Systematic Entomology

Systematic Entomology (2020), **45**, 1–19 DOI: 10.1111/syen.12371

Inverse dispersal patterns in a group of ant parasitoids (Hymenoptera: Eucharitidae: Oraseminae) and their ant hosts

AUSTIN J. BAKER[1], JOHN M. HERATY[1], JASON MOTTERN[1], JUNXIA ZHANG[1], HEATHER M. HINES[2], ALAN R. LEMMON[3] and EMILY MORIARTY LEMMON[4]

[1]Department of Entomology, University of California, Riverside, CA, U.S.A., [2]Department of Biology, Pennsylvania State University, University Park, PA, U.S.A., [3]Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL, U.S.A. and [4]Department of Biological Science, Florida State University, Tallahassee, FL, U.S.A.

Abstract. When postulating evolutionary hypotheses for diverse groups of taxa using molecular data, there is a tradeoff between sampling large numbers of taxa with a few Sanger-sequenced genes or sampling fewer taxa with hundreds to thousands of next-generation-sequenced genes. High taxon sampling enables the testing of evolutionary hypotheses that are sensitive to sampling bias (i.e. dating, biogeography and diversification analyses), whereas high character sampling improves resolution of critical nodes. In a group of ant parasitoids (Hymenoptera: Eucharitidae: Oraseminae), we analyse both of these types of datasets independently (203 taxa with five Sanger loci, 92 taxa with 348 anchored hybrid enrichment loci) and in combination (229 taxa, 353 loci) to explore divergence dating, biogeography, host relationships and differential rates of diversification. Oraseminae specialize as parasitoids of the immature stages of ants in the subfamily Myrmicinae (Hymenoptera: Formicidae), with ants in the genus *Pheidole* being their most common and presumed ancestral host. A general assumption is that the distribution of the parasite must be limited by any range contraction or expansion of its host. Recent studies support a single New World to Old World dispersal pattern for *Pheidole* at *c*. 11–27 Ma. Using multiple phylogenetic inference methods (parsimony, maximum likelihood, dated Bayesian and coalescent analyses), we provide a robust phylogeny showing that Oraseminae dispersed in the opposite direction, from Old World to New World, *c*. 24–33 Ma, which implies that they existed in the Old World before their presumed ancestral hosts. Their dispersal into the New World appears to have promoted an increased diversification rate. Both the host and parasitoid show single unidirectional dispersals in accordance with the presence of the Beringian Land Bridge during the Oligocene, a time when the changing northern climate probably limited the dispersal ability of such tropically adapted groups.

Introduction

Next-generation sequencing (NGS) methods in phylogenetics are becoming increasingly popular for resolving contentious relationships among taxa (Maddison, 2016). Anchored hybrid enrichment (AHE) has been developed as an efficient and low-cost means to sequence hundreds of orthologues that provide phylogenetic signal at both deep and shallow scale analyses (Lemmon *et al.*, 2012; Lemmon & Lemmon, 2013). In comparison, traditional Sanger sequencing is an affordable, though less efficient, means of sequencing a few loci for many individuals. Having both types of datasets provides the opportunity to combine taxon-rich and character-rich datasets to create a phylogeny that is well sampled with high backbone support; however, this combination results in large amounts of

Correspondence: Austin J. Baker, Department of Entomology, University of California, Riverside, CA 92521, U.S.A. E-mail: abake005@ucr.edu

missing data. To test the efficacy of combining these two types of datasets, we compare the results from analyses using separate (Sanger-only and AHE-only) and combined (Sanger+AHE) datasets analysed under multiple phylogenetic frameworks [parsimony, maximum likelihood (ML), Bayesian inference, coalescence] to thoroughly explore the phylogenetics, dating, ancestral host associations and biogeography of a group of ant parasitoids.

In the northern hemisphere, biogeographic hypotheses are complicated by the on-and-off connections between continents and high variability in climatic conditions (Sanmartín *et al.*, 2001; Milne, 2006). The North Atlantic Land Bridges (NALBs) were important for faunal exchange before *c*. 50 Ma (Thulean) and 39 Ma (De Geer) (McKenna, 1983; Tiffney, 1985), while the Bering Land Bridge (BLB), which connected Western North America to Eastern Eurasia from *c*. 65 to 5.5 Ma (Sanmartín *et al.*, 2001; Milne, 2006), is considered to be a more likely route for more recently dispersed groups. The change in climate and associated flora and fauna in the northern hemisphere during the Tertiary allows the BLB hypothesis to explain distributions of both tropical and temperate biota. Groups with disjunct tropical or subtropical distributions probably crossed the BLB before *c*. 15 Ma, when the northern climate was warm with high equability (i.e. low variation between summer and winter temperatures) (Wolfe, 1993; Milne, 2006), a pattern commonly seen in dated phylogenies [e.g. lizards (Brandley *et al.*, 2011; Gamble *et al.*, 2011; Townsend *et al.*, 2011), snakes (Burbrink & Lawson, 2007; Wüster *et al.* 2008; Guo *et al.*, 2012; Chen *et al.*, 2013), butterflies (Condamine *et al.*, 2013), ambush bugs (Masonick *et al.*, 2017) and ants (Blaimer *et al.*, 2016)]. We investigate the timing and patterns of dispersal relative to these events in a parasitoid lineage and its ant hosts.

Parasitoids are a specialized subset of holometabolous insect parasites whose larvae feed on and kill a single arthropod host (Askew, 1971; Eggleton & Belshaw, 1992). The wasp family, Eucharitidae, is composed of obligate ant parasitoids and is the only insect family known to specialize on attacking ant brood (Clausen, 1941; Heraty, 1994b; Heraty, 2000; Heraty, 2002). They exhibit both host specificity and a high degree of endemism, which makes them an excellent candidate for studies of ancient host associations and modelling of biogeographic events (Heraty, 1994b; Heraty, 2000).

Oraseminae (Hymenoptera: Eucharitidae) parasitoid wasps are common throughout tropical and subtropical environments worldwide, with only a few widespread species found in temperate environments in the Nearctic. Oraseminae are specialized parasitoids of the ant subfamily Myrmicinae (Clausen, 1941; Heraty, 1994b; Heraty, 2000; Heraty, 2002; Lachaud & Pérez-Lachaud, 2012), which includes some of the world's most invasive ant genera: *Pheidole* Westwood, *Solenopsis* Westwood, *Wasmannia* Forel and *Monomorium* Mayr (GISD, 2015). The most common host based on the vast majority of host records (Table S1) is *Pheidole*, a genus with a worldwide distribution and a well-documented single dispersal event from the New World to the Old World (Moreau, 2008), estimated to have occurred in the Middle Miocene (11–13 Ma) (Ward *et al.*, 2015; Economo *et al.*, 2019) or Late Oligocene (22–27 Ma) (Economo *et al.*, 2015a,b). If the ancestral host of Oraseminae is *Pheidole*, it would be expected that they would have dispersed intercontinentally together or at least Oraseminae would have been absent from the Old World before the arrival of *Pheidole*. However, a prior phylogenetic analysis of Eucharitidae (with limited sampling within Oraseminae) using five gene regions suggested that the New World clade within Oraseminae was derived from the Old World taxa *c*. 20–23 Ma (Murray *et al.*, 2013). This intercontinental inverse dispersal between a parasitoid and its host seems counterintuitive and thus needs to be rigorously tested with increased taxon and gene sampling.

Oraseminae have a dynamic and unconventional life history whose evolution can be better understood within a phylogenetic framework (Fig. 1). These wasps use a specialized (expanded) ovipositor to hollow a cavity in plant tissue in which one or more eggs are deposited (Clausen, 1940b; Heraty, 2000). The first-instar larvae (planidia, < 0.15 mm in size) emerge on the plant and are responsible for gaining access to the ant brood. The larvae are hypermetamorphic (Pinto, 2009), meaning that the active sclerotized planidia do not closely resemble the later larval instars, which are sessile and hymenopteriform (Wheeler, 1907; Clausen, 1940a; Heraty, 2000). The exact means by which the planidia infiltrate the ant colony have never been directly observed; however, it is presumed that planidia gain access by targeting one of the ants' food sources. Proposed mechanisms include random attachment to foraging ants, phoretic attachment to an intermediate host (prey item) such as immature leafhoppers or thrips, or aggregating near or in extrafloral nectaries (EFNs) of plants and being picked up by feeding ant workers (Clausen, 1941; Das, 1963; Wilson & Cooley, 1972; Johnson *et al.*, 1986; Carey *et al.*, 2012; Herreid & Heraty, 2017). In the latter case (and perhaps all cases), the planidia are proposed to be transported by the adult ants within the infrabuccal pocket in their mouthparts and passed to the ant larvae via trophallaxis (Herreid & Heraty, 2017). With any of these nest infiltration methods, a tight tritrophic association is required, in which the plant has to be available and a suitable host for both the wasp and foraging ants.

Oraseminae species tend to specialize on certain plant structures for oviposition. The preferences can be partitioned into those that deposit eggs into leaves or green stems (or banana skins) and those that deposit into involucral bracts of unopened flower buds and near EFNs (Heraty, 1994a,b, 2000; Carey *et al.*, 2012; Herreid & Heraty, 2017). Some species have very specific plant hosts [e.g. *Orasema simulatrix* Gahan only oviposits near leaf EFNs of *Chilopsis linearis* (Cav.) Sweet (desert willow)], whereas for others, plant hosts can be very broad [e.g. *Orasema simplex* Heraty oviposits onto the leaves of at least eight plant families (Varone & Briano, 2009)]. Generalist species probably focus more on a particular plant structure than on a plant species. None of the Old World species has been associated with oviposition into the involucral bracts of flower buds, which is probably derived behaviour within the New World *Orasema* Cameron, instead choosing either leaves or green stems (Das, 1963; Kerrich, 1963; Heraty, 2000). Because of a general focus on plant structure over plant species, we expect that plant host is not likely to be a limiting factor for establishment and spread of many of the species.

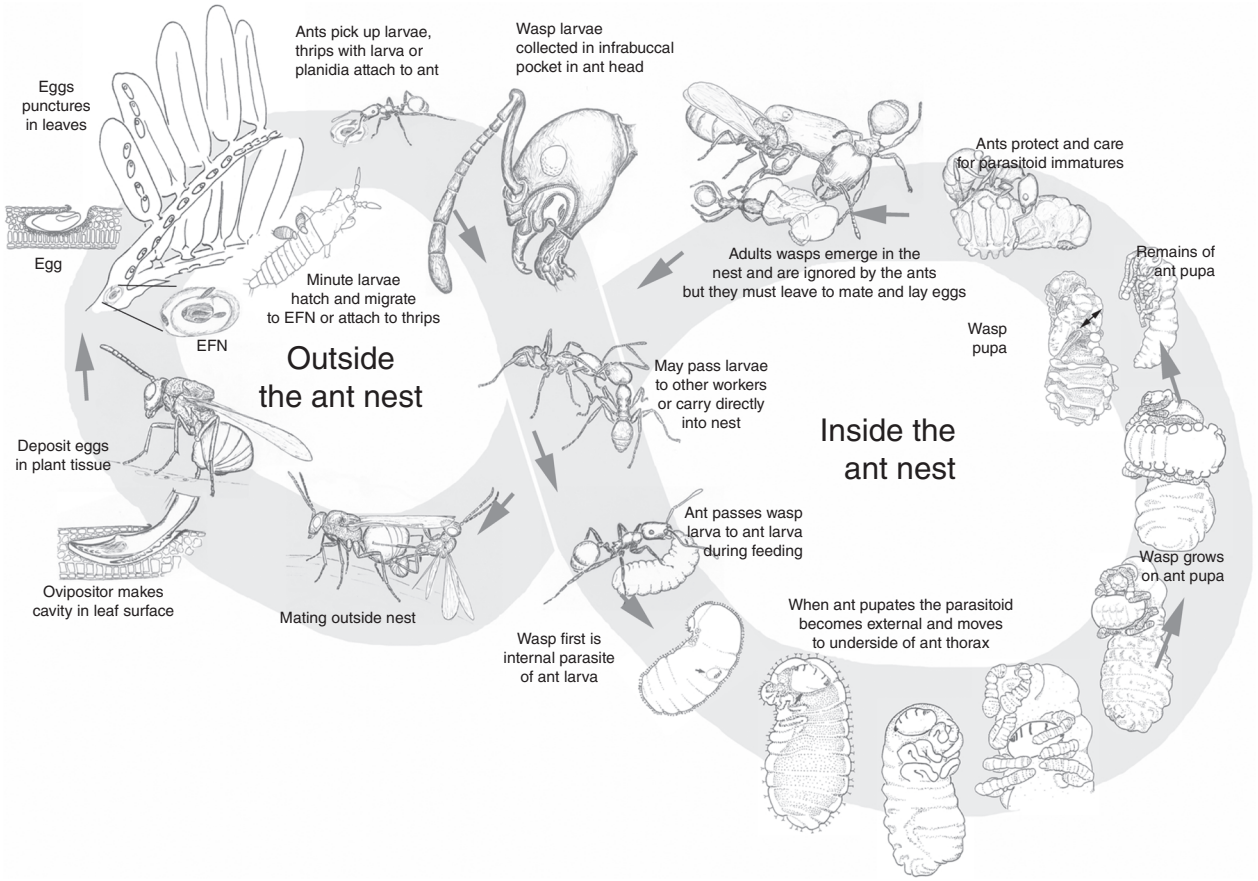


Fig. 1. The life cycle of a typical orasemine wasp. On the left are shown behaviours occurring outside the ant nest: mating, oviposition and planidial acquisition by the ants (EFN, extrafloral nectary). On the right are shown behaviours occurring within the ant nest: attachment to the ant larva, development on the ant pupa, pupation and eclosion.

Once the planidium is carried back to the nest and attaches to an ant larva, it will feed in a state of semi-arrested development until the host ant pupates (Wheeler, 1907; Clausen, 1940a; Clausen, 1941). In all known Oraseminae, the planidium burrows just under the larval cuticle, usually on the dorsal thoracic region, and the larva then feeds and expands (sometimes over 100 times its original size) without changing instars (Wheeler, 1907; Heraty, 2000). After the ant pupates, the wasp migrates to the ventral region of the thorax and develops through its second and third instars, which results in a desiccated living ant pupa that cannot continue its development (Wheeler, 1907; Clausen, 1940a; Heraty & Murray, 2013). All myrmicine ants have naked pupae (lacking cocoons), and thus parasitoid larvae are exposed throughout their development. *Formica* L. (Formicinae) (Johnson *et al.*, 1986) and *Eciton* Latreille (Dorylinae) (Heraty 1990, 1994b) have been proposed as potential Oraseminae hosts based on indirect associations in the field (they were not observed developing in nests). However, the complete pupal cocoons of these genera make them unlikely associations for an orasemine parasitoid. We are confident that neither *Eciton* nor *Formica* is a valid host association, leaving only Myrmicinae as a host for the entire subfamily.

Pheidole comprises the majority of host records, both numerically and in all biogeographic regions where Oraseminae is found (Table S1). Among the other credible host records, *Solenopsis* is a known host for species in five New World species groups of *Orasema* (Wheeler, 1907; Wheeler & Wheeler, 1937; Heraty *et al.*, 1993; Varone *et al.*, 2010). In a few cases, *Pheidole* is one of several hosts; for example, *Orasema minutissima* Howard attacks both *Wasmannia* and *Pheidole* in the Caribbean (Mann, 1918; Heraty, 1994a; Burks *et al.*, 2018), *Orasema minuta* Ashmead attacks both *Temnothorax* Mayr and *Pheidole* in Florida (Heraty, 1990), and an undescribed species of *Orasema* in Mexico (sp. 2 near *bakeri*) attacks both *Tetramorium* Mayr and *Pheidole* (Heraty, 1990). *Monomorium* is the host of *Zuparka monomoria* (Heraty) from Madagascar (Heraty, 2000), and it is the only known host in the Old World other than *Pheidole*. Thus, use of a host other than *Pheidole* is both rare and geographically differentiated.

Oraseminae is one of four subfamilies within Eucharitidae (Burks, 1979; Bouček, 1988; Heraty, 2002; Heraty *et al.*, 2004). It is the second most diverse subfamily (after Eucharitinae), including 13 genera and 89 described species (Heraty, 2017). Until recently, Oraseminae included four genera: *Orasema*

sensu lato (worldwide), *Indosema* (Afrotropical/Indian), *Timioderus* (Afrotropical) and *Orasemorpha* (Australasian) (Heraty 1994b, 2000, 2002). In all previous phylogenetic analyses of this subfamily using molecular data, *Orasema* sensu lato was polyphyletic (Heraty *et al.*, 2004; Heraty & Darling, 2009; Murray *et al.*, 2013). Based largely on preliminary analyses with the data presented herein, the Old World Oraseminae were revised and are now represented by 12 genera (Burks *et al.*, 2017). *Orasema* sensu stricto is now recognized as an exclusively monophyletic New World genus and is by far the most diverse in terms of morphology, life history and geographic distribution, ranging as far north as southern Canada (50.2°N) and as far south as central Argentina (40.5°S), and containing nine *Orasema* sensu stricto species groups (Heraty, 2000). Our analyses herein constitute the most comprehensive attempt at determining the relationships among the genera and species groups of Oraseminae, in terms of both taxon and gene sampling. Taken together, a taxon- and sequence-rich phylogeny of Oraseminae obtained from combining Sanger and AHE datasets enables examination of their complex life histories, host associations, biogeography and timing relative to their hosts, and serves as a framework for future classification.

Materials and methods

Georeferencing

We documented the worldwide distribution and abundance of Oraseminae using a total of 12 673 specimens of Oraseminae that were databased and georeferenced in FILEMAKER PRO 11.0v.3 using material borrowed from over 70 museum collections. Unless exact coordinates were provided on the label, coordinates were estimated using Google Earth (earth.google.com). Collection localities were plotted using Google Maps (maps.google.com).

Taxon sampling

Analyses were performed on the following three datasets:

1 The 'Sanger sequencing dataset' (Sanger) contains 203 taxa in the family Eucharitidae with 164 specimens of Oraseminae and 39 outgroup taxa from three subfamilies: Akapalinae (one specimen), Gollumiellinae (six) and Eucharitinae (32). All genera in Oraseminae except *Matantas* from New Caledonia are represented. All described species groups, as well as some newly proposed species groups and unplaced taxa within *Orasema* are represented (Heraty 1994b, 2000) (Table S2). The previously proposed *smithi* and *costaricensis* species groups have been redefined as the *stramineipes* species group (Burks *et al.*, 2018), and the *tolteca* group has been taken into the *cockerelli* group. We also recognize two new morphologically distinct species groups, the *sixaolae* (previously in *smithi* species group) and *susanae* species groups. Several species cannot as yet be placed confidently into defined species groups and may require recognition of additional species groups; however, these do not include any described species, which precludes their definition at this time.
2 The 'anchored hybrid enrichment dataset' (AHE) contains 92 eucharitid taxa, with 85 specimens of Oraseminae and seven outgroup taxa representing all three previously mentioned subfamilies. All of the AHE taxa were included in the Sanger dataset with the exception of *Akapala* (explained later). Eight of the 12 Old World genera (12 specimens) and nine out of 10 species groups (described and new) of *Orasema* (73 specimens) are represented. *Orasema* was intensively sampled for future species delimitation in particularly difficult species groups: the *coloradensis* species group (14 specimens), *cockerelli* species group (18 specimens), and *stramineipes* species group (19 specimens) (Table S2).
3 The 'Sanger + AHE dataset' (combined) contains 229 taxa. In this combined taxon set over half of the samples are represented by Sanger data only, thus allowing us to provide a combined phylogeny but potentially being complicated by missing data. The taxa are the same as those in the Sanger set with the following exceptions. *Akapala astriaticeps* (Girault) is a chimera of two specimens from the same collection event (D4286, AHE data; and D0360a, Sanger data), *Orasema yaax* Burks *et al.* is a chimera of two species from the same collection event (D4079, AHE; and D3733, Sanger), and *Orasema coloradensis* (D4891) used the Sanger data of another specimen from the same collection locality (D3756) for the BEAST analysis. The following taxa were included in the combined dataset but excluded from the Sanger dataset to reduce redundancy because they matched other specimens with more complete Sanger data: *Colocharis napoana* Heraty (D4288), *Orasema coloradensis* (D3148, D3227, D4213, D4219, D4221, D4520, D4653, D4891), *O.* near *coloradensis* (D3937), *O. costaricensis* Wheeler & Wheeler (D4628), *O. evansi* Burks *et al.* (D3799, D3765, D4648), *O. minutissima* (D4207, 3810, 2766, 2808), *O. rapo* (Walker) (D3808), *O. simulatrix* (transcriptome), *O. sixaolae* Wheeler & Wheeler (D2683, D4694), *O. stramineipes* Cameron (D4705), *Orasema* sp. (D4230) and *Zuparka fisheri* Heacox & Dominguez (D0638b).

Specimens were collected from malaise traps, yellow pan traps, ant nests, or sweep netting vegetation and stored in 95% ethanol at −80°C for tissue preservation. Voucher specimens were imaged, databased and point-mounted, and specimens were deposited with identification numbers in the museums listed in Table S2. The majority of extracted specimens are primary DNA vouchers (used for sequencing), but in the few cases where the primary voucher was destroyed, we created secondary DNA vouchers from specimens included in the same collecting event.

Gene sampling

Extraction. Most specimens were extracted using DNeasy Blood and Tissue Kit manufactured by Qiagen (Valencia, CA,

U.S.A.) with 1 μL RNAse A added after incubation, but some older extractions were performed with phenol-chloroform or Chelex (Bio-Rad, Richmond, CA, U.S.A.) following the methods of Heraty *et al.* (2004). PCR products were purified with DNA Clean & Concentrator-5 kits by Zymo Research (Irvine, CA, U.S.A.). The PCR product concentrations were determined using Nanodrop 2000c (Thermo Scientific, Waltham, MA, U.S.A.) or Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, U.S.A.).

Sanger sequencing. The Sanger sequencing dataset includes five gene regions: 18S ribosomal DNA, 28S D2 rDNA, 28S D3–5 rDNA, COI barcoding mitochondrial DNA and COI NJ mtDNA. Each gene was PCR amplified individually and Sanger-sequenced from both primers. Samples for Sanger sequencing were sent to Retrogen Inc. (San Diego, CA, U.S.A.) for sequencing on an Applied Biosystems 3730xl DNA Analyzer (Waltham, MA, U.S.A.). Chromatograms were inspected for base calling errors and edited in MESQUITE v.3.31 (Maddison & Maddison, 2017b) using CHROMASEQ v.1.2 (Maddison & Maddison, 2017a). In four cases where Sanger sequencing failed for specimens with AHE data, partial sequences of the genes were recovered from unmapped AHE reads using aTRAM (Allen *et al.*, 2015) and closely related reference sequences from the Sanger dataset. Sequences were uploaded to GenBank with the accession numbers listed in Table S2. The concatenated alignment is 3046 characters. The primers, alignment lengths and thermocycler protocols are listed in Table S3.

Anchored hybrid enrichment locus selection and probe design. The AHE probe set that we initially used was developed for Ichneumonoidea (Hymenoptera) (Sharanowski *et al.*, in preparation). We refer to this probe set as *Hym_Ich*. As a starting point, this probe set leveraged the 941-locus Coleoptera alignments developed by Haddad *et al.* (2017). These 941 loci were derived by identifying homologous coding regions across 13 insect species, including representatives of Hymenoptera, Diptera, Lepidoptera, Strepsiptera, Coleoptera, Anoplura and Hemiptera, which were aligned and used to target conserved, single-copy exons compared with Coleoptera reference genomes for a Coleoptera-specific probe set. Using the same approach, a Hymenoptera-specific enrichment kit was developed using references from Hymenoptera. Methods for AHE locus selection and probe design followed Hamilton *et al.* (2016). For the 941 loci, the assembled reference genome of the red flour beetle (*Tribolium castaneum*) was used to identify and extract the best matching region in each of the two Hymenoptera genomes, the bumblebee (*Bombus impatiens*) and a parasitoid wasp (*Nasonia vitripennis*). The extracted regions were then aligned using MAFFT (v.7.023b, with –genafpair and –maxiterate 1000 flags; Katoh & Standley, 2013). To increase representation of diverse lineages across Hymenoptera in the *Hym_Ich* probe set, 12 assembled hymenopteran genomes [*Apis mellifera* (The Honeybee Genome Sequencing Weinstock *et al.*, 2006); *Bombus impatiens* (Sadd *et al.*, 2015); *N. vitripennis* (Werren *et al.*, 2010); seven ant species (Gadau *et al.*, 2012); and two braconid genomes, including *Microplitis demolitor* (Burke *et al.*, submitted) and *Diachasma alleoleum* (NCBI GCA_001412515.1 unpublished i5K genome, Hugh Robertson)] and 11 Illumina sequenced unassembled ichneumonoid genomes covering both Braconidae and Ichneumonidae (Sharanowski *et al.*, in preparation) were scanned for anchor regions using the *N. vitripennis* sequences as a reference from the pairwise alignments generated earlier. For the 12 assembled genomes, 4000 bp surrounding the region that best matched the reference were isolated, and each sequence was used for probe design. For the unassembled genomes, reads were merged and trimmed following Rokyta *et al.* (2012) and then mapped to the reads of the *N. vitripennis* sequences. The consensus sequences from the resulting assemblies were then extended into the flanks, producing up to a 4000 bp sequence for each species at each locus. Alignments were generated for each locus from each of the 24 target species using MAFFT. After visual inspection in GENEIOUS (Kearse *et al.*, 2012), the alignments were trimmed and masked such that flanking regions not containing all of the reference sequences were trimmed out and sequences obviously misaligned in internal regions were masked to produce well-aligned regions containing no poorly aligned sequences. To ensure sufficient enrichment efficiency, we removed loci that contained < 50% of the taxa. Kmer profiles were analysed to identify and mask repetitive alignment regions in each locus, following Hamilton *et al.* (2016). This process resulted in 528 anchor loci and 13 loci targeted for their function (total target size is 212 392 bp). These loci are represented with 120 bp probes tiled at 2× density across each of the 24 sequences to produce 57 066 probes.

For Chalcidoidea, we aimed to improve the enrichment efficiency of the *Hym_Ich* target loci through comprehensive representation of the superfamily and related outgroups in Proctotrupomorpha using chalcid sequences from 47 assembled 1KITE transcriptomes and two published genomes (see Table S6 for details) (Peters *et al.*, 2018). We refer to this new probe set as *Hym_Cha*. The same general procedure was followed for probe design as with *Hym_Ich*. More specifically, following Hamilton *et al.* (2016), we scanned each genome and transcriptome for the target anchor regions using single exon *N. vitripennis* probe region sequences as references. For each locus, we then isolated up to 4000 bp containing the best matching region for each species, which in many cases now spanned multiple exon regions, and then aligned the resulting sequences across species using MAFFT. Following visual inspection in GENEIOUS, we trimmed the alignments down to well-aligned regions and masked poorly aligned regions. After removing overlapping regions of adjacent loci, we identified and masked repetitive regions in the alignments [see Hamilton *et al.* (2016) for methodological details]. The final alignments resulting from this procedure contained 421 012 sites distributed across 441 loci. We tiled 120 bp probes at 1.8× density across all sequences in each alignment. After removing redundant probes, the final probe set contained 171 070 probes.

Our AHE dataset includes specimens sequenced from the *Hym_Ich* (42 taxa) and *Hym_Cha* (49 taxa) probe sets (Table S2) and the *O. simulatrix* transcriptome. The *Hym_Cha*

probe set includes the same but fewer loci than *Hym_Ich*, but with extended probe regions for included loci. Using the additional reference taxa, we reassembled, aligned and masked the read data for our final taxon set using the same methods described earlier, and trimmed the alignments to exclude data not present in taxa sequenced from the *Hym_Ich* probe set.

Anchored hybrid enrichment data collection. Anchored hybrid enrichment data were collected at the Center for Anchored Phylogenomics (www.anchoredphylogeny.com). Extracted DNA was used for 8 bp single-indexed library preparation following Lemmon *et al.* (2012) and Prum *et al.* (2015). The indexes were chosen so that at least two differences existed between all combinations. During demultiplexing, reads with indexes that did not match exactly to one of the expected indexes were removed (no tolerance for mismatched indexes). Libraries were combined equally into *c*.16-sample pools prior to enrichment using the *Hym_Ich* or *Hym_Cha* probe sets contained within an Agilent XT SureSelect kit (Santa Clara, CA, U.S.A.). Enriched library pools were further pooled and sequenced on 2 PE150 Illumina 2500 (San Diego, CA, U.S.A.) lanes (*c*. 46 samples per lane, 96 Gb of total sequencing effort). After removing poor-quality reads using the Casava high-chastity filter, paired reads were demultiplexed (with no mismatches tolerated) then merged following Rokyta *et al.* (2012). Merged and unmerged reads were assembled with a quasi-de-novo assembler (Hamilton *et al.*, 2016) using *O. simulatrix*, *Eucharis adscendens*, *Sycophila biguttata* and *Eurytoma brunniventris* as references. To reduce the effects of possible low levels of misindexing and sample contamination, the resulting assembly clusters containing < 25 reads were removed from downstream analyses. Consensus sequences from clusters passing the filter were then compared to generate a pairwise sequence similarity matrix that was used to determine orthology (see Hamilton *et al.*, 2016 for details). After aligning orthologous sequence sets in MAFFT (v.7.023b, with genafpair and –maxiterate 1000 flags; Katoh & Standley, 2013), alignments were auto-trimmed as follows: 60% identity was required to designate a site as conserved, sequence regions containing < 10 of 20 common bases in conserved sites were masked, and sites containing < 38 unmasked, unambiguous characters were removed (details of the auto-trimmer given in Hamilton *et al.*, 2016). Finally, alignments were inspected in GENEIOUS v.9 (Kearse *et al.*, 2012), and all remaining misaligned regions were removed. The final AHE alignments contained 92 taxa, 348 loci, and 279 468 characters with 16.6% missing data. There is very little variation in AT/GC content among taxa in these datasets (Table S7).

Combining datasets. The combined dataset contains the AHE dataset concatenated with the Sanger dataset for a total of 353 loci and 282 514 characters with *c*. 66% missing data (*c*. 51% of the missing data is from Sanger specimens with no AHE data).

Deposition

Aligned matrices, partition files, probe sequences and program files are deposited in Dryad (doi:10.5061/dryad.df66th2). Sanger sequences and AHE read data are available in GenBank.

Phylogenetic analyses

Alignments for Sanger sequencing data were performed in MAFFT v.7 with the E-INS-i option selected (Katoh & Standley, 2013). Individual gene alignments were concatenated with SEQUENCEMATRIX (Vaidya *et al.*, 2011). All final alignments were spot checked in GENEIOUS v.10.2.3 (Kearse *et al.*, 2012).

The Sanger matrix was partitioned with the following scheme: 18S, 28S D2, 28S D3–5, COI BC/NJ (first and second codon positions), and COI BC/NJ (third codon positions). Two different partitioning schemes were tested for the AHE data matrix: (i) partitioning by gene region (348 partitions); and (ii) partitions resulting from PARTITIONFINDER 2 (Lanfear *et al.*, 2017) using the 'rcluster' search (Lanfear *et al.*, 2014) with the '–RAXML' option (Stamatakis, 2006) (151 partitions). Both schemes resulted in identical RAXML phylogenies with negligible differences in bootstrap support values (BS), and further analyses were partitioned only by gene region. The combined dataset was a combination of the five Sanger partitions plus the 348 AHE partitions, resulting in 353 partitions. The Sanger, AHE, and combined datasets were analysed with parsimony and ML to check for consistency and the influence of a model-based approach on inferences from the dataset. Parsimony analyses were performed with TNT v.1.1 (Goloboff *et al.*, 2008) and PAUP* v.4.0a (Swofford, 2002) using equal weights. TNT analyses were run using the New Technology Search and default settings for Sectorial Search, Ratchet, Drift, and Tree Fusing, and finding the minimum length 10 times. PAUP* analyses were performed using Stepwise Addition Search and 100 replicates, and in cases where TNT found a shorter tree, PAUP* was used to verify the TNT trees and to assess tree statistics. Maximum likelihood analyses were performed in RAXML-HPC2 v.8.2.10 on XSEDE (Stamatakis, 2014) via the CIPRES Science Gateway (Miller *et al.*, 2010) using default parameters (GTRCAT for rapid bootstraps and GAMMA for ML optimization) from two different starting seeds (12 345 and 98 765) to check that the trees converged on the same topology.

Dating analyses

All dating analyses were performed in BEAST v.2.4.4 (Drummond & Rambaut, 2007) on XSEDE via the CIPRES Science Gateway (Miller *et al.*, 2010) or through the University of California, Riverside computing cluster. For the Sanger dataset, we included all five gene regions with no a priori constraints on the topology. In the combined dataset, there is a large amount of missing data for taxa without AHE. Because of this, our BEAST analysis of the combined data used only the five Sanger genes

to calibrate dates on a fixed tree topology from the combined RAXML analysis, which was made ultrametric using nonparametric rate smoothing in MESQUITE.

Dates were calibrated for these analyses using a relaxed lognormal molecular clock using the Baltic amber (44.1 ± 1.1 Ma; Ritzkowski, 1997) fossil taxon *Palaeocharis rex* (Eucharitidae: Eucharitinae) (Heraty & Darling, 2009) to constrain the stem of *Psilocharis* following the methods of Murray *et al.* (2013). Although there is contention about the exact date of Baltic amber (37.7 ± 3 Ma; Kaplan *et al.*, 1977; Perkovsky *et al.*, 2007), we chose to use the dates from Ritzkowski (1997) to keep results comparable to Murray *et al.* (2013). The prior for this node was set to lognormal with the mean (in real space) at 8.08, SD at 1.0, and offset at 39.2, with dates estimated from two independent runs with 500 million generations each, sampling every 50 000 for the Sanger analysis and one run with 50 million generations, sampling every 50 000 for the combined analysis. TRACER v.1.6.0 (Rambaut *et al.*, 2014) was used to verify that the chains reached an appropriate effective sample size (ESS) and converged. LOGCOMBINER v.2.4.4 and TREEANNOTATOR v.2.4.4 were used to remove 10% of the trees as burn-in and compute the maximum clade credibility tree.

The AHE dataset was too large to be effectively sampled in BEAST, so subsets of 25 loci were selected for branch length calibration on a fixed tree topology (AHE RAXML tree) that was made ultrametric with nonparametric rate smoothing in MESQUITE. Three independent subsets were selected based on their phylogenetic informativeness profiles resulting from PHYDESIGN (Lopez-Giraldez & Townsend, 2011) with a JC model: the first subset contained the 25 highest ranked loci (first to 25th) for phylogenetic informativeness over the time period encompassing the diversification of the ingroup (from the crown to terminal bifurcations of Oraseminae), the second subset contained the next 25 most informative loci for that time period (26th–50th), and the third subset followed the same pattern (51st–75th) (Table S4). Alternative evolutionary models were tested in PHYDESIGN, but had no effect on the ranking of loci. Only one specimen of *Psilocharis* was sampled for AHE; as the *Palaeocharis* fossil calibration could no longer be used at the stem of *Psilocharis* (lack of a *Psilocharis* node), we chose instead to use a secondary calibration date for the crown of Eucharitinae from Murray *et al.* (2013). The crown node age of Eucharitinae was set to a normal distribution with the mean at 52.0, sigma at 7.8, and offset at 0.95, which gives a 95% probability range of 40.1–65.8 Ma. A uniform root age prior of 45–100 Ma was used to constrain the maximum age, and 100 Ma was conservatively chosen for the maximum based on oldest fossil Chalcidoidea known: Mymaridae in Cretaceous amber from upper Albian deposits (Poinar Jr. & Huber, 2011). The three data subsets were each analysed from a single run using the fixed tree AHE topology for 150 million generations sampled every 20 000 to test for congruence amongst dates. Maximum clade credibility trees were computed using the same methods as the other datasets.

Host associations

Ancestral host mapping was performed on the RAXML tree from the combined dataset using unordered parsimony reconstruction in MESQUITE. This tree was chosen because it is the most taxon-inclusive dataset and is also in accordance with the topology of the AHE results. Outgroups were coded for host ant subfamily (Ectatomminae, Formicinae, Myrmeciinae, and Ponerinae), whereas all Oraseminae only parasitize species in the subfamily Myrmicinae. Because the host ant subfamilies do not overlap between our outgroups and ingroup, we coded Oraseminae for host ant genus (*Monomorium*, *Pheidole*, *Solenopsis*, *Temnothorax* and *Wasmannia*). Taxa with unknown hosts were left ambiguous if no closely related taxa (same genus or species group of *Orasema*) had known hosts, but if closely related taxa had a host record, then it was carried over to other species in the same clade. See Table S1 for complete host information.

Biogeography

Biogeographic analyses were run on the combined BEAST chronogram as it was the most taxon-inclusive. Dispersal–extinction–cladogenesis (DEC) with and without jump dispersal (DEC + J) models (Matzke, 2014) were tested using BIOGEOBEARS (Matzke, 2013). Biogeographic regions included Nearctic, Neotropical, Australasian, Oriental, Afrotropical, Madagascan, Indian and Palearctic. Dispersal probability multipliers between regions were established for each historical epoch (Walker *et al.*, 2012) in the dated phylogeny (Table S5). These values were chosen subjectively using the methods of Condamine *et al.* (2013) (based on the principles of Ree & Smith, 2008) and based on palaeogeographic reconstructions of the continents through time (Blakey, 2008). The dispersal probability for travelling from any biogeographic region to itself was set to 1; regions that were connected without a significant barrier were given a rate of 0.5; regions that were separated by a narrow barrier, such as an ocean strait splitting two relatively close landmasses, were given a rate of 0.25; any combination of these scenarios had a probability that was the product of multiplying the other probabilities (e.g. region 1 is connected to region 2, which is narrowly separated from region 3; the probability of going from 1 to 3 is $0.5 \times 0.25 = 0.125$); and any dispersal separated by more than two unconnected regions was given a probability multiplier of 0.01.

Diversification analysis

Diversification rates were assessed in the program BAMM (Bayesian Analysis of Macroevolutionary Mixtures) (Rabosky *et al.*, 2014) using the BEAST tree for the combined dataset to maximize taxon inclusion. Taxon sampling percentages were calculated for Old World genera and species groups in *Orasema* (or estimated for groups currently under revision) and used in the analysis to account for missing taxa. *Matantas*, the only

unsampled genus, was accounted for in the global sampling fraction. The Markov chain Monte Carlo analysis was run for 1 million generations, reaching stationarity, and sampled every 1000. Ten percent of the chain was discarded as burn-in. Statistics for the analysis were examined in the R package BAMMTOOLS (Rabosky *et al.*, 2014).

Coalescence

To assess the impact of coalescence on the tree topology, we ran a coalescent analysis in ASTRAL-II v.4.10.12 (Mirarab & Warnow, 2015). We generated 348 gene trees in RAXML for the AHE dataset with 100 rapid bootstraps per tree. Gene trees were visually inspected for anomalous relationships as a measure of quality control. Distantly related taxa appearing identical (contamination) and ingroup taxa on excessively long branches (sequencing error) were removed. Branches in gene trees with bootstrap scores ≤ 10 were contracted into polytomies with the software NEWICK UTILITIES (Junier & Zdobnov, 2010).

Likelihood mapping

To rigorously assess the support of Old World Oraseminae paraphyly, support for the sister relationship between *Leiosema* and the remaining orasemine taxa in the AHE dataset (RAXML AHE tree) was further tested using quartet likelihood mapping in the software TREE-PUZZLE v.5.3 (Schmidt *et al.*, 2002). Taxa were binned into four groups: (i) outgroups (seven taxa); (ii) *Leiosema* (two taxa); (iii) *Timioderus* (two taxa); and (iv) remaining Oraseminae (81 taxa). These four groups represent the four connection points to the internal branch separating *Leiosema* from the remaining Oraseminae on an unrooted tree. Each sampled quartet selects one random taxon from each of the four groups. Support is established by comparing how many times each of the three possible topologies (four taxon unrooted trees) is sampled. We sampled 10 000 random quartets.

Results

Georeferencing

The map of Oraseminae collection localities (Fig. 2) is a comprehensive summary of the distributions of each genus, as well as an indication of the relative abundances of those genera. *Orasema* are noticeably more common and diverse in the New World, when compared with Oraseminae in the Old World. Although this is probably impacted by collecting bias, extensive collecting by JMH in both hemispheres supports the conclusion that the Old World genera of Oraseminae (700 specimens) are far less plentiful than *Orasema* in the New World (11 845 specimens).

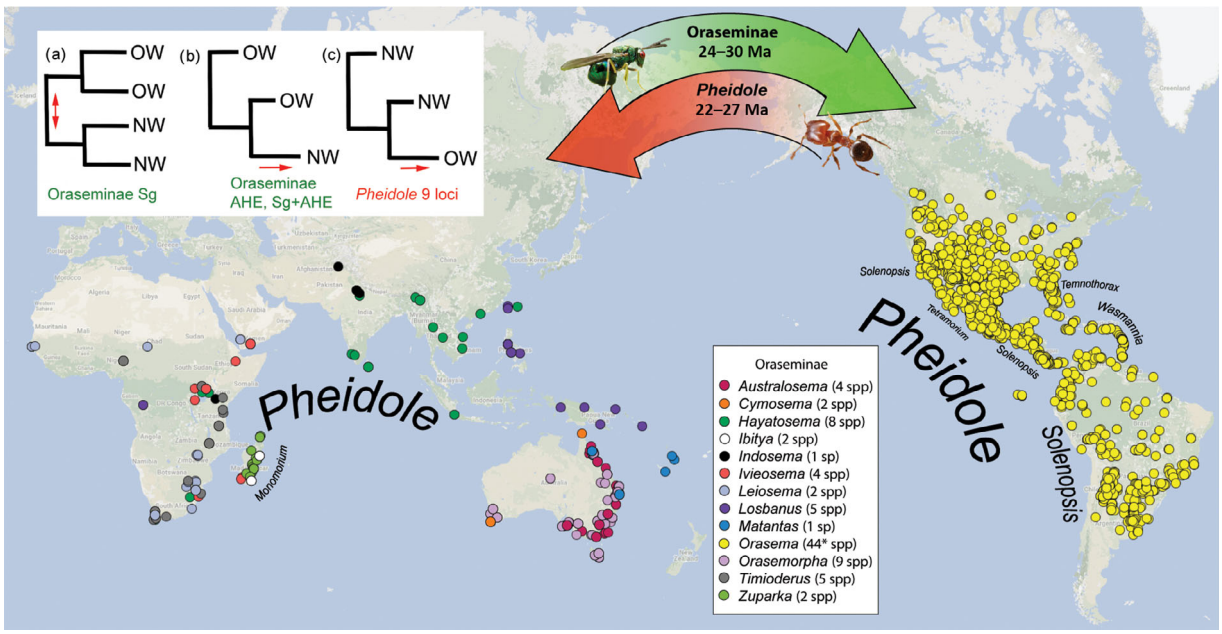


Fig. 2. The distribution of genera in Oraseminae. Host ant names shown on the map have their size relative to the abundance of records on that genus. The direction of dispersal with date ranges from the combined BEAST analysis (crown to stem) for Oraseminae and analysis of *Pheidole* (crown to stem) (Economo *et al.*, 2015b) are shown at the top. The number of described species per genus is shown in the middle (several groups within *Orasema* are currently under revision). Upper left: (a) an equivocal biogeographic hypothesis lacks directionality, seen in analyses of the Sanger dataset in Oraseminae; (b) an uneqivocal Old World to New World hypothesis, seen in analyses of the anchored hybrid enrichment (AHE) and combined datasets in Oraseminae; (c) an unequivocal New World to Old World hypothesis, seen in molecular analyses of *Pheidole* (Economo *et al.*, 2015a,b). [Colour figure can be viewed at wileyonlinelibrary.com].

Sanger sequencing

For the 203 specimens included in the Sanger dataset, 18S was successfully sampled for 173 specimens (complete or partial sequences), 28S D2 for 203 specimens, 28S D3–5 for 193 specimens, COI BC for 74 specimens, and COI NJ for 144 specimens, with a total of 28.1% missing data (specimen data and GenBank numbers in Table S2; alignment lengths in Table S3). COI sequences were manually checked for stop codons to identify and remove nuclear mitochondrial DNA (NUMTs), which were only found in COI BC of a single species, *O. minutissima*.

Anchored hybrid enrichment

The AHE dataset contains 348 gene regions with a total alignment length of 279 468 nucleotides for 92 taxa. An average of 83 taxa were successfully captured per locus (SD = 8.1), with 16.5% missing data overall, and the average locus length was 803 bp (SD = 325.8).

Combining data matrices

Having a large proportion of missing data is a necessary consequence of concatenating a taxon-rich Sanger dataset with a character-rich AHE dataset. Taxa that only have Sanger sequence data (missing AHE) comprise 52% (119/229) of the sampled taxa, which accounts for at most *c*. 1% of the total data (3046/282 514 characters). Thus, 51% of the data in the combined matrix is missing because of concatenating the Sanger and AHE matrices, resulting in a total of 66.1% missing data.

Phylogenetic analyses

Sanger dataset. Parsimony (Fig. S1) and ML (Fig. S4) analyses of the Sanger sequencing dataset produced slightly different tree topologies; both support the monophyly of Oraseminae and the reciprocal monophyly of the Old World Oraseminae and New World *Orasema*. A strict consensus of the 20 most parsimonious trees [4161 steps, retention index (RI) = 0.73] was poorly resolved, although Eucharitinae and Eucharitini (excluding *Neolosbanus* and *Psilocharis*) were monophyletic, and Oraseminae (bootstrap score = 94), Old World Oraseminae (BS < 50), and *Orasema* sensu stricto (BS = 65) were each monophyletic. The best-scoring ML analysis (Fig. S4) had outgroup relationships that were in accord with Murray *et al.* (2013), including monophyly of Gollumiellinae and a sister-group relationship between *Psilocharis* (the fossil reference taxon) and the remaining Eucharitinae. Oraseminae (BS < 50), Old World Oraseminae (BS = 72) and *Orasema* sensu stricto (BS = 95) were each monophyletic. The ML tree supports the monophyly of all of the Old World genera, with the exception of the genus *Hayatosema* from Africa and Southeast Asia, which was polyphyletic. The species groups of *Orasema* in the ML analysis are mostly monophyletic, with the exception of the *cockerelli* and *xanthopus* species groups.

Anchored hybrid enrichment dataset. Parsimony analysis of the AHE dataset (Fig. S2) produced a single tree with 412 925 steps (RI = 0.77), which was nearly identical to the ML tree (Fig. S5). All ML analyses for this dataset produced identical trees regardless of starting seed or partitioning scheme. The two differences between the parsimony and ML trees include a paraphyletic Gollumiellinae (outgroup) in parsimony versus being monophyletic in ML and slightly different positions of *Cymosema waterworthae* Burks & Mottern and *Ivieosema confluens* Burks within the Old World Oraseminae; neither hypothesis is highly supported by bootstrap scores in either analysis. Otherwise, phylogenies resulting from both analyses are highly supported at both deep and shallow nodes. Oraseminae (BS 100 MP/100 ML) and *Orasema s.s.* (BS 100/100) were each monophyletic, but in contrast to the Sanger results, the Old World Oraseminae were always paraphyletic with *Leiosema* monophyletic and sister to the remaining Oraseminae (BS 100/100). All of the species groups of *Orasema s.s.* were monophyletic and strongly supported (BS 100/100) in both analyses. Relationships between species groups and the *incertae sedis* species were identical in both analyses with high support (BS 100/100) for all but the single *bakeri* group species. The coalescent analysis in ASTRAL produced a species tree that was highly congruent with the tree topologies of concatenated analyses with the position of *Timioderus* and *Ibitya* shifted slightly within the Old World group (Fig. S13).

To further assess support for the *Leiosema* sister group relationship, which is important to our biogeographic hypotheses, all gene trees were scored based on the position of *Leiosema*: 42.5% unequivocally support *Leiosema* as sister to the remaining Oraseminae, 36.5% do not support this relationship, and 21% are ambiguous (often because of missing outgroups). Additionally, likelihood mapping resulted in 64.1% of sampled quartets supporting the relationship: (outgroup, *Leiosema*), (*Timioderus*, other Oraseminae); 35.5% of quartets support: (outgroup, other Oraseminae), (*Leiosema*, *Timioderus*); and support for other hypothesis is negligible (Fig. 5).

Combined dataset. Parsimony and ML analyses of the Combined dataset produced trees with similar topologies but with much lower support on the parsimony tree (Figs S3, S6, respectively). In accordance with other analyses, Oraseminae [BS < 50, maximum parsimony (MP); 99, ML] and *Orasema* sensu stricto (BS = 95/100) were each monophyletic, and in accordance with the AHE results, the Old World Oraseminae were paraphyletic, with *Leiosema* monophyletic (BS 55/100) and sister to the remaining Oraseminae (BS < 50/86). Within *Orasema* sensu stricto, species groups were monophyletic, with only the *xanthopus* species group not monophyletic in both analyses (Figs S3, S6), and the *cockerelli* species group not monophyletic in the parsimony analysis (Fig. S3). The relationships between species groups in the combined ML tree were the same as in the AHE trees, whereas the parsimony tree had low support and slightly different relationships. The *susanae* species group was not sequenced for AHE and was placed between the *festiva* and *stramineipes* groups in both results.

Dating analyses

The BEAST analysis of the Sanger dataset (Fig. S7) was the only dating analysis not constrained by a fixed tree topology because the size of the dataset was within the limits of the computational power of BEAST. Even so, this dataset was run on a computing cluster for increasingly long generation times (up to 500 million), and with two runs combined only reached a moderate ESS (ESS posterior = 135). Because this is the only BEAST analysis generated without a fixed tree topology, it is the only tree where we report posterior probability values. The topology of the BEAST tree is nearly identical to the ML tree with all of the genera and species groups recovered and only a few changes in species group relationships within *Orasema* (*bakeri*, *simulatrix* and *wayqecha* groups, and some *incertae sedis*). Oraseminae and *Orasema* sensu stricto are fully supported (posterior = 100). The mean crown age of Oraseminae is 34 Ma [95% highest posterior density (HPD) = 23–45] and the mean crown age of *Orasema* is 25 Ma (95% HPD = 17–34).

Three BEAST analyses run on different subsets of the AHE dataset were performed on the fixed AHE ML tree topology (Fig. S5). The 'Top25' (Fig. 3) and '26–50' (Fig. S8) subsets reached reasonable posterior ESS values at 150 million generations (posterior ESS = 284 and 158, respectively); however, the analysis for '51–75' subset (Fig. S9) ended prematurely around 6.1 million generations but still achieved high ESS values (posterior ESS = 319). Because of the fixed tree topology, other ESS values across the three analyses were very high, typically ranging from 1000 to 6000, with the exception of the uncorrelated log-normal relaxed molecular clock mean and SD, which were < 200. The variation in dates among the analyses is reasonably small [e.g. the mean crown age for Oraseminae is 43 (95% HPD: 26–58), 48 (31–67) and 45 (28–60) Ma, and the mean crown age of *Orasema* is 28 (17–39), 33 (20–46) and 29 (18–39) Ma for 'Top25', '26–50' and '51–75', respectively] (Figs 3, S8, S9).

Our dated analysis of the combined dataset (Fig. S10) used the fixed combined ML tree topology (Fig. S6) and only Sanger data, which are available for all taxa (30% missing data), to calibrate branch lengths. The mean crown age of Oraseminae is 35 (26–47) Ma, and the mean crown age of *Orasema* is 24 (17–33) Ma.

Taken together, these dates place the mean crown age of *Orasema* between 20 and 33 Ma, which spans the early Miocene to the early Oligocene, and the mean crown age of Oraseminae between 30 and 48 Ma, which spans the early Oligocene to the early Eocene (Fig. 3).

Biogeography

Using the dated combined tree topology and estimated dispersal rates (Table S5), both DEC (Fig. S11) and DEC + J (Figs 4, S12) analyses support an African origin for Oraseminae and a North American origin for *Orasema* in the New World. DEC + J shows higher support for these same conclusions and is a better-fitting model with a lower AIC score. This is expected because the j parameter allows for founder-event speciation, in which dispersal can occur at cladogenesis (chance dispersal) rather than being restricted to anagenesis (range expansion) (Matzke, 2014). This may have happened with *Orasema*, where there was a single dispersal event into the New World followed by a relatively rapid isolation event with the loss of favourable conditions for dispersal back across the BLB.

In the Old World, both DEC and DEC + J analyses support transitions from Africa through Oriental/Australasian regions leading up to a single New World dispersal event. Although the support values along the Old World backbone are not strong across all phylogenetic analyses (Figs S1–S7), there is strong support in all analyses for a single New World clade. Two independent dispersals to Australia are proposed, one in *Cymosema* and one in the *Orasemorpha*/*Australosema* clade. There are also two independent dispersals to Madagascar, once in the *Ibitya*/*Zuparka* clade and once in *Ivieosema*. All of these proposed dispersals occurred more recently than 30 Ma, long after Madagascar and Australia were separated from Africa (Sanmartín & Ronquist, 2004).

Ancestral host mapping

Parsimony reconstruction on the dated combined ML tree topology (Fig. 5) inferred *Pheidole* to be the ancestral host along the backbone of the tree. As many as four separate shifts onto *Solenopsis* are inferred in the New World. Shifts onto other host genera, including *Monomorium*, *Wasmannia* and *Temnothorax*, each happened once in *Z. monomoria*, *O. minutissima* and *O. minuta*, respectively. The other host shift to *Tetramorium* is not shown because the taxon is missing from the phylogeny (*bakeri* group: *O*. sp. 2 nr. *bakeri* from Mexico) (Heraty, 1990).

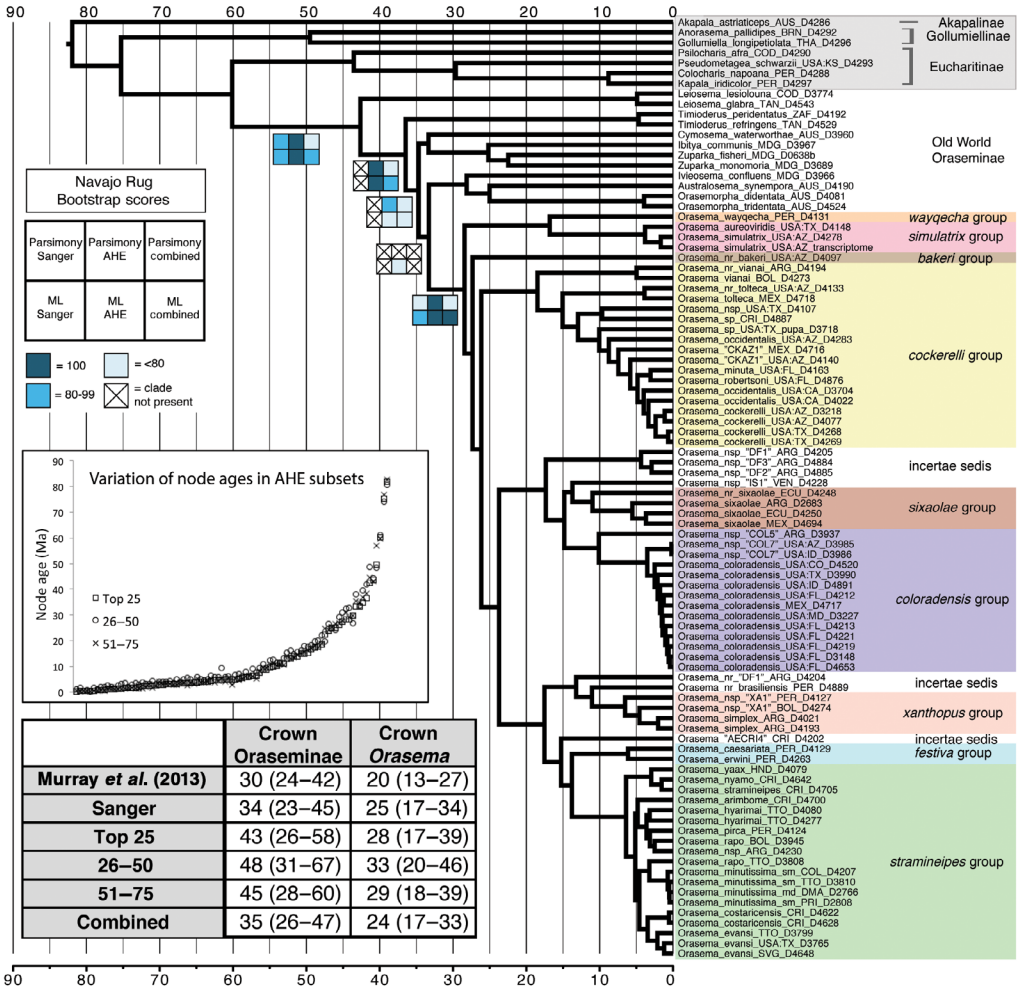
Diversification analysis

The diversification rate analysis from BAMM (Fig. 5) supported a single rate shift between the Old World and New World taxa with the highest probability (posterior probability = 0.3, the highest out of 42 credible shift sets). In other words, an increase in diversification rate in Oraseminae was most likely correlated to dispersing to the New World. The posterior probabilities of alternative credible shift sets drop off precipitously (e.g. second most supported = 0.12, third = 0.11, fourth = 0.034).

Discussion

Dating and biogeography

The biogeographic history of Oraseminae and their primary host ants, *Pheidole*, is unique, to the best of our knowledge, because of their intercontinental dispersals in opposite directions around the same time period (Fig. 2). Sanger data support monophyly of both Old World and New World lineages, and thus no direction of dispersal can be inferred (Figs S1, S4,



	Crown Oraseminae	Crown *Orasema*
Murray *et al.* (2013)	30 (24–42)	20 (13–27)
Sanger	34 (23–45)	25 (17–34)
Top 25	43 (26–58)	28 (17–39)
26–50	48 (31–67)	33 (20–46)
51–75	45 (28–60)	29 (18–39)
Combined	35 (26–47)	24 (17–33)

Fig. 3. BEAST chronogram of top 25 anchored hybrid enrichment (AHE) loci using a fixed topology (AHE maximum likelihood tree). Species groups of *Orasema* are shown in colour . Bootstrap support values are summarized as a Navajo rug pattern, with darker squares showing higher support (values shown on Figs S1–S6). The variation in age estimates at each node for the AHE data subsets is shown in a scatterplot; the same node in each analysis is plotted on the same *x*-axis. Comparison of all dating analyses, including Murray *et al.* (2013), for the crown age of Oraseminae (minimum estimated date for specialization on Myrmicinae) and the crown age of *Orasema* (minimum estimated date for dispersing into the New World). Mean dates shown first with 95% highest posterior density range in brackets. [Colour figure can be viewed at wileyonlinelibrary.com].

S7). All of the analyses that incorporate AHE sequence data (AHE and combined datasets), however, unequivocally support an Old World to New World dispersal event indicated by a grade of Old World taxa leading to a single New World clade (Figs S2, S3, S5, S6). Given the magnitude of difference in the quantity of sampled loci and characters between Sanger (five loci: 3046 characters) and AHE (348: 279 468) datasets, the stability of the AHE dataset (similar results in parsimony and ML; concatenation and coalescence), and additional support measures (the majority of gene trees and sampled quartets favouring paraphyly in the AHE dataset), we conclude Old World paraphyly to be more likely. The opposite hypothesis (New World to Old World) is never supported in any analysis of Oraseminae, including analyses of independent loci. The dated analyses of Oraseminae (Figs 3, S7 – S10) place the mean crown age of *Orasema* between 24 and 33 Ma (Oligocene), which can be used as an estimate of the minimum age of dispersal to the New World. This is slightly older than the estimate of Murray *et al.* (2013), which is around 20 Ma. This range of dates is too recent for Gondwanan vicariance or dispersal across NALBs, leaving dispersal across a BLB or chance oceanic dispersal as more likely hypotheses. By examining their host ant biogeography, we can further refine our biogeographic hypotheses.

Pheidole is the most parsimonious ancestral host for Oraseminae (Fig. S13); however, the host of *Leiosema*, which is sister to the rest of Oraseminae in all analyses incorporating AHE data, is unknown. The *Pheidole* dispersal from the New World to the Old World is highly supported by all molecular phylogenetic analyses of this genus (Moreau, 2008; Ward *et al.*, 2015; Economo *et al.*, 2015a; Economo *et al.*, 2015b; Economo *et al.*, 2019), but the age of the Old World crown group varies

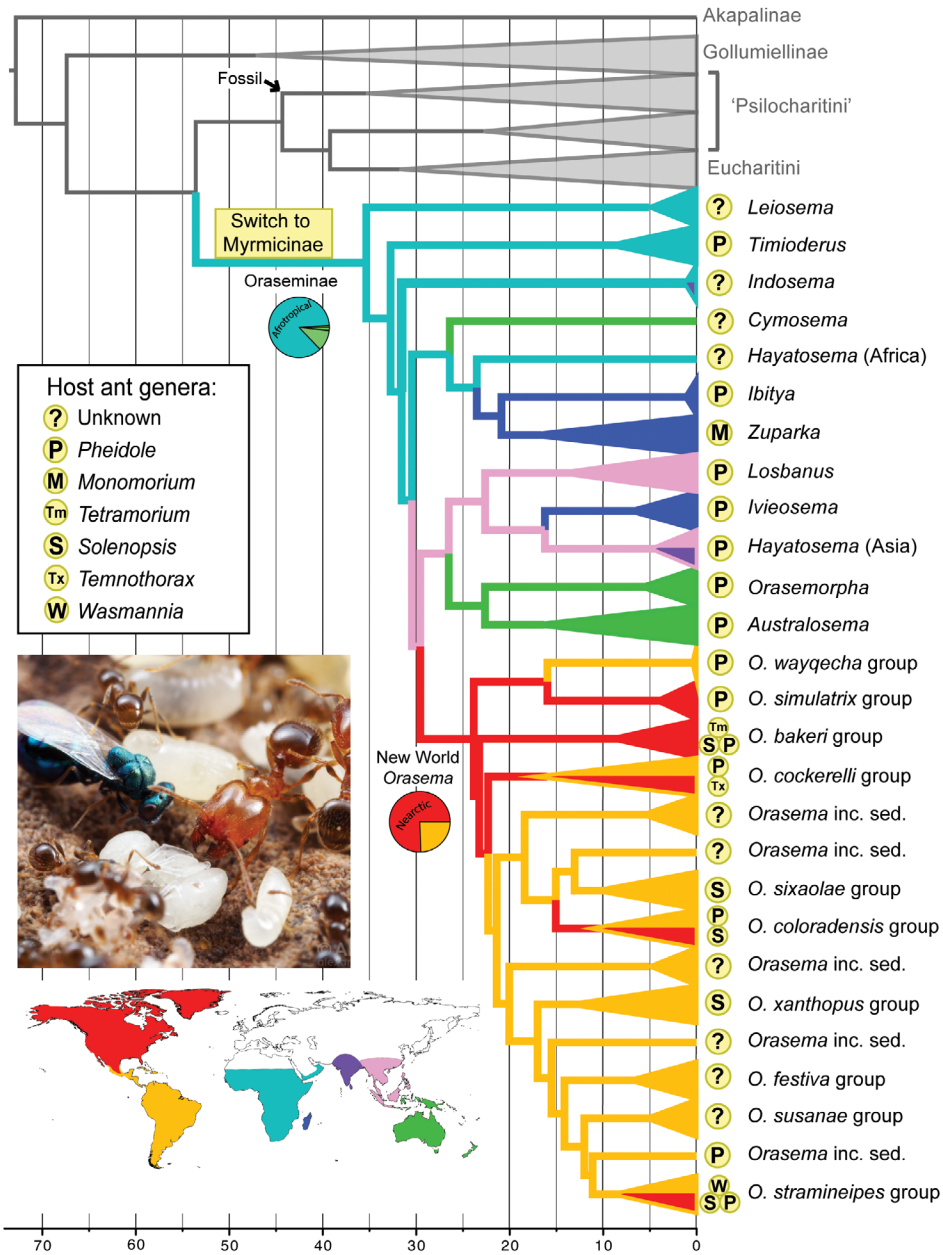


Fig. 4. BEAST analysis of the combined dataset with species collapsed into higher taxa (outgroup Eucharitidae to subfamily or tribe, Old World Oraseminae to genus, and *Orasema* to species group). Clades and branches are coloured by biogeographic area relating to the map. Biogeographic reconstruction is summarized from dispersal–extinction–cladogenesis with jump dispersal (DEC + J) analysis with the two pie charts summarizing probabilities for the stem of Oraseminae and *Orasema*. Host ant records for Oraseminae are mapped to the tips of the tree. (Photograph by Alexander Wild.) [Colour figure can be viewed at wileyonlinelibrary.com].

between analyses. The phylogeny of Ward *et al.* (2015) places the Old World crown age at *c*. 12 Ma (Middle Miocene). This analysis used 27 fossil calibration points (20 within Myrmicinae, seven outgroups), but their sampling of *Pheidole* was sparse (four New World taxa, three Old World taxa). By comparison, the phylogeny of Economo *et al.* (2015b) placed the Old World crown age at *c*. 22 Ma (Early Miocene) using an extensive sampling of *Pheidole* (two New World, 177 Old World) selected from an earlier undated analysis of 285 species (107 New World, 173 Old World, five widespread) (Economo *et al.*, 2015a), but this analysis lacks any calibration points except for a minimum age constraint on the root node (crown *Pheidole*) at 59.8 Ma (Economo *et al.*, 2015a) based on the phylogeny of Ward *et al.* (2015). The lack of calibration points in their analysis was due to the difficulty of assigning presumed *Pheidole* fossils to clades within the genus (Economo *et al.*, 2015a).

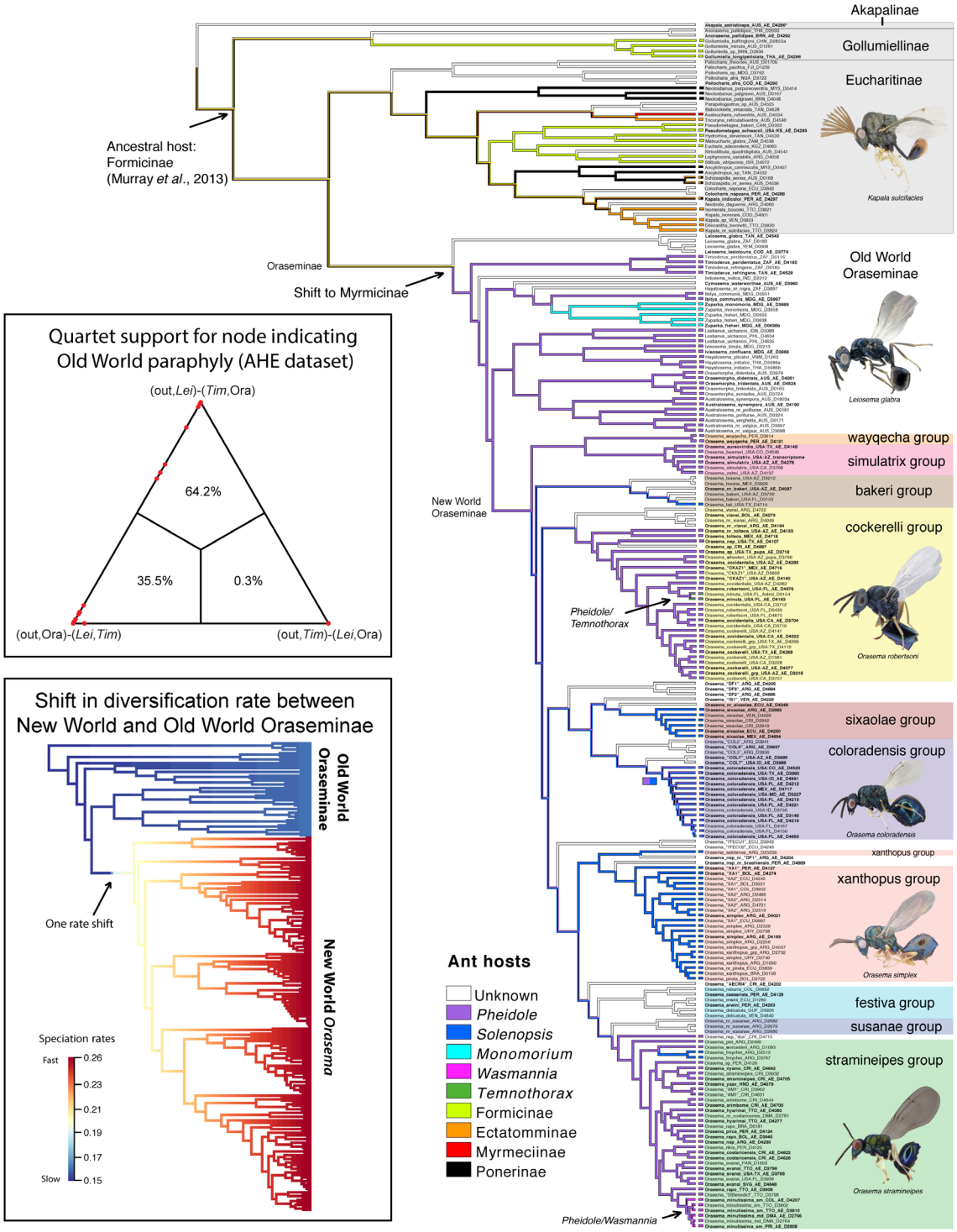


Fig. 5. Additional phylogenetic comparative methods. Main figure: parsimony reconstruction of ancestral ant hosts on the combined dated tree (Fig. S10) using MESQUITE. Top inset: quartet support for the *Leiosema* sister relationship to the remaining Oraseminae using the anchored hybrid enrichment dataset from TREE-PUZZLE. Bottom inset: diversification analysis of Oraseminae (outgroups pruned off) using the combined dated tree from BAMM showing the single most highly supported rate shift configuration from the Bayesian credible set (42 configurations) with a posterior probability of 0.3. [Colour figure can be viewed at wileyonlinelibrary.com].

The most recent analysis (Economo *et al.*, 2019) places the Old World crown age at 11–13 Ma based on 449 ingroup taxa and eight nuclear loci using the same calibration as Economo *et al.* (2015b). A *Pheidole* specimen assumed to be in Baltic amber (44.1 Ma; Dubovikoff, 2011) would have complicated the biogeographic scenario by preceding the earliest possible date for Old World *Pheidole*; however, this specimen is now considered to be from copal ('immature amber' from a much more recent time; Perkovsky, 2016). Regardless of the difference between dates, both phylogenies suggest that the dispersal and distribution of extant *Pheidole* are too recent to be explained by Gondwanan vicariance or dispersal across the NALBs.

The monophyly of *Pheidole* in the Old World and Oraseminae in the New World suggests the dispersal and establishment of a single ancestor, which could have been accomplished over a land-based filter bridge (sensu Simpson, 1940) or by chance dispersal through over-water rafting (sensu de Queiroz, 2005). We argue against chance over-water dispersal for Oraseminae because it would require a parasitized nest of *Pheidole* to have dispersed back into the New World after establishment in the Old World, and there is no phylogenetic evidence to suggest that either group dispersed back into their native range. The probability of Oraseminae surviving outside a host nest on a transoceanic voyage combined with the probability of successfully developing on a novel host upon arriving in the New World would make dispersing without their host highly unlikely. Instead, the dispersal of these wasps would probably require a land bridge tightly correlated with the presence and establishment of *Pheidole*, which would allow each group to cross in opposite unidirectional patterns. Thus, dispersing across the BLB during a period of high temperature and equability (> 15 Ma; Wolfe, 1993; Milne, 2006) is the best-supported and most logical biogeographic hypothesis for both Oraseminae and *Pheidole*, which are predominantly tropical groups. The northern climate would have experienced some cooling between 50 and 35 Ma, followed by a period of fluctuation in the Oligocene to Early Miocene with decreased cold-month temperatures and increased aridity and then a period of progressive cooling after 15 Ma (Milne, 2006; Eldrett *et al.*, 2009). The climatic conditions of the Oligocene may have facilitated a filter across Beringia, such that a limited number of species of *Pheidole* and Oraseminae were able to occupy and cross the land bridge but with changing climatic conditions limiting any further backcrossing as seen in more temperate groups [e.g. ants (Branstetter, 2009; Jansen *et al.*, 2010), bumble bees (Hines, 2008) and blue butterflies (Vila *et al.*, 2011)]. An open, more climatically favourable corridor would have allowed many crossings in both directions, creating a mixture of Old World and New World clades, which is not seen in either group. The mean crown age for Oraseminae is between 34 and 48 Ma (Figs 3, S7 – S10), which represents a minimum age for shifting to Myrmicinae as hosts (assuming *Leiosema* parasitizes *Pheidole*) and is significantly earlier than the proposed invasion and spread of *Pheidole* in the Old World (Ward *et al.*, 2015; Economo *et al.*, 2015a; Economo *et al.*, 2015b; Economo *et al.*, 2019). This could be explained by at least two hypotheses: (i) Oraseminae was present in the Old World on some unknown ancestral host and underwent multiple host shifts to *Pheidole* after its arrival; or (ii) the invasion of New World *Pheidole* predates the diversification of Oraseminae in the Old World despite the results from the dating analyses. Either way, the Old World Oraseminae underwent a host shift to *Pheidole* of New World origin that allowed for a favourable crossing into the New World.

Dispersal into the New World by Oraseminae is accompanied by a tremendous increase in diversification and abundance within *Orasema*. This diversification is complemented by a wider array of oviposition strategies, morphologies and host preferences than are found in the Old World genera. This pattern may indicate that shifting onto *Pheidole* or dispersing into a new area with a diversity of potential *Pheidole* host species increased the diversification rate for Oraseminae. This is supported by our diversification rates analysis from BAMM (Fig. 5), which shows that the single most likely rate shift configuration is a single shift between Old World and New World taxa.

Combining datasets

Missing data have long been a contentious issue in phylogenetics, and it has resurfaced in the era of phylogenomics (Wiens, 2003; Lemmon *et al.*, 2009; Wiens & Morrill, 2011; Streicher *et al.*, 2016). Our combined dataset, with *c.* 51% missing data (artifact of concatenation; not including missing data from independent datasets), is just over the threshold found to maximize branch support at 50% (Streicher *et al.*, 2016). However, it should be noted that the highly uneven distribution of missing data makes our combined matrix substantially different from that of Streicher *et al.* (2016). Although we did not assess the impact of missing data with simulated datasets, our empirical analyses show no signs of taxa with AHE data (or without) clustering together (Figs S3, S6), which might be expected if missing data are affecting tree topologies. Additionally, the AHE ML tree (Fig. S5) and the combined ML tree (Fig. S6) have identical topologies for the 92 AHE taxa, indicating that the addition of Sanger data and taxa did not change the relationships inferred from AHE data. Missing data can have an impact on branch length estimates as well (Lemmon *et al.*, 2009), but we avoided this problem in our dating analysis of the combined dataset by using only the Sanger data, which were available for every taxon. Using similar concatenation methods, Peloso *et al.* (2015) found that increased character sampling for a subset of taxa (i.e. adding AHE data to some but not all taxa with Sanger data) generally lead to increased tree resolution (parsimony consensus) and increased support values (ML), while increasing taxon sampling (i.e. adding taxa with Sanger data to a matrix with Sanger + AHE data) increased support in shallow nodes but not in deeper nodes.

In our study, the Sanger dataset alone, using either parsimony or likelihood, resulted in monophyly of the Old World Oraseminae, albeit with little or no support. The AHE data alone strongly supported Old World paraphyly; however, the AHE dataset alone would not have thoroughly sampled across Oraseminae for either our host association or biogeographic studies. By combining datasets, we were able to produce comprehensive phylogenies with strong backbone support driven by the wealth of character information from AHE and the numerous

taxa available through the Sanger dataset. For dated trees, having a large dataset with a high percentage of missing data presented a hurdle for some analyses, but subsampling data by prioritizing the genes with the highest phylogenetic informativeness (AHE) and using a fixed tree topology based on the complete datasets (AHE and combined) were effective methods for efficiently analysing next-generation sequence data in programs like BEAST. The consistency of the dates produced by the three independent subsets of the AHE dataset (no overlapping gene regions among them) showed that this sampling method did not misrepresent the signal in the total AHE dataset; however, these three analyses did produce older dates than either the Sanger or combined analyses, which relied entirely on Sanger data (Fig. 3).

Phylogenetics of Oraseminae

This study has the most comprehensive sampling of orasemine taxa and genetic data compared with any previous phylogenetic analyses of this group (Heraty 1990, 1994b, 2000; Heraty *et al.*, 2004; Murray *et al.*, 2013). All genera except *Matantas* are represented, and all species groups in *Orasema* are represented, although there are still many species that we were unable to include either because they were rare or because specimens were too old to obtain quality DNA.

Results of all analyses support the monophyly of Oraseminae and its sister-group relationship to Eucharitinae. Gollumiellinae is most often sister to Eucharitinae + Oraseminae, except in the parsimony analyses of the AHE and combined datasets (Figs S2, S3), in which it is paraphyletic. This paraphyly is probably an artifact of the outgroup sampling and did not impact any of our results.

The generic-, species group-, and species-level classifications in Oraseminae are broadly supported across analyses with one major exception in the genus *Hayatosema*, where the African species never grouped with the species from Southeast Asia. The proposed reorganization of genera resulting from our analyses was addressed in the recent revision of the Old World genera (Burks *et al.*, 2017). Appendix S1 provides a detailed description of relationships across analyses for Oraseminae.

Georeferencing

Oraseminae are more diverse in the number of species and morphological variability in the New World (Fig. 2). Although the southern limits of their distribution are similar between the Old World and New World, the New World taxa range much farther north. Oraseminae are far more diverse in the tropics but with a number of Nearctic species in the *O. bakeri*, *cockerelli*, *coloradensis* and *simulatrix* species groups found predominately in desert environments. Most of the northerly records are from a single widespread species, *O. coloradensis*, which is known to oviposit on a wide variety of plants, occurs in a wide variety of habitats (mostly open scrub) from Florida to British Columbia and is a confirmed parasitoid of both *Pheidole* and *Solenopsis*. While not a candidate to be the sister group of the New World *Orasema*, these behaviours are what we might envision in a generalist species able to move easily into a new host/habitat niche (i.e. capable of crossing a temperate BLB).

Life-history evolution

Life-history traits beyond ant host association (e.g. oviposition strategy, plant hosts, etc.) are difficult to analyse within a phylogenetic context in Oraseminae because of the scarcity of information for Old World taxa, especially *Leiosema*, and some New World taxa. Ant host associations, more than any other life-history character, are far better understood and seem to be more of a limiting factor for dispersal and diversification in Oraseminae. This phylogeny provides a framework for future analyses of these traits.

Conclusions

By analysing taxon-rich and character-rich datasets independently and combined, we have shown a most likely scenario whereby Oraseminae originated in the Old World and dispersed once into the New World during the Oligocene. Based on climatic conditions and biogeographic analyses, the most likely route of dispersal was across the BLB. Their primary host ant, *Pheidole*, originated in the New World and dispersed into the Old World around the same time and probably across the same land bridge. Our results suggest that Oraseminae were well established in the Old World before the invasion of New World *Pheidole*, and multiple lineages may have shifted to this novel host. The difference in diversity and abundance of Oraseminae in the Old World and New World indicates that shifting onto *Pheidole* as a new host or dispersing into a new geographic area allowed the parasitoids to diversify more rapidly.

Our study shows how both phylogenetic and biogeographic information of a group of relatively specialized parasitoids can potentially inform us about both themselves and their hosts. No formal biogeographic hypothesis has ever been proposed to explain how *Pheidole* dispersed into the Old World, but using the context of Oraseminae biogeography and dating analyses, we strongly believe that the only way the parasitoids could enter the New World would have been across a host-occupied land bridge during suitable climatic conditions, and therefore, propose that *Pheidole* probably dispersed across the BLB.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Strict consensus of 20 most parsimonious trees of the Sanger dataset from TNT (4161 steps). Bootstrap support values are indicated unless lower than 50. Taxa in bold have been sampled for AHE genes.

Figure S2. Single most parsimonious tree of the AHE dataset from PAUP* (412 925 steps). Bootstrap support values are indicated unless they are < 50.

Figure S3. Strict consensus of 25 most parsimonious trees of the combined dataset from TNT (417 143 steps). Bootstrap support values are indicated unless they are < 50. Taxa in bold have been sampled for AHE genes.

Figure S4. Maximum likelihood tree of the Sanger dataset from RAXML. Ribosomal genes partitioned by sequenced region and mitochondrial genes partitioned by codon position. Bootstrap support values are indicated unless they are < 50. Taxa in bold have been sampled for AHE genes.

Figure S5. Maximum likelihood tree of the AHE dataset from RAXML. Genes partitioned by locus. Bootstrap support values are indicated unless they are < 50.

Figure S6. Maximum likelihood tree of the combined dataset from RAXML. Sanger and AHE gene regions partitioned the same way as in Figs S4, S5, respectively. Bootstrap support values are indicated unless they are < 50. Taxa in bold have been sampled for AHE genes.

Figure S7. Maximum clade credibility chronogram of the Sanger dataset from BEAST2. Posterior probability support values are indicated unless they are < 50. Taxa in bold have been sampled for AHE genes. Ninety five percent HPD is indicated by blue bars on nodes. The time scale is in Ma.

Figure S8. Maximum clade credibility chronogram for loci ranked 26–50 of the AHE dataset from BEAST2 using a fixed topology (RAXML tree for all AHE loci; Fig. S5). Rankings are based on a PHYDESIGN analysis of the phylogenetic informativeness of the AHE loci (Table S4). Ninety five percent HPD is indicated by blue bars on nodes. The time scale is in Ma.

Figure S9. Maximum clade credibility chronogram for loci ranked 51–75 of the AHE dataset from BEAST2 using a fixed topology (RAXML tree for all AHE loci; Fig. S5). Rankings are based on a PHYDESIGN analysis of the phylogenetic informativeness of the AHE loci (Table S4). Ninety five percent HPD is indicated by blue bars on nodes. The time scale is in Ma.

Figure S10. Maximum clade credibility chronogram of the combined dataset from BEAST2 using a fixed topology (RAXML tree of combined dataset; Fig. S6) and using only the Sanger genes. Ninety five percent HPD is indicated by blue bars on nodes. The time scale is in Ma.

Figure S11. Dispersal–extinction–cladogenesis (DEC) analysis of the biogeography of the combined dated tree (Fig. S10). Areas include Nearctic, Neotropical, Australasian, Oriental, Afrotropical, Madagascan, Indian and Palearctic. Dispersal rates were set by the position of the continents during historical epochs (Table S5).

Figure S12. Dispersal–extinction–cladogenesis with jump dispersal (DEC + J) analysis of the biogeography of the combined dated tree (Fig. S10). Areas include Nearctic, Neotropical, Australasian, Oriental, Afrotropical, Madagascan, Indian and Palearctic. Dispersal rates were set by the position of the continents during historical epochs (Table S5).

Figure S13. Coalescence species tree of Oraseminae from the AHE dataset using the program ASTRAL. In all, 348 input trees (one for each AHE locus) were created in RAXML with clades supported by bootstrap values of 10 or less collapsed into polytomies. Support values are shown as local posterior probabilities (Sayyari & Mirarab, 2016); only values < 100 are shown.

Table S1. Oraseminae ant and plant host records.

Table S2. List of voucher specimens and GenBank accession numbers.

Table S3. Primers and protocols for Sanger sequencing.

Table S4. AHE phylogenetic informativeness profiles.

Table S5. Dispersal rates for biogeographic analyses in DEC and DEC + J.

Table S6. Genomic resources for probe design (*Hym_Cha*), 7) AHE assembly data.

Appendix S1. Summary of relationships within Oraseminae and references for Table S1.

Acknowledgements

We thank Roger Burks, Scott Heacox and Krissy Dominguez for their contributions to this research. We thank Sean Holland and Michelle Kortyna at the Center for Anchored Phylogenomics for their assistance in collecting and analysing the AHE data. Many individuals have contributed specimens and biological data in support of this research, but we would like to extend special thanks to Wendy Porras, Javier Torréns, Laura Varone and Eric Yabar for their assistance with collecting in Central and South America. Support was provided by the Harry Scott Smith Biological Control Award to AJB, the Robert and Peggy van den Bosch Memorial Scholarship to AJB, USDA National Institute of Food and Agriculture Hatch project 1015803 to JMH, and National Science Foundation grants (DEB 1257733 and 1555808) to JMH. The authors declare there are no conflicts of interest.

References

Allen, J.M., Huang, D.I., Cronk, Q.C. & Johnson, K.P. (2015) aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics*, **16**, 98. https://doi.org/10.1186/s12859-015-0515-2.

Askew, R.R. (1971) *Parasitic Insects*. American Elsevier, New York, New York.

Blaimer, B.B., LaPolla, J.S., Branstetter, M.G., Lloyd, M.W. & Brady, S.G. (2016) Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. *Molecular Phylogenetics and Evolution*, **102**, 20–29.

Blakey, R.C. (2008) Gondwana paleogeography from assembly to breakup – a 500 million year odyssey. *Resolving the Late Paleozoic Ice Age in Time and Space* (Ed. by R.C. Fielding, T.D. Frank & J.L. Isbell), pp. 1–28. Geological Society of America Special Paper, Boulder, CO.

Bouček, Z. (1988) *Australasian Chalcidoidea (Hymenoptera): A Biosystematic Revision of Genera of Fourteen Families, with a Reclassification of Species*. CAB International, Wallingford.

Brandley, M.C., Wang, Y., Guo, X., de Oca, A.N., Feria-Ortiz, M., Hikida, T. & Ota, H. (2011) Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic Biology*, **60**, 3–15.

Branstetter, M.G. (2009) The ant genus *Stenamma* Westwood (Hymenoptera: Formicidae) redefined, with a description of a new genus *Propodilobus*. *Zootaxa*, **2221**, 41–57.

Burbrink, F.T. & Lawson, R. (2007) How and when did Old World ratsnakes disperse into the New World? *Molecular Phylogenetics and Evolution*, **43**, 173–189.

Burks, B.D. (1979) Family Eucharitidae. *Catalog of Hymenoptera in America North of Mexico* (ed. by K.V. Krombein, P.D. Hurd, Jr., D.R. Smith & B.D. Burks), pp. 875–878. Smithsonian Institution Press, Washington, D.C.

Burks, R.A., Heraty, J.M., Dominguez, C. & Mottern, J.L. (2018) Complex diversity in a mainly tropical group of ant parasitoids: revision of the *Orasema stramineipes* species group (Hymenoptera: Chalcidoidea: Eucharitidae). *Zootaxa*, **4401**, 107.

Burks, R.A., Heraty, J.M., Mottern, J., Dominguez, C. & Heacox, S. (2017) Biting the bullet: revisionary notes on the Oraseminae of the Old World (Hymenoptera, Chalcidoidea, Eucharitidae). *Journal of Hymenoptera Research*, **55**, 139–188.

Carey, B., Visscher, K. & Heraty, J.M. (2012) Nectary use for gaining access to an ant host by the parasitoid *Orasema simulatrix* (Hymenoptera, Eucharitidae). *Journal of Hymenoptera Research*, **27**, 47–65.

Chen, X., Huang, S., Guo, P. *et al.* (2013) Understanding the formation of ancient intertropical disjunct distributions using Asian and Neotropical hinged-teeth snakes (*Sibynophis* and *Scaphiodontophis*: Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*, **66**, 254–261.

Clausen, C.P. (1940a) The immature stages of the Eucharidae. *Proceedings of the Entomological Society of Washington*, **42**, 161–170.

Clausen, C.P. (1940b) The oviposition habits of the Eucharidae (Hymenoptera). *Journal of the Washington Academy of Sciences*, **30**, 504–516.

Clausen, C.P. (1941) The habits of the Eucharidæ. *Psyche (Cambridge)*, **48**, 57–69.

Condamine, F.L., Sperling, F.A.H. & Kergoat, G.J. (2013) Global biogeographical pattern of swallowtail diversification demonstrates alternative colonization routes in the Northern and Southern hemispheres. *Journal of Biogeography*, **40**, 9–23.

Das, G.M. (1963) Preliminary studies on the biology of *Orasema assectator* Kerrich (Hym., Eucharitidae), parasitic on *Pheidole* and causing damage to leaves of tea in Assam. *Bulletin of Entomological Research*, **54**, 373–378.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Dubovikoff, D. (2011) The first record of the genus *Pheidole* Westood, 1839 (Hymenoptera: Formicidae) from the Baltic amber. *Russian Entomological Journal*, **20**, 255–257.

Economo, E.P., Huang, J.P., Fischer, G. *et al.* (2019) Evolution of the latitudinal diversity gradient in the hyperdiverse ant genus *Pheidole*. *Global Ecology and Biogeography*, **00**, 1–15.

Economo, E.P., Klimov, P., Sarnat, E.M., Guenard, B., Weiser, M.D., Lecroq, B. & Knowles, L.L. (2015a) Global phylogenetic structure of the hyperdiverse ant genus *Pheidole* reveals the repeated evolution of macroecological patterns. *Proceedings of the Royal Society Biological Sciences Series B*, **282**, 20141416.

Economo, E.P., Sarnat, E.M., Janda, M. *et al.* (2015b) Breaking out of biogeographical modules: range expansion and taxon cycles in the hyperdiverse ant genus *Pheidole*. *Journal of Biogeography*, **42**, 2289–2301.

Eggleton, P. & Belshaw, R. (1992) Insect parasitoids: an evolutionary overview. *Philosophical Transactions: Biological Sciences*, **337**, 1–20.

Eldrett, J.S., Greenwood, D.R., Harding, I.C. & Huber, M. (2009) Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. *Nature*, **459**, 969–973.

Gadau, J., Helmkampf, M., Nygaard, S. *et al.* (2012) The genomic impact of 100 million years of social evolution in seven ant species. *Trends in Genetics*, **28**, 14–21.

Gamble, T., Bauer, A.M., Colli, G.R., Greenbaum, E., Jackman, T.R., Vitt, L.J. & Simons, A.M. (2011) Coming to America: multiple origins of New World geckos. *Journal of Evolutionary Biology*, **24**, 231–244.

GISD (2015) *Global Invasive Species Database* [WWW document]. URL http://www.iucngisd.org/gisd/. [accessed on 10 January 2017].

Goloboff, P.A., Farris, J.S. & Nixon, K.C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, **24**, 1–13.

Guo, P., Liu, Q., Xu, Y. *et al.* (2012) Out of Asia: natricine snakes support the Cenozoic Beringian dispersal hypothesis. *Molecular Phylogenetics and Evolution*, **63**, 825–833.

Haddad, S., Shin, S., Lemmon, A.R. *et al.* (2017) Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). *Systematic Entomology*, **43**, 68–89.

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M. & Bond, J.E. (2016) Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, **16**, 212. https://doi.org/10.1186/s12862-016-0769-y.

Heraty, J.M. (1990) *Classification and evolution of the Oraseminae (Hymenoptera: Eucharitidae)*. PhD Thesis, Texas A&M University, Entomology.

Heraty, J.M. (1994a) Biology and importance of two eucharitid parasites of *Wasmannia* and *Solenopsis*. *Exotic Ants: Biology, Impact and Control of Introduced Species* (ed. by D.F. Williams), pp. 104–120. Westview Press, Oxford.

Heraty, J.M. (1994b) Classification and evolution of the Oraseminae in the Old World, including revisions of two closely related genera of Eucharitinae (Hymenoptera: Eucharitidae). *Life Sciences Contributions*, **157**, 1–174.

Heraty, J.M. (2000) Phylogenetic relationships of Oraseminae (Hymenoptera: Eucharitidae). *Annals of the Entomological Society of America*, **93**, 374–390.

Heraty, J.M. (2002) *A Revision of the Genera of Eucharitidae (Hymenoptera: Chalcidoidea) of the World*, Vol. **68**, pp. 1–359. *Memoirs of the American Entomological Institute*, Gainesville, Florida.

Heraty, J.M. (2017) *Catalog of World Eucharitidae, 2017* [WWW document]. URL https://hymenoptera.ucr.edu/EucharitidaeCatalog2017.pdf UC Riverside [accessed on 28 April 2018].

Heraty, J.M. & Darling, D.C. (2009) Fossil Eucharitidae and Perilampidae (Hymenoptera: Chalcidoidea) from Baltic amber. *Zootaxa*, **2306**, 1–16.

Heraty, J.M., Hawks, D., Kostecki, J.S. & Carmichael, A. (2004) Phylogeny and behaviour of the Gollumiellinae, a new subfamily of the ant-parasitic Eucharitidae (Hymenoptera: Chalcidoidea). *Systematic Entomology*, **29**, 544–559.

Heraty, J.M. & Murray, E. (2013) The life history of *Pseudometagea schwarzii*, with a discussion of the evolution of endoparasitism and koinobiosis in Eucharitidae and Perilampidae (Chalcidoidea). *Journal of Hymenoptera Research*, **35**, 1–15.

Heraty, J.M., Wojcik, D.P. & Jouvenaz, D.P. (1993) Species of *Orasema* parasitic on the *Solenopsis saevissima*-complex in South America (Hymenoptera: Eucharitidae, Formicidae). *Journal of Hymenoptera Research*, **2**, 169–182.

Herreid, J.S. & Heraty, J.M. (2017) Hitchhikers at the dinner table: a revisionary study of a group of ant parasitoids (Hymenoptera: Eucharitidae) specializing in the use of extrafloral nectaries for host access. *Systematic Entomology*, **42**, 204–229.

Hines, H.M. (2008) Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). *Systematic Biology*, **57**, 58–75.

Jansen, G., Savolainen, R. & Vepsalainen, K. (2010) Phylogeny, divergence-time estimation, biogeography and social parasite-host relationships of the Holarctic ant genus *Myrmica* (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, **56**, 294–304.

Johnson, J.B., Miller, T.D., Heraty, J.M. & Merickel, F.W. (1986) Observations on the biology of two species of *Orasema* (Hymenoptera: Eucharitidae). *Proceedings of the Entomological Society of Washington*, **88**, 542–549.

Junier, T. & Zdobnov, E.M. (2010) The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, **26**, 1669–1670.

Kaplan, A.A., Grigelis, A.A., Strelnikova, N.I. & Glikman, L.S. (1977) Stratigraphy and correlation of Palaeogene deposits of south-western cis-Baltic region. *Soviet Geology*, **4**, 30–43.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kearse, M., Moir, R., Wilson, A. *et al.* (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kerrich, G.J. (1963) Descriptions of two species of Eucharitidae damaging tea, with comparative notes on other species (Hym., Chalcidoidea). *Bulletin of Entomological Research*, **54**, 365–371.

Lachaud, J.-P. & Pérez-Lachaud, G. (2012) Diversity of species and behavior of hymenopteran parasitoids of ants: a review. *Psyche*, **2012**, 1–24. https://doi.org/10.1155/2012/134746.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, **14**, 82. https://doi.org/10.1186/1471-2148-14-82.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.

Lemmon, A.R., Brown, J.M., Stanger-Hall, K. & Lemmon, E.M. (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology*, **58**, 130–145.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.

Lemmon, E.M. & Lemmon, A.R. (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Lopez-Giraldez, F. & Townsend, J.P. (2011) PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology*, **11**, 1–4.

Maddison, D.R. (2016) The rapidly changing landscape of insect phylogenetics. *Current Opinion in Insect Science*, **18**, 77–82.

Maddison, D.R. & Maddison, W.P. (2017a) *Chromaseq: a Mesquite package for analyzing sequence chromatograms. Version 1.3* [WWW document]. URL http://mesquiteproject.org/packages/chromaseq [accessed on 28 April 2018].

Maddison, W.P. & Maddison, D.R. (2017b) *Mesquite: a modular system for evolutionary analysis. Version 3.31* [WWW document]. URL http://mesquiteproject.org [accessed on 28 April 2018].

Mann, W.M. (1918) Myrmecophilous insects from Mexico. *Psyche*, **25**, 104–106.

Masonick, P., Michael, A., Frankenberg, S., Rabitsch, W. & Weirauch, C. (2017) Molecular phylogenetics and biogeography of the ambush bugs (Hemiptera: Reduviidae: Phymatinae). *Molecular Phylogenetics and Evolution*, **114**, 225–233.

Matzke, N.J. (2013) *BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R scripts. R package, version 0.2.1, published July 27, 2013* [WWW document]. URL http://CRAN.R-project.org/packages=BioGeoBEARS [accessed on 28 April 2018].

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.

McKenna, M.C. (1983) Holarctic landmass rearrangement, cosmic events, and Cenozoic terrestrial organisms. *Annals of the Missouri Botanical Garden*, **70**, 459–489.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA*, 1–8.

Milne, R.I. (2006) Northern hemisphere plant disjunctions: a window on tertiary land bridges and climate change? *Annals of Botany*, **98**, 465–472.

Mirarab, S. & Warnow, T. (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, **31**, i44–i52.

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

Murray, E.A., Carmichael, A.E. & Heraty, J.M. (2013) Ancient host shifts followed by host conservatism in a group of ant parasitoids. *Proceedings of the Royal Society Biological Sciences Series B*, **280**, 20130495.

Peloso, P.L.V., Frost, D.R., Richards, S.J. *et al.* (2015) The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). *Cladistics*, **32**, 1–28.

Perkovsky, E.E. (2016) Tropical and holarctic ants in Late Eocene ambers. *Vestnik Zoologii*, **50**, 111–122.

Perkovsky, E.E., Rasnitsyn, A.P., Vlaskin, A.P. & Taraschuk, M.V. (2007) A comparative analysis of the Baltic and Rovno amber arthropod faunas: representative samples. *African Invertebrates*, **48**, 229–245.

Peters, R.S., Niehuis, O., Gunkel, S. *et al.* (2018) Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Molecular Phylogenetics and Evolution*, **120**, 286–296.

Pinto, J.D. (2009) Hypermetamorphosis. *Encyclopedia of Insects* (ed. by V.H. Resh & R.T. Carde), pp. 484–486. Academic Press, Cambridge, Massachusetts.

Poinar, G. Jr & Huber, J.T. (2011) A new genus of fossil Mymaridae (Hymenoptera) from Cretaceous amber and key to Cretaceous mymarid genera. *Zookeys*, **130**, 461–472.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. & Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **526**, 569–573.

de Queiroz, A. (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution*, **20**, 68–73.

Rabosky, D.L., Grundler, M., Anderson, C. *et al.* (2014) BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, **5**, 701–707.

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. (2014) *Tracer v1.6* [WWW document]. URL http://beast.bio.ed.ac.uk/Tracer [accessed on 28 April 2018].

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.

Ritzkowski, S. (1997) K-Ar-Altersbestimmungen der bernsteinführenden sedimente des Samlandes (Paläogen, Bezirk Kaliningrad). *Metalla, Sonderheft*, **66**, 19–23.

Rokyta, D.R., Lemmon, A.R., Margres, M.J. & Aronow, K. (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*, **13**, 312. https://doi.org/10.1186/1471-2164-13-312.

Sadd, B.M., Barribeau, S.M., Bloch, G. *et al.* (2015) The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology*, **16**, 76. https://doi.org/10.1186/s13059-015-0623-3.

Sanmartín, I., Enghoff, H. & Ronquist, F. (2001) Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society*, **73**, 345–390.

Sanmartín, I. & Ronquist, F. (2004) Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*, **53**, 216–243.

Schmidt, H.A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**, 502–504.

Simpson, G.G. (1940) Paleontology: mammals and land bridges. *Journal of the Washington Academy of Sciences*, **30**, 137–163.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Streicher, J.W., Schulte, J.A. II & Wiens, J.J. (2016) How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology*, **65**, 128–145.

Swofford, D.L. (2002) *PAUP* v.4.0: Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Sinauer, Sunderland, Massachusetts.

Tiffney, B.H. (1985) The Eocene north Atlantic land bridge: its importance in tertiary and modern phytogeography of the northern hemisphere. *Journal of the Arnold Arboretum*, **66**, 243–273.

Townsend, T.M., Leavitt, D.H. & Reeder, T.W. (2011) Intercontinental dispersal by a microendemic burrowing reptile (Dibamidae). *Proceedings of the Royal Society Biological Sciences Series B*, **278**, 2568–2574.

Vaidya, G., Lohman, D.J. & Meier, R. (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, **27**, 171–180.

Varone, L. & Briano, J. (2009) Bionomics of *Orasema simplex* (Hymenoptera: Eucharitidae), a parasitoid of *Solenopsis* fire ants (Hymenoptera: Formicidae) in Argentina. *Biological Control*, **48**, 204–209.

Varone, L., Heraty, J.M. & Calcaterra, L.A. (2010) Distribution, abundance and persistence of species of *Orasema* (Hym: Eucharitidae) parasitic on fire ants in South America. *Biological Control*, **55**, 72–78.

Vila, R., Bell, C.D., Macniven, R. *et al.* (2011) Phylogeny and palaeoecology of *Polyommatus* blue butterflies show Beringia was a climate-regulated gateway to the New World. *Proceedings of the Royal Society Biological Sciences Series B*, **278**, 2737–2744.

Walker, J.D., Geissman, J.W., Bowring, S.A. & Babcock, L.E. (2012) *Geologic Time Scale v.4.0*. Geological Society of America, Boulder, Colorado.

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

Weinstock, G.M., Robinson, G.E., Gibbs, R.A. *et al.* (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, **443**, 931–949.

Werren, J.H., Richards, S., Desjardins, C.A. *et al.* (2010) Functional and evolutionary insights from the genome of three parasitoid *Nasonia* species. *Science*, **327**, 343–348.

Wheeler, G.C. & Wheeler, E.W. (1937) New hymenopterous parasites of ants (Chalcidoidea: Eucharidae). *Annals of the Entomological Society of America*, **30**, 163–175.

Wheeler, W.M. (1907) The polymorphism of ants, with an account of some singular abnormalities due to parasitism. *Bulletin of the American Museum of Natural History*, **23**, 1–93.

Wiens, J.J. (2003) Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology*, **52**, 528–538.

Wiens, J.J. & Morrill, M.C. (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology*, **60**, 719–731.

Wilson, T.H. & Cooley, T.A. (1972) A chalcidoid planidium and an entomophilic nematode associated with the western flower thrips. *Annals of the Entomological Society of America*, **65**, 414–418.

Wolfe, J.A. (1993) An analysis of Neogene climates in Beringia. *Palaeogeography Palaeoclimatology Palaeoecology*, **108**, 207–216.

Wüster, W., Peppin, L., Pook, C.E. & Walker, D.E. (2008) A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Molecular Phylogenetics and Evolution*, **49**, 445–459.

Accepted 30 April 2019
First published online 9 June 2019