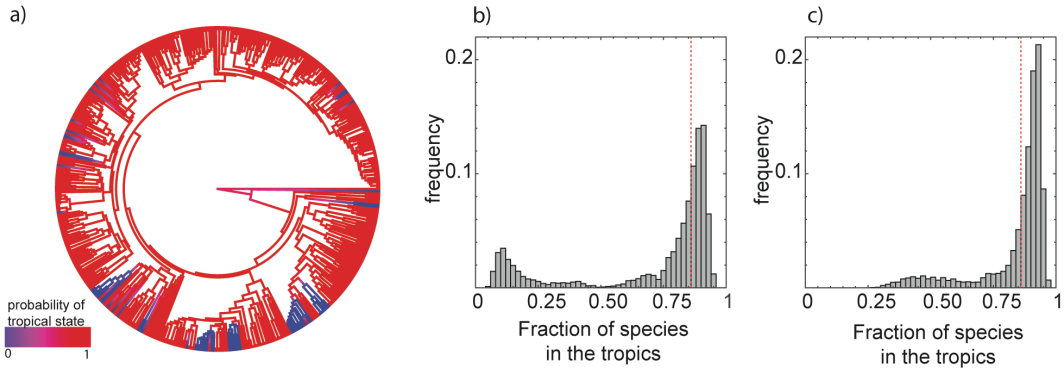

Figure 5 | Evolution of latitudinal affinity in *Pheidole*. a) Branch-wise probability of ancestral tropical state inferred from stochastic character mapping. b-c) Histograms of latitudinal richness differences beween tropics and extratropics simulated with stochastic character mapping on the empirical phylogeny assuming a tropical ancestor and the inferred degree of niche conservation using symmetric (b) or asymmetric (c) models of character evolution. The vertical dashed line is the empirical richness fraction.

DISCUSSION

Our analysis of *Pheidole* macroevolution sheds light on the mechanisms responsible for the evolution of the latitudinal diversity gradient. By focusing on the recent evolutionary dynamics of a single large radiation, our study complements taxon-wide studies that focus on differences among highly divergent clades and deeper timescales.

We find no evidence of higher diversification rate for tropical *Pheidole* lineages across any of our analyses (Figs. 2-4, S1), as would be predicted by the diversification rate hypothesis. For some analyses, there was a weak *positive* relationship between latitude and net diversification rate (Fig. 4), but this result was not robust to different model structures (e.g. BAMM vs. GeoSSE) and assumptions about latitudinal sampling biases.

While there is no relationship between diversification rate and latitude, the pattern of diversification suggests *Pheidole* evolution is being shaped by diversity regulation. Even if one assumes that there are over 1300 undescribed *Pheidole* species (the higher end of our estimates) in addition to the 1175 currently described species, our analysis found that diversification is still decelerating in the genus. Moreover, there is evidence that diversification accelerated after colonization of new areas, specifically when *Pheidole* colonized the Old World, and again after it colonized Australia and New Guinea. This lends further support to the idea that there are ecological limits to *Pheidole* diversity, because when new continents are colonized, ecological opportunity is high. However, it is more difficult to determine if these limits vary with latitude in a way that is causally responsible for the richness gradient. For example, the limits to diversity could be similar at all latitudes, but phylogenetic conservatism could be causing the higher latitudes to lag behind tropical latitudes in reaching their steady state. Here we might expect a positive trend of *Pheidole* diversification rate running counter to the richness gradient, because

temperate zones would be further from the equilibrium number of species. While there were some hints of a positive latitude-diversification correlation (e.g. Fig. 4), there was not a robust relationship.

The results match the predictions of the tropical conservatism hypothesis (TCH) well overall. We found that latitudinal affinity is moderately conserved in *Pheidole*. While there have been a number of transitions from the tropics to the temperate zone, latitudinal affinity evolves slowly enough to make a richness gradient the most likely outcome simply due its tropical ancestry and phylogenetic inertia. A notable point is that dispersal rates were inferred to be higher from the extratropics back into the tropics, rather than from the tropics to the temperate zone. This runs counter to the general expectation of the "out-of-the-tropics" model, that posits higher dispersal from the tropics out to the extratropics. However, for ectothermal organisms like ants, it makes sense under a niche conservatism scenario that transitions should be more difficult from warmer to colder environments than the reverse. Ants need specialized behavioral and physiological adaptations to survive high-latitude winters, and presumably it is easier to lose those adaptations than it is to evolve them in the first place. This could lead to a "leaky temperate zone" effect whereby gene flow from the lower latitude end of ranges drags the species toward the tropics, reinforcing the gradient further. Note, however, that niche conservatism was high enough to make the diversity gradient a likely outcome under either asymmetric or symmetric models of latitudinal range evolution (Figure 5).

These results for *Pheidole* evolution over the last 30my connect well to results on ant diversification on deeper timescales[27], and together tell a coherent story about the evolution of latitudinal gradients in ants. Most ant lineages older than 34mya are reconstructed to be tropical, including the *Pheidole* stem lineage. Around 15 mya, *Pheidole* exhibited a many-fold

acceleration in diversification rate and began a massive radiation. The reason for this initial acceleration, such as evolution of a key innovation, remains unknown. It took time for some *Pheidole* lineages to evolve the requisite traits for colonization of high latitudes, however once colonization of cold biomes occurred, diversification was not detectibly slower. In their analysis across all ant clades, Economo et al. also found no evidence for elevated net diversification rates among clades centered in the tropics relative to those in the temperate zone, although clades are quite heterogeneous in rate, probably due to other latent biological and historical differences. Within *Pheidole*, diversification rate is much less heterogeneous, but there is still no hint of a negative latitudinal correlation. This implies that biological variation among clades is not masking a latitudinal correlation within clades. *Pheidole* provides additional insight that diversity regulation is a prominent feature of the global evolution of the genus, although it is unclear if it is causal in the gradient. Finally, since we have a better handle on sampling biases within *Pheidole* than we do for ants as a whole, we can be more confident that latitudinal sampling biases are not masking latitudinal diversification rate variation.

While thus far the evidence is consistent with both phylogenetic niche conservatism (TCH) and diversity regulation (ERH) playing a role in ant diversification, determining whether one alone or both together are responsible for the diversity gradient remains a challenge for future work. Moreover, as these are "umbrella" hypotheses, each individual hypothesis could encompass a range of different mechanisms. One way forward is a hierarchical, systematic approach, where broad categories of hypotheses are evaluated (e.g. like these in this study), followed then by more targeted studies devised to tease apart the mechanisms within the larger classes of hypothesis that fit the data well. We also agree with the approach advocated by Hurlbert and Stegen[34], toward a quantitative formulation of multiple competing and intersecting

hypotheses, combined with a simulation-based approach to identify their key predictions. We felt initial efforts in this direction were not yet mature enough to use as a basis for the current study, but look forward to further development of the approach in the future. Finally, we need further work to resolve and analyze other hyperdiverse ant radiations (e.g. *Camponotus, Strumigenys, Tetramorium*) that also exhibit strong latitudinal gradients.

Despite the high level of research effort directed toward understanding the latitudinal gradient, the matter is far from resolved. Studies have differed in their conclusions about the origins of the gradient, probably due to both differences in conceptual and methodological approaches, along with real variation in process and history across taxonomic groups. The former should continue to improve as we develop more penetrating quantitative methods that make use of more diverse data types. Variability across taxonomic groups is best assessed and understood by examining more of them. The vast majority of studies on the diversity gradient have focused on vertebrates. While of obvious intrinsic interest, vertebrates may not be good surrogates for understanding general patterns across the rest of the tree of life. For example, mammals have been impacted by human activities so dramatically that it can affect large-scale macroecological patterns[35,36]. With development of global invertebrate datasets like the one analyzed here, we stand to broaden our perspective on large-scale biological patterns and their origins.

REFERENCES:

1 Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* **10**, 315-331 (2007).

2 Willig, M. R., Kaufman, D. M. & Stevens, R. D. Latitudinal gradients of biodiversity:

pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273-309 (2003).

3 Fine, P. V. A. Ecological and Evolutionary Drivers of Geographic Variation in Species Diversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 369-392, doi:doi:10.1146/annurev-ecolsys-112414-054102 (2015).

4 Pianka, E. R. Latitudinal Gradients in Species Diversity - a Review of Concepts. *Am. Nat.* **100**, 33-&, doi:Doi 10.1086/282398 (1966).

5 Weir, J. T. & Schluter, D. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574-1576 (2007).

6 Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444-448, doi:10.1038/nature11631 (2012).

7 Cardillo, M., Orme, C. D. L. & Owens, I. P. Testing for latitudinal bias in diversification rates: an example using New World birds. *Ecology* **86**, 2278-2287 (2005).

8 Rolland, J., Condamine, F. L., Jiguet, F. & Morlon, H. Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLoS Biol* **12**, e1001775 (2014).

9 Pigot, A. L., Tobias, J. A. & Jetz, W. Energetic Constraints on Species Coexistence in Birds. *PLoS ONE* **11**, doi:ARTN e1002407 10.1371/journal.pbio.1002407 (2016).

10 Duchêne, D. A. & Cardillo, M. Phylogenetic patterns in the geographic distributions of birds support the tropical conservatism hypothesis. *Glob. Ecol. Biogeogr.* **24**, 1261-1268 (2015).

11 Pyron, R. A. & Wiens, J. J. Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proc. R. Soc. Lond.* **280**, 20131622 (2013).

12 Kerkhoff, A. J., Moriarty, P. E. & Weiser, M. D. The latitudinal species richness gradient

in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8125-8130, doi:10.1073/pnas.1308932111 (2014).

13 McKenna, D. D. & Farrell, B. D. Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10947-10951 (2006).

14 Condamine, F. L., Sperling, F. A. H., Wahlberg, N., Rasplus, J. Y. & Kergoat, G. J. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecol. Lett.* **15**, 267-277, doi:10.1111/j.1461-0248.2011.01737.x (2012).

15 Moreau, C. S. & Bell, C. D. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* **67**, 2240-2257 (2013).

16 Pie, M. R. The macroevolution of climatic niches and its role in ant diversification. *Ecol. Entomol.* **41**, 301-307, doi:10.1111/een.12306 (2016).

17 Guénard, B., Weiser, M. D., Gomez, K., Narula, N. & Economo, E. P. The Global Ant Biodiversity Informatics (GABI) database: synthesizing data on the geographic distribution of ant species (Hymenoptera: Formicidae). *Myrmecol. News* **24**, 83-89 (2017).

18 Moreau, C. S., Bell, C. D., Vila, R., Archibald, S. B. & Pierce, N. E. Phylogeny of the ants: diversification in the age of angiosperms. *Science* **312**, 101-104 (2006).

19 Ward, P. S., Brady, S. G., Fisher, B. L. & Schultz, T. R. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). *Syst. Ent.* **40**, 61-81 (2015).

20 Brady, S. G., Schultz, T. R., Fisher, B. L. & Ward, P. S. Evaluating alternative hypotheses for the early evolution and diversification of ants. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18172-18177 (2006).

21 Rohde, K. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos*, 514-527 (1992).

22 Wiens, J. J. & Donoghue, M. J. Historical biogeography, ecology and species richness. *Trends Ecol. Evolut.* **19**, 639-644 (2004).

23 Losos, J. B. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* **11**, 995-1003 (2008).

24 Guénard, B., Perrichot, V. & Economo, E. P. Integration of global fossil and modern biodiversity data reveals dynamism and stasis in ant macroecological patterns. *J. Biogeogr.* **42**, 2302-2312 (2015).

25 Rabosky, D. L. & Hurlbert, A. H. Species Richness at Continental Scales Is Dominated by Ecological Limits. *Am. Nat.* **185**, 572-583, doi:10.1086/680850 (2015).

26 Hurlbert, A. H. & Stegen, J. C. When should species richness be energy limited, and how would we know? *Ecol. Lett.* **17**, 401-413, doi:10.1111/ele.12240 (2014).

27 Economo, E.P., Narula, N., Friedman, N.R., Weiser, M.D., Guénard, B. Macroecology and macroevolution of the latitudinal diversity gradient in ants. (*In Revision*).

28 Buckley, L. B. & Jetz, W. Environmental and historical constraints on global patterns of amphibian richness. *Proc. Natl. Acad. Sci. U.S.A.* **274**, 1167-1173 (2007).

29 Blanchard, B. D. & Moreau, C. S. Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. *Evolution* **71**, 315-328, doi:10.1111/evo.13117 (2017).

30 Moreau, C. S. Unraveling the evolutionary history of the hyperdiverse ant genus Pheidole (Hymenoptera : Formicidae). *Mol. Phylogenet. Evol.* **48**, 224-239, doi:10.1016/j.ympev.2008.02.020 (2008).

31 Economo, E. P. *et al.* Global phylogenetic structure of the hyperdiverse ant genus Pheidole reveals the repeated evolution of macroecological patterns. *Proc. R. Soc. Lond.* **282**, doi:UNSP 20141416 10.1098/rspb.2014.1416 (2015).

32 Economo, E. P. *et al.* Breaking out of biogeographical modules: range expansion and taxon cycles in the hyperdiverse ant genus Pheidole. *J. Biogeo.* **42**, 2289-2301, doi:10.1111/jbi.12592 (2015).

33 Sarnat, E. M. & Moreau, C. S. Biogeography and morphological evolution in a Pacific island ant radiation. *Mol. Ecol.* **20**, 114-130, doi:10.1111/j.1365-294X.2010.04916.x (2011).

34 Hurlbert, A. H. & Stegen, J. C. On the processes generating latitudinal richness gradients: identifying diagnostic patterns and predictions. *Front. Genet.* **5**, doi:ARTN 420 10.3389/fgene.2014.00420 (2014).

35 Turvey, S. T. & Fritz, S. A. The ghosts of mammals past: biological and geographical patterns of global mammalian extinction across the Holocene. *Phil. Trans. R. Soc. Lond. B. Bio.l Sci.* **366**, 2564-2576, doi:10.1098/rstb.2011.0020 (2011).

36 Santini, L., González-Suárez, M., Rondinini, C., Di Marco, M. Shifting baseline in macroecology? Unravelling the influence of human impact on mammalian body mass. *Divers. Distrib.* **23**, 640-659 (2017).

37 Janicki, J., Narula, N., Ziegler, M., Guenard, B. & Economo, E. P. Visualizing and interacting with large-volume biodiversity data using client-server web-mapping

applications: The design and implementation of antmaps.org. *Ecol. Inform.* **32**, 185-193, doi:10.1016/j.ecoinf.2016.02.006 (2016).

38 Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294-299 (1994).

39 Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67-D72, doi:10.1093/nar/gkv1276 (2016).

40 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

41 Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLoS ONE* **6**, doi:ARTN e22594 10.1371/journal.pone.0022594 (2011).

42 Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol.* **29**, 1695-1701, doi:10.1093/molbev/mss020 (2012).

43 Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537, doi:10.1371/journal.pcbi.1003537 (2014).

44 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

45 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290, doi:10.1093/bioinformatics/btg412 (2004).

46 Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543, doi:10.1371/journal.pone.0089543 (2014).

47 Goldberg, E. E., Lancaster, L. T. & Ree, R. H. Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Syst. Biol.***60**, 451-465, doi:10.1093/sysbio/syr046 (2011).

48 Rabosky, D. L. & Huang, H. Minimal effects of latitude on present-day speciation rates in New World birds. *Proc. R. Soc. Lond.* **282**, 20142889 (2015).

49 Rabosky, D. L. & Goldberg, E. E. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* **64**, 340-355 (2015).

50 Rabosky, D. L. *et al.* BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Meth. Ecol. Evol.* **5**, 701-707 (2014).

51 FitzJohn, R. G. Diversitree: comparative phylogenetic analyses of diversification in R. *Meth. Ecol. Evol.* **3**, 1084-1092, doi:10.1111/j.2041-210X.2012.00234.x (2012).

52 Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717-745, doi:DOI 10.1111/j.0014-3820.2003.tb00285.x (2003).

53 Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877-884, doi:Doi 10.1038/44766 (1999).

54 Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Meth. Ecol. Evol.* **3**, 217-223, doi:10.1111/j.2041-210X.2011.00169.x (2012).

55 Pennell, M. W. *et al.* geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216-2218,

doi:10.1093/bioinformatics/btu181 (2014).

Electronic Supplementary Information

Figure S1: Expanded version of figure 2, including phylogeny, latitudinal range, and inferred diversification rate. (included in preprint pdf)

Figure S2: Probable locations of diversification rate shifts on the phylogeny**.** (included in preprint pdf)

Supplementary Methods: Calculation of sampling completeness. (included in preprint pdf)

Table S1: Specimen information and GenBank accession codes.

Table S2: Calculation of sampling completeness. (included in preprint pdf)

Table S3: Partitions and nucleotide substitution models. (included in preprint pdf)

Table S4: Full results of the GeoSSE analyses.

Acknowledgements: This work was supported by NSF (DEB-1145989 to EPE and LLK), a Japan Society for the Promotion of Science KAKENHI grant (JP17K15180) to EPE, and by subsidy funding to OIST. We thank P.S. Ward and B.L. Fisher for sharing specimens used in this study.

Author Contributions: EPE and LLK designed the research. EPE, EMS, GF, JTL, ML, and BG contributed specimens and/or data. EPE, EMS, GF, ML, and JTL performed the specimen curation and identification. JPH and LLK performed the DNA sequencing. EPE, JPH, and NN analyzed the data. EPE led the writing, with contributions from all co-authors.

Competing Financial Interests: None

Materials & Correspondence: Evan P. Economo, evaneconomo@gmail.com.

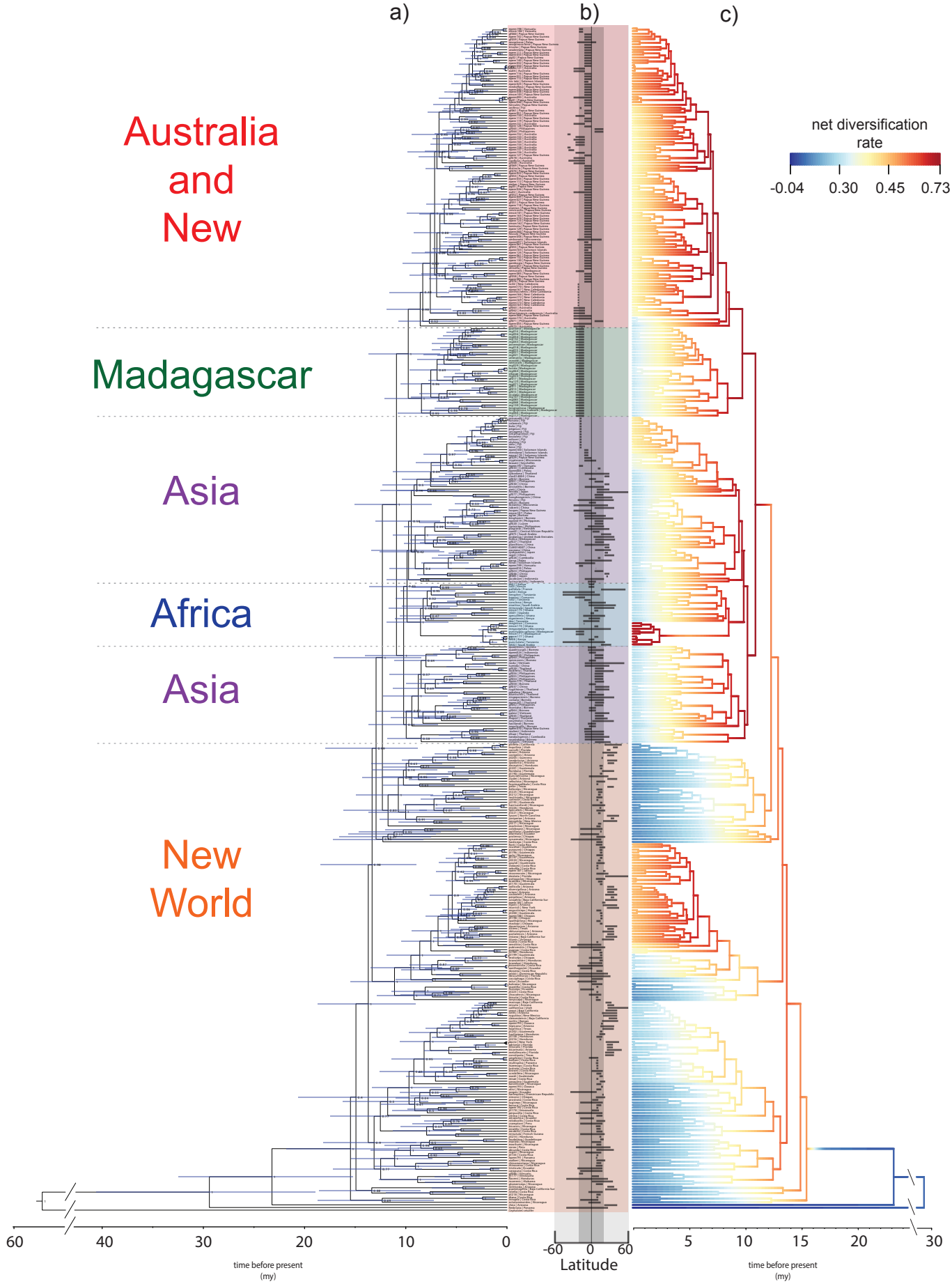

Figure S1 | The global *Pheidole* phylogeny annotated with inferred diversification rate and latitudinal distribution. a) The maximum clade credibility (mcc) phylogeny inferred with BEAST, with nodes set to median ages. Node labels are posterior supports, node bars are 95% age distributions. b) Latitudinal extent of each *Pheidole* species and morphospecies on the tree. c) Median diversification rate inferred with BAMM across the same phylogeny (M sampling fraction assumption).

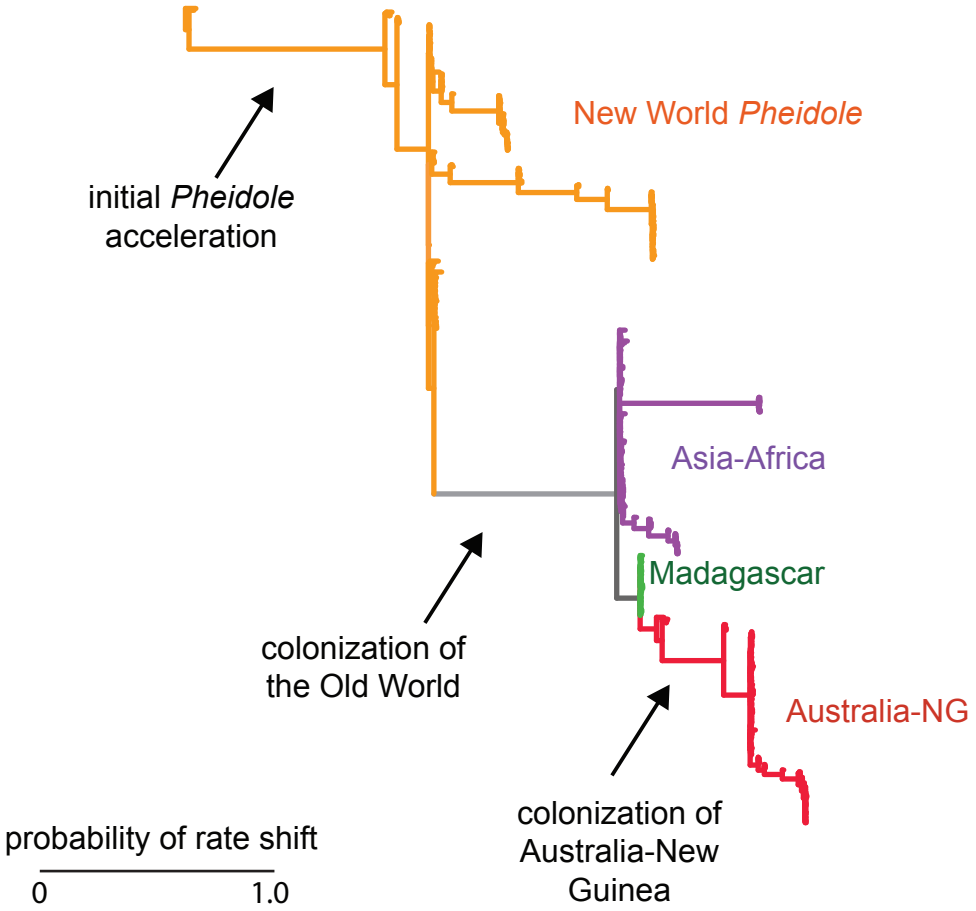

Figure S2 | Probable locations of rate shifts in *Pheidole* on the phylogeny inferred with BAMM. Here, branch length indicates probability of a rate shift. The colonization of new biogeographic regions generally corresponds to rate shifts, which tended to be accelerations as seen in Figure 1. Several other rate shifts occurred that are not obviously explainable by geographic transitions.

Supplementary Methods: Calculation of sampling completeness across geography and clade.

One challenge to working with relatively poorly known taxa is the possible influence of unknown and unsampled taxa. While there are 1159 described *Pheidole* species and subspecies globally, there are many more yet to be described. Moreover, the distribution of this sampling completeness is not random. Tropical faunas may be less documented than temperate faunas, and different bioregions have received different amounts of attention from researchers. For example, it is likely that New World *Pheidole* species are much better documented than the Old World *Pheidole*, due to a major monograph pursued by Wilson that described a large number of species [30].

We devised a method to account for this potential bias that, while coarse and imperfect, should address the big picture of whether our overall sampling reflects the distribution of overall *Pheidole* richness. Over the last years, our project has sought material from *Pheidole* faunas around the world, both described and undescribed. We made a vigorous effort to identify all the taxa we included in the phylogenetic study. We are thus quite confident that the morphospecies on our tree reflect undescribed, rather than unidentified, taxa. For the purposes of this calculation, we made the *assumption* that the fraction of described species (FD) within each clade/region on our tree reflects the true fraction of undescribed species in the region. For example, on our tree there is a higher fraction of morphospecies from the Australia-New Guinea clade (undescribed: 93, described: 21, FD ~ 0.18) compared to the New World clade (described: 143, undescribed: 34, FD ~ 0.80). From the GABI data, we know the number of described species occuring in each region (and thus clade, given the tight correspondence between region and clade), which we call ND. Using the number described from each region and the estimated fraction described we can calculate the estimated total richness in the region/clade (ER=ND/FD). For example, for Australia-New Guinea there are 112 described taxa, while in the New World there are 649 described taxa. Using the FD from our sampling, we estimate there to be ER~608 species in the Australia-New Guinea clade ($ER_{AUSNG}=ND_{AUSNG}/FD_{AUSNG}=93/0.18=608$) and 803 species in the New World clade ($ER_{NEWWORLD}=ND_{NEWWORLD}/FD_{NEWWORLD}=649/0.39=803$). From this we can compare the sampling representation of taxa on our tree to the estimated

distribution of richness in *Pheidole* (Table 1).

We performed the calculation two ways, first by estimating sampling fraction for each clade, and second by estimating sampling fraction for the tropical and extratropical species of each hemisphere (since this latitudinal distinction is important for our analysis). Using clades, we estimated there to be 2175 *Pheidole* species globally using the clade-wise or latitude-wise calculation with about 55% of them described. Our sampling of different clades broadly matched the estimated proportion of total *Pheidole* richness in those clades (Table 1). Using latitude and hemisphere, we estimated 2127 species globally (54% described). We detected a degree of oversampling of New World temperate taxa on our tree (11% actual vs. 6% estimated) and undersampling of Old World temperate taxa (3.3% actual vs. 12% estimated). Overall, we estimated global fraction of temperate species to be 19%, which was reasonably close to the fraction on our tree (14%).

For both the GeoSSE and BAMM analyses, we chose to explore a range of global sampling fractions (i.e. the fraction of total *Pheidole* species that are on the tree) that broadly encompass the estimated value, and evaluate whether our conclusions are stable within that range. We chose a range of 1500, 2000, 2500 total *Pheidole* species representing the minimum and maximum plausible *Pheidole* richness, and set sampling fraction accordingly at 0.3, 0.22, or 0.18, which we denote L, M, H, respectively. This range enclosed our estimated values (2127 or 2179).

Overall, the distribution across clades and latitudes of our sampling intensity was reasonably similar to the estimated richnesses for *Pheidole* (Table S1). Thus, for the most part we did not use clade-specific sampling correction factors. However, the oversampling of taxa on our tree in the Nearctic region relative to the Neotropical region is of some concern, particularly as we performed analyses on the New World data alone. The latter was performed because the New World is less broken up by biogeographic barriers than the Old World and most likely to clearly reflect diversification-latitude correlations (if they exist). Thus, we performed two additional GeoSSE analyses to account for differential sampling of the tropics and extratropics, which we denote M*. First, we set state-specific sampling fractions to be different for extratropical and tropical taxa. For global *Pheidole*, we calculated the extratropics to be slightly undersampled on our tree (tropics=0.22, extratropics=0.16), and for New World *Pheidole* the reverse was true (tropics=0.185, extratropics=0.35), using our estimated total richness (2127). We used these

values to set the state-specific sampling fraction in the GeoSSE analysis (AB range was set to middle of the two). Second, we performed an additional set of analyses on 10 trees where the number of Nearctic taxa had been reduced by randomly pruning taxa from the tree to lower the number from 48 to 27. The latter adjusts the fraction of Nearctic species to be consistent with the estimated fraction using our calculation. We also ran BAMM on these trees using the M assumption for the global sampling fraction.

Table S2: Calculation of sampling completeness across regions and clades

	Taxa on tree	Estimated fraction described	Current described richness	Estimated richness (described / fraction described)	Fraction of total taxa on tree	Target fraction of total richness (estimated fraction of total richness)
By Clade						
New World	177	0.81	649	803	0.39	0.38
Africa-Asia	124	0.60	356	597	0.28	0.28
Australia-New Guinea	114	0.18	112	608	0.25	0.29
Madagascar	34	0.29	35	119	0.08	0.06
Total	449	0.55	1138	2126		
By Latitude/Hemisphere						
New World tropical	129	0.73	512	695	0.29	0.32
New World extratropical	48	1.00	137	137	0.11	0.06
Old World tropical	257	0.39	424	1089	0.57	0.50
Old World extratropical	15	0.33	86	258	0.03	0.12
Total				2179		

Table S3: Partitions and nucleotide substitution models, used in the phylogenetic analysis, as optimized with PartitionFinder.

Partition	Best Model	Subset Partitions	Size
p1	K80+Γ	CAD_pos3, EF1AF2_pos3	357
p2	K80+Γ	CAD_pos1, LOP1_pos2, unc-4_pos2, TOP1_pos1	780
p3	HKY+Γ	CAD_pos2 ,TOP1_pos2	417
p4	HKY+Γ	CAD_introns, LOC15, LOP1_introns	2007
p5	HKY+Γ	GRIK2	960
p6	SYM+Γ	COX1_pos1	353
p7	TrN+Γ	COX1_pos2	353
p8	GTR+Γ	COX1_pos3	353
p9	TrN+Γ	EF1AF2AF2_pos1, HIS3.3B_pos1, unc-4_pos1	433
p10	JC+Γ	EF1AF2_pos2, HIS3.3B_pos2	349
p11	HKY+Γ	EF1AF2_introns, unc-4_introns	1210
p12	GTR+Γ	HIS3.3B_pos3	143
p13	GTR+Γ	HIS3.3B_introns	213
p14	SYM+Γ	LOP1_pos1	280
p15	GTR+Γ	LOP1_pos3, UNC-4_pos3	363
p16	TrN+Γ	TOP1_pos3	268