A global database of ant species abundances


1Department of Ecology, Environment and Evolution, La Trobe University, Melbourne 3086, Victoria, Australia.
2Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695, USA.
3Center for Macroecology, Evolution, and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark.
4Department of Environmental Science, University of Girona, Montilivi Campus s/n, 17071 Girona, Spain.
5CSIRO Ecosystem Sciences, Tropical Ecosystems Research Centre, PMB 44 Winnellie, Northern Territory 0822, Australia.
29 Department of Environment and Agriculture, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia.

30 Instituto de Biologia, Universidade Federal de Uberlândia (UFU) Rua Ceara, Uberlândia, Minas Gerais 38400-902, Brazil.

31 School of Biological Sciences, The University of Hong Kong, Pok Fu Lam Road, Hong Kong SAR, China.

32 University of Guanajuato, Department of Biology, Noria Alta sn. Guanajuato, Mexico.

33 IPÊ-Instituto de Pesquisas Ecológicas, Nazaré Paulista, São Paulo 12960-000, Brazil.

34 Department of Biology, 730 Van Vleet Oval, Rm 314, University of Oklahoma, Norman OK, 73019, USA.

35 New Guinea Binatang Research Center, P.O. Box 604, Madang, Papua New Guinea.

36 Centre for Tropical Biology and Climate Change, School of Marine and Tropical Biology, James Cook University, PO Box 6811, Cairns, Queensland 4870, Australia.

37 Saarland University, Germany.

38 Departamento de Zoologia, Universidade Federal do Paraná, Caixa Postal 19020, 81531-980 Curitiba, PR Brazil.

39 Royal Belgian Institute of Natural Sciences, Section of Biological Evaluation, Rue Vautier, 29, Brussels 1000, Belgium.

40 Department of Biology, Concordia University, Montreal, QC, H4B-1R6.

41 Entomology and Nematology Department, University of Florida, 970 Natural Area Drive, Gainesville, FL 32611-0620, USA.

42 School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

43 Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

44 School of Plant Biology, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia.

45 Department of Biology, California State University Dominguez Hills, 1000 E. Victoria Street, Carson, CA 90747 USA.

46 Department of Entomology, Natural History Museum of Los Angeles County, USA.

47 Department of Biology, Lake Forest College, 555 North Sheridan Road, Lake Forest, IL 60045, USA.

48 Field Museum of Natural History, Department of Zoology, Division of Insects, Moreau Lab, 1400 South Lake Shore Drive, Chicago, IL 60605, USA.

49 School of Life Sciences, College of Agriculture Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, 3209, South Africa.
Institute of Animal Ecology and Cell Biology, TiHo Hannover, Bünteweg 17d, Hannover 30559, Germany.

Department of Ecology, National University of Mongolia, Baga toiruu 47, PO Box 377, Ulaanbaatar 210646, Mongolia.

Environmental Studies Department, University of California, 1156 High Street, Santa Cruz, CA 95060, USA.

The Department of Ecology and Evolutionary Biology, University of Colorado UCB 334, Boulder 80309, USA.

Universitat Autònoma Barcelona, Cerdanyola del Vallès 08193, Spain.

Museu Paraense Emílio Goeldi, Coordenação de Ciências da Terra e Ecologia, Belém, PA, Brazil.

Coordenação de Biodiversidade, National Institute of Amazonian Research, Manaus, AM, Brazil.

Department of Entomology, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA.

Department of Tropical Ecology and Animal Biodiversity, University of Vienna, Rennweg 14, Vienna 1030, Austria.

Department of Biology, Stanford University, Stanford, CA 94305-5020, USA.

Centre for Behavioural and Physiological Ecology, Zoology, University of New England, Armidale, NSW, Australia.
Abstract

What forces structure ecological assemblages? A key limitation to general insights about assemblage structure is the availability of data that are collected at a small spatial grain (local assemblages) and a large spatial extent (global coverage). Here, we present published and unpublished data from 51,388 ant abundance and occurrence records of more than 2693 species and 7953 morphospecies from local assemblages collected at 4212 locations around the world. Ants were selected because they are diverse and abundant globally, comprise a large fraction of animal biomass in most terrestrial communities, and are key contributors to a range of ecosystem functions. Data were collected between 1949 and 2014, and include, for each geo-referenced sampling site, both the identity of the ants collected and details of sampling design, habitat type and degree of disturbance. The aim of compiling this dataset was to provide comprehensive species abundance data in order to test relationships between assemblage structure and environmental and biogeographic factors. Data were collected using a variety of standardised methods, such as pitfall and Winkler traps, and will be valuable for studies investigating large-scale forces structuring local assemblages. Understanding such relationships is particularly critical under current rates of global change. We encourage authors holding additional data on systematically collected ant assemblages, especially those in dry and cold, and remote areas, to contact us and contribute their data to this growing dataset.

Keywords: Abundance, ants, database, disturbance, Formicidae, geo-referenced, habitat, local assemblage, occurrence, pitfall trap, Winkler trap

* E-mail: h.gibb@latrobe.edu.au
Introduction

General aims of database

Questions concerning communities, or sets of co-occurring species, are among the most challenging in ecology (Sutherland et al. 2013). Ecologists have focussed largely on the role of local-scale biotic and abiotic processes in determining the diversity and composition of communities. More recent macroecological work has highlighted that factors operating at larger scales, such as biogeography (e.g., Violle et al. 2014) and evolutionary history (e.g., Ricklefs 2008) also play a role (Sutherland 2013). In addition, chance is important (e.g., Hubbell 2001). Recognition of increasing anthropogenic pressures on assemblages at multiple scales has also focussed research on understanding multi-scalar effects of factors such as anthropogenic disturbance, climate change and species invasions (e.g., Mishra et al. 2004, Woodward et al. 2010, Pacciardi et al. 2011). Important insights into the relative importance of different forces in determining assemblage structure and composition at local scales can be obtained by examining how their influence changes across broader scales. For example, Gibb et al. (2015) showed that climate regulates the impact of disturbance on species richness and evenness. However, very few publicly available datasets exist that allow researchers to test the effects of broad-scale drivers on local assemblages. Although plot-based data on plant assemblages are relatively common (Weiser et al. 2007, Swenson et al. 2012), they often cover only limited geographic or climatic scales (e.g., Andersen-Teixeira et al. 2015). Analogous data are usually missing for animals, particularly invertebrates. Terrestrial animal assemblage datasets often contain only presence-absence information collected at a coarse spatial grain, precluding meaningful analysis of species co-occurrence (e.g., PanTHERIA, Jones et al. 2009) or cover only limited geographic extent (e.g., Atlas of Living Australia, www.ala.org.au, Carabids.org, Homburg et al. 2014). Other databases (e.g., PREDICTS, Hudson et al. 2014) focus primarily on human impacts on biodiversity and are not yet publicly available.

Why ants?

Ants (Hymenoptera: Formicidae) were selected as the target taxon because: (1) ants comprise a large fraction of animal biomass in most terrestrial communities (King et al. 2013); (2) ants perform a range of important ecosystem functions (Folgarait 1998, Del Toro et al. 2012); (3) ant sampling uses standard methods of sampling (i.e., pitfall traps and Winkler samplers), making inter-site comparisons possible (Agosti et al. 2000); (4) ant workers are abundant, so
they are likely to be trapped if present; (5) ants are diverse, but more manageable than some insect groups; (6) data on ant morphology is obtainable from museums and other collections worldwide; (7) ants are well described relative to other easily trapped groups and are well documented online through a unique digital resource of images (antweb.org), catalogue and taxonomic literature (antbase.org, antcat.org, plazi.org); and (8) a robust molecular phylogeny of ants to the genus level exists (Ward et al. 2010, 2015, 2016a, 2016b, Moreau and Bell 2013, Brady et al. 2014).

History of the database

The database was originally assembled by Dunn et al. (2007), who focused on total species richness and abundance of georeferenced ant assemblages. Several papers using that dataset and investigating species richness and abundance responses to climate have been published (Dunn et al. 2009, Weiser et al. 2010, Jenkins et al. 2011). More recently, the database was significantly expanded to include more studies and data on the abundance of individual species within each assemblage. This has allowed collaborators to explore questions related to the composition of species within local assemblages, for example by investigating the influence of the interactive effects of climate and disturbance on species richness and evenness (Gibb et al. 2015). Other research currently in progress using these data includes an examination of global drivers of the dominance-impoverishment rule (Arnan et al. in prep.), impacts of invasive species on assemblage structure and composition, and whether climate differentially affects ants belonging to different trophic groups (Sagata 2016).

Suggested links

Questions

The database will allow researchers to ask questions about the drivers of co-existence and diversity of ants at local scales and to separate the effects of different drivers of global change on species assemblages. Some questions worthy of investigation include:

- What are the key environmental drivers of assemblage structure and composition?
- Does the effect of global change drivers (climate, disturbance and invasive species) on species depend on trophic position, taxonomy or another trait?
- How do patterns of ant diversity and distribution compare with other better-studied taxa, such as plants and terrestrial vertebrates?
- What are the global-scale drivers of beta diversity for ants?
- What are the multi-scalar effects of spatial and temporal environmental heterogeneity on ant diversity?

Getting started

The database is presented as three separate csv files. These are the “Source”, “Localities” and “Observations” files, the contents of which are detailed in the Metadata section of this paper. Broadly, the Source file contains information about where the data came from, including the details of publications from which it was drawn or whether it is unpublished data, the Localities file includes a range of information about the studies, such as georeferencing, site descriptions, including disturbance status, details on trapping methods and simple measures of total abundance and total species richness for each locality, and the Observations file includes lists of species or morphospecies collected in each study and a measure of their abundance.

Files can be linked using the shared terms: Source and Localities files are linked by the “Source_ID” term, while Locality and Observation files are linked using the “Locality_ID” or “Source_ID” term, depending on the resolution required. All Observations data sets can be linked to Locations data by the “Source_ID” term, and most can be linked with the “Locality_ID” term. There are a few datasets for which ant assemblage details were not provided at the local scale and therefore cannot be linked with the “Locality_ID” term. Using the “Source_ID” link gives the assemblage for an entire study (several localities combined), while using the “Locality_ID” link gives the assemblage for individual localities, so is preferable for most questions. Not all sites given in the Locality file link to Observations data: for these sites, the data were extracted from papers that reported only total species
richness and abundance (but not species composition). Limitations of the data are detailed in the “Data limitations” section of this paper.

Metadata

Class I. Data set descriptors

A. Data set identity

Title: A global database of ant species abundances.

B. Data set and metadata identification codes

Suggested data set identity codes: global_ants_sources.csv, global_ants_localities.csv, global_ants_observations.csv.

C. Data set description

Investigators: same names and addresses as above.

Abstract: same as above.

D. Key words

Abundance, ants, database, disturbance, Formicidae, geo-referenced, habitat, local assemblage, occurrence, pitfall trap, Winkler trap

Class II. Research origin descriptors

A. Overall project description

Identity: local assemblage composition of ants

Originators: same names and addresses as above. Data were extracted from the literature or provided by co-authors; this project was part of an Australian Research Council grant (DP120100781) to HG, CLP, NJS and RRD.

Period of Study: 1949-2014
**Objective:** the aim of the study was to compile data detailing the abundance of ant species in local assemblages, i.e., co-occurring species. This data is ideal as a basis for studies investigating global drivers of the local–scale determinants of community structure.

**Abstract:** same as above.

**Sources of funding:**

Elena Angulo, Raphaël Boulay, and Xim Cerdá: Regional Government of Andalusia (Consejería de Medio Ambiente, 863/2004/M/00), and the Spanish MINECO and FEDER (CGL2012-36181).

Inge Armbrecht: Wildlife Conservation Society, GEA, Financiera Electrica Nacional and Universidad del Valle; International Institute of the University of Michigan, Universidad del Valle and Colciencias grant code project 1106-12-11693.

Xavier Arnan: CICYT project REN2001-2500/GLO to Anselm Rodrigo.

Tom R. Bishop: NERC (UK) and the CIB DST-NRF (SA).


Thibaut Delsinne & Maurice Leponce: Fonds pour la Formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA, Belgium) and the “Fonds National de la Recherche Scientifique” (FNRS, Belgium).

David A. Donoso: Institutional support and internal grants from the Instituto de Ciencias Biológicas and Museo de Historia Natural Gustavo Orcés at Escuela Politécnica Nacional (EPN).

Aaron M. Ellison: the US National Science Foundation (awards 0452254, 0541680, 1003938, and 1136646), the US Department of Energy (award DE-FG02-
08ER64510), the Massachusetts Natural Heritage and Endangered Species Program, the Nantucket Biodiversity Initiative, the Conservation Research Foundation, and the Museum of Comparative Zoology at Harvard University.

Tom M. Fayle: Czech Science Foundation Centrum of Excellence Grant (14-36098G) and Yayasan Sime Darby.


Matthew C. Fitzpatrick: The University of Maryland Center for Environmental Science.

Heloise Gibb: Swedish Research Council (Formas), Commonwealth Scientific and Industrial Research Organisation Office of the Chief Executive Postdoctoral Fellowship, Australian Research Council DP120100781.

Nicholas J. Gotelli: U.S. National Science Foundation Grants DEB-1257625, DEB-1144055, DEB-1136644.

Nihara Gunawardene: Curtin University Endeavour International Postgraduate Research Scholarship, the International Foundation for Science Grant No. D/3929-1 and the T.L. Higginson, Jr. Fund managed under the auspices of the CTFS-AA Asia Program.

Milan Janda: Czech Science Foundation (P505/12/2467), Marie Curie Fellowship (PIOFGA2009–25448).

Clinton Jenkins: NASA Biodiversity Grant (ROSES-NNX09AK22G)

Petr Klimes and Jimmy Moses: Czech Science Foundation PK (P505/12/P875), JM (14-36098G).
Lori Lach: EPA Science to Achieve Results (STAR) fellowship.


Sarah H. Luke: The UK Natural Environment Research Council (NERC), the University of East Anglia, Sir Philip Reckitt Educational Trust, and the Sime Darby Foundation.

Sean Menke: National Estuarine Research Reserve System Graduate Research Fellowship grant from NOAA (NA04NOS4200146); Pest Management Foundation; McHenry County Conservation District.

Dirk Mezger: Travel Grant, University of Wuerzburg, Germany

Dirk Mezger and Martin Pfeiffer: DFG (German Research Foundation) grant to Martin Pfeiffer PF441/3-1 and PF441/3-3.


Thinandavha Caswell Munyai: DST-NRF Center of Excellence for Invasion Biology and the University of Venda.

Stacy Philpott: Smithsonian Migratory Bird Center, University Research Award Fellowship Program of the University of Toledo.

Julian Resasco: NSF Graduate Research Fellowship (NSF-GRFP); USDA Forest Service-Savannah River, under Interagency Agreement with DOE (DE-AI09-00SR22188).
Javier Retana: Spanish ‘Ministerio de Ciencia e Innovacion’ (project Consolider MONTES, CSD 2008-00040).

Nathan J. Sanders: National Science Foundation Dimensions of Biodiversity grant (NSF-1136703) and acknowledges the support of the Danish National Research Foundation for its support of the Center for Macroecology, Evolution and Climate.

Rogerio R. Silva: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grants No. 2010/51194-1 and 2010/20570-8.

Jorge Souza & Fabricio B. Baccaro: PIPT/FAPEAM 1750/08; PNPD/CAPES 03017/19-05; FAPEAM 062.01325/ 2014; CNPq via PRONEX 16/2006, the Program for Biodiversity Research (PPBio) 558318/2009-6, 457545/2012-7; the Long Term Ecological Research (PELD) 403764/2012-2 and the Center for Integrated Studies of the Amazonian Biodiversity (CENBAM).

Andrew Suarez: National Science Foundation, NSF DEB-9610306.

Heraldo L. Vasconcelos: Brazilian Council of Research and Scientific Development (CNPq).

B. Specific subproject description

Site description: this data set comprises georeferenced local assemblage data on ant abundances from publications (i.e., 291 peer-reviewed manuscripts, 4 books, 16 theses and 11 reports or technical papers) and co-authors’ unpublished data (47 studies).

Experimental/Sampling design: all data were obtained from the literature (articles, books, theses, reports) and ecologists/myrmecologists. Sampling methods differed among studies and the details for each study are included in the
global_ants_localities.csv file. Examples of methods used are aerial nets, artificial nests, baits, Berlese, emergence, excavation, fogging/ insecticide spraying, foliage beating, hand collecting, interception, light traps, malaise, mini Winkler, pan traps, pitfall traps, soil sample, sticky traps, subterranean, sweeping, tray sifting, vacuum, window and Winkler. Pitfall trapping was the most common method used (3142 out of 4212 localities) but was often combined with other trapping methods.

**Research methods:** Dunn et al. (2007) compiled a global database of ant biodiversity studies, including details of species richness and abundance for local ant assemblages. The original database has not been made publicly available. This database extends Dunn et al. (2007) by including additional studies and by adding an ‘observations’ (global_ants_observations.csv) page, which details the abundance of individual species whereas the original database provided only abundance and species richness for the entire assemblage. The original database commenced with compilation of data from North American ant assemblages, followed by a larger-scale collaboration initiated at the International Union for the Study of Social Insects meeting in Washington D.C. in 2006. Collaborators compiled data from their own work and other studies. Further, we searched the Web of Science, Google Scholar and Formis (an ant literature database covering the 1800s through to 2003) using key words including FORMICIDAE + PITFALL, FORMICIDAE + LITTER, FORMICIDAE + TRANSECT and FORMICIDAE + DIVERSITY. Although we focussed on searches including the terms ‘pitfall’ (pitfall trapping) and ‘Winkler’ (Winkler litter sampling), we also included studies that used other methods, e.g., baiting and hand collection. We did not include studies that were limited to specific trophic groups, i.e., only studies sampling whole assemblages within a habitat were considered. For a limited
of datasets, repeated sampling at different times was provided as separate datasets for the same localities.

Class III. Data set status and accessibility

A. Status

Latest updates: the formal literature search for trait information ended in December 2010. Since that date, data has been collected through contributions from collaborators on an ongoing basis.

Latest archive date: current.

Metadata status: current.

Data Verification: data were double-checked for accuracy.

B. Accessibility

Storage location and medium: the original data files are held by the authors. This data set, published in Ecology, is the first public release of this data.

Contact person: queries about the entire data set or individual specific studies can be initially directed to Heloise Gibb, email: h.gibb@latrobe.edu.au or directly to the authors of individual studies (co-authors of this dataset).

Copyright and proprietary restrictions: none. When using the dataset, please cite this article to recognize the effort involved in gathering and collating the data and the willingness of the authors to make it publicly available.

Costs: none.

Class IV. Data structural descriptors

A. (1) Source file

Identity: global_ants_sources.csv

Size: 369 lines of source data, excluding header row.
**Format and storage mode:** comma-delimited, no compression.

**Header information:** column headers contain character formatted labels for source data. Each column has a unique column header e.g. Source ID, Author, Contributor, Source Type, Year of Publication and Source citation.

**Alphanumeric attributes:** mixed.

**Special characters/fields:** none.

(2) **Localities data set file**

**Identity:** global_ants_localities.csv

**Size:** 4455 lines of data, excluding header row. Overall, this data set contains 4212 georeferenced locations where abundance or occurrence and species richness of ant assemblages were recorded. Not all locality data has associated observed data as data that was extracted from manuscripts sometimes contained only abundance and species richness for the entire assemblage and no data on the abundance/occurrence of individual species. Additional lines of data for the same locality appear if different trapping methods or data from different sampling times were kept separate.

**Format and storage mode:** comma-delimited, no compression.

**Header information:** column headers contain character formatted labels for locality data. See ‘locality table’ below for details.

**Alphanumeric attributes:** mixed.

**Special characters/fields:** none.

(3) **Observations data set file**

**Identity:** global_ants_observations.csv
Size: 51,388 lines of data, excluding header row. Overall, this data set contains 51,388 individual observations of abundance or occurrence for 2693 species and 7953 morphospecies.

Format and storage mode: comma-delimited, no compression.

Header information: column headers contain character formatted labels for observation data. See ‘observations table’ below for details.

Alphanumeric attributes: mixed.

Special characters/fields: none.
<table>
<thead>
<tr>
<th>Column header</th>
<th>Variable definition</th>
<th>Unit</th>
<th>Data storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source ID</td>
<td>Unique identification name used to link the associated locality, observed and traits data with the source</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Authors</td>
<td>Author/s of the article or researchers responsible for the collection of the data (for unpublished data)</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Institution of primary author</td>
<td>The institution (university, museum, government department) associated with the primary author of the source</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Contributor</td>
<td>Person/ people who submitted the data for entry into the Global Ants Database</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Source Type</td>
<td>Location of original data e.g. Published manuscript, Unpublished data, Book chapter, Master's thesis, PhD thesis, Report</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Year of Publication</td>
<td>The year the data was published (&quot;no year&quot; for unpublished data)</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Source Citation</td>
<td>Complete citation of data source e.g. APA format for published manuscripts, books, reports and theses, and author and date of data collection for unpublished data</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Column header</td>
<td>Variable definition</td>
<td>Unit</td>
<td>Data storage</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Source ID</td>
<td>Link to source data</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Locality ID</td>
<td>Unique identification name used to link the locality data with a specific location and to a source</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Contributor</td>
<td>Person/people who submitted the locality data for entry into the Global Ants Database</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Continent</td>
<td>Continent where the data was collected e.g. Oceania</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Political Region 1</td>
<td>Country where the data was collected e.g. Australia</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Political Region 2</td>
<td>State/region where the data was collected e.g. Victoria</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Political Region 3</td>
<td>Name of local area where the data was collected e.g. Wilson's Promontory National Park</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Locality Name</td>
<td>Locality name assigned by researcher/as described in publication, e.g. Wilson's Promontory Closed Forest J</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Locality Description</td>
<td>Description of the location e.g. Grazed pasture, Savanna, Secondary forest</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Elevation</td>
<td>Elevation</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Latitude</td>
<td>Coordinates</td>
<td>decimal degrees</td>
<td>Numerical</td>
</tr>
<tr>
<td>Longitude</td>
<td>Coordinates</td>
<td>decimal degrees</td>
<td>Numerical</td>
</tr>
<tr>
<td>Location Source</td>
<td>Source of locality position information (elevation, latitude, longitude)</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Accuracy of coordinates</td>
<td>Accuracy of the provided coordinates in metres</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Total Ant Abundance</td>
<td>Total ant abundance per locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Total Ant Species Richness</td>
<td>Total ant species richness per locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Total Native Species Richness</td>
<td>Total native ant species richness per locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Column header</td>
<td>Variable definition</td>
<td>Unit</td>
<td>Data storage</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Total Non Native Species Richness</td>
<td>Total non native ant species richness per locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Number Of Sampling Events</td>
<td>The number of separate data collection events e.g. 2, if sampled over two separate seasons</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Start Date (dd)</td>
<td>Day of the month when data collection began (dd)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Start Date (mm)</td>
<td>Month when data collection began (mm)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Start Date (yyyy)</td>
<td>Year when data collection began (yyyy)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>End Date (dd)</td>
<td>Day of the month when data collection ended (dd)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>End Date (mm)</td>
<td>Month when data collection ended (mm)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>End Date (yyyy)</td>
<td>Year when data collection ended (yyyy)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Total Number Of Transects</td>
<td>Total number of transects or plots per locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Length</td>
<td>Length of transects/plots</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Width</td>
<td>Width of transects/plots</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Plot Transect Separation</td>
<td>The distance between plots or transects</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Method</td>
<td>Method used to collect ant specimens e.g. Pitfall, Baits, Winkler, Hand collecting, Vacuum, Arboreal baits</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Method Description</td>
<td>Written description of the method/procedure used to collect ant specimens</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Pitfall Number</td>
<td>Number of pitfall traps (total for location per sampling event)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Pitfall Surface Area</td>
<td>Surface area of a single pitfall trap open face</td>
<td>centimetres$^2$</td>
<td>Numerical</td>
</tr>
<tr>
<td>Pitfall Spacing</td>
<td>Distance between pitfalls</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Pitfall Duration</td>
<td>Total duration pitfall traps were exposed i.e. combined total of all sampling events</td>
<td>hours</td>
<td>Numerical</td>
</tr>
<tr>
<td>Baits Number</td>
<td>Number of baits (total for location per sampling event)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Bait Stations Number</td>
<td>Number of bait stations (total for location per sampling event)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Baits Spacing</td>
<td>Distance between baits/bait stations</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Column header</td>
<td>Variable definition</td>
<td>Unit</td>
<td>Data storage</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>--------------</td>
</tr>
<tr>
<td>Baits Duration</td>
<td>Total duration baits/bait stations were exposed i.e. combined total of all sampling events</td>
<td>hours</td>
<td>Numerical</td>
</tr>
<tr>
<td>Winkler Number</td>
<td>Number of Winkler traps (total for location per sampling event)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Winkler Spacing</td>
<td>Distance between Winkler traps</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Berlese Number</td>
<td>Number of Berlese traps (total for location per sampling event)</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Berlese Spacing</td>
<td>Distance between Berlese traps</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Total Litter Sample Volume</td>
<td>Volume of individual litter samples</td>
<td>litres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Total Litter Sample Area</td>
<td>Area of individual litter samples</td>
<td>metres</td>
<td>Numerical</td>
</tr>
<tr>
<td>Habitat Type</td>
<td>Habitat type e.g. closed canopy forest</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Disturbance</td>
<td>Category of habitat disturbance e.g. Transformed, Disturbed, Undisturbed</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Disturbance Type 1</td>
<td>Description of type of habitat disturbance e.g. Agriculture, Cropping, Fire, Mining</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Disturbance Type 2</td>
<td>Description of second type of habitat disturbance e.g. Agriculture, Cropping, Fire, Mining</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Habitat Description</td>
<td>Written description of the habitat where data was collected</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Notes</td>
<td>Additional relevant information that is not appropriate for the determined column headers</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td><strong>Observed table</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Column header</strong></td>
<td><strong>Variable definition</strong></td>
<td><strong>Unit</strong></td>
<td><strong>Storage type</strong></td>
</tr>
<tr>
<td>Observed ID</td>
<td>Unique identifier for each species observed at each Locality ID</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Source ID</td>
<td>Links the Locality data with the Source data</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Locality ID</td>
<td>Links the observed data to the Locality data</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Contributor</td>
<td>Name of the person/people who submitted the observed data to the Global Ants Database</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Genus</td>
<td>Genus name</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Species</td>
<td>Species name</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Morphospecies</td>
<td>Morphospecies name as designated by contributor</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Measurement Type</td>
<td>Type of abundance measure e.g. Abundance, Occurrence, Other</td>
<td>N/A</td>
<td>Character</td>
</tr>
<tr>
<td>Abundance</td>
<td>The number of individuals of each species per locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Occurrence</td>
<td>The number of traps in which a species occurred at each locality</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Other</td>
<td>Data not in the form of abundance or occurrence e.g. frequency</td>
<td>N/A</td>
<td>Numerical</td>
</tr>
<tr>
<td>Notes</td>
<td>Additional relevant information that does not fit into a determined column header</td>
<td>N/A</td>
<td>Character</td>
</tr>
</tbody>
</table>
B. Data limitations

Studies were conducted over varying timeframes, at varying scales and with a range of sampling methods and efforts, all of which differ in the completeness in which they sample assemblages. These elements of the dataset are documented so it is possible to use measures of scale and trapping effort as covariates in analyses. In some cases different trapping methods were pooled, but we have made an effort to keep different trapping methods separate as often as possible. Similarly, different sampling events at the same locality were sometimes kept separate.

There are limitations associated with sampling workers of colonial insects. As noted previously (Gotelli et al. 2011), estimating ant abundance from the number of workers collected in traps can be problematic. Ants vary in colony size, foraging and recruitment behaviour and these (along with other factors) will influence the number of individuals found in a pitfall trap or Winkler sample. While the ideal ecological estimate of abundance would include the number of nests or colonies per unit area or sampling effort, the number of workers collected still holds value as a measure of relative abundance or indication of species-specific activity during the period of sampling. For guidelines on how to analyse abundance or occurrence data based on worker capture, see Gotelli et al. (2011).

Methods of recording abundances include mainly the total count of individuals (abundance) and the count of traps in which a species occurs (occurrence), but some studies also report abundances using less common metrics, including: mean abundance, proportion of occurrence, presence/absence, adjusted abundance, frequency of occurrence, number of nests (nest excavation studies), and categorised occurrence, e.g., present in 5 or fewer traps, present in more than 5 traps. Spatial coverage is somewhat unbalanced, with very few studies from Asia and north Africa.
Additionally, ants are poorly described in some parts of the world and in those areas, morphospecies, rather than species, are commonly used in community analyses. Limitations associated with the use of morphospecies should be considered when using the dataset. These sources of variation may limit the analyses that the data can be used for.

Class V. Data set references

These are provided in the global_ants_sources.csv file, either as publications or as authors and institutions for unpublished data.

Acknowledgements

We thank the Australian Research Council for funding this work (DP120100781 to HG, CLP, NJS, RDD). Additional support was provided by US Department of Energy PER (DE-FG02-08ER64510) and US National Science Foundation (NSF 1136703) to NJS and RRD.

Funders for collection of individual data sets are listed in Class II A above.

Literature cited


Sagata, K. 2016. Climate effects on ant-Hemiptera interactions, ant richness and morphology. PhD thesis, La Trobe University, Melbourne, Australia.


