

Databases and ontologies

Traitpedia: a collaborative effort to gather species traits

Pablo Mier * and Miguel A. Andrade-Navarro

Faculty of Biology, Institute of Organismic and Molecular Evolution, Johannes Gutenberg University, Mainz, Germany

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on June 5, 2018; revised on August 17, 2018; editorial decision on August 20, 2018; accepted on August 23, 2018

Abstract

Summary: Traitpedia is a collaborative database aimed to collect binary traits in a tabular form for a growing number of species.

Availability and implementation: Traitpedia can be accessed from <http://cbdm-01.zdv.uni-mainz.de/~munoz/traitpedia>.

Contact: munoz@uni-mainz.de

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Background

Species can be univocally defined by their genotypes and phenotypes. Both are vastly intertwined, with the additional environmental component complicating the broad comprehension of this connection. Phenotypes, or traits, depend to some or most extent on the genetic information of the organism, and thus they are usually taxonomically driven. Gene evolution events (e.g. gene loss/duplication, horizontal transfer, etc.) (Koonin, 2005) and traits arisen from convergent evolution (Stayton, 2015) complicate the inference of these phenotypic/taxonomic associations. One fascinating example is the evolution of multicellularity in fungi, which might have happened no less than 11 times in different lineages (Nagy *et al.*, 2018). Such examples allow correlating molecular features with phenotypes; these correlations can give insights about mechanisms of evolutionary convergence and about molecular functions associated to complex biological processes.

One may think that traits such as whether a fungal organism is multicellular or not are very simple to define. However, this is not the case. Take as example the species *Dothistroma septosporum*, a fungus that causes the red band needle blight disease in conifers. Its genome is completely sequenced (de Wit *et al.*, 2012), it is covered in the EnsemblFungi database (<https://fungi.ensembl.org>), and selected in the set of fungal reference proteomes in UniProt (<https://www.uniprot.org/proteomes/>). Surprisingly, information about whether *D. septosporum* is unicellular, multicellular or colonial is not found in any database or in literature. We strongly believe that researchers working with this organism have not thought about

reporting this feature because they did not think that it could be of interest to anyone.

The absence of trait information may also be true for model organisms. Even though in most of the cases trait information can be mined from literature, retrieving this information for some species might not be an easy task. Databases like Encyclopedia of Life (<http://www.eol.org>), to some extent the Tree of Life Web Project (<http://tolweb.org>) and species-specific databases were developed for this purpose. What they all lack is a binary classification of traits in a table- and parsing-friendly format. And they are not prepared to be easily mined to look for trait associations, or trait comparisons between species.

Here we describe Traitpedia, a collaborative repository to gather species traits from an increasing number of species. The information is presented in tabular format with simple (mostly binary) values.

2 Database

The Traitpedia currently contains trait information for 181 eukaryotic species (Supplementary File S1). There are 15 traits per species, distributed in four categories:

1. General, miscellaneous features of the species.
2. Individual (phenotypical), traits inherent to each individual from a species.
3. Intraspecific (behavioral), traits related to how individuals from a species interact between themselves.

Traitpedia

Current species' general information

Tax ID | 7460
 Scientific name | *Apis mellifera*
 Common name | Honeybee
 UniProt Proteome ID | UP000005203
 Related databases | BeeBase
 Taxonomy | cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Holometabola; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apinae; Apini; Apis

Former species' general information

Tax ID | 7159
 Scientific name | *Aedes aegypti*
 Common name | Yellow fever mosquito
 UniProt Proteome ID | UP000008820
 Related databases | VectorBase
 Taxonomy | cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Holometabola; Diptera; Nematocera; Culicomorpha; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia

Additional execution

We offer you the additional possibility to compare the traits in the current entry to a different entry. To do so, just write here the second TaxID, Species name, or Common name.

GENERAL	VALUE [APIS MELLIFERA]	VALUE [AEDES AEGYPTI]
Sequenced	Yes	Yes

INDIVIDUAL [PHENOTYPICAL]	VALUE [APIS MELLIFERA]	VALUE [AEDES AEGYPTI]
Multicellularity	Multicellular	Multicellular
Winged	Yes	Yes
With tail	No	No
Hair	Yes	Yes
Photosynthetic	No	No
Hypoxia resistant	No	No
Torpor	No	No
Migratory	No	No
Environment	Cosmopolitan	Tropical

INTRASPECIFIC [BEHAVIOURAL]	VALUE [APIS MELLIFERA]	VALUE [AEDES AEGYPTI]
Reproduction	Sexual	Sexual
Social	Yes	No
Communication	Dance, vibration	Acoustic

INTERSPECIFIC	VALUE [APIS MELLIFERA]	VALUE [AEDES AEGYPTI]
Parasitic	No	Yes
Poisonous	No	No

Fig. 1. Example of use. Results obtained in Traitpedia when comparing the traits of *Apis mellifera* and *Aedes aegypti*

4. Interspecific, traits describing the relation between individuals from different species.

To look for the set of traits of an organism, the user can use either the NCBI Tax ID, species name or common name of said species to univocally identify it. Assuming that we have an entry for the requested species, a section with general information about the species will be shown, followed by information about its traits. Otherwise, a page will indicate that we do not have yet information about that species in our database. Alternatively, the user can query the database by trait, to display all the available trait values for the current set of species.

A novelty we introduce is the possibility to compare the traits of two species of interest. Given the example of the honeybee *Apis mellifera* and the yellow fever mosquito *Aedes aegypti* (Fig. 1), one can compare their set of traits in two easy steps. First, look for the entry of one organism; then, select the second organism to be compared in the section *Additional execution*. Below the general information of the two selected species, a simplified table with all available traits for both species is shown. The table can be downloaded from the results page. The pairwise comparison can be iterated by selecting a new species in the aforementioned section; the last species that was selected from the original pair would be then compared with the new one.

We have established Traitpedia as a platform that should develop into a comprehensive trait database: to reach this goal we rely on researchers to help us append additional traits that may be relevant to report to phenotypically characterize a species. We have prepared a file for anyone to fill in to let us know about new species or traits not currently covered in the database. It is hosted in the Traitpedia webpage, within the *Contact information* section. By submitting it to us we will dynamically include the received information in the Traitpedia. Trait information may be supported by references and clarified by comments.

We have strived to create Traitpedia as the simplest possible resource. For example, the underlying data is a table that can be downloaded at once. This simplicity should facilitate the future maintenance of the dataset, and, eventually, its integration or migration into future resources. In addition, simplicity is one of the factors that should facilitate contributions from the research community, which ultimately are crucial for the success of Traitpedia.

For simplicity, we considered most of the traits in Traitpedia as binary. However, some have more than two values. For example, the trait “Communication” has currently the values: sonorous, visual, pheromones, dance vibration, and none. Following feedback from users, we might increase the granularity of this and other traits such as infectivity, host range and virulence.

We foresee the possibility of supporting the growth of Traitpedia by developing automated data mining mechanisms, for example, taking annotations from the PubMed records of the biomedical literature associated to species by the NCBI Taxonomy Database (Federhen, 2012).

3 Conclusion

Finding molecular mechanisms responsible of the emergence of traits, such as limb formation in vertebrates and arthropods and its relation to the expression of particular developmental genes (Pueyo and Couso, 2005; Zhang *et al.*, 2010), creates valuable hypothesis for evolutionary and molecular studies of gene and protein function. To facilitate this exploratory research, we established Traitpedia, a resource to deposit simply formatted trait information. This idea comes from the realization that, although researchers working with a species may have a comprehensive knowledge about its traits, this information needs to be translated into a simplified yet illustrative resource. We believe the Traitpedia is a necessary integrative effort to be used both in research and in education. Once it grows above

threshold in the number of species covered, its tabular form and the limited trait values (for most of them, binary) will help in the mining of information and the extraction of trait correlations.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [AN 735/4-1 to M.A.A.N].

Conflict of Interest: none declared.

References

- de Wit,P.J. *et al.* (2012) The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLoS Genet.*, 8, e1003088.
- Federhen,S. *et al.* (2012) The NCBI taxonomy database. *Nucleic Acids Res.*, 40, D136–D143.
- Koonin,E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.*, 39, 309–338.
- Nagy,L.G. *et al.* (2018) Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biol. Rev. Camb. Philos. Soc.*, doi: 10.1111/brv.1241.
- Pueyo,J.I., and Couso,J.P. (2005) Parallels between the proximal-distal development of vertebrate and arthropod appendages: homology without an ancestor? *Curr. Opin. Genet. Dev.*, 15, 439–446.
- Stayton,C.T. (2015) What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface Focus*, 5, 20150039.
- Zhang,J. *et al.* (2010) Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature*, 466, 234–237.