Digital supplementary material to

GUÉNARD, B., WEISER, M.D., GÓMEZ, K., NARULA, N. & ECONOMO, E.P. 2017: The Global Ant Biodiversity Informatics (GABI) database: synthesizing data on the geographic distribution of ant species (Hymenoptera: Formicidae). – Myrmecological News 24: 83-89.

Supplement S1: Supplementary information on data compilation and structure.

In this section we present more details on data compilation and structure of the database. Finally, an example is taken to illustrate the different taxonomic levels and use of the database to map taxa distribution.

Compilation of data from literature and databases: The literature search has been particularly exhaustive and represents the core contribution of GABI. The literature search on ant records started in 2007, as part of work on ant genera distribution (GUÉNARD & al. 2009, 2012) which allowed the gathering of thousands of publications on ant taxonomy and distribution. The publications retrieved were first compiled into the GABI database. Yet, those publications mainly focused on generic records did not cover the entire spectrum necessary for a species-level database. Then, a second literature search from 2012 to 2016 was performed extracting all articles with species distribution data using several search engines including Google Scholar, Scopus and CNKI with one or a combination of the following keywords: "Formicidae" or "ants" or "hormigas" or "fourmis" or "formigas" or "蚂蚁" (Chinese) or "アリ" (Japanese) or "муравьи" (Russian) alone and in combination with one of the 592 geographic regions (relevant to the language) used in our study. Names of genera were also used as keywords to retrieve publications. Those are extremely efficient, as they do not conform to a particular language. Furthermore to cover all the publications produced in taxonomy, all the publications known for each author referenced in BOLTON (2015) and ANTWIKI (2015) were searched through.

To ensure a complete coverage of the published literature, a separate database with all the publications included was built to verify the completion status for each author publishing on ants. For authors with a personal webpage, or with an account on ResearchGate.net, or Academia.edu presenting their publications list, we verified that all their relevant publications had been entered into our database. When the publication could not be gathered directly, the authors were contacted to obtain copies. Hundreds of e-mails have been sent and we warmly thank all the scientists who responded positively to our request (see also acknowledgements section).

Entire journal publications in entomology or myrmecology have been searched one volume at a time to extract all the data on ant distribution contained. Those, for instance, include all the journals presented in Scielo.br, and Ari (Japanese), Asian Myrmecology, Boletin Sociedad Entomologica Aragonesa, CheckList, Dugesiana, Florida Entomologist, Myrmecological News, Zookeys, Zootaxa. Finally, several books were also purchased and used to include all published records. Information from some pertinent websites, such as ANTAREA (2015), was also extracted directly.

Besides publications on taxonomy, biogeography, and ecology, other publications in behavioral or chemical ecology were compiled when judged relevant and accurate. For publications that could not be accessed through academic libraries, authors were contacted directly. However, despite these efforts, a list of 106 publications (or 1.2%) could not be obtained. Language barriers impose a second important limitation. For this project publications in English (the majority), Chinese, Czech, Dutch, French, German, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, Thai, Turkish, and Ukrainian have been incorporated. Particular attention has been focused on the translation of the information from the abundant Japanese and Chinese literature with the help of native speakers. At this point though the main remaining gaps are the older Russian and German literature which have only been partially processed and should gain more attention in the future. Literature published in specific languages such as Russian or German could benefit from volunteers willing to include it into the database.

Data from ANTWEB (2015) were extracted on February 2014 using the "Advanced Search" function for each subfamily; with a specific search by biogeographic regions for Formicinae and Myrmicinae due to the high number of data available. For the data from external databases, only the information needed in GABI (see Supplement S2) was kept and formatted to conform its specificity.

Compilation of morphospecies information: Numerous species records only indicate a generic identification with a more or less specific morphospecies code associated. The morphospecies code as presented in the original publication was preserved in order to keep the possibility of future updates once new species are being described or revised. For instance, records reported as *Pristomyrmex* sp. 13 SKY in EGUCHI & al. (2005) were later identified as *Pristomyrmex sulcatus* EMERY, 1895 by EGUCHI & al. (2011) and can thus be updated in the database. When an additional taxonomic level such as subgenus, species group or species complex was provided, then these information was kept but as a morphospe-

cies name entry; e.g., *Aphaenogaster* sp. (*rudis* complex). The database includes a total of 346,935 morphospecies entries. While those morphospecies records can be difficult to assign to a specific described species (at the exception of the few cases described above), they still hold some taxonomic values. First, they provide information about the distribution and presence of a genus in a particular location (and associated variables such as elevation). However for some recently revised and "split" genera (e.g., *Amblyopone*, *Aphaenogaster*, *Monomorium*, *Pachycondyla*, *Paratrechina* …), the direct assignment to a specific genus might be impossible for some regions. For instance, *Paratrechina* sp. in a region of South East Asia could belong to the genus *Paratrechina*, *Paraparatrechina* or *Nylanderia*. As a positive counterpart, morphospecies from "lumped" genera can be resolved. For example, morphospecies presented as *Pyramica* sp. or *Smithistruma* sp. can be transferred to *Strumigenys* confidently. Knowledge on the taxonomic history of the different genera is thus required before the use of those data. A second reason to keep these records relate to the species richness observed in a given genus at a specific locality. Even with incomplete identification, those records still hold ecological value on the expected number of species that could be retrieved in a region. A good example is provided with the genus *Pheidole*, for which species are often left unidentified in many ecological studies conducted in tropical regions, where diversity of this genus peaks (ECONOMO & al. 2015). In those studies, it is common to observed more than 10 species of *Pheidole* being reported in a single study, but only two or three species identified. The exclusion of the morphospecies for macroecological studies will thus largely underestimate species richness. It should be kept in mind that for most cases, morphospecies codes used in one study cannot be directly transposed for a different study (e.g., *Pheidole* sp. 1 in the study "A" can represent a totally different species than *Pheidole* sp. 1 in a study "B" conducted in the same location). Thus, with the limitation of keeping each study separated, the number of morphospecies is still informative on local assemblages' composition and richness.

Geographic scale used for data display / analyses: The collection and sampling of ants over history was an idiosyncratic process that resulted in large disparities in the sampling effort (GUÉNARD & al. 2012) and variation in methods for recording geographic location (varying from a country name to a full georeferenced locality, see above). As a consequence, the sum of records collected over the past 200 years has to be displayed within coherent but non-equivalent units that reflect this idiosyncrasy but mediate some of its biases (e.g., differential sampling effort). This approach was previously considered in the geographic unit used by GUÉNARD & al. (2012), with the use of regions that both reflect human-limitations (e.g., country level; first administrative level), geology (islands) and scientific knowledge (clumping or split of specific political entities such as in Borneo [lumping Brunei with Malaysian and Indonesian Borneo] or Western Australia [splitting into North Western Australia and South Western Australia at 26.55°S]). The geographic units used for displaying the information or for analysis are also project-specific: we named those B e n t i t y (plural: B e n t i t i e s).

Those units correspond to constraints inherent to the literature as well as the geographic features of certain countries or islands. When possible, large countries have been divided according to their first administrative level (e.g., China, Japan, USA). Ideally, the regions considered should tend to use the smallest geographical unit possible, however the sampling effort and taxonomic knowledge for each region represent the main limitations. Furthermore, for global comparison, it is important to limit the size difference between the largest and the smallest regions. We present two maps showing all the regions considered according to our B e n t i t y system 2 and 3 in our study (Supplement S5), but it should be kept in mind that the database, through the m i n i m u m g e o g r a p h i c l e v e l o f i n f o r m a t i o n p r o v i d e d, allows the separation of existing regions into smaller ones or at the opposite to clump several regions together. The B e n t i t y system 2 and 3 include 592 and 417 regions respectively (versus 353 regions in GUÉNARD & al., 2012). The political scale, while not optimal for biological surveys, has for long been used for large scales studies for other groups of insects, such as Carabid and Cicindelidae beetles (CASSOLA & PEARSON 2000; SCHULDT & ASSMANN 2011) or mosquitoes (FOLEY & al., 2007); but also mosses (MUTKE & BARTHLOTT 2005) or liverwort (VON KONRAT & al. 2008) and is also useful for comparison between different taxonomic groups for which the scale at which the data are available can present important variation (e.g., JENKINS & al. 2013).

The major differences between the B e n t i t y system 2 and 3 lies in the consideration of Colombia, Mexico and India at their first administrative division in the B e n t i t y system 2, versus a clumping method between neighboring regions of first administrative divisions in the B e n t i t y system 3 to form larger more inclusive regions. Similarly, smaller groups of islands considered individually in B e n t i t y 2 have been clustered together in B e n t i t y 3.

Here we briefly discuss a few specific cases as shown on Supplement S5a, b. Colombia was divided in five regions (the Amazonian, Andean, Caribbean, Orinoquio & Pacific regions) which represent the clumping of several departments and tend to match the geographic regions of the country with relative similar separations used previously (e.g., CHACON DE ULLOA & ABADIA 2014); Indonesia was separated based on the main islands (Java, Sulawesi & Sumatra) and islands groups (Maluku islands, Lesser Sunda Islands), while finally regions present on islands but shared with other countries were considered as a single island (e.g., Borneo and New Guinea). Finally in both systems Western Australia was divided in a northern and southern part.

The geographic information retained in the GABI database includes above t h e m i n i m u m g e o g r a p h i c l e v e l o f i n f o r m a t i o n p r o v i d e d (see above): country, 1st administrative level, 2nd administrative level, latitude and longitude, decimal latitude and longitude (all records are ultimately converted in this format), and for records from insular regions: island name (e.g., Tanegashima Island), island archipelago (e.g., Satsunan Islands), and island group (e.g., Osumi Islands), and elevation (when provided) in the unit used within the publication (feet or meters) and then ultimately converted in meters (in a separate column) (more details provided in Supplement S2).

An example

An example is presented in Supplement S6 as well as on antmaps.org (http://antmaps.org/?mode=species&species=Pheidole.megacephala). Several levels of information related to the record history are presented in these maps. Here, we chose to introduce an example with *Pheidole megacephala* (FABRICIUS, 1793), which represents all the different levels of "complexity" in the history of its records. *Pheidole megacephala* is an exotic species introduced in many regions of the world but suspected to be native to the Afrotropical bioregion without a known specific region being defined (WETTERER 2012). Here, we considered the Malagasy region as part of its introduced range (BLAIMER & al. 2015), but the species could potentially be native there as well (WETTERER 2012; FISCHER & FISHER 2013). As such, its recorded native distribution in the Afrotropical region is presented in blue. This species has been recorded in many different continental or insular regions around the world outside of its native range. Some of these records have been reported from quarantine or from heated buildings (e.g., greenhouses) without any evidence of populations established outside. Those "indoor or interception" records are presented in green on the map. Other populations of this species are known from established outdoor populations and are presented on the maps as red. Finally, historically, some records for this species have been reported for specific regions (e.g., Portugal & Spain), but reexamination of the specimens has revealed misidentifications (ESPADALER & COLLINGWOOD 2000). Those records, called "dubious", are presented in brown. A last type of record, not presented here on Supplement S6, exists in our database for records presenting an anomalous distribution from the rest of the known native range (e.g., disjunction) but for which we did not find specific reference arguing against it and for which the climatic envelope could still support populations based on other known records.

As illustrated by the example above, the taxonomic and biogeographic knowledge of a given taxon is changing over time as information and comprehension of its biology are accumulating. As a result, keeping track of the published literature as well as communication with regional and/or taxonomic experts is key to maintain the data as accurate as possible. These problems have been illustrated well recently for large and inclusive databases such as GBIF (2016) that does not provide a critical curation of the records gathered. As a result, an important percentage of erroneous taxa and biogeographic records have accumulated over time weakening the interest for such large databases and their use in research (GOODWIN & al. 2015, MALDONADO & al. 2015).

References

ANTAREA 2015: AntArea. Etude, identification, répartition, localisation des fourmis françaises métropolitaines. – <http://antarea.fr/fourmi>, retrieved on 4 February 2015.

ANTWEB 2015: AntWeb. – <http://www.antweb.org>, retrieved on 16 May 2015.

ANTWIKI 2015: AntWiki. – <http://www.antwiki.org/wiki/Welcome_to_AntWiki>, retrieved on 16 May 2015.

BLAIMER, B.B., BRADY, S.G., SCHULTZ, T.R. & FISHER, B.L. 2015: Functional and phylogenetic approaches reveal the evolution of diversity in a hyper diverse biota. – Ecography 38: 901-912.

BOLTON, B. 2015: An online catalog of the ants of the world. – <http://www.antcat.org>, retrieved on 17 May 2015.

CASSOLA, F. & PEARSON, D.L. 2000: Global patterns of tiger beetle species richness (Coleoptera: Cicindelidae): their use in conservation planning. – Biological Conservation 95: 197-208.

CHACON DE ULLOA, P. & ABADIA, J.C. 2014: Two decades of study of the diversity of ants in Colombia. – Revista de la Academia Colombiana de Ciencias Exactas, Fisicas y Naturales 38: 250-260.

EGUCHI, K., BUI, T.V., YAMANE, SK., OKIDO, H. & OGATA, K. 2005: Ant fauna of Ba Vi and Tam Dao, N. Vietnam (Insecta: Hymenoptera: Formicidae). – Bulletin of the Institute of Tropical Agriculture, Kyushu University 27: 77-98.

EGUCHI, K., VIET, B.T. & YAMANE, SK. 2011: Generic synopsis of the Formicidae of Vietnam (Insecta: Hymenoptera), part I – Myrmicinae and Pseudomyrmecinae. – Zootaxa 2878: 1-61.

ESPADALER, X. & COLLINGWOOD, C.A. 2000: Transferred ants in the Iberian Peninsula (Hymenoptera, Formicidae). – Nouvelle Revue d'Entomologie 17: 257-263.

FISCHER, G. & FISHER, B.L. 2013: A revision of *Pheidole* WESTWOOD (Hymenoptera: Formicidae) in the islands of the Southwest Indian Ocean and designation of a neotype for the invasive *Pheidole megacephala*. – Zootaxa 3683: 301-356.

FOLEY, D.H., RUEDA, L.M. & WILKERSON, R.C. 2007: Insight into global mosquito biogeography from country species records. – Journal of Medical Entomology 44: 554-567.

GOODWIN, Z.A., HARRIS, D.J., FILLER, D., WOOD, J.R.I. & SCOTLAND, R.W. 2015: Widespread mistaken identity in tropical plant collections. – Current Biology 25: 1057-1069.

GUÉNARD, B., WEISER, M.D. & DUNN, R.R. 2009: Ant genera of the world. – < http://robdunnlab.com/antgenera/ >, retrieved on 19 May 2015.

GUÉNARD, B., WEISER, M.D. & DUNN, R.R. 2012: Global models of ant diversity suggest regions where new discoveries are most likely are under disproportionate deforestation threat. – Proceedings of the National Academy of Sciences of the United States of America 109: 7368-7373.

JENKINS, C., GUÉNARD, B., DIAMOND, S.E., WEISER, M.D. & DUNN, R.R. 2013: Conservation implications of divergent global patterns of ant and vertebrate diversity. – Diversity and Distribution 19: 1084-1092.

MALDONADO, C., MOLINA, C.I., ZIZKA, A., PERSSON, C., TAYLOR, C.M., ALBÁN, J., CHILQUILLO, E., RØNSTED, N. & ANTONELLI, A. 2015: Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? – Global Ecology and Biogeography 24: 973-984.

MUTKE, J. & BARTHLOTT, W. 2005: Patterns of vascular plant diversity at continental to global scales. – Biologiske Skrifter 55: 521-537.

SCHULDT, A. & ASSMANN, T. 2011: Patterns and hotspots of carabid beetle diversity in the Palaearctic: insights from a hyperdiverse invertebrate taxon. In: ZACHOS, F.E. & HABEL, J.C. (Eds.): Biodiversity hotspots: distribution and protection of conservation priority areas. – Springer-Verlag, Berlin, Heidelberg, pp. 175-188.

VON KONRAT, M., HAGBORG, A., SÖDERSTRÖM, L., MUTKE, J., RENNER, M., GRADSTEIN, S.R., ENGEL, J., ZHU, R.-L. & PICKERING, J. 2008: Early land plants today: global patterns of liverwort diversity, distribution, and floristic knowledge. In: MOHAMED, H., BAKI, B.B., NASRULHAQ-BOYCE, A. & LEE, P.K.Y. (Eds.): Bryology in the new millennium. – University of Malaya, Kuala Lumpur, pp. 425-438.

WETTERER, J.K. 2012: Worldwide spread of the African big-headed ant, *Pheidole megacephala* (Hymenoptera: Formicidae). – Myrmecological News 17: 51-62.

Supplement S2: Different categorical variables available in the GABI database. Categories with * represent the mandatory information to a record to be included in the database.

Category's name	Function
Genus name in publication*	Genus name as presented in publication
Species name in publication*	Species name as presented in publication
Valid genus*	Current valid genus name
Valid species*	Current valid species name
Exotic	If species record is presented as exotic in the given region in the publication
Locality*	This represents the minimum geographic level of information provided (see text)
Island	For localities located on islands (with the exclusion of freshwater, lake and river islands)
Source*	Reference to the source of a record (e.g., full publication records, database)
Latitude	Latitude value for a locality expressed in degrees
Decimal Latitude	Latitude value for a locality expressed in decimal degrees
Longitude	Longitude value for a locality expressed in degrees
Decimal Longitude	Longitude value for a locality expressed in decimal degrees
Country	Name of the country where the record was collected (if available)
First administrative level	Name of the first level of administrative division of the country where the record was collected (if available)
Second administrative level	Name of the second level of administrative division of the country where the record was collected (if available)
Island division 1	For insular records, name of a larger group of islands that includes the island where the record was collected (e.g., Satsunan Islands)
Island division 2	For insular records, name of a smaller group of islands, included within the island division 1, that includes the island where the record was collected (e.g., Tokara Islands)
Elevation	Elevation at which the species was collected and as presented in publication or database
Notes	Any notes relative to the identification of a record
Notes-2	Notes relative to the change of status of a record
Subfamily	Subfamily to which the species belongs
Accession number*	Each entry received a specific and unique record within the GABI database which includes its origin if from a database (e.g., AntWeb) in addition to the existing accession number within the database; or if entered from publication: the name of the person who entered the data in addition to a unique number.
Type of data*	Indicates if the record has a publication (Literature), a database, or a collection for source
Dubious record	Indicates if a record is dubious followed in parenthesis of the justification of this status with a specific reference or information on personal communication.
Bentity-2	Geographic units used to map individual species distribution
Bentity-3	Geographic units used to compile species richness for all regions of the globe
Exotic species	Final status on introduced species. This can be Exotic for non-native populations established in nature, or Indoor Introduced if known only from populations collected within buildings or quarantine but not from wild populations.
Altitude (in m)	Elevation of the record expressed in meters

Supplement S3: List of databases created and extracted with sources of access and number of records included in GABI.

Data name	Access	# records
GABI – Literature records	GABI – Literature (BG)	761,956
AntWeb	http://www.antweb.org/	475,798
iDigBio	www.idigbio.org	120,669
INBio Collection	http://www.gbif.org/ (extracted from GBIF)	50,644
Australian National Insect Collection (ANIC)	Provided by Dr. Steve Shattuck	45,140
Bees, Wasps and Ants Recording Society	http://www.gbif.org/ (extracted from GBIF)	43,294
Dr. William MacKay database	Provided by Dr. William MacKay	30,732
Formidabel Data	http://ipt.biodiversity.be/resource.do?r=formidabel	27,234
Museum of Comparative Zoology, Harvard University	http://www.gbif.org/ (extracted from GBIF)	26,179
Australian Museum	OZCAM, http://ozcam.org.au/	13,150
Zoological Museum, Natural History Museum of Denmark	http://www.gbif.org/ (extracted from GBIF)	12,147
Instituto de Ciencias Naturales de la Universidad Nacional de Colombia	http://www.gbif.org/ (extracted from GBIF)	12,127
Triplehorn Insect Collection (OSUC), Ohio State University	http://www.gbif.org/ (extracted from GBIF)	11,578
Instituto Nacional de Pesquisas de Amazonia	http://www.gbif.org/ (extracted from GBIF)	11,776
Dr. Corrie Moreau Database	provided by Dr. Corrie Moreau	10,973
Field Museum of Natural History	http://www.gbif.org/ (extracted from GBIF)	10,060
Paraguayan dry Chaco, Royal Belgium Institute of Natural Sciences	http://www.gbif.org/ (extracted from GBIF)	8,846
Zoologisches Forschungsinstitut und Museum Alexander Koenig	http://www.gbif.org/ (extracted from GBIF)	7,435
Zoologisches Forschungsinstitut und Museum Alexander Koenig Hymenoptera collection	http://www.gbif.org/ (extracted from GBIF)	7,173
Dr. Simon Robson database	Provided by Dr. Simon Robson	6,762
Dr. Robert Johnson Database	Provided by Dr. Robert Johnson	5,877
ArtDatabanken Bugs	http://www.gbif.org/ (extracted from GBIF)	4,603
Museo de Entomología de la Universidad del Valle	http://www.gbif.org/ (extracted from GBIF)	3,929
UAM Entomology Collection (Arctos)	http://www.gbif.org/ (extracted from GBIF)	3,760
Dr. Legakis Ant Collection	Provided by Dr. Christos Georgiadis	2,365
Dr. Alberto Tinaut Database	http://www.gbif.org/ (extracted from GBIF)	1,416
CANADENSYS Data	http://data.canadensys.net/	1,185

Supplement S4: Complete reference list for data extracted from the Global Biodiversity Information Facility system (GBIF 2016).

Museum of Comparative Zoology, Harvard University (downloaded on 6 February 2014)	Museum of Comparative Zoology, Harvard University (2012): Museum of Comparative Zoology, Harvard University. Dataset/Occurrence. http://digir.mcz.harvard.edu/ipt/resource?r=mczbase doi:10.15468/p5rupv
INBio Collection (downloaded on 6 February 2014)	National Biodiversity Institute (INBio) of Costa Rica. (2001 -). Insecta occurrence data of Costa Rica. 3276500 records, Online, http://atta2.inbio.ac.cr
Bees, Wasps and Ants Recording Society (downloaded on 6 February 2014)	UK National Biodiversity Network: Bees, Wasps and Ants Recording Society – Bees, Wasps and Ants Recording Society – Trial Dataset. doi:10.15468/vfqcwq Accessed via http://www.gbif.org/dataset/4392dba3-3e44-42aa-8878-19d7b9819bdf
Zoological Museum, Natural History Museum of Denmark (downloaded on 6 February 2014)	Zoological Museum, Natural History Museum of Denmark: Entomology at ZMUC, Natural History Museum of Denmark. doi:10.15468/nnobcm Accessed via http://www.gbif.org/dataset/e0459621-92c2-40ac-9934-72b3b1384dc3
Instituto de Ciencias Naturales de la Universidad Nacional de Colombia (downloaded on 6 February 2014)	Instituto de Ciencias Naturales: Instituto de ciencias naturales. doi:10.15468/ddtrfz Accessed via http://www.gbif.org/dataset/dce681d3-4bc6-4d78-bb66-b4dd73aec081
Triplehorn Insect Collection (OSUC), Ohio State University (downloaded on 6 February 2014)	Museum of Biological Diversity, The Ohio State University: C.A. Triplehorn Insect Collection (OSUC), Ohio State University. doi:10.15468/efb17f Accessed via http://www.gbif.org/dataset/84ab7b76-f762-11e1-a439-00145eb45e9a
Instituto Nacional de Pesquisas de Amazonia (INPA) (downloaded on 23 December 2013)	Instituto Nacional de Pesquisas da Amazônia - INPA: Hymenoptera Collection – Instituto Nacional de Pesquisas da Amazônia (INPA). doi:10.15468/eupzne Accessed via http://www.gbif.org/dataset/c001a28f-6f37-40e9-8e34-e0ada1772e4d
Field Museum of Natural History (Zoology) Insect, Arachnid and Myriapod Collection (downloaded on 6 February 2014)	Field Museum: Field Museum of Natural History (Zoology) Insect, Arachnid and Myriapod Collection. doi:10.15468/0ywfpc Accessed via http://www.gbif.org/dataset/7931dcab-94f1-46ce-8092-56e4335423de
Paraguayan dry Chaco, Royal Belgium Institute of Natural Sciences (downloaded on 6 February 2014)	Belgium Biodiversity Platform: Ants from the Paraguayan dry Chaco. doi:10.15468/c0zp7k Accessed via http://www.gbif.org/dataset/0b0bc0f0-bfad-11dd-aa16-b8a03c50a862
Zoologisches Forschungsinstitut und Museum Alexander Koenig (downloaded on 7 February 2014)	Zoologisches Forschungsinstitut und Museum Alexander Koenig: Hymenoptera. doi:10.15468/dsd416 Accessed via http://www.gbif.org/dataset/d45b8e8a-1f32-40b0-a132-bd690bd90b30
ArtDatabanken Bugs (downloaded on 6 February 2014)	ArtDatabanken: Artdata. doi:10.15468/kllkyl Accessed via http://www.gbif.org/dataset/38b4c89f-584c-41bb-bd8f-cd1def33e92f
Museo de Entomología de la Universidad del Valle (downloaded on 6 February 2014)	Base de datos de la colección de insectos del Museo de Entomología de la Universidad del Valle, 30171 Registros, Contribuidos por Posso-Gomez CE (Publicador, Creador del Recurso, Proveedor de los Metadatos), Salas C (Procesador), Canacuán DE (Procesador), En linea, http://ipt.sibcolombia.net/valle/resource.do?r=insectos-universidad-del-valle, Versión 2.0 (actualizado por última vez el 11/02/2013), http://ipt.sibcolombia.net/valle/
UAM Entomology Collection (Arctos) (downloaded on 6 February 2014)	University of Alaska Museum Insect Collection (UAM), University of Alaska Fairbanks
Dr. Tinaut Database (downloaded on 4 February 2014)	Dept. of Zoology, Faculty of Science, University of Granada: Colección Alberto Tinaut (Formicidae). doi:10.15468/tvk5ua Accessed via http://www.gbif.org/dataset/7a5996ee-f762-11e1-a439-00145eb45e9a

Supplement S5: Global map of the different regions used for the characterization of B e n t i t y 2 (a) and B e n t i t y 3 (b) presenting the 592 and 417 geographic (e.g., islands) or political regions respectively used for mapping ant taxa distribution. Differences between B e n t i t y 2 and 3 are mostly in Colombia, India and within islands groups (not shown here).

Supplement S6: Known global distribution of *Pheidole megacephala*. Native = species considered native in the region; Country rec = species recorded only at the country level and without information at the first level of administrative division (for Colombia or India only); Need verification = uncertainty about the validity of the record, Dubious = presence considered as erroneous (e.g. misidentification); Exotic = species considered as non-native with established populations outdoor; Indoor introduced = exotic record known only from indoor structures (e.g., buildings or greenhouses) or from quarantine record. Details are provided in the text.