The Tardigrada Register: a comprehensive online data repository for tardigrade taxonomy

Łukasz MICHALCZYK,1* Łukasz KACZMAREK2

¹Department of Entomology, Institute of Zoology, Jagiellonian University, Gronostajowa 9, 30-387 Kraków, Poland; ²Department of Animal Taxonomy and Ecology, A. Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland *Corresponding author: LM@tardigrada.net

ABSTRACT

In this paper we present the Tardigrada Register (www.tardigrada.net/register): a free, comprehensive, and standardised online data repository for tardigrade taxonomy. We outline key problems of the modern tardigrade systematics and we propose the Register as a potential solution to some of them. We then describe the idea, structure and works of the service and discuss challenges it may face. However, most importantly, we hope to convince fellow Tardigradologists that sharing their data via the Register will benefit the entire community of the contemporary and future tardigrade researchers.

Key words: Tardigrada Register, Tardigrada Newsletter, free online data repository, taxonomy, water bears.

INTRODUCTION

Modern tardigrade taxonomy, despite its continuous growth, is still very much in the 20th century regarding the standards of species descriptions and results dissemination. Molecular data are rarely included in species descriptions, morphometric variability is often neglected, and few studies capitalise on powerful optical and photographic equipment or graphic software to present in detail the full range of variability in tardigrade traits. Also, it would benefit the community of tardigradologists to make data sets available to the public. Taxonomic journals are largely to blame for this, as their chief editors have potentially the greatest power to enforce data-sharing policies so that the data underlying peer-reviewed tardigrade studies are available to scientists and the tax-paying public. For instance, some forms of data, such as DNA sequences, have been uploaded onto public data repositories (e.g. GenBank, http://www. ncbi.nlm.nih.gov/genbank/) for many years. More recently, prestigious scientific journals (e.g. Science, Evolution) have additionally made it a requirement for publication to upload all raw data onto online data repositories such as DRYAD (http://datadryad.org). Also, many US and UK funding agencies require or encourage authors to deposit data in widely accessible archives. Taxonomic journals should not remain behind this trend, and authors themselves can help to speed up the progress.

Why is it so important to report entire variability and also share it with others?

Due to the practical difficulties of rearing and breeding tardigrades, and because many species reproduce parthenogenetically, tardigrade taxonomy has been built almost entirely on the morphological (typological) species concept (Ruse, 1969; Pilato and Binda, 2010); only recently, and with limited impact, has molecular data been incorporated (phylogenetic species concept) (Coyne and Orr, 2004; Guidetti *et al.*, 2009). The analyses of variance, whether morphological, morphometric or molecular, within and between populations are needed to reveal discrete clusters of specimens we call separate species. However, proper and meaningful analyses can only be performed when we are equipped with as much knowledge on variability as possible. In other words, more detailed information about variability in taxonomic traits results in higher quality species descriptions and in more confident species identification.

Unfortunately, many papers still contain only measurements of what is considered a typical form, neglecting the remaining majority of variance that constitutes valuable information about variation in nature. Still, even if authors provide basic statistics in their papers, data sets used to calculate them hold much more information than these statistics can reveal. It is also important to keep in mind that in the future these data could be used for more powerful analyses than those available today. Also, among modern tardigradologists there are two methods of comparing species, either by contrasting statistics (usually ranges and means) for the populations of interest or by choosing a single specimen of a similar body size to the one described as a typical in an earlier paper. Having morphometric data sets for different species allows both types of comparisons. Moreover, it increases the chances finding specimens that match in body size.

However, it is not possible to publish raw data in most journals and, if it is, the data would rarely be in an



easy to read spreadsheet format. Similarly, images in a standard taxonomic paper are very often limited to typical specimens, whereas the authors usually have additional material. Being limited by restrictions on the quantity of data that is acceptable for publication has the potential to create difficulties in species identification or even lead to descriptions of synonymous taxa (e.g. when a non-typical form that was not described in the original report is interpreted as a new species). Another problem is the difficulty of obtaining type material for direct comparison with new or similar species. Although digitising type series through uploading both morphometric and image data into an on-line taxon database will always be second best and would never substitute a direct observation of real specimens, it can surely be helpful in situations when access to the type material is difficult or impossible (e.g. when sending slides risks them being lost or destroyed).

The mother all of invention

Given the aforementioned reasons for depositing data in online repositories, we are convinced that tardigrade taxonomy with no further delay should embrace digital data archiving to enable tardigradologists to routinely share raw data with their peers. However, the existing repositories (e.g. Morphbank, http://www.morphbank.net/ or the Barcode of Life Database, http://www.barcodinglife.org/) are constructed to accept various types of data concerning wide ranges of taxa. Thus, the form of data sets deposited in such general repositories inevitably varies between authors making comparisons of different populations or species far from straightforward. Also, data deposited in numerous places are difficult to locate. Finally, many repositories, due to the lack of taxonomic specialists in their teams, are not able to properly verify submissions. Erroneous records may result in misinformation, which could be more detrimental to taxonomy than the lack of available data sets. For example, many sequences in the GenBank labelled as belonging to a given tardigrade genus clearly represent other tardigrade families or, in extreme cases, even a different kingdom (fungi, most likely a contamination sequenced by (too) universal primers). Also, some entries in the Encyclopaedia of Life (http://eol.org) labelled as trusted contain photos with misidentified families. Such errors would probably not happen if taxonomic specialists were involved in data verification.

Taking into account all abovementioned needs and challenges, we decided to create the Tardigrada Register (TR), a comprehensive online data repository devoted exclusively to tardigrade taxonomy and run by tardigradologists. In this paper we would like to introduce the idea, structure and the working of the service. Importantly, we would also like to convince fellow Tardigradologists to participate in the making of the Register, as this is the only way the project can succeed.

The only other online database devoted entirely to tardigrades is the Tardigrade Barcoding Project (http://www.tardigradebarcoding.org). The service has somewhat similar goals to those of the TR, but with different priorities and structure. The Tardigrade Barcoding Project, as its name suggests, is focused on barcoding, although it permits conventional taxonomic information. The difference between the Barcoding Project and the TR is that the Register makes it a requirement to always provide morphometric data for every record (i.e. barcodes cannot be uploaded without the corresponding morphometric data). Also, all species in the Register must have data sets for the type series, against which non-typical records can be compared and verified by the users. Finally, the TR is interlinked with the Tardigrada Newsletter, which lists currently published tardigrade papers. Nevertheless, the two services can and should co-exist as they are likely to be complimentary and as such both will be useful for tardigrade researchers.

THE TARDIGRADA REGISTER (www.tardigrada.net/register)

The idea

The idea behind this project is simple: provide a free online data repository exclusive to tardigrade taxonomy (both terrestrial and marine). The service ought to be available to anyone with Internet access, easy to use, should store data in a standardised format and must ensure that contributors and original sources are properly acknowledged. By providing accurate scientific information, the repository should complement papers in which the original findings were described and interpreted. Data stored in a single place and in a unified format will allow quick and easy taxonomic comparisons. Digitised collections (especially type series) in the form of high resolution photographs and morphometric measurements should make species identification and description both faster and more accurate. Such register should focus on currently described species, but the ultimate goal ought to be to collect data on as many type series (and consecutive records, if available) as possible. Finally, the service should be open to tardigradologists willing to collaborate in its creation.

The structure

The TR comprises nine main sections, which we briefly describe below.

- i. Home. The Register's front page shortly explains the idea and the purpose of the TR. It also provides an overview of all sections of the service.
- ii. Register. This is the heart of the service. The species

register contains an alphabetic list of species that have their files uploaded onto the TR. Next to each species name there are five icons corresponding with the main parts of a species file (i.e. verbal description, imaging, morphometry, molecular data, and geographic distribution). Black icons indicate that a given piece of information is available, grey icons mean that it has not been uploaded. Every species in the Register has its own unique Uniform Resource Locator (URL): http://www. tardigrada. net/register/XXXX.htm, where XXXX is the species file number (from 0001 to 9999). The URL is stable, meaning that even if the genus or species name or the server on which the TR is placed change, the URL remains the same. Every species file in the Register starts with the important notice about citation of the information stored in the TR (apart from the Register, the original sources must be cited) and with the date stamp indicating when the file was last changed. The important notice is followed by seven subsections (see Fig. 1 for an example of a species file):

- *Taxonomy*. Taxonomic account of the species (from phylum to the genus/subgenus, with authorities and dates).
- *Description*. A concise verbal description divided into six parts (habitus, cuticle, buccal apparatus, claws, eggs, and remarks). Thanks to a unified form of the description, all taxonomically important traits are covered.
- Images. Images (photographs and drawings) are grouped into five columns corresponding to the specific parts of the verbal description (*i.e.* habitus, cuticle, buccal apparatus, claws, eggs). All photographs are raw (not processed by any imaging software) and with scale bars. If a structure is too deep for all its details to be visible in focus on a single photo, multiple images (layers) of the structure should be provided. Images are named as follows: Genus.species_population_type.of.individual_stru cture.photo.number.layer_magnification_microscope.type, for example: Milnesium.tardigradum_ neotype.series female(neotype) claws.IV.1b ×60

_PCM.jpg means that the phase contrast photo shows the second layer of the hind claws of a *Milnesium tardigradum* Doyère, 1840 neotype (which is a female). Thus, all vital information is described by the file name. Images are listed by sample (by clicking the sample code, the user is taken to the sample description in the section *Distribution*, see below).

 Morphometry. Measurements are stored in special TR templates in the form of MS Excel files (see below), named as follows: Genus.species_population. Having all measurements in a standardised format makes it easy to compare (also statistically) different populations and species. Morphometric data are listed by sample (by clicking the sample code, the user is taken to the sample description in the section *Distribution*, see below).

- *Molecular data*. DNA sequences are provided as GenBank entries, with direct links to the service from which they can be downloaded. Similarly to morphometric data, DNA sequences are listed by sample (by clicking the sample code, the user is taken to the sample description in the section *Distribution*, see below).
- Distribution. In this section detailed information on samples and collection localities is provided. The first column contains sample codes, which comprise of a country name followed by a number. The type or the neotype sample is marked as a 0and any other samples in a given country are denoted with a 1, 2, 3, etc. If specimens from the same locality were sampled more than once, the number in such cases is followed by a lowercase Latin alphabet character (e.g. Portugal.1a). This way every sample has a unique code, which is bookmarked and used as a grouping variable in sections Images, Morphometry and Molecular data. The next column contains geographic coordinates in the degree/minute/second format, linked to the exact location in Google™ Maps. Following columns comprise altitude (in m asl), location name (country, administrative unit/s), substrate type, and collection date.
- *Sources*. A list of references and other sources used to create the species file.
- iii. Submit. In this section an explanation on how to submit data to the Register is provided. Also, the section contains MS Excel morphometric templates. The templates automatically calculate relative indices widely used in tardigrade taxonomy such as the pt (Pilato, 1981) and the sc ratio (Fontoura et al., 2008), as well as basic statistics (sample size, range, mean and standard deviation). Additional data sheets also automatically arrange data in a format suitable for the majority of statistical software, making betweenpopulation and between-species comparisons easy. Thanks to the templates, morphometric data for all species are stored in the TR in a standardised format. Last but not least, summary tables with statistics can be copied and pasted directly to manuscripts (i.e. authors using the templates for their papers do not need to spend any extra time on arranging their morphometric data for the Register).
- iv. Taxonomy. In this section an up-to-date taxonomy down to the genus/subgenus level with authorities and

dates is provided. Also, a link to a tardigrade species checklist is available (Guidetti and Bertolani, 2005; Degma and Guidetti, 2007; Degma *et al.*, 2009-2013).

v. Methodology. This section contains an overview of methods used in tardigrade taxonomy: from specimen collection, isolation and preservation to microscope

techniques, taxonomic terminology, imaging, morphometry and finally DNA barcoding. We hope that freely available methods will subsequently help to improve standards in tardigrade taxonomic descriptions.

vi. Collections. Users will find links to tardigrade collections in this section. We anticipate this should fa-



Fig. 1. An example of a species file in the Tardigrada Register – a screenshot of the *Milnesium tardigradum* Doyère, 1840 (http://www.tardigrada.net/register/0001.htm, accessed 04.02.2013). All seven subsections of the species file are visible (for the detailed description of these sections please see paragraphe The structure). Most data are available for two populations, one in Germany (neotype population) and one in Poland (a subsequent record). Raw photographic images are linked to the image section under the relevant descriptor and raw morphometric data are provided in spreadsheets linked to the Morphometry section. DNA sequences are provided as links to GenBank, whereas the geographic coordinates are linked to Google MapsTM. *To be continued on next page*

Fig. 1. Continued from previous page.

Population	Habitus	Cuticle	Buccal apparat	us Claws	Egg	s	Contributor	
Neotype series	[holotype]♀ ventral view.1a x40 PCM] [holotype]♀ ventral view.1b]x40 PCM]	7	[paratype]≩[vent view.2a ×100]PCN [paratype]≩[vent view.2b ×100]PCN [paratype]≦]vent view.2c ×100]PCN	[] II.1a ×60 PCM] ral [holotype ♀ claw [] II.1b ×60 PCM] ral [holotype ♀ claw	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		Ł. MICHALCZYK, Ł. KACZMAREK	
Poland.1	[♀ ventral view.1 ×20 PCM]	7	[♀ ventral view.1a ×60 PCM] [♀ ventral view.1b ×100 PCN	[9]claws I.1c ×60	0[PCM] 0[PCM] 00[PCM] 00[PCM] 00[PCM]		K. ZAWIERUCHA	
. Morphometry								
Population	Data		So	Source		Contributor		
leotype series	[152]		100	MICHALCZYK et al. (2012)		Ł. MICHALCZYK, Ł. KACZMAREK		
oland.1	[15¢]		Ŭn	Ünpublished		K. ZAWIERUCHA		
. Molecular data								
Population	Sequence		GenBank	S	Source		Contributor	
leotype series	COI ITS2		JN664950 JF951049	N	MICHALCZYK et al. (2012)		Ł. MICHALCZYK, Ł. KACZMAREI	
Poland.1	?		n/a	r	/a		n/a	
. Distribution								
ample	Coordinates	Altitude	Location name	Substrate	Collection date	Source	Contribute	
ermany.0 <i>locus neotypicus</i>)	52°16'53"N 13°38'23"E	37 m asl	Germany, Wildau, Zeesen	moss and lichen from roof	02.2011.	MICHALCZYK et al. (2012)	Ł. MICHALCZY Ł. KACZMARE	
oland.1	52°24'11"N 16°53'19"E	88 m asl	Poland, Poznań	moss from stone wall	06.2012.	Unpublished	K. ZAWIERUC	

cilitate the exchange of specimens between researchers and institutions holding tardigrade collections.

- vii. Links. In this section a list of other online data repositories is provided. The majority of listed services contain some tardigrade taxonomic and biogeographical data.
- viii. Contributors. A very important section of the TR in which names, contact details and specific contributions made by the TR contributors are listed. We expect that the majority of entries will be done by the authors of species descriptions, but anyone having access to type material can be a TR Contributor. By acknowledging everyone's input, this section makes

the Register a community effort and work. Also, TR users may easily identify those who provided data and contact them if in need of further information. Contributors' names are also displayed in species files next to the pieces of information that they have provided. The names are linked directly to the *Contributors* section, where all contributions done by specific researches can be viewed. When citing information stored in the Register, contributor's names should also be mentioned.

ix. Secretariat. The last section provides the names and contact information of people who are responsible for running the service, *i.e.* for processing, verifying and uploading data sent in by researchers. Currently, the authors of this article are the only secretaries for the Register; however it would be desirable if more people from the tardigrade community join the TR in the future. That would speed up the process of uploading data onto the Register.

Data submission procedure

Any type series can be added to the Register at any time. The only requirement is the availability of raw data. However, it would be very desirable for papers, in which species are described, to add linking information about files in the Register (this way the reader of the description is informed about the associated TR file). This can be organised by authors (or editors) requesting a unique URL for their species from the Register's TR Secretariat at the manuscript proofing stage (i.e. when the species description has been accepted but changes to the manuscript are still possible). We suggest a sentence in the material and methods section of a manuscript along the lines of: Raw data underlying the description of Genus (Subgenus) species are deposited in the Tardigrada Register (Michalczyk and Kaczmarek, 2013; i.e. the current paper.) under http://www.tardigrada.net/register/XXXX.htm.Either before or just after requesting a URL the authors should submit their data to the Register following the instructions available from the Submit section of the TR. Importantly, the species file will not be released until the paper with the description is published.

Non-typical records are allowed only if there already are data for the type series deposited in the Register. Also, if a non-typical record is to be submitted, it must be supported at least by morphometric and sample data. Preferably, other data categories (*i.e.* imagery and DNA sequences) should too be provided for any given population (sample). Thanks to these two restrictions all records in the TR can be taxonomically verified by users themselves. Such verification is vital and is one of the reasons why the TR is different from any other existing repositories that store tardigrade data. In the longer run, files stored in the TR should help to better understand and identify species by providing data-based species delimitation procedures.

Copyright

Given that data (*i.e.* facts) are not copyrighted in most countries and since authors by submitting their data agree for them being freely available to the public, there should be no issues with copyright. The rules by which data are available are conforming to those described in the Creative Commons Zero Licence (CC0 1.0, http://creativecommons.org/publicdomain/zero/1.0). In short this means that the contributor has dedicated their work to the public domain by waiving all of their rights to the work worldwide under copyright law, including all related and neighbouring rights, to the extent allowed by law. This means that any user of the Register may copy, use, modify and distribute the data (even for commercial purposes) without asking permission, only on condition the sources of the original material are cited. Many established services such as BioMed Central (http://biomedcentral.com), PLoS (http://www.plos.org) and DRYAD (http:// datadryad.org) provide their resources under the Creative Commons Licence.

It is important to point out that scientific papers contain much more information than can be uploaded to the TR. For example, unlike in the Register, data in a paper are put in a context, interpreted and discussed. The TR does not contain differential diagnoses or any other form of data interpretation. Similarly, papers do not usually contain raw data and the morphometric data are typically presented in descriptive statistical form, while photographs are processed and arranged in plates (figures). In the TR only raw data and unprocessed pictures are deposited. Finally, depositing data in an accessible archive increases the chances of them being used and cited, which is in the interest of both the author and the journal.

Funding and long-term storage

After the presentation of the TR at the XII International Symposium on Tardigrada some concerns about the preservation of the data were raised by several tardigradologists. Here we would like to address these issues and explain how we envision the development of the service.

At the moment, the TR is in a very young initiative state and funded solely by the authors of this article [as is the Tardigrada Newsletter (www.tardigrada.net/newsletter)], which has served the community of tardigradologists for the last eight years). We hereby commit ourselves to run the TR by our own means, and concurrently seek storage at an established public institution, such as a government-funded university, which will guarantee continuous storage for the service. In our opinion, finding external funding will be more feasible after several years from the launch of the service, *i.e.* when we have proved that the TR works. In other words, the more support from tardigradologists we will get, the higher the chances of securing the Register's future. Given that our community of tardigrade researchers is relatively small and produces only *ca.* 20 new species and 3 species redescriptions a year (statistics based on the first decade of the current century, http://www.tardigrada.net/newsletter/archives.htm), the TR would not require considerable funds or staff in order to function. Therefore, the success of the Register seems plausible.

We hope that with time more scientists will join the Register's Secretariat and when we are no longer able to run the service, the next generation of tardigradologists will take the lead. However, if this were not to happen, at least data collected when the TR was functional would be preserved thanks to placing the TR in a recognised public institution. Last but not least, the Register is and should remain a non-profit service.

AN APPEAL TO FELLOW TARDIGRADOLOGISTS

We hope this presentation of the TR will convince our Colleagues that this initiative is worth considering. Nearly a decade of experience with the Tardigrada Newsletter, a service providing mainly references of currently published tardigrade papers and informing the community about important events, has shown that such projects are possible even without a formal society that would organise the community of tardigradologists.

Naturally, submitting data to the Register is an additional effort to that of writing a manuscript. However, the extra work needed for a TR submission is only a small fraction of that already devoted to the publication process, and at the same time this additional effort translates into a much greater and long-term benefit for science. It is important to recognise that virtually everyone gains from higher standards in taxonomy as these lead to more accurate species identifications, which in turn translate directly into grater quality of not only the taxonomy itself but also of any branch of science that relies on confident species identification (*e.g.* empirical ecology). Finally, digitised type series should reduce the risk of damage or loss of precious specimens, and freely accessible data of all sorts should stimulate the development of tardigrade taxonomy in general.

To conclude, we are convinced that sharing data via the TR will benefit the entire community of the contemporary and future tardigradologists.

ACKNOWLEDGMENTS

We are very grateful to Nathan W. Bailey (University of St. Andrews, UK), Sandra McInnes (British Antarctic Survey, UK) and two anonymous reviewers for their valuable comments on the manuscript. ŁM is supported by the Foundation for Polish Science (FNP), co-financed by the EU Regional Development Fund (grant no. Homing Plus/2012-5/8).

REFERENCES

- Coyne JA, Orr HA, 2004. Speciation. Sinauer Associates Inc., Sunderland: 545 pp.
- Degma P, Bertolani R, Guidetti R, 2009-2013. Actual checklist of Tardigrada species. Ver. 22: 01-03-2013. Available from: http://www.tardigrada.modena.unimo.it/miscellanea/Actual%20checklist%20of%20Tardigrada.pdf
- Degma P, Guidetti R, 2007. Notes to the current checklist of Tardigrada. Zootaxa 1579:41-53.
- Fontoura P, Pilato G, Lisi O, 2008. Echiniscidae (Tardigrada, Heterotardigrada) from Faial and Pico islands, the Azores, with the description of two new species. Zootaxa 1693:49-61.
- Guidetti R, Bertolani R, 2005. Tardigrade taxonomy: an updated check list of the taxa and a list of characters for their identification. Zootaxa 845:1-46.
- Guidetti R, Schill RO, Bertolani R, Dandekar T, Wolf M, 2009. New molecular data for tardigrade phylogeny, with erection of *Paramacrobiotus* gen. n. J. Zool. Syst. Evol. Res. 47:315-321.
- Pilato G, 1981. [Analisi di nuovi caratteri nello studio degli eutardigradi]. [Article in Italian]. Animalia 8:51-57.
- Pilato G, Binda MG, 2010. Definition of families, subfamilies, genera and subgenera of the Eutardigrada, and keys to their identification. Zootaxa 2404:1-54.
- Ruse M, 1969. Definitions of species in biology. Brit. J. Phil. Sci. 20:97-119.