ORIGINAL
ARTICLE

Breaking out of biogeographical modules: range expansion and taxon cycles in the hyperdiverse ant genus *Pheidole*

Evan P. Economo[1,2]*, Eli M. Sarnat[3], Milan Janda[4,5], Ronald Clouse[6], Pavel B. Klimov[2,7], Georg Fischer[1], Benjamin D. Blanchard[2,8], Lizette N. Ramirez[2], Alan N. Andersen[9], Maia Berman[9], Benoit Guénard[1], Andrea Lucky[10], Christian Rabeling[11], Edward O. Wilson[12] and L. Lacey Knowles[2]

[1]*Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan,* [2]*Department of Ecology & Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI, USA,* [3]*Department of Entomology, University of Illinois at Urbana-Champaign, Chicago, IL, USA,* [4]*Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic,* [5]*Department of Biology, University of Guanajuato, Guanajuato, Mexico,* [6]*Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA,* [7]*Department of Zoology, Tyumen State University, Tyumen, Russia,* [8]*Committee on Evolutionary Biology, University of Chicago, Chicago, IL, USA,* [9]*CSIRO Land & Water Flagship, Darwin, NT, Australia,* [10]*Department of Entomology & Nematology, University of Florida, Gainesville, FL, USA,* [11]*Department of Biology, University of Rochester, Rochester, NY, USA,* [12]*Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA*

*Corresponding author: Evan P. Economo, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan.
E-mail: evaneconomo@gmail.com

ABSTRACT

Aim We sought to reconstruct the biogeographical structure and dynamics of a hyperdiverse ant genus, *Pheidole,* and to test several predictions of the taxon cycle hypothesis. Using large datasets on *Pheidole* geographical distributions and phylogeny, we (1) inferred patterns of biogeographical modularity (clusters of areas with similar faunal composition), (2) tested whether species in open habitats are more likely to be expanding their range beyond module boundaries, and (3) tested whether there is a bias of lineage flow from high- to low-diversity areas.

Location The Old World.

Methods We compiled and jointly analysed a comprehensive database of *Pheidole* geographical distributions, the ecological affinities of different species, and a multilocus phylogeny of the Old World radiation. We used network modularity methods to infer biogeographical structure in the genus and comparative methods to evaluate the hypotheses.

Results The network analysis identified eight biogeographical modules, and a suite of species with anomalous ranges that are statistically more likely to occur in open habitat, supporting the hypothesis that open habitats promote range expansion. Phylogenetic analysis shows evidence for a cascade pattern of colonization from Asia to New Guinea to the Pacific, but no 'upstream' colonization in the reverse direction.

Main conclusions The distributions of *Pheidole* lineages in the Old World are highly modular, with modules generally corresponding to biogeographical regions inferred in other groups of organisms. However, some lineages have expanded their ranges across module boundaries, and these species are more likely to be adapted to open habitats rather than interior forest. In addition, there is a cascade pattern of dispersal from higher to lower diversity areas during these range expansions. Our findings are consistent with the taxon cycle hypothesis, although they do not rule out alternative interpretations.

Keywords

ants, colonization, dispersal, diversification, Formicidae, island biogeography, phylogeny, radiation, range expansion, taxon cycle

INTRODUCTION

The disproportionately high richness of some taxonomic groups remains one of the central puzzles of biodiversity science. Such 'hyperdiverse' taxa, as extreme cases, present opportunities to learn about the evolutionary and ecological processes regulating biodiversity. In addition to any broader lessons they may offer, these groups are of intrinsic interest for the simple reason that they constitute major components of Earth's ecosystems. Even the idiosyncrasies of their

evolutionary histories have implications for our understanding of the history, nature and organization of the biosphere. This is certainly the case for the globally distributed and hyperdiverse ant genus *Pheidole* Westwood (Wilson, 2003). Despite being only one of 324 currently recognized ant genera, 8% of described ant species belong to the genus. *Pheidole* species can be found in nearly every continent and biogeographical realm but are particularly ecologically dominant and species-rich in tropical wet forests.

How taxa such as *Pheidole* disperse across different geographical areas, and their evolutionary dynamics after colonization, touches upon many basic issues in ecology, evolution, and biogeography. One such question is whether geographical expansions are due to chance dispersal events by otherwise unremarkable lineages, or whether colonizing lineages have ecological and/or phenotypic characteristics that promote range expansion (Van Bocxlaer *et al.*, 2010). Dispersal and colonization ability can be under direct selection, such as in patchy, dynamic environments (Levin *et al.*, 1984; Johnson & Gaines, 1990), or dispersal ability can be a by-product of other ecological or phenotypic adaptations that are not related to dispersal ability *per se*. For example, lineages adapted to low-elevation coastal habitats and forest edges may be more successful overseas colonists than their montane forest relatives, not because colonization ability is under selection, but simply because of greater connectivity among coastal habitats than among montane habitats. Wilson elaborated on this general idea with his taxon cycle hypothesis (Wilson, 1959, 1961). The taxon cycle hypothesis proposes that lineages pass through different phases of range expansion and contraction associated with ecological niche shifts, and it represents a markedly deterministic view of biogeographical dynamics (Wilson, 1961; Ricklefs & Cox, 1972; Ricklefs & Bermingham, 2002). Recent work has brought new interest in the idea both in ants (Sarnat & Moreau, 2011; Economo & Sarnat, 2012; Clouse *et al.*, 2015) and other organisms (Cook *et al.*, 2008; Ricklefs & Bermingham, 2008; Jønsson *et al.*, 2014), but testing such complex hypotheses remains a major challenge even with modern methods.

This current study has five objectives. First, we seek to provide phylogenetic structure to the large *Pheidole* fauna by reconstructing a time-calibrated phylogeny of 177 Old World species. Second, we use a recently assembled comprehensive database on ant geographical distributions to statistically identify biogeographical modules – clusters of areas and species that are associated with each other (Carstensen *et al.*, 2012, 2013). Third, we use the modularity analysis to identify a subset of species with anomalous ranges; those distributed across modules indicating range expansion. Fourth, we use our phylogeny to evaluate the phylogenetic structure underlying the biogeographical modularity, including where anomalous range-expanding species originate. Finally, we test taxon cycle predictions that range expansion occurs in a cascade pattern from high- to low-diversity areas, and that dispersal ability is related to ecological habitat affinity. We first

provide background of the state of *Pheidole* phylogenetic systematics and review Wilson's taxon cycle hypothesis.

Phylogenetic systematics and biogeography of Old World *Pheidole*

From local to global scales, documenting and organizing the hyperdiversity of *Pheidole* poses a major challenge. The genus is among the most taxonomically difficult ant groups, and specimens are often left as unidentified morphospecies in community ecology studies. Nevertheless, researchers have been steadily accumulating information about the genus. Notably, Wilson completed a major monograph on New World *Pheidole* (Wilson, 2003), and subsequent authors have continued this work (Longino, 2009; Longino & Cox, 2009). In the Old World, the morphological taxonomy of *Pheidole* has been advanced by Eguchi in Asia (Eguchi, 2001a,b, 2006; Eguchi *et al.*, 2007; Eguchi, 2008), Sarnat in Fiji (Sarnat, 2008; Sarnat & Economo, 2012), and Fischer in the Afrotropics (Fischer *et al.*, 2012; Fischer & Fisher, 2013). The first molecular phylogeny of the genus *Pheidole* presented the remarkable finding that most Old World species belong to a single monophyletic group nested within the New World taxa (Moreau, 2008). This phylogeny was subsequently extended by Sarnat & Moreau (2011) who focused on the Fijian radiation in the context of a limited sampling of other Indo-Pacific *Pheidole* lineages.

Recently, Economo *et al.* (2015) reconstructed a new global phylogeny building upon pre-existing datasets (Moreau, 2008; Sarnat & Moreau, 2011) and adding new loci and several hundred new species. In particular, the new data increase taxonomic sampling from Old World *Pheidole*, providing a basis for evolutionary, biogeographical and systematic analyses of the clade. This analysis confirmed that the entire extant fauna native to the Old World is monophyletic and descended from a single New World colonist and found that most Old World taxa were in one of two clades, one including most Asian-African species and a second including most Australian-New Guinean species. Here, we follow these efforts with more concentrated attention on the phylogenetic patterns and biogeographical distributions of Old World *Pheidole*, pairing the phylogenetic data with biogeographical data on individual species distributions.

Pheidole and the taxon cycle

The taxon cycle hypothesis (Wilson, 1961; Ricklefs & Bermingham, 2002) provides a framework for generating predictions about biogeographical dynamics in the region. In synthesizing his taxonomic work on Indo-Pacific ants, Wilson proposed a dynamic process in which species expand and contract their ranges through phases of dispersal, niche shift and speciation (Wilson, 1959, 1961). A taxon cycle begins when a species in a highly diverse, mainland source area adapts to low-elevation 'marginal' habitats including coastlines, riverbanks and forest edges, which tend to be

more open and disturbed. These habitats facilitate dispersal and range expansion across the region, including colonization of remote island archipelagos. After range expansion, niche shifts occur in the recipient regions as the newly arrived colonists shift away from low-elevation marginal areas towards interior primary, and high-elevation forests. This precipitates local speciation and range fragmentation. Ultimately, lineages derived from widespread margin-adapted species transform and/or diversify into locally distributed interior forest species. In other words, the taxon cycle predicts that endemic interior forest faunas in the more remote Pacific archipelagos, such as Fiji and Micronesia, are derived from wide-ranging margin-adapted species rather than from other interior forest species directly colonizing from source areas. More comprehensive details of the taxon cycle are discussed in previous publications (Ricklefs & Bermingham, 2002; Economo & Sarnat, 2012).

In Wilson's view, central to this dynamic is a deterministic dominance hierarchy in which species evolving in source areas are better competitors than species resident in more isolated and species-poor recipient areas. This also implies a polarity of lineage flow; colonization occurs from mainland areas to islands, and not the reverse. Indeed, on the basis of morphological analyses of species groups, Wilson perceived that groups centred in Asia could expand to New Guinea and the Pacific, but groups centred in New Guinea or more remote Oceania could not invade 'upstream'. However, as this work was conducted in the 1950s, molecular phylogenetic analyses were not possible, and morphologically defined species groups have proven to be unreliable predictors of phylogeny in *Pheidole* (Moreau, 2008). More recent work has challenged the unidirectional hypothesis that islands are only lineage sinks. Notably, a case of upstream colonization has been demonstrated in monarch flycatchers, where a clade originated on remote Pacific islands and subsequently invaded and radiated in New Guinea (Filardi & Moyle, 2005), and there is also increasing evidence for east–west colonization and invasion of Asia by some insect groups (Balke *et al.*, 2009; Tanzler *et al.*, 2014). This prediction has not been tested with ants using phylogenetic methods until recently. In their molecular analysis of a widespread *Camponotus* species complex, Clouse *et al.* (2015) found evidence of complicated patterns of dispersal and colonization including both upstream and downstream movements of widespread species across the Indo-Pacific.

The taxon cycle is a complex narrative encompassing several constituent hypotheses, many of which are mutually independent. Economo & Sarnat (2012) tested several of these hypotheses by analysing ecological distributions of 183 ant species, including 17 *Pheidole*, across different levels of endemism in the Fiji. Their results supported several predictions including correlations between range size and habitat affinity, and evidence for post-colonization niche shift and radiation in *Pheidole*. However, given the sparse molecular sampling of Indo-Pacific *Pheidole* species at the time, their

data had little bearing on whether colonization was unidirectional in the region from high- to low-diversity regions. With phylogenetic reconstructions for a more comprehensive sampling of the region, our study allows us to better resolve the origins of the Fijian and other Pacific Island endemic radiations. Furthermore, it provides the requisite framework for inferring the origins of the several widespread 'expanding' species in the region, allowing us to test for a polarity to lineage movement.

MATERIALS AND METHODS

Biogeographical data

Our biogeographical data are taken from the Global Ant Biodiversity Informatics (GABI) project, a database consolidating literature, museum, and biodiversity database records on ant species distributions (Guénard *et al.*, in review). At the time of our analysis here, the database constitutes 1.3 million species occurrence records, including information from > 7500 published papers. From this source, we aggregated the occurrence records for all 484 valid Old World *Pheidole* species and subspecies assigned to 355 geographical areas. To this dataset, we added 109 morphospecies delineated during our DNA sequencing project. Despite the volume of data, some geographical areas remain undersampled. Although the modularity analysis performed here should be robust to variations in sampling effort among regions, two tropical regions were excluded for being severely undersampled (the Indian state of Manipur and the Maluku Archipelago, which, despite tropical climates, had fewer than five *Pheidole* records each). After removing regions with zero *Pheidole* richness, our incidence matrix contained 587 taxa distributed across 232 regions. See Appendix S1 in Supporting Information for specimens included in this study and associated GenBank accession numbers.

Network modularity analysis

Identifying clusters of geographical areas with high biotic affinities, and the borders between these clusters, has long been a goal of biogeographical research. Carstensen *et al.* (2012, 2013) recently proposed an interesting use of network theory to identify such 'modules', extending methods that are used more generally in network science (Guimera & Amaral, 2005) and network ecology (Olesen *et al.*, 2007). A set of areas and species can be considered as a bipartite graph, in which both geographical areas and species are nodes connected by links if the species occurs in that area. A network is considered 'modular' if the graph can be broken into sub-networks that are highly connected within groups but more weakly connected across groups, a concept appealingly similar to the idea of biogeographical regions. The modularity of a graph grouped into n modules can be summarized by a network modularity statistic, Q (also sometimes called M). We used the program MODULAR (Marquitti *et al.*,

2014) to find module configurations that maximize the *Q* statistic using simulated annealing with the following parameters (initial temperature factor: 2.0, cooling: 1.01, number of iterations: 1). We repeated analyses multiple times to check for consistency of results, and performed 100 replicates of the two different null models available in Modular to check for significance of the identified module partition. Null model 1 is an Erdõs–Rényi random graph and null model 2 randomizes network connections while preserving node degrees of both species and areas (i.e. occupancy of species and richness of areas, this model is equivalent to null model 2 in Bascompte *et al.* (2003)).

One attractive feature of the network method is that, beyond identifying modules of areas and species, it provides a convenient framework for summarizing the 'roles' of both areas and species in the network. In other words, it provides a basis to quantify how widespread species are within modules and the degree to which they are spread across modules. Although the modularity of *Pheidole* faunas is of intrinsic interest, our main purpose of using this method is to identify species that have anomalous distributions. The expanding phase of the taxon cycle (denoted 'Stage I' in Wilson's original hypothesis) implies an ability to jump biogeographical barriers, a characteristic which is not simply captured by comparing range sizes across species. In large, contiguous regions, such as mainland Asia or Africa, it may be typical for even poorly dispersive species to have large ranges. In contrast, in regions fragmented by ocean barriers, such as the Indo-Pacific, it may be very atypical for species to be spread across great distances. Carstensen *et al.* (2013) first made the connection between taxon cycle-stage and the 'role' that species may play in the network (in a statistical sense). Expanding Stage-I taxa ought to be species that are widespread *and* distributed across typical biogeographical barriers (i.e. are spread across modules); their ranges should be aberrant relative to other species. For each species we calculated two statistics: the *normalized degree* (z) and the *among module connectivity* (r). The *normalized degree* ($z_i = \frac{k_i}{k_m}$) is the total number of occurrences of species *i* across all modules (the degree, k_i, in network terminology) divided by the mean degree (k_m) of all species assigned to the same module as species *i*. The z_i statistic is an index of the range extent considering all occurrences (regardless of module), but adjusted for the fact that modules are different sizes and 'typical' species in different modules on average have different range extents. This differs slightly from Carstensen *et al.* (2012, 2013) who used normalized 'within module degree' which does not consider out of module connections. The second statistic, *among module connectivity,* is the same used by Carstensen and other authors, and measures how evenly a species' connections are distributed across modules. This is calculated as $r_i = 1 - \sum_m (p_{im})^2$ where *m* indexes over all modules and p_{im} is the fraction of occurrences of species *i* in module *m*. This statistic is zero when a species only occurs in one module and increases as a species' connections are more evenly spread across many modules.

Network visualization

Visualizing patterns in large, complex biodiversity datasets can be challenging, but the network representation facilitates the use of a wide variety of tools and algorithms for visualization. Recently, Misue (2008) proposed a method for visualization of bipartite networks (networks having two classes of nodes) by optimally arranging one class of nodes in circular form, fixing them in place, and allowing the other class of nodes to be arranged by a force-directed algorithm. This method was extended to hierarchical bipartite graphs using a multi-circular layout method that uses circular layouts for each module, but then lets nodes of the other class be arranged within and between modules by force-directed method (Ito *et al.*, 2010). At the time of this analysis, Ito's full procedure was not implemented in any available software package, thus we used a combination of programs to implement a simplified version of the procedure. In our network, the two classes of nodes are areas and species. We first used GraphDraw to perform a Misue layout of each module individually, using areas as the circularly arranged 'fixed' nodes and species as the 'free' nodes. This determines the optimal ordering of geographical area nodes around a circle. We then arranged the circular layouts of each module into a larger circle. Finally, with area nodes fixed in position in the multi-circular layout, we performed a force-directed layout using the Fruchterman–Reingold algorithm with the 'igraph' package (Csardi & Nepusz, 2006) in R to optimally arrange the free nodes (species).

Ecological data and analysis

We sought to characterize the habitat affinities of *Pheidole* species in order to test for correlations between range size, module breakout, and affinity with marginal habitat, a necessary (but not sufficient) component of the taxon cycle. A full characterization of the ecological niche of all species is not possible with available data, but we looked for general evidence that open/edge habitat-tolerant species have larger, anomalous ranges. Our main approach was to rely on previous taxonomic studies of *Pheidole* species, where the author(s) had characterized the ecology of the species. Secondarily, we used a few studies that performed intensive sampling on a particular region and provided a characterization of the spread of species across habitats (see Appendix S2). We did not attempt to make *de novo* characterizations of habitat affinities with specimen data, a much more complicated task that is beyond the scope of this paper, particularly in the light of taxonomic uncertainties in the genus. Although not perfect, we felt this approach relying on expert study of individual species is most reliable given the data available.

We sought to make a distinction between open habitats, disturbed forest and closed-canopy primary habitats. 'Open' environments include urban, agricultural fields, pastures, grassland and beaches. In the 'Disturbed forest' category, we

included descriptions that mentioned forest that was some-how impacted by human activities such as logging, partial or historical clearing, forest edges or other anthropogenic disturbance, in addition to forest under natural disturbance regimes such as littoral or riverbank forest. Habitats either directly identified by the authors as primary or not identified as disturbed were considered 'Primary Forest'.

The current analysis, like both Wilson's original contributions and subsequent analyses, is centred on the moist tropical biomes widespread in Southeast Asia, Melanesia, and the Pacific islands. It is unclear how the ideas behind the taxon cycle in terms of the connection between dispersal and open habitat apply or do not apply in naturally open habitats such as xeric and semi-arid environments. For this reason, although the phylogenetic analysis geographically covers the entire Old World, the ecological analyses in this paper are concentrated in the tropical moist areas of Southeast Asia and the Pacific Islands where the taxon cycle has been proposed. A better consideration of how species evolved in naturally open habitats, such as savannas and deserts, do or do not fit into the framework is an interesting direction for future work.

We first assigned species into two categories, those that had been recorded as occurring in open/disturbed habitats (but may or may not also occur in interior forest) and those that are only known to occur in interior primary forest (Analysis 1). Additionally, we performed a second analysis by dividing the groups an alternate way; by those that have been known to occur in open habitat, and those that are not (Analysis 2). The difference between the two versions of the analysis is that in the first, species need to be only in disturbed forest or forest edge to be in the 'open-tolerant' group, while in the second they must be in true non-forest open habitat to be in the 'open-tolerant' group. We performed a one-tailed Wilcoxon rank-sum nonparametric test to evaluate the hypothesis that species inhabiting open/disturbed habitats are more likely to have large/anomalous ranges as quantified by the network statistics (z_i, r_i). We performed these analyses with and without known human-dispersed species included.

Phylogenetic reconstruction

We inferred a time-calibrated phylogeny of Old World *Pheidole* in a Bayesian framework. We used the molecular dataset generated by Economo *et al.* (2015), which in turn added new sequencing to datasets generated by Moreau (2008) and Sarnat & Moreau (2011). The study by Economo *et al.* established the main topological features of the Old World *Pheidole* tree, and our focus here is a time-calibrated analysis of Old World lineages that can be used for biogeographical analysis. We selected 177 Old World *Pheidole* species for our analysis, and as Economo *et al.* demonstrated with high confidence that the Old World *Pheidole* were monophyletic, we used two New World species (*Pheidole rhea* and *P. hyatti*) and one taxon from a related genus (*Cephalotes setulifer*) as

outgroups. The species in the dataset were delineated through traditional morphology-based taxonomic analysis, but subsequently revised and informed by preliminary rounds of sequencing. We analysed sequences from nine loci [CO1 (cytochrome oxidase 1), 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), CAD1 (conserved ATPase domain protein gene), EF-1α F2 (elongation factor 1-alpha F2) and Top1 (DNA topoisomerase 1); total of 8820 base pairs, aligned], with one substitution model partition for each gene. Data from three loci (CO1, His3.3B, Lop1) are available for nearly all taxa, while the full nine loci are available for a representative subset of species dispersed across the different clades. We used the program BEAST 2.0 (Bouckaert *et al.*, 2014) to perform a time-calibrated Bayesian phylogenetic analysis. We used a relaxed lognormally distributed molecular clock unlinked across all partitions, Yule prior, and a GTR+G+I model of molecular evolution for all loci except for two (unc_4, GRIK2) which used a GTR+G model. We used the nucleotide models found by Economo *et al.* (2015), who used PARTITIONFINDER (Lanfear *et al.*, 2012) for model selection on the same dataset. We ran 10 independent runs for 40 million generations, used the program TRACER to determine the length of burnin for individual runs, and compared posteriors of parameters across multiple runs. On the basis of this, we removed the trees sampled from the first 20 million states and thinned the combined posterior tree sample to approximately 16,540 trees using LOGCOMBINER. From this posterior distribution, we calculated the maximum clade credibility tree using TREEANNOTATOR and used this tree for our analysis. We ran the analysis in relative time and subsequently dated the tree with the mean Old World crown node age inferred by Economo *et al.* (2015) in a genus-wide species-tree analysis.

Biogeographical ancestral range reconstruction

We sought to determine whether there is bias to the flow of lineages across biogeographical regions as predicted by the taxon cycle. We used a maximum likelihood method for inferring geographical range evolution implemented in the R package BioGeoBears using the maximum clade credibility tree from the BEAST analysis (after pruning all outgroups). BioGeoBears (Matzke, 2013) fits a variety of different models describing rates of dispersal, extinction, cladogenesis, vicariance and founder event speciation. We fitted all six models available in the package (DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J), and compared their fit with the Akaike information criterion (AIC). Ranges for each species were coded as present/absent in each of the eight modules detected by the network analysis, with a limitation that species cannot occupy more than five modules at once (necessary for reducing the state space). Using the best model selected from the six options, we found the marginal

probability of each possible range configuration at each internal node. Finally, we summed over these to find the marginal probability that a lineage occupied each module (irrespective of whether it occupies other modules), and mapped those onto the tree to provide a general picture of the biogeographical origins of each clade.

RESULTS

Biogeographical structure of Old World *Pheidole*

The network analysis consistently identified eight modules in multiple independent runs. The network had a modularity of $Q = 0.73$ which was much higher than null distributions from 100 randomizations under both null models (Null 1: $Q = 0.51$, min = 0.50, max = 0.53, Null 2: $Q = 0.49$, min = 0.47, max = 0.52), implying that the spatial organization of *Pheidole* faunas is significantly modular. The eight modules (Figs 1 & 2) are generally consistent with known biogeographical regions in the Old World, with modules including the Malagasy, Afrotropical, Saharo-Arabian and Mediterranean, New Guinean, Australian and Oceanian regions. Most of Asia was split between two modules, one including India, China and more northern regions (which we call 'Indian-Chinese'), and the other including parts of Southeast Asia and Sundaland (which we call 'Malayan').

Considering that the network method does not use spatial information to cluster areas, the modules are remarkably contiguous. Of the 587 species included in the analysis, 527 are exclusively found within one module, although some, such as *Pheidole pallidula*, are wide-ranging within the module (Figs 2 & 3). The species with the most anomalous range is *P. megacephala* (Figs 2 & 3), which is unsurprising considering that the species is one of the five worst invasive ant species in the world as recognized by IUCN and has been spread around the world through human commerce (McGlynn, 1999; Lowe *et al.*, 2000). The remaining anomalous species include both 'borderline' species that are generally not very widespread but are found in more than one module, and those that are both widespread and distributed across modules.

We collected published habitat affinities for 85 Asian–Pacific species (Appendix S2). The one-tailed Wilcoxon tests found that species occupying open/disturbed habitats had a significantly higher r_i and are thus more likely to be spread across modules and have a higher z_i, meaning they are more widespread in general relative to other species in their module (Table 1), and these results were robust to different ways of grouping the data and the inclusion or exclusion of human-transported species.

Phylogenetic structure and biogeographical history of Old World *Pheidole*

Our phylogenetic analyses recovered a well-supported topology that was broadly consistent with previous results (Fig. 4, Appendix S3). Most Old World species are divided between two major *Pheidole* clades, one composed of mostly New Guinean and Australian taxa, and the other composed of mostly African and Asian *Pheidole* species. The Pacific *Pheidole* faunas consist of smaller radiations nested within these larger clades as well as widespread lineages. Many of the endemic species occurring in Fiji, Vanuatu, Micronesia and the Solomon Islands belong to a single clade nested within Asian groups. In contrast, all endemic New Caledonian species were monophyletic and nested within the New Guinea-Australia clade.

The AIC-minimizing model was DEC+J with an AIC of 480.2, while the next closest model (DIVALIKE+J) had an AIC of 491.9, and all the rest were > 500. A AIC of > 2 is typically considered meaningful, so we interpreted these results as clear support for the DEC+J model and used it for our analysis. The DEC+J model is the dispersal–extinction–cladogenesis model implemented in LAGRANGE (Ree & Smith, 2008), but with an extra process of founder event speciation captured by the 'J' parameter (Matzke, 2014).
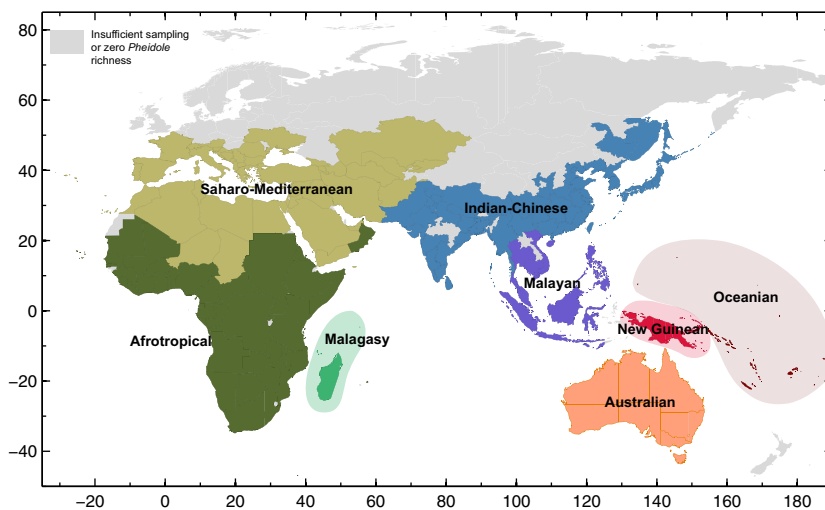

Figure 1 The areas assigned to each of the eight biogeographical modules identified by network modularity analysis of Old World *Pheidole* species distributions.
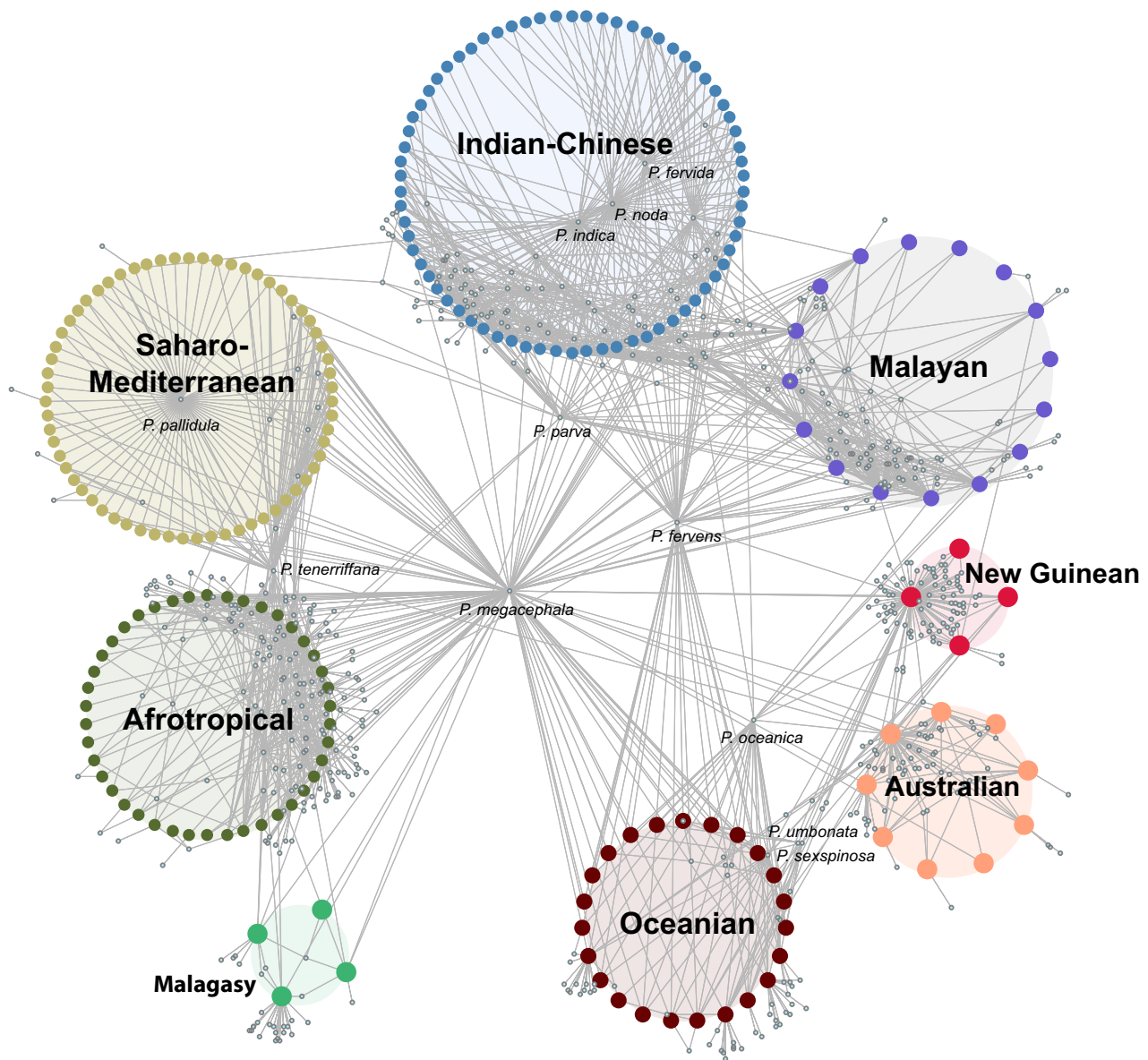
Figure 2 Visualization of the network of areas (larger dots) and *Pheidole* species (small dots), including the modules inferred by the network modularity analysis. The areas correspond to the regions depicted in Fig. 1, and widespread species discussed in the text are identified by name. The network layout is generated following the method for hierarchical bipartite graphs described by Ito *et al.* (2010).

Origins and spread of widespread species

The phylogeny allows us to ascertain the likely geographical origins of the widespread 'expanding' *Pheidole* species in the Pacific (*P. oceanica, P. fervens, P. umbonata, P. sexspinosa, P. parva*) and biogeographical data provide a record of their distributions across the regions. The taxon cycle predicts a faunal cascade from Asia to New Guinea and to the Pacific Islands, but no colonization 'upstream' in the reverse direction. Interestingly, there is evidence for this pattern in our data (Figs 4 & 5). *Pheidole oceanica* and *P. fervens* are related species from the same Asia-centred species group, but have spread across the region and successfully colonized nearly

every archipelago in the Pacific region. *Pheidole parva* has also expanded from Asia into the eastern Indo-Pacific, but has not reached Melanesia, Micronesia east of Palau, or Polynesia. Two widespread species, *P. umbonata* and *P. sexspinosa*, evolved from New Guinean lineages and have colonized Melanesia, Micronesia and the more remote Pacific, but have not colonized Asia (including nearby islands of Wallacea).

In addition to these species that are currently widespread in the region, there are a number of other more spatially restricted lineages that are likely descended from older Asian colonists. Notably, *P. hospes* and (it is possible conspecific) *P. impressiceps* in the *fervens* clade are restricted to – but
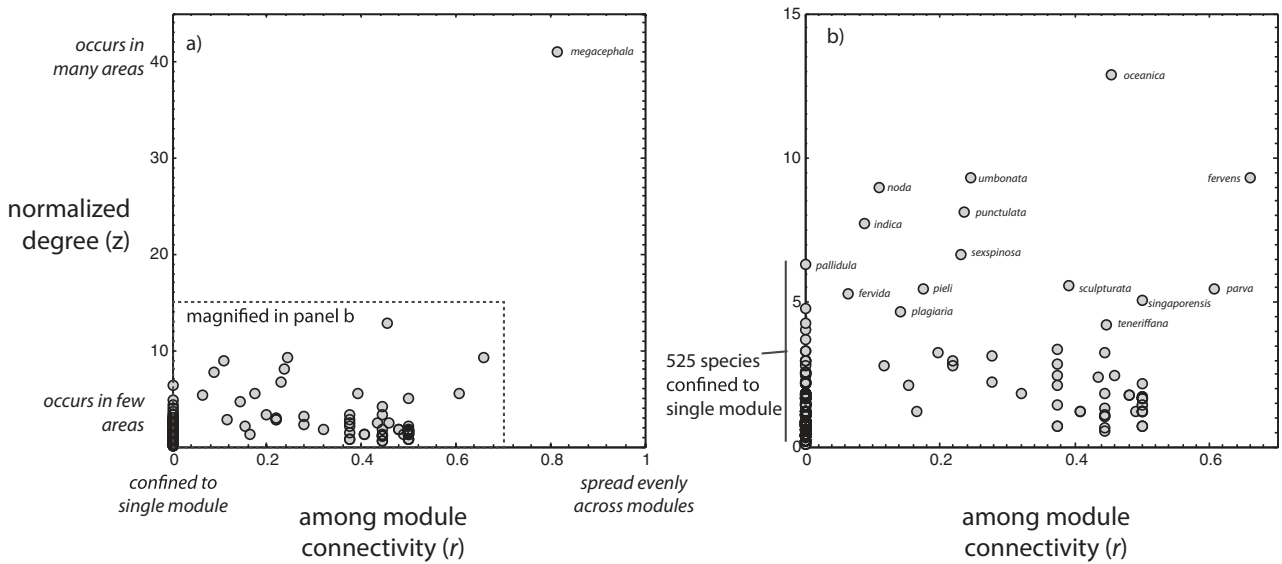
Figure 3 *Pheidole* species network 'role' plotted in two dimensions. The *x*-axis reflects the spread of a species across modules while the *y*-axis reflects the overall geographical spread of the species.

Table 1 One-tailed Wilcoxon test of two hypotheses relating *Pheidole* species' range to open-tolerance. In Analysis 1, we classified species known from forest edges/disturbed forest as 'open tolerant', whereas in Analysis 2, only species found in true open habitat were classified as 'open tolerant'. Tramps refer to species known to be spread by humans.

	Number of open-tolerant species	Number of open-intolerant species	H: Open-tolerant species have higher z_i; more widespread in general relative to other species in module		H: Open-tolerant species have higher r_i; species are more spread across different modules	
			Z	P	Z	P
Analysis 1, with tramps	36	49	5.95	$<3 \times 10^{-9}$	8.51	$<9 \times 10^{-18}$
Analysis 1, without tramps	31	49	5.36	$<5 \times 10^{-8}$	8.42	$<2 \times 10^{-17}$
Analysis 2, with tramps	20	65	9.16	$<1 \times 10^{-19}$	4.81	$<8 \times 10^{-7}$
Analysis 2, without tramps	15	65	8.88	$<1 \times 10^{-19}$	4.47	$<4 \times 10^{-6}$

abundant in – New Guinea and Australia. In the Asian *parva* group, there are several species restricted to Melanesia and Micronesia (e.g. *Pheidole nindi*, *Pheidole* epem199). The clade including the Oceania-centred *cryptocera*, *mendanai*, *knowlesi* and *roosevelti* groups is part of a broader African-Asian clade and not the Australasian clade, implying an old colonization of the Pacific. Yet, no lineages from the large radiations in Australia and New Guinea have managed to colonize 'upstream' into Asia (i.e. the areas contained in the Malayan and Indian-Chinese modules), although several eastward colonizations to the Pacific have occurred including *P. sexspinosa* and *P. umbonata* (Fig. 5).

DISCUSSION

In the last 20 million years, a single lineage of *Pheidole* colonized the Old World and spread across the hemisphere as it diversified into over 500 species. Reconstructing this history and understanding the dynamics shaping it provides insight

into the evolution and success of one of the most dominant lineages on the globe. In this study, we used a macroecological perspective to characterize the species-level biogeographical structure of *Pheidole*, identifying eight biogeographical modules of interconnected species and areas. We identified species that had anomalous geographical distributions; those that crossed typical biogeographical boundaries and thus show evidence of 'Stage-I' expanding status. Second, we investigated the phylogenetic structure underlying these modules and the origins of these widespread species.

Our results suggest that widespread expanding species are more likely to occupy marginal habitats, supporting the hypothesis that this ecological niche facilitates range expansion, a key component of the taxon cycle. The widespread species are distributed across the phylogeny, originating in clades centred in both Southeast Asia (*oceanica*, *fervens*, *parva*) and New Guinea (*umbonata*, *sexspinosa*). Interestingly, we also find support for a bias in lineage movement. Widespread species originating in Asia colonize New Guinea, Aus-
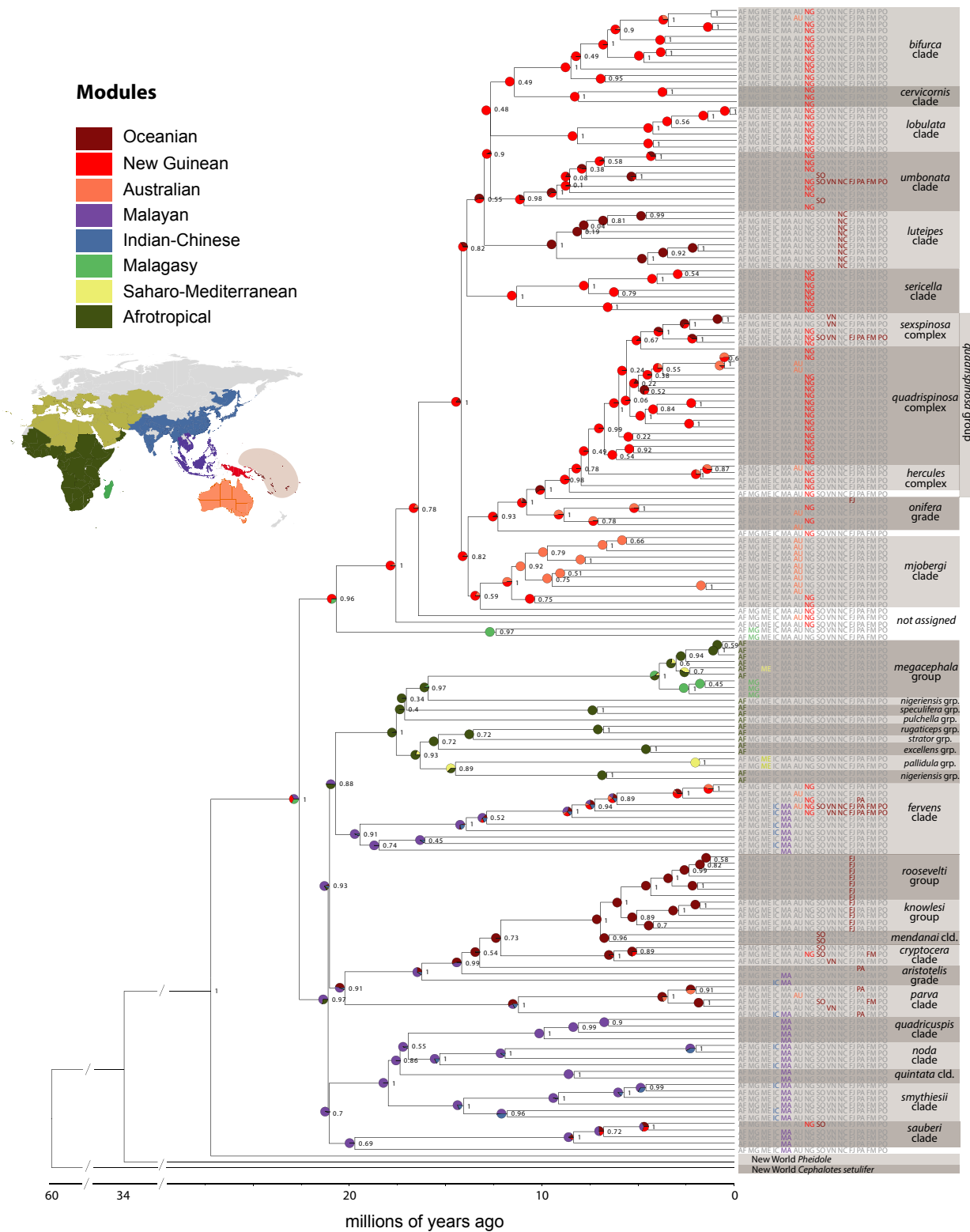
Figure 4 Maximum clade credibility phylogeny of Old World *Pheidole* inferred with BEAST 2.0 and dated with the crown age inferred by Economo *et al.* (2015). Nodes are labelled with Bayesian posterior probabilities and marginal probabilities of range occupancy inferred with BioGeoBears. The ranges of extant species are denoted with abbreviations for modules in Fig. 2 as Afrotropical (AF), Malagasy (MG), Saharo-Mediterranean (ME), Indian-Chinese (IC), Malayan (M), Australian (AU) and New Guinean (NG). The island groups of the Oceanian module are denoted individually as Solomon Islands (SO), Vanuatu (VN), New Caledonia (NC), Fiji (FJ), Palau (PA), Federated States of Micronesia (FM) and Polynesia east of Fiji (PO). A high-resolution version of this figure, with taxon names, is available online in Appendix S3.
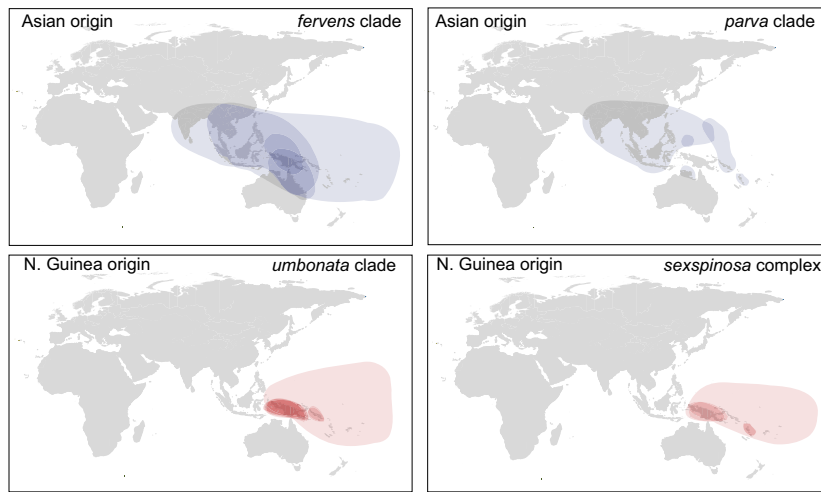
Figure 5 Approximate ranges of widespread Indo-Pacific *Pheidole* species and their close relatives. The panels depict groups that are inferred to have originated in Asia (top row) and New Guinea/Australia (bottom row). Expanding taxa from New Guinea have spread eastward, but have not colonized westward to nearby 'upstream' areas.

tralia and the Pacific Islands, while those originating in New Guinea apparently cannot colonize Asian regions. For example, *P. umbonata* and *P. sexspinosa*, two species originating in New Guinea, have colonized eastward to most of the remote Pacific, yet we have no record of their presence in 'upstream' areas relatively close to New Guinea, such as the archipelagos of Sundaland or mainland Asia. Wilson (1959, 1961) hypothesized this was due to competitive asymmetry: species evolving in highly diverse areas are more competitively dominant, and diversity is reduced as one moves further from mainland Asia, resulting in successful colonizations downstream but not upstream. This could explain a lack of back-colonization for the Pacific islands such as Fiji, but there is no evidence that New Guinea or Australia are actually less diverse than the Southeast Asian tropics for ants.

While there is evidence for a connection between ecological niche and range expansion, and evidence for a polarity to lineage movement, both in accordance with taxon cycle predictions, other aspects of biogeographical pattern do not conform as clearly to taxon cycle predictions. Many of the endemic *Pheidole* species in Fiji, Micronesia, the Solomon Islands and Vanuatu (including the *cryptocera*, *mendanai*, *roosevelti* and *knowlesi* groups) belong to a single clade descended from a colonist from Asia and centred in eastern Melanesia and Micronesia. Endemic New Caledonian *Pheidole* mainly belong a single radiation nested within the NG-Australian clades. While subsequent colonizations have led to the formation of endemic species in these recipient areas (e.g. the species *P. onifera* in Fiji), there is not clear evidence of cladogenesis by any of these colonizing lineages. Thus, there is no evidence that after the first *Pheidole* lineage initially colonizes an area and populates it with species through cladogenesis, that new colonizations can subsequently radiate and displace the residents. An alternative hypothesis to the taxon cycle that could explain this pattern

is that initial colonization leads to radiation and filling of niche space on the areas, and later expanding widespread lineages may gain a foothold on an island and possibly evolve into an endemic species, but the widespread lineages do not radiate and replace the incumbent fauna descended from the older colonization.

Another question that arises from the current analysis is the extent to which the cycle is a species- or broader clade level-phenomenon. The taxon cycle involves a niche shift into marginal habitats and range expansion, followed by range fragmentation and localized niche shifts out of disturbed habitats across the range. Generally, in previous studies the range expansion state has been discussed as arising in a single species otherwise nested in a clade of interior forest species. It is possible that the traits that facilitate range expansion may persist for longer evolutionary time-scales, resulting in entire clades that more readily disperse across barriers (or more readily produce expanding species). Indeed, there do seem to be some hints of this in the *Pheidole* biogeographical patterns. The *fervens* clade includes two widely distributed species, *P. oceanica* and *P. fervens* (which are also sometimes spread by human commerce), in addition to several other lineages localized in different parts of Southeast Asia and Oceania. Other clades, such as the *parva* and *sauberi* groups, contain closely related species that are disjunct geographically (and in the case of the *parva* group, also containing a widespread species). This is also true in the Afrotropical and Malagasy regions. The widespread invasive *P. megacephala* is part of a clade that includes other geographically widespread species, such as *P. punctulata*. Moreover, lineages of the *megacephala* group have successfully colonized Madagascar (in addition to the human-mediated introduction of *P. megacephala*), a feat that has not been achieved by any other Afrotropical species groups in the region (Fischer & Fisher, 2013). These cases seem to imply

that entire clades may share traits or ecological habits that promote dispersal ability and range expansion. If the traits that promote range expansion ability are conserved over macroevolutionary time-scales, this could lead to pulses of expansion, radiation, and contraction that are similar to taxon cycles, but on a deeper phylogenetic scale.

Additionally, the data suggest that *Pheidole* species that are also spread by human commerce are likely to be members of these same clades. We hasten to add that in some cases (in particular, *P. oceanica* and *P. parva*) it is difficult to delineate the actual native and introduced ranges of these species. Indeed, many of the patterns suggestive of a natural taxon cycle might be influenced by human activities and introductions. Further molecular phylogeographical studies of widespread species complexes are needed to determine with more precision the native and introduced ranges in the region.

Our analysis provides both biogeographical and phylogenetic structure to the massive Old World *Pheidole* fauna, which, combined with taxonomic efforts, will hopefully facilitate future progress on understanding the evolution of this interesting and anomalous genus. However, although our sampling was extensive from New Guinea and the Pacific Islands, and included representatives from other areas, major gaps in our knowledge remain. While the faunas of all areas could benefit from additional phylogenetic attention, the large fauna of Madagascar is only represented by three species in our molecular dataset, the Afrotropics in general could be better sampled and our sampling lacks any species at all from the Middle East, Central Asia or the Indian subcontinent. In all tropical regions a huge number of *Pheidole* species remain undescribed. For all the undescribed and most of the described taxa, we have sparse or no ecological information, which provides a fundamental limitation to our understanding of the interaction between evolutionary, ecological and biogeographical processes in the genus.

Aside from the usual need for more empirical data, we sorely need further development of quantitative methods that are capable of statistically rigorous tests of complex biogeographical hypotheses such as the taxon cycle. As with previous studies on the topic, our current analysis relies heavily on qualitative interpretation of patterns. Such efforts have historically paved the way for insight in evolutionary biology, but to persist in the modern era, complex integrative hypotheses should motivate development of sophisticated quantitative methods that we can use to challenge our interpretations.

ACKNOWLEDGEMENTS

REFERENCES

Balke, M., Ribera, I., Hendrich, L., Miller, M.A., Sagata, K., Posman, A., Vogler, A.P. & Meier, R. (2009) New Guinea highland origin of a widespread arthropod supertramp. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2359–2367.

Bascompte, J., Jordano, P., Melian, C.J. & Olesen, J.M. (2003) The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences USA*, **100**, 9383–9387.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Carstensen, D.W., Dalsgaard, B., Svenning, J.C., Rahbek, C., Fjeldså, J., Sutherland, W.J. & Olesen, J.M. (2012) Biogeographical modules and island roles: a comparison of Wallacea and the West Indies. *Journal of Biogeography*, **39**, 739–749.

Carstensen, D.W., Dalsgaard, B., Svenning, J.C., Rahbek, C., Fjeldså, J., Sutherland, W.J. & Olesen, J.M. (2013) The functional biogeography of species: biogeographical species roles of birds in Wallacea and the West Indies. *Ecography*, **36**, 1097–1105.

Clouse, R.M., Janda, M., Blanchard, B., Sharma, P., Hoffmann, B.D., Andersen, A.N., Czekanski-Moir, J.E., Krushelnycky, P., Rabeling, C., Wilson, E.O., Economo, E.P., Sarnat, E.M., General, D.M., Alpert, G.D. & Wheeler, W.C. (2015) Molecular phylogeny of Indo-Pacific carpenter ants (Hymenoptera: Formicidae, *Camponotus*) reveals waves of dispersal and colonization from diverse source areas. *Cladistics*, **31**, 424–437.

Cook, B.D., Pringle, C.M. & Hughes, J.M. (2008) Molecular evidence for sequential colonization and taxon cycling in freshwater decapod shrimps on a Caribbean island. *Molecular Ecology*, **17**, 1066–1075.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, Complex Systems, 1695.

Economo, E.P. & Sarnat, E.M. (2012) Revisiting the ants of Melanesia and the taxon cycle: historical and human-mediated invasions of a tropical archipelago. *The American Naturalist*, **180**, E1–E16.

Economo, E.P., Klimov, P., Sarnat, E.M., Guénard, B., Weiser, M.D., Lecroq, B. & Knowles, L.L. (2015) Global phylogenetic structure of the hyperdiverse ant genus *Pheidole* reveals the repeated evolution of macroecological patterns.

Proceedings of the Royal Society B: Biological Sciences, **282**, 20141416.

Eguchi, K. (2001a) A revision of the Bornean species of the ant genus *Pheidole* (Insecta: Hymenoptera: Formicidae: Myrmicinae). *TROPICS Monograph Series*, **2**, 1–154.

Eguchi, K. (2001b) A taxonomic study on Asian *Pheidole* (Hymenoptera: Formicidae): new synonymy, rank changes, lectotype designations and redescriptions. *Insecta Koreana*, **18**, 1–35.

Eguchi, K. (2006) Six new species of *Pheidole* Westwood from North Vietnam (Hymenoptera, Formicidae). *Revue Suisse de Zoologie*, **113**, 115–131.

Eguchi, K. (2008) A revision of Northern Vietnamese species of the ant genus *Pheidole* (Insecta: Hymenoptera: Formicidae: Myrmicinae). *Zootaxa*, **1902**, 1–118.

Eguchi, K., Yamane, S. & Zhou, S.Y. (2007) Taxonomic revision of the *Pheidole rinae* Emery complex. *Sociobiology*, **50**, 257–284.

Filardi, C.E. & Moyle, R.G. (2005) Single origin of a pan-Pacific bird group and upstream colonization of Australasia. *Nature*, **438**, 216–9.

Fischer, G. & Fisher, B.L. (2013) A revision of *Pheidole* Westwood (Hymenoptera: Formicidae) in the islands of the southwest Indian Ocean and designation of a neotype for the invasive *Pheidole megacephala*. *Zootaxa*, **3683**, 301–356.

Fischer, G., Garcia, F.H. & Peters, M.K. (2012) Taxonomy of the ant genus *Pheidole* Westwood (Hymenoptera: Formicidae) in the Afrotropical zoogeographic region: definition of species groups and systematic revision of the *Pheidole pulchella* group. *Zootaxa*, **3232**, 1–43.

Guimera, R. & Amaral, L.A.N. (2005) Functional cartography of complex metabolic networks. *Nature*, **433**, 895–900.

Ito, T., Misue, K. & Tanaka, J. (2010) Drawing clustered bipartite graphs in multi-circular style. *14th International Conference on Information Visualisation*, **IV**, 23–28.

Johnson, M.L. & Gaines, M.S. (1990) Evolution of dispersal: theoretical models and empirical tests using birds and mammals. *Annual Review of Ecology and Systematics*, **21**, 449–480.

Jønsson, K.A., Irestedt, M., Christidis, L., Clegg, S.M., Holt, B.G. & Fjeldså, J. (2014) Evidence of taxon cycles in an Indo-Pacific passerine bird radiation (Aves: Pachycephala). *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20131727.

Lanfear, R., Calcott, B., Ho, S.Y. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Levin, S.A., Cohen, D. & Hastings, A. (1984) Dispersal strategies in patchy environments. *Theoretical Population Biology*, **26**, 165–191.

Longino, J.T. (2009) Additions to the taxonomy of New World *Pheidole* (Hymenoptera: Formicidae). *Zootaxa*, **2181**, 1–90.

Longino, J.T. & Cox, D.J. (2009) *Pheidole bilimeki* reconsidered (Hymenoptera: Formicidae). *Zootaxa*, **1985**, 34–42.

Lowe, S., Browne, M., Boudjelas, M. & De Poorter, M. (2000) *100 of the world's worst invasive alien species: a selection from the Global Invasive Species Database*. Invasive Species Specialist Group, Auckland, New Zealand.

Marquitti, F.M.D., Guimaraes, P.R., Pires, M.M. & Bittencourt, L.F. (2014) MODULAR: software for the autonomous computation of modularity in large network sets. *Ecography*, **37**, 221–224.

Matzke, N.J. (2013) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, **5**, 242–248.

Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, **63**, 951–970.

McGlynn, T.P. (1999) The worldwide transfer of ants: geographical distribution and ecological invasions. *Journal of Biogeography*, **26**, 535–548.

Misue, K. (2008) Anchored map: graph drawing technique to support network mining. *IEICE Transactions on Information and Systems*, **E91d**, 2599–2606.

Moreau, C.S. (2008) Unraveling the evolutionary history of the hyperdiverse ant genus *Pheidole* (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, **48**, 224–239.

Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. (2007) The modularity of pollination networks. *Proceedings of the National Academy of Sciences USA*, **104**, 19891–19896.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

Ricklefs, R.E. & Bermingham, E. (2002) The concept of the taxon cycle in biogeography. *Global Ecology and Biogeography*, **11**, 353–361.

Ricklefs, R. & Bermingham, E. (2008) The West Indies as a laboratory of biogeography and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2393–2413.

Ricklefs, R.E. & Cox, G.W. (1972) Taxon cycles in West-Indian avifauna. *The American Naturalist*, **106**, 195–219.

Sarnat, E.M. (2008) A taxonomic revision of the *Pheidole roosevelti*-group (Hymenoptera: Formicidae) in Fiji. *Zootaxa*, **1767**, 1–36.

Sarnat, E.M. & Economo, E.P. (2012) *The ants of Fiji*. University of California Press, Berkeley, CA.

Sarnat, E.M. & Moreau, C.S. (2011) Biogeography and morphological evolution in a Pacific island ant radiation. *Molecular Ecology*, **20**, 114–130.

Tanzler, R., Toussaint, E.F.A., Suhardjono, Y.R., Balke, M. & Riedel, A. (2014) Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132528.

Van Bocxlaer, I., Loader, S.P., Roelants, K., Biju, S.D., Menegon, M. & Bossuyt, F. (2010) Gradual adaptation toward a range-expansion phenotype initiated the global radiation of toads. *Science*, **327**, 679–682.

Wilson, E.O. (1959) Adaptive shift and dispersal in a tropical ant fauna. *Evolution*, **13**, 122–144.

Wilson, E.O. (1961) The nature of the taxon cycle in the Melanesian ant fauna. *The American Naturalist*, **95**, 169–193.

Wilson, E.O. (2003) Pheidole *in the New World: a dominant, hyperdiverse ant genus*. Harvard University Press, Cambridge, MA.

SUPPORTING INFORMATION

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

Appendix S1 Specimen data and GenBank accession numbers for sequences used in the phylogenetic analyses.

Appendix S2 Supplemental methods, data and references related to the ecological habitat analyses.

Appendix S3 High-resolution version of the *Pheidole* phylogeny (Fig. 4) presented with additional annotation.

BIOSKETCH

Evan P. Economo is interested in documenting and understanding biodiversity patterns and processes from a theoretical and empirical perspective.

Author contributions: E.P.E., E.M.S. and L.L.K. designed the research. E.P.E., E.M.S., M.J., R.C., B.G., A.N.A., A.L., C.R., G.F., E.O.W. and M.B. collected and contributed specimens and helped interpret/check molecular results. E.P.E., E.M.S., G.F. and B.D.B. performed the specimen work. L.N.R., P.K. and L.L.K. performed the molecular work. E.P.E. analysed the data and led the writing, with input from all co-authors.

Editor: Kate Parr