



HAL
open science

An introduction to the World Asellidae Database

Florian Malard, Ph. Grison, Louis Duchemin, Maud Ferrer, Julian Lewis,
Lara Konecny-Dupré, Tristan Lefébure, Christophe J Douady

► **To cite this version:**

Florian Malard, Ph. Grison, Louis Duchemin, Maud Ferrer, Julian Lewis, et al.. An introduction to the World Asellidae Database. 18th International Congress of Speleology, Symposium 7, Biology, Subterranean life, Jul 2022, Chambéry, France. pp.301-304. hal-03873933

HAL Id: hal-03873933

<https://hal.science/hal-03873933v1>

Submitted on 13 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An introduction to the world Asellidae database

Florian Malard (1), Philippe Grison (2), Louis Duchemin (3), Maud Ferrer (1), Julian Lewis (4), Lara Konecny-Dupré (1), Tristan Lefébure (1) & Christophe J. Douady (1)

(1) Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France, Florian.Malard@univ-lyon1.fr (corresponding author)

(2) UMS 3468 BBEES, Muséum national d'Histoire naturelle / CNRS, Paris, France

(3) UMR CNRS 5558 Biométrie et Biologie Evolutive, Univ. Lyon 1 / CNRS / Univ. de Lyon, France

(4) Department of Invertebrate Zoology, U.S. National Museum of Natural History, Smithsonian Institution, USA

Abstract

The waterlice (Asellidae, Isopoda) are one of the rare family of metazoans containing both a great number of surface aquatic species and subterranean aquatic species. This makes it an ideal case study for understanding eco-evolutionary processes taking place during the transition to groundwater habitats. Comparative studies can be performed on multiple independent pairs of surface and subterranean species which provide true replicates of the ecological transition from surface water to groundwater. However, such studies require new collaborative research tools to assemble taxonomic, phylogenetic, distributional, and biological data for a species-rich family which is distributed across North America, Europe, North Africa and Asia. Here, we introduce the World Asellidae database, a tool that is being developed to allow multiple laboratories to jointly produce, visualize and query morphology-based and DNA-based species occurrence data and their associated biological materials. Access to metadata and data will be granted to users long before data are publicly released, thereby promoting collaborative projects among laboratories and reducing duplicate sampling and sequencing works.

1. Introduction

The understanding of eco-evolutionary processes taking place during the transition of organisms from surface aquatic habitats to groundwater habitats is one of the pillars of subterranean biology (CULVER *et al.*, 1995). This understanding has long been based on studies conducted at population level, in which the traits of surface aquatic and groundwater populations of the same species are compared. Key comparative studies of that kind include those conducted on populations of the cavefish *Astyanax mexicanus* and the waterlouse *Asellus aquaticus*. (JEFFERY, 2001; PROTAS & JEFFERY, 2012). Another comparative approach consists of comparing traits of multiple surface aquatic and groundwater species while taking into consideration their phylogenetic relationships (LEFÉBURE *et al.*, 2017). This phylogenetic comparative approach can complement intraspecific comparisons by considering multiple independent evolutionary events, for example multiple pairs of sister surface and groundwater species. It also enables consideration of evolutionary processes taking place over long time periods, at least longer than the lifespan of natural populations. However, comparative studies of multiple surface aquatic and groundwater species have remained scarce for two major reasons. First, many groundwater species-rich taxa lack surface relatives,

thereby precluding their use in comparative analyses of surface aquatic and groundwater species. For example, *Niphargus* is a species-rich genus of amphipods, but most of its over 300 described species inhabit groundwater (FISER *et al.*, 2008). Second, species-rich taxa containing both a number of surface aquatic and groundwater species have a wide distribution spanning several continents. Performing comparative analyses using these taxa requires new collaborative research tools to assemble the necessary taxonomic, phylogenetic, distributional and biological data. Here, we introduce the World Asellidae database (WAD), a tool that is being developed to allow multiple laboratories to jointly manage and query morphology-based and DNA-based species occurrence data and their associated biological vouchers. The Asellidae (Isopoda, Pancrustacea) are one of the rare family of metazoans containing both a great number of surface aquatic species and subterranean aquatic species. This makes it an ideal case study for understanding eco-evolutionary processes taking place during the transition from surface aquatic to groundwater habitats. Yet, developing a multiple-user collaborative tool is crucial to facilitate data acquisition and information sharing because the family is distributed across four continents.

2. Materials and methods

The structure of WAD (Figure 1) and its web interface (Figure 2) were first described and released as a free and open-source laboratory information management software in September 2019 (at [https://github.com/GOTIT-](https://github.com/GOTIT-DEV/GOTIT/releases/tag/v1.1.1)

[DEV/GOTIT/releases/tag/v1.1.1](https://github.com/GOTIT-DEV/GOTIT/releases/tag/v1.1.1); MALARD *et al.*, 2020). The database structure and web interface were developed using the system PostgreSQL and PHP Symfony framework, respectively.

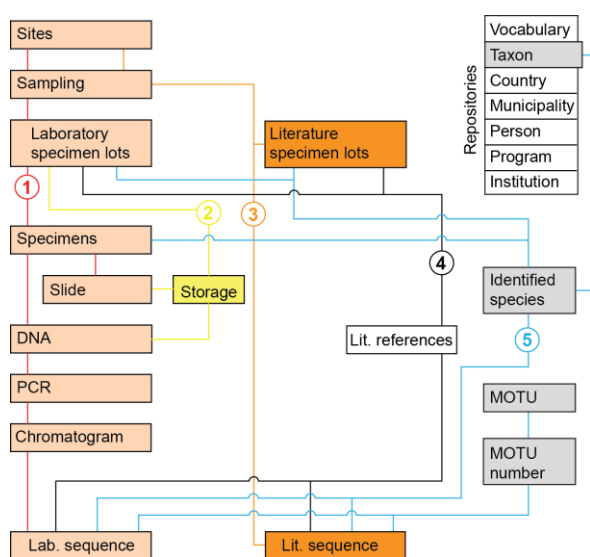


Figure 1: Simplified schematic view of the World Asellidae database (WAD). Numbered and colorized lines show five distinct pathways. (1) red: hosting laboratory data; (2) yellow: storage of biological vouchers; (3) orange: Literature data; (4) black: literature references; (5) blue: species assignment to specimens and DNA sequences. A lot is a collection of specimens of a single species taken at a site and date. MOTU: Molecular taxonomic units; Lab.: laboratory; Lit.: literature. Modified from MALARD *et al.*, 2020.

Briefly (see details in MALARD *et al.*, 2020), the database has five distinct pathways for managing the hosting laboratory data (from sampled site description to DNA sequence metadata), biological vouchers (storage of specimens and DNA), literature data (specimen lots and DNA sequence metadata), literature references, and the assignment of species to specimens and DNA sequences (Figure 1). DNA sequences, either from the hosting laboratory or the literature, can also be assigned to

3. Results

The database contains distributional data for all the 23 genera and 428 species and subspecies of asellids known to date. Of importance for studying evolution during transition of organisms from surface aquatic to groundwater habitats, the database contains data and metadata for 150 oculated and pigmented species and 278 depigmented and eyeless species. It also contains distributional data for approximately 40 species and subspecies of Stenasellidae (Isopoda, Pancrustacea), primarily European ones. The literature reference table has over 600 bibliographic references containing information on the original description and distribution of species. The database dashboard provides a snapshot of the distributional data and their origin in date of November 2020 (Figure 2). The distribution of species occurrence data stored in the database is shown in Figure 3A. Asellids are widespread in North America, Europe, northern Africa and

molecular taxonomic units (MOTU). MOTUs are delimited by the users using whatever molecular species delimitation methods and they are linked to taxa. The full database contains 45 tables and 450 fields: The full database contains 45 tables and 450 fields: taxonomic curation and data quality control are achieved by scientists of LEHNA (Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés) as part of their every-day laboratory activity. The web interface has been conceived to ease the every-day management and inputs of data and metadata in the database. A demonstration version of the interface is available at <https://gotit.cnrs.fr> with test data and full documentation. Each heading of the menu contains a number of subheadings which give access to a content form (Figure 2). Content forms list the records contained within tables and provide buttons to delete records and access record forms for creating, reading and updating individual records. The interface also comprises 29 tools for uploading records in batches (Figure 2). They allow to upload csv file templates containing records. Prior to uploading data in the database, a procedure checks and reports potential errors in the file template, such as identifier duplicates and incorrect data format. A number of ready-to-use queries have been implemented in the 'Species search' add-on of the interface. Among other functionalities, they allow the user to map the distribution of species, to explore the number and geographic distribution of MOTU within species, and to assess the exhaustiveness of molecular sampling within species.

The WAD project was first launched in 2015 by laboratories LEHNA and BBEES (Bases de données Biodiversité, Ecologie, Environnements Sociétés). Benchmark data were first migrated to the database in October 2017 and since August 2018 the web interface has been used as a laboratory information management software to download and manage Asellidae species occurrence data and molecular metadata produced on a day-by-day basis by scientists of LEHNA.

most probably also in Asia, although sampling effort has been less in the latter continent, except in Japan.

Species occurrence data are from 9403 samples collected at 8114 sites (Figure 2). A total of 7669 species occurrence data are from published literature (Literature data: specimen lots in Figure 2). Another set of 1863 species occurrence data are from sampling of specimen lots carried out by staff of LEHNA and multiple providers who kindly accepted to leg their biological material (Biological material: specimen lots in Figure 2). These specimen lots are labelled and stored in the LEHNA collections at -20°C. The dashboard also provides a summary of molecular work achieved at LEHNA in date of November 2020 (Figure 2). A total of 5490 individual specimens were extracted from specimen lots for morphological and/or molecular analyses, giving rise to 1702 microscope slides (dissection) and 4811 DNA extracts. Treatment of these DNA extracts

resulted in 18560 polymerase chain reactions (PCR), 11529 electrophoregrams (chromatograms in figure 2) and 8335 DNA sequences. DNA extracts are stored in the LEHNA collections at -80°C. The DNA sequences essentially target

two mitochondrial genes, the mitochondrial cytochrome oxidase subunit I (COI) and the 16S mitochondrial rDNA genes, and two nuclear genes, the 28S nuclear rDNA and FAST-2 genes (FAST kinase-like protein, subdomain 2).

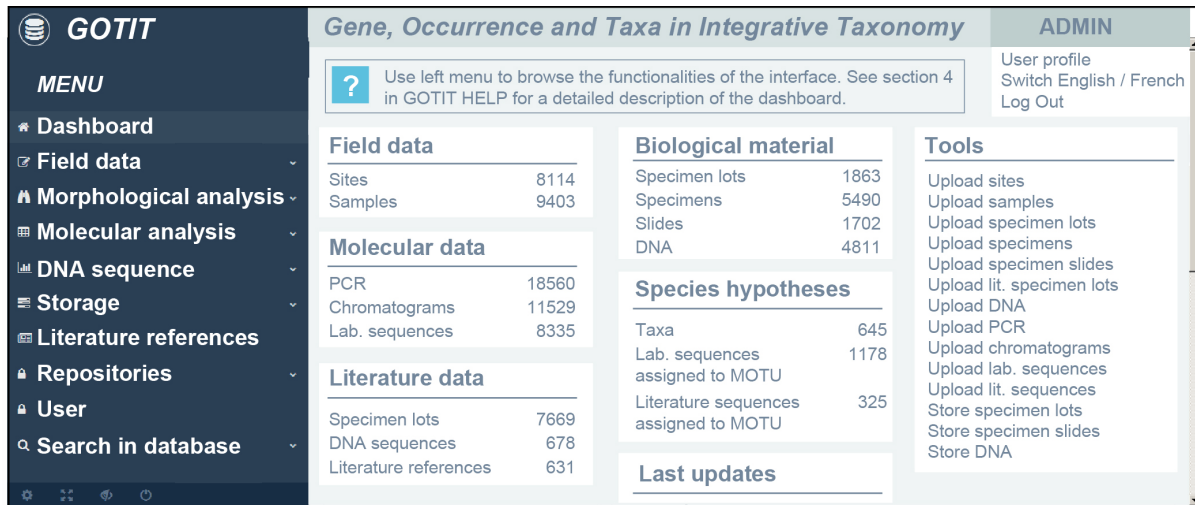


Figure 2: Dashboard of the web interface used for managing data and metadata stored in the World Asellidae database. Headings of the menu give access to lists of records contained in the database tables and to record forms for creating, reading and updating individual records. Tools are for uploading records in batches.

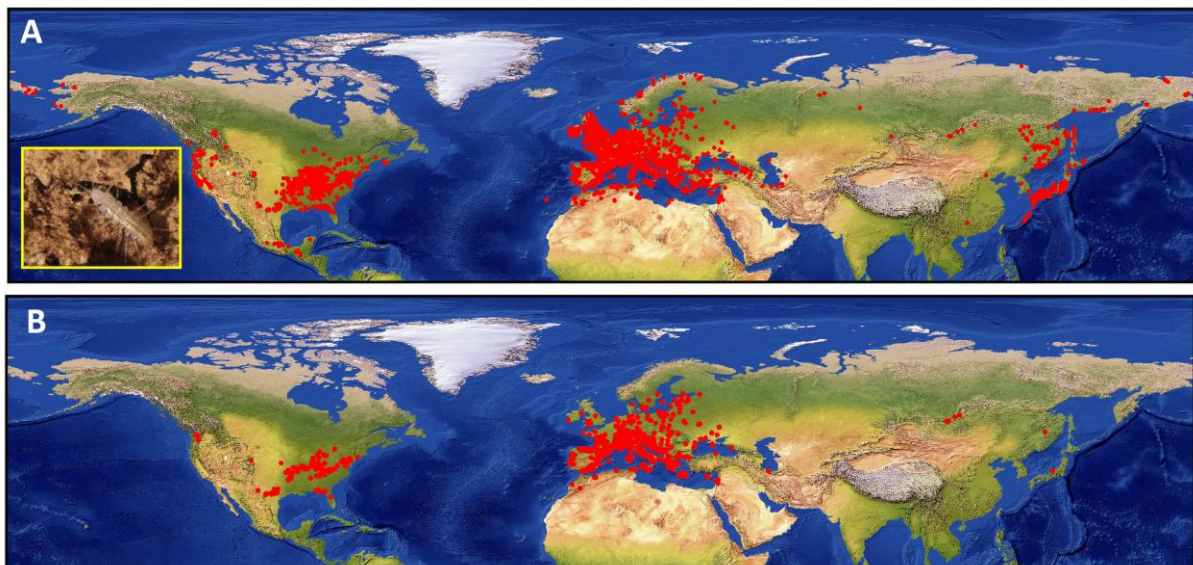


Figure 3: Worldwide distribution of species occurrence (red dots in A) and DNA sequence data (B) contained in the world Asellidae database. Photo of *Proasellus valdensis* (Asellidae, size: 8 mm) in Planches Cave, Jura, France (from R. Le Pennec).

These four genes are used to build phylogenies of Asellidae. Among the available COI sequences, 1178 were assigned to MOTU using four different molecular species delimitation methods. Figure 3B shows the distribution of DNA sequences registered in the database. Comparison

with the map of species occurrence data show that the sequencing effort has focused on Europe and to a lesser extent North America. However, there is a gap in DNA sequences for North-Africa and Asia.

4. Discussion

WAD provides one of the biggest data sets available to date for phylogenetic comparative studies of surface aquatic

and groundwater organisms. Recent published studies focused on the comparison of molecular traits, namely

genome size (LEFÉBURE *et al.*, 2017) and the rate of molecular evolution (SACLIER *et al.*, 2019), using a dozen of surface/subterranean species pairs, for which transcriptomic and genomic data were assembled. Yet, comparative studies can potentially be extended to a variety of traits and a much larger number of species by combining family-wide phylogenies (MORVAN *et al.*, 2013) with morphometric / biological measurements on specimens stored in collections. Candidate traits include body size and shape, number and size of eggs, and the size of species' geographic ranges. Incorporating such traits into the database represents a next important step.

As experienced by a dozen of regular LEHNA users, WAD and its web interface enhance key laboratory activities including research project design, sampling implementation and gap analysis, species identification, DNA sequencing, and contributing information to global biodiversity facilities. On-going developments are directed

towards opening database access to biodiversity professionals and promoting collaborative projects among laboratories, while promoting the FAIR-data and TRUST principles (WILKINSON *et al.*, 2016; LIN *et al.*, 2020). User access would be granted either in a read-only mode or in a collaborative mode for data contributors. Meantime, efforts are being made to provide a set of tools to retrieve data and metadata from the database. Expert tools will provide the users with the possibility to retrieve data using SQL queries. Non-expert tools will include the possibility to download data directly from the content forms of the web interface as well as using a friendly-user query builder. Efforts are being made to increase interoperability of WAD with other biodiversity facilities. Links to GenBank are already available for published sequences and developments can easily be implemented to export occurrence data in a Darwin Core format (WIECZOREK *et al.*, 2012).

5. Conclusion

Comparative phylogenetic studies using the Asellidae as a case-study hold much promise for understanding evolution during the ecological transition from surface water to groundwater. These data-intensive studies rely on close cooperation both between research laboratories as well as between laboratories and speleologists who provide many

specimen samples. In the near future, we hope that opening access to WAD strengthens cooperation between research laboratories and speleologists by providing them easier access to biodiversity data to which they continuously contribute.

Acknowledgments

We thank all collectors, including the many speleologists, who kindly provided specimens of asellids to LEHNA: their donation together with their names are gratefully acknowledged in the database. We thank G. Lapetoule and A. Bréchet for their help in entering records into the database. F.M., L.K.D., T.L. and C.J.J. were supported by the French National Research Agency projects CONVERGENOMICS (ANR-15-CE32-0005) and EUR H2O'Lyon (ANR-17-EURE-0018). WAD is hosted by the CNRS/IN2P3 Computing Center (Villeurbanne, France). We thank two anonymous reviewers for providing constructive comments.

References

- CULVER D.C., KANE T.C., FONG D.W. (1995) Adaptation and natural selection in caves: the evolution of *Gammarus minus*. Harvard University Press, Cambridge, US, 223 p.
- FISER C., SKET B., TRONTELI P. (2008) A phylogenetic perspective on 160 years of troubled taxonomy of *Niphargus* (Crustacea: Amphipoda). *Zoologica Scripta*, 37(6), 665-680.
- JEFFERY W.R. (2001) Cavefish as a model system in evolutionary developmental biology. *Developmental Biology*, Mar 1;231(1):1-12. doi: 10.1006/dbio.2000.0121.
- LEFÉBURE T., MORVAN C., MALARD F., FRANÇOIS C., KONECNY-DUPRÉ L., ..., DOUADY C.J. (2017) Less effective selection leads to larger genomes. *Genome Research*, 27, 1016-1028.
- LIN D., CRABTREE J., DILLO I. *et al.* (2020) The TRUST Principles for digital repositories. *Sci Data* 7, 144.
- MALARD F., GRISON P., DUCHEMIN L., KONECNY-DUPRÉ L., LEFÉBURE T., ..., DOUADY C.J. (2020) GOTIT: A laboratory application software for optimizing multi-criteria species-based research. *Methods in Ecology & Evolution.*, 11, 159-167.
- MORVAN C., MALARD F., PARADIS E., LEFÉBURE T., KONECNY-DUPRÉ L., DOUADY C.J. (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, 62(4), 512-522.
- PROTAS M., JEFFERY W.R. (2012) Evolution and development in cave animals: from fish to crustaceans. *Wiley interdisciplinary reviews. Developmental biology*, 1: 823-845.
- SACLIER N., FRANCOIS C.M., KONECNY-DUPRÉ L., LARTILLOT N., GUÉGUEN L., ..., LEFÉBURE T. (2019) Life history traits impact the nuclear rate of substitution but not the mitochondrial rate in isopods. *Molecular Biology & Evolution*, 36(3), 641.
- WIECZOREK J., BLOOM D., GURALNICK R., BLUM S., DÖRING M., ..., VIEGLAIS D. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715.

WILKINSON M.D. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*

3:160018.