Global ant (Hymenoptera: Formicidae) biodiversity and biogeography
– a new database and its possibilities


Abstract

Despite several centuries of research, the global patterns of species diversity, individual abundance and community composition and their drivers and subtleties remain poorly resolved. We have developed a global database for the diversity of ants, perhaps the best-studied of ecologically important insect taxa. We describe the database and aspects of its limitations and, at more length, possibilities. The database offers the possibility of testing general macroecological theory with an ecologically important group of insects. The database will also allow us to understand ways in which the global diversity, distribution and biogeography of ants differs from the more often studied vertebrates and plants and some of the consequences of those differences.

Key words: Biodiversity, macroecology, distribution, Formicidae, database, biogeography.
Introduction

Nearly all analyses of patterns of diversity, abundance or composition of ecological communities have been confined to regions, or, at best, continents. Global scale analyses offer the hope of adding additional nuance to biogeographic and macroecological questions, in large part because global datasets allow comparisons among continents and biogeographic regions (Gaston 2000, Gaston & al. 2003, Turner 2004, Lamoreux & al. 2006). Such comparisons allow very general tests of putative patterns and mechanism. One approach to generating global diversity datasets is to combine the results of the many community and diversity studies done around the world. This approach has recently been used for birds (Pautasso & Gaston 2005), but ideally one would like to be able to compare results for birds with those of other taxa more representative of the global fauna, which consists primarily of insects.

Ants represent an ideal group for which to compile a global database. Ants are ecologically important in nearly every terrestrial biome. In addition, ants run the spectrum in terms of abundance and distribution. Ants are overly represented as global introduced or invasive species (McGlynn 1999, Suárez & al. 2005). At the same time, ants, like all social insects, are disproportionately prone to endangerment (Chapman & Bourke 2001) and, at least within regions, respond in consistent and predictable ways to disturbance (Andersen 1997b). In addition, ants pose practical advantages for a global database. Studies of ant communities are common and tend to use a relatively consistent set of sampling methods, at least when compared to studies of other vertebrates and other factors of concern. With this global ant database, we seek to address many fundamental questions in macroecology and conservation that have thus far been largely neglected due to lack of appropriate data. Some of these questions can be addressed immediately, others will require that we add additional components to the database (for example, data on distributions of individual species for areas where taxonomy is relatively stable). In Box 1, we list some of the questions we hope to address with the proposed database.

Ultimately, we plan to address a suite of broad questions using this database. But before we can address any of these questions, it is worth asking more general questions about the extent to which ant biodiversity has been adequately assessed globally. In this paper we consider how well existing quantitative ant biodiversity studies sample across continents, biomes and regions.

Methods

Beginning in 2006, Dunn, Sanders, and colleagues began to compile a global database of ant biodiversity studies. This collaboration began after the development of a separate database focusing on North American ants by Dunn (with collaboration from Lessard, Sanders, Laurent, and Fitzpatrick). Once the North American database included nearly all of the published studies of ant communities in North America, Dunn and Sanders announced the initiation of the global database at the International Union for the Study of Social Insect meetings in Washington, D.C. in 2006. At the meetings (more specifically in a bar at the hotel), those wishing to contribute data and insight to the project were invited to participate. Additional participants were later invited to increase coverage for poorly represented regions. Participants were asked to compile data from their own work and other studies. In addition to the studies identified through this network approach, we also searched Web of Science, Formis (a nearly exhaustive ant literature database for the 1800s through 2003) and Google Scholar using a variety of key words, including (formicidae + pitfall, formicidae + litter, formicidae + transect, and formicidae + diversity) related to ant communities (the same approach initially used for the North American database). We focused on searches that included the terms "pitfall" (for pitfall traps) and "Winkler" (for the Winkler litter sampling device), but also included studies in the database that used other methods such as baiting or hand collection. In an effort to also capture those studies that sampled ants without using pitfall traps or Winkler samples, we also searched using the terms "Formicidae" and "diversity". Ultimately, the coauthors on this study were as productive in compiling studies of ant communities from their personal libraries as were more standardized searches. As an index of the relative completeness of the final database, we counted the proportion of studies found on Google Scholar under the search terms "Formicidae + pitfall" and "Formicidae + Winkler".

All studies recovered from these searches, regardless of sampling methods and including studies that focused on sampling ant nests and studies that focused on forest canopies, were included in the database. We did not include studies that are tropically limited (e.g., focused only on granivores) or those that focused on arboreal ant communities (because of the problems associated with quantitatively sampling arboreal communities). We consider in depth here the subset of those studies that meets the following criteria: attempt to take a standardized sample of the ground-foraging and / or litter ant community of an area of roughly a hectare in size, and used one of the most
Box 1: Here we highlight some of the kinds of questions our global database will be able to address.

- Recent reviews (HAWKINS & al. 2003) have indicated the central importance of both solar energy and water availability as fundamental drivers of diversity gradients. But via what mechanism? Is it simply that energy limits abundance which in turn mediates extinction rates (SRIVASTAVA & LAWTON 1998), or alternatively does energy drive speciation rates (ALLEN & al. 2002, KIER & al. 2005)?

- Metabolic and food web theories (e.g., BROWN & al. 2004) predict that higher trophic levels increase in diversity disproportionately with increases in productivity. Do these predictions hold up at large scales?

- How do the effects of disturbances interact with energy availability and habitat complexity, and how does the effect of disturbance on ant communities vary among habitat types (RATCHFORD & al. 2005)?

- How do spatial gradients of global ant diversity compare to those of ecosystem processes, such as decomposition or litter fall rates?

- How do tests of global concordance of diversity patterns among taxa have been completely limited to comparisons among vertebrate taxa (LAMOREUX & al. 2006). How do the patterns of ant diversity and distribution compare to those of other taxa, especially insectivorous birds, mammals, and reptiles?

- Conservation plans at global, regional, and local scale rely invariably on a subset of taxa to act as surrogates for overall diversity (CEBALLOS & EHRLICH 2006). How well do conservation plans designed for other taxa capture rare ant species or ant diversity?

- Where have ants been undersampled? Which of those undersampled regions are most likely to have a high diversity of species?

- How will climate change affect global patterns of ant diversity?

- What are the correlates of ant invasion globally? Does ant community composition determine susceptibility to invasion?

quantitative standardized ant sampling methods, pitfall trapping, litter sifting or baiting. Because of changes in ant taxonomy through time, we include here only studies published in the last twenty years, though earlier studies are included in the full database. Where the same site was surveyed multiple years, we did not treat the multiple sampling events as independent samples but instead took the mean estimate of richness. Very few studies, mostly from Australia, included multi-year data. Because samples from different years or times of year do not necessarily yield identical or even similar results (STEINER & al. 2005), we are considering including data from each sampling event separately in future versions of the database. From the original studies, we extracted data on the number of samples, the kinds of sampling techniques used, the area sampled, latitude, longitude, elevation, and number of individuals sampled, all as potential correlates of ant species richness. Because we were interested in characterizing local communities, we do not include here studies that provide species lists for regions, but not estimates of site richness (e.g., county or province lists).

**Sampling grains**

In nearly all studies of ant communities, one grain size that researchers attempt to sample is an area over which colonies might interact with one another, a grain size we call the "site". Sample "sites" vary from a 10 × 10 m plot to a 100 × 100 m plot. Each such sampling of a site yields a measure of site richness (the total number of species collected in the site). A second grain size that we discern is what we call the "landscape", a roughly 5 × 5 km or smaller area that includes multiple sites and, except in regional studies, tends to include most of the sites within a study (regardless of habitat type). In cases where the landscape was sampled at multiple sites, a measure of landscape ant richness can be derived from the total number of species sampled across the sites within the landscape. Additional measures of richness (such as rarefied and / or estimated asymptotic richness) could be derived from the dataset but are not our focus here. Note, if other grain sizes were recorded, we tried to include those grains in the database. However, we focus on just two grain sizes herein.

We include as pitfall traps any method that involves leaving cups of some sort buried in the ground in order to intercept ants that walk over the cups. We use the term Winkler as short hand for any method that involves sampling leaf-litter and extracting ants from that leaf-litter by allowing the litter to dry and the ants to walk out of the drying litter. Nearly all of the "Winkler" studies in the database actually used a standardized Winkler bag for processing samples (AGOSTI & al. 2000). Baiting studies include a variety of bait types but are dominated by Tuna and Sardine baits and cookies / biscuits, but also other food sources meant to attract a broad suite of ground-foraging ants.

**The state of the database**

To date, the full database includes data from more than 2700 sample sites and 225 studies.
How exhaustive is the database to date?

Our database includes > 90% of papers on ant communities retrieved using the search term "Formicidae + Pitfall" and more than 95% of those papers retrieved using "Winkler + Formicidae as of February, 2007". See "Adding data" in the Discussion for more about papers not yet in the database. Although the lead PI’s on the project have an English language bias, we do not suspect that this bias is strong in most regions. For example, if we search on Formicidae and "trampa de caída" (pitfall trap) in Google Scholar we find only twelve studies, all of which were retrieved when the search was performed in English. However, in Europe and Asia, our English language bias has undoubtedly led us to miss some studies that could be usefully included in the overall database. We are in the process of additional searches in Korean, Japanese, Swedish, Bulgarian, French and several other languages to supplement our sampling from Europe and Asia.

Certainly studies exist that we might have overlooked (if you know of any obscure ones, please contact the lead author, rob_dunn@ncsu.edu). A full list of studies included in the database as of the publication of this paper is available online at:
http://www4.ncsu.edu/~rrdunn/AntMacroecology.html

Missing regions

To consider the relative completeness of sampling of different continents, we tabulated the number of sites in the study from different continents. Overall, North and South + Central America and Australia have been much better sampled than Asia and Africa and Europe (Tab. 1, Fig. 1). Europe is better represented in the full database than it is represented for the sampling methods considered here because a number of European studies have used standardized nest sampling methods (Steiner & al. 2005) that are rare elsewhere (Johnson 1992). Nonetheless, the overall rank order of the number of sites by region would hold if we were to include hand sampling of nests. Similarly, while we know of a handful of studies from Africa that have yet to be added, addition of these studies will still leave most of the continent unsampled in the last twenty years. While the presence of a quantitative sample does not necessarily correlate with the study of ants more generally (for example, ants in many European countries are well-studied, but quantitative studies of ant communities are just more rare) we believe the absence of studies that could be usefully included from Europe and Asia to be representative of a more general lack of empirical work on ants in these regions.

Sampling methods and their use

Of the methods considered here, the two primary sampling methods were Winkler samples and pitfall traps. In another recent study (Schlick-Steiner & al. 2006), authors similarly found that pitfall studies were the most common. Schlick-Steiner also found that in studies of ant species or assemblages that nest counting was also very common (second only to pitfall trapping), but such nest sampling is more poorly represented in our database, perhaps in part because we focus only on studies of entire communities or assemblages, where such intensive methods are more rare.

The tendency was for pitfall traps to be used in low productivity environments, whereas Winkler samples were used in high productivity environments. The mean productivity of sites sampled by pitfall trapping was 720 g C m² y⁻¹ vs. 1005 g C m² y⁻¹ for Winkler samples. In contrast to the results for productivity, the mean annual temperature of sites where ants were sampled with pitfalls was slightly lower (17.9 °C) but similar to the mean temperature of sites sampled with Winkler samples (19.1 °C). To a great extent, this pattern was due to the use of pitfalls in the many desert studies included in the database (where litter is sparse and hard to sample) and litter sampling in tropical forests.

As has been pointed out elsewhere, on their own all methods of sampling ant communities have biases and only sample a subset of the entire ant community (Schlick-Steiner & al. 2006). Because the component of the ant community sampled by different methods differs, we will need to carefully consider differences among sampling methods when examining broad patterns of diversity. This is particularly true if which sampling method is used is nonrandom with respect to environmental gradients, as appears to be the case.

Missing conditions and biomes

Knowing how environmental conditions shape ant communities depends on data on ant communities across a full range of existing conditions. When we plotted the mean annual temperature of sites where ants were sampled against the productivity of those same sites, both very hot and very cold conditions tended to be underrepresented (e.g., Fig. 2). We overlaid the temperature by precipita-
Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.

Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.

Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.

Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.

Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.

Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.

Fig. 2: Distribution of sampling points across temperature by rainfall space. Points are overlaid on a diagram showing the major biomes from WHITTAKER (1975). Note that most temperature by rainfall space is well sampled, exceptions being temperate rainforests, tundra, the hottest subtropical deserts and to a lesser extent taiga. Importantly, the areas that are most poorly sampled (really hot, really cold) are those thought based on nonstandardized sampling to be the most species poor (HEATWOLE 1991, 1996). Temperate woodland/shrubland (red), temperate deciduous forests (forest green) and tropical seasonal forests (lime green), on the other hand, are very well sampled.

Fig. 3: Total richness of species in a study landscape (where a landscape includes 3 or more study sites and is typically of < 5 × 5 km). Larger circles represent more diverse landscapes, with circle area scaled to diversity. Points in this figure represent the subset of points in Fig. 2 for which more than 3 samples were taken.

Fig. 4: An example of the relationship between number of samples and richness for Winkler samples in tropical (red circles) and temperate forests (white circles). Note, one tropical forest point falls off the scale of the figure, but is included in the regression lines. Best-fit lines are lognormal. The regression for the tropical forest, but not the temperate forest is significant at p < 0.05.
for temperate forest sites, there is a correlation between the number of Winkler samples taken and the number of species collected. Comparison of small samples from tropical and temperate forest sites could lead to erroneous conclusions about the relative diversity of communities in those two biomes. However, once 15 or 20 Winkler samples have been taken for tropical sites, the increase in diversity sample number slows. These results compare well with one of the most detailed studies of sampling at a single site to date (Lepont & al. 2004). Lepont & al. (2004) found that 20 samples (in this case Winkler) captured slightly less than half of the species they judged to be capturable species in a subtropical forest. Similarly, 20 samples in the among site comparisons captured 40 - 50 % of the mean number of species captured after 60 or 70 samples.

**Taxonomic limitations**

One final potential limitation of this database, like any database of its kind, has to do with taxonomy. For ants, like nearly every (or perhaps every) other insect taxon, many species remain to be described. In addition, many described species will, on closer scrutiny, likely prove to hold additional cryptic species. This taxonomic impediment is well known and could limit our analyses in two ways. First, because studies often include either unnamed or simply unidentified species, we cannot consider patterns of beta diversity for most parts of the world since comparisons of beta diversity depend upon the ability to compare the species at multiple sites. One way around this problem may be to focus on taxonomically well-studied regions for such comparisons. For the moment, we avoid this problem by focusing only on alpha diversity patterns.

Second, within studies the presence of cryptic species in some but not other sites may lead to underestimates of diversity in some locals (Delabie & Fowler 1990). A related problem is that myrmecologists in some regions may split species more so than do myrmecologists in other regions. For example, in Australia biologists working in Western Australia tend to be more conservative in splitting species than those working in the Northern Territory (B. Hetterick & A. Andersen, pers. comm.). It is our suspicion that the biases introduced by both of these problems are small relative to the biases among studies due to differences in sampling. Nonetheless, it is worth being cognizant of this potential problem, one which is not unique to ants.

**The future of the database**

We think the existing ant literature represents an exciting resource to address questions about the ecology, conservation biology, biogeography and macroecology of ant diversity. Sampling biases exist in what we know about ant communities (bias against Africa and Asia, bias toward litter samples in the tropics) as does important variation among studies in sampling methods and completeness. However, the broad trends in the patterns of diversity, for example (Fig. 3), indicate that the data are sufficient to recover an informative signal. Our next step will be to use the database to characterize broad patterns of diversity among ant communities at site and landscape grains globally. However, we will also be adding to the database for future analyses, with a particular focus on underrepresented biomes.

In addition to the short term goals of the database, in the long-term we hope to see the database used in a wide diversity of contexts. We have already begun discussions to bridge the work on the database and ongoing ant phylogenetic work (with C. Moreau and B. Fisher), to compare patterns of ant diversity to patterns of diversity for other taxa (with W. Jetz) and to better link our existing focus on species richness to species and specimen data in those regions where taxonomy permits.

Ultimately, however, we hope that the greatest successes of the database will come from the use of the tool by many researchers with interests more diverse than we have been able to anticipate here. Once several initial papers are published on the database, the database will be freely available online. Anyone will be able to add data to the database and also to analyze data from the database. This more distributed analysis of ant communities will ultimately, we hope, yield far more interesting results than what we could generate on our own.

**Acknowledgements**

We gratefully acknowledge support from the Department of Energy (NICCR), the Australian Research Council, and the many other funding agencies that funded the original data collection in the studies that when combined constitute the database.

**Zusammenfassung**


**References**


