

Metadata for GABI data release 1.0

Benoit Guénard & Evan Economo
January 2020

Description: This file (GABI_Data_release1.0_18012020.csv) reflects occurrence data for valid ant species and subspecies data compiled as part of the Global Ant Biodiversity Informatics (GABI) Project, as of January 18, 2020. Some data visualized on Antmaps.org have been excluded from this release, notably the Antweb data, because they should be retrieved directly from Antweb.org. We also have excluded records that couldn't be assigned to a valid species, couldn't be assigned to a bentity_2 polygon, and those that we believe are outside the native ranges of the species.

Methods: The methods for compiling the data are included in the following publication.

Guénard, B., Weiser, M., Gomez, K., Narula, N., Economo, E.P. (2017) The Global Ant Biodiversity Informatics (GABI) database: synthesizing data on the geographic distributions of ant species. *Myrmecological News* 24: 83-89.

Citation: The above paper (Guénard et al. 2017) can be used as a citation for the data, however, we strongly encourage you to cite original data sources as well wherever possible.

Disclaimers: These data are a compilation of many different original sources, from individual papers to museum databases to private collections. As part of our research, we have performed extensive work to clean and improve the data to make it suitable for analysis, while putting disparate sources in a common spatial framework. These changes reflect our personal opinions and we do not intend to supersede the original authors, and they should be used at your own risk. Most of our analyses thus far have been on polygon level checklists (using a system we call bentity2, described below), which is cleaner and more accurate than the point-level data. We generally recommend the polygon level data for analysis, as the point data have many errors. We have a major effort to vet and georeference the point data underway, and we will release a cleaner and expanded dataset of point data when this is finished.

We will be updating these data in future releases.

Description of Data Fields

GABI_acc_number: A unique code we use to organize our data.

accession_number: Additional codes that are inherited from or refer to the original data source.

citation: The source of the record.

genus_name_pub: The ant genus name included as reported in the original source.

species_name_pub: The ant species name as reported in the original source.

valid_species_name: The valid species name based on current taxonomy, determined by us to the best of our ability.

type_of_data: Refers to the basis of the record (Literature, Museum Specimen, Personal collection)

dubious: Indicates whether the record has been flagged as potentially invalid. These should be filtered out before analysis.

note_dubious: A note for more info on why a record is considered dubious, if available.

locality: Text description of the locality.

country: The country.

adm1: The 1st level administrative area.

adm2: The 2nd level administrative area.

dec_lat: The latitude in decimal degrees.

dec_long: The longitude in decimal degrees.

bentity2_name: Since many points do not have latitude-longitude point data, we assigned each record to a common system of 546 polygons for visualization and analysis. These mostly reflect political boundaries, but can be on the country level (for small and medium-sized countries) or adm1 levels (for larger countries). Some polygons reflect spatial units that are biogeographic in nature rather than political, for example island boundaries. The bentity2 polygons are included in the accompanying shapefile.

insert_date: The date the record was added to our PostGRES database, although it may have been originally recorded by other sources earlier.

update_date: The date the record was last changed.