

## OPINION

# A plea for digital reference collections and other science-based digitization initiatives in taxonomy: Sepsidnet as exemplar

YUCHEN ANG<sup>1</sup>, JAYANTHI PUNIAMOORTHY<sup>1</sup>, ADRIAN C. PONT<sup>2,3</sup>, MIROSLAV BARTAK<sup>4</sup>, WOLF U. BLANCKENHORN<sup>5</sup>, WILLIAM G. EBERHARD<sup>6,7</sup>, NALINI PUNIAMOORTHY<sup>1,5</sup>, VERA C. SILVA<sup>8</sup>, LORENZO MUNARI<sup>9</sup> and RUDOLF MEIER<sup>1,10</sup>

<sup>1</sup>Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, <sup>2</sup>Oxford University Museum of Natural History, Oxford, England, <sup>3</sup>Natural History Museum, London, England, <sup>4</sup>Department of Zoology and Fisheries, Faculty of Agrobiological Sciences, University of Life Sciences Prague, Prague, Czech Republic, <sup>5</sup>Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, <sup>6</sup>Escuela de Biología, Universidad de Costa Rica, Ciudad Universitaria, San Pedro, Costa Rica, <sup>7</sup>Smithsonian Tropical Research Institute, Balboa, Panama, <sup>8</sup>Departamento de Morfología e Fisiología Animal, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, São Paulo, Brazil, <sup>9</sup>Natural History Museum, Entomology Section, Venice, Italy and <sup>10</sup>University Scholars Programme, National University of Singapore, Singapore

## Introduction

Few disciplines in science are as obsessed with ‘going digital’ as taxonomy. This is not surprising given that ‘taxonomy is made for the web: it is an information-rich subject, often requiring copious illustrations’ (Godfray, 2002: 17). However, obsessions are not healthy [Merriam Webster defines them as ‘persistent and disturbing intrusion(s) of or anxious and inescapable preoccupation(s) with an idea or feeling especially if known to be unreasonable’] and not every cybertaxonomy initiative is equally well motivated. Similarly, not all pre-digital practices in taxonomy were fatally flawed. It would be better to adopt the more measured attitude toward digitization shown by other disciplines. They calmly went digital, used the new opportunities when they enhanced scientific research, and stayed traditional when there was little incentive for change: overall digital innovations were judged based on scientific merit and not automatically considered superior. For example, evolutionary biologists adopted digital publication formats and embraced supplementary materials that allowed for the publication of primary data in various formats, but there was neither a frenzied search for a digital revolution nor over-reaching proposals to digitize all organisms. Here we argue for the same measured approach in taxonomy; i.e. digital initiatives should go ahead if they (i) enhance taxonomic output and quality, (ii) are feasible, and (iii) have favourable cost–benefit

ratios. Such assessment avoids the waste that comes with poorly-reasoned activism that can lead to unrealistic visions, dismissive attitudes toward existing practices and further reductions in taxonomic capacity by investing in ventures with modest scientific returns.

A recent special issue of *Zookeys* [‘No specimen left behind: mass digitization of natural history collections’ (Blagoderov & Smith, 2012)] provides an overview over collection-based cybertaxonomy initiatives. The most extreme initiatives aim to digitize just about anything that is found under a museum’s roof: field notebooks (Hill *et al.*, 2012; Thomer *et al.*, 2012), images and other media (Berendsohn & Güntsch, 2012), and, of course, specimens and their label data despite the fact that this requires huge amounts of funding. For example, Tegelberg *et al.* (2012) estimates that ‘[o]n average, a staff member has been able to produce about 40 images or data entries per day. The cost of digitisation is currently 3.99 € per image and 5.61 € for data entry of a specimen, which makes a total of 9.60 € for a fully processed sample’. Yet, the scientific value of wholesale digitization is not obvious. For example, only specimens that are correctly identified should be digitized, because digitizing unidentified specimens is wasteful and digitizing misidentified specimens is harmful. After all, digitized specimen records are captured by data aggregators (e.g. GBIF: <http://www.gbif.org/>) and then used by macroecologists in meta-analyses. Relatively little is known about the proportion of correctly identified, unidentified and misidentified specimens in collections, but Meier & Dikow (2004) described that for the 1383 specimens included in

Correspondence: Rudolf Meier, Department of Biological Sciences, National University of Singapore, 14 Science Dr 4, 117543 Singapore, Singapore. E-mail: meier@nus.edu.sg

a revision of the predominantly Afrotropical asilid genus *Euscelidia* Westwood, 1850, 74% were unidentified, 12% were misidentified and/or carried an incorrect name, and only 4% were correctly identified; obviously only the latter should have been digitized prior to a revision of the genus. The proportion of correctly identified specimens may have been particularly small in this case, but misidentifications are not uncommon in museums and the proportion of unidentified specimens generally is large.

Nonetheless, many papers in the aforementioned *Zookeys* issue still promote wall-to-wall digitization of collections. This often comes in the form of drawer digitization with or without assigning physical or virtual unique specimen identifiers (e.g. Bertone *et al.*, 2012; Blagoderov *et al.*, 2012; Schmidt *et al.*, 2012; van Oever & Gofferje, 2012). There is anecdotal evidence that this encourages loan activity (Bertone *et al.*, 2012; Mantle *et al.*, 2012), but given that the digitization of collections takes many people-years, a rigorous cost/benefit analysis would reveal that a few emails would probably be more effective in encouraging loans. Of course, wholesale collecting of data without (much) quality control is found not only in museums. Instead it is also the technique *du jour* of data aggregators (e.g. GBIF, EOL <http://eol.org/>) and also has become common in DNA barcoding. Here it is now so rampant that the majority of DNA barcodes in Genbank had to be withdrawn because they were not identified to species [423 188 records: Kwong *et al.* (2012)].

Fortunately, many museums have recognized that only well-curated parts of a collection should be digitized (Haston *et al.*, 2012; Mantle *et al.*, 2012; Tulig *et al.*, 2012). In these institutions specimens destined for digitization are pre-selected and/or curated prior to digitization. The downside is that this requires considerable curatorial time that arguably should be spent on activities that generate new scientific insights. Indeed, it appears that in some museums the various digitization activities now employ more personnel than research; i.e. wholesale digitization is in the process of changing the institutions' character without helping to overcome the taxonomic impediment. It appears to us that healthier approaches are based on a mixture of on-demand digitization (see Mantle *et al.*, 2012) and integrating science and digitization (Meier & Dikow, 2004; Schuh, 2012; Wheeler *et al.*, 2012). Such approaches often yield high-quality data that are more likely to justify the legacy costs (e.g. see Haston *et al.*, 2012) associated with digital databases. These costs should be carefully considered in all digitization projects given that special funding may be available for the initial digitization of specimens/drawers but is unlikely to be available for the maintenance of databases. Cost–benefit analyses appear particularly lacking for drawer-digitization projects where maintenance is particularly labour-intensive. After all, when a specimen from a digitized drawer is (re)moved, the drawer has to be re-digitized and/or the specimen movement has to be tracked (Bertone *et al.*, 2012; Dietrich *et al.*, 2012).

## Digital reference collections

Instead of selling a grand vision and discarding current taxonomic practice, we here promote a more modest cybertaxonomy tool that may have received little attention because of its lack of grandeur: the digital reference collection. Physical reference collections are important tools in taxonomy; they are used by many systematists for confirming identifications by comparing undetermined specimens with specimens that have been identified by taxonomic experts. Here we argue for digital reference collections which have the advantage that they are accessible to virtually all entomologists and can cover more species diversity than physical collections. We demonstrate the value of such a digital collection by presenting one for Sepsidae (Diptera) that includes high-quality images for c. 40% of the species-level diversity. Note that for ease of use, we believe that a digital reference collection should be a clean, simple, purpose-built tool for confirming species identification and not a collection of species pages that are information-rich but make it difficult to compare the images for multiple species. For example, in the otherwise excellent and information rich Antweb (<http://www.antweb.org/>), a user browsing for images of a particular species has to navigate three URLs before arriving at the species page. Images are all provided as thumbnails. Clicking on one opens a pop-up window with a high-resolution image. To view the next image or species, the user has to return to the previous species page; all subsequent high-resolution images open in the same pop-up window and override the previous image.

Online tools are particularly useful and needed for identification given that traditional tools can suffer from serious problems (Gotelli, 2004). As documented for sepsid flies below, these traditional tools are generally (i) scattered across many journals thus creating an access problem, (ii) of variable quality, (iii) become outdated quickly (e.g. through the addition of new species and emended nomenclature including synonymization of species names), and (iv) suffer from ontology problems when different authors do not use the same terms for the same structure. Given these problems, it is not uncommon for biologists to have little confidence that a particular specimen has been identified correctly even if the most up-to-date identification literature has been consulted. Hence, many specialists still maintain reference collections consisting of reliably identified specimens. These reference collections are used for confirming identifications through direct comparison of unidentified with identified material. However, few biologists have access to reference collections and most contain predominantly species from one region. This is where digital reference collections can make a difference. Comparing an unidentified specimen under the microscope with a high-quality image on a nearby computer screen is in most cases sufficient and more convenient than finding and retrieving a specimen from a traditional reference collection and juggling simultaneously multiple pinned and/or ethanol-preserved vouchers. Digital reference collections also are the best long-term solution for the problem of rarity. Many species have been collected only once and/or described based on a single specimen (Lim *et al.*, 2012). Many of these species

are missing in physical reference collections while they can be included easily in digital collections. Note that having identification tools for these species is important because collectively they can constitute a large proportion of the species-level diversity in biodiversity samples (Longino *et al.*, 2002).

Based on cost–benefit arguments, digital reference collections are useful for taxa that need to be identified routinely by biologists from different backgrounds. A good example is Sepsidae (Diptera) whose conspicuous sexual dimorphism, short generation times and tolerance for laboratory conditions have attracted the attention of biologists from a variety of disciplines. For instance, researchers have used Sepsidae to address theoretical issues in systematics (Meier, 1997; Su *et al.*, 2008; Tan *et al.*, 2010), to study the behaviour of these flies (Eberhard & Huber, 1998; Eberhard, 1999, 2000, 2001a,c, 2003, 2005; Puniamoorthy *et al.*, 2008, 2009; Tan *et al.*, 2011), their development (Bowsher & Nijhout, 2007, 2009; Bowsher *et al.*, 2013), genetics (Reusch & Blanckenhorn, 1998; Mühlhauser & Blanckenhorn, 2004; Hare *et al.*, 2008) and evolution (Eberhard & Pereira, 1996; Blanckenhorn *et al.*, 1998, 1999, 2000, 2002, 2004, 2007; Eberhard, 2001b, 2002a,b,c; Mühlhauser & Blanckenhorn, 2002, 2004; Teuschl & Blanckenhorn, 2007; Teuschl *et al.*, 2007; Ingram *et al.*, 2008; Puniamoorthy *et al.*, 2010; Teuschl *et al.*, 2010a,b; Wild *et al.*, 2011; Puniamoorthy *et al.*, 2012). Consequently, there is a large community of biologists with an interest in and need for up-to-date identification tools for Sepsidae.

Digital reference collections can be created for any group of insects, but arguably they are suited best for those that require relatively few images for capturing the critical structures that are used for identification. In addition, the species diversity should be fairly well known overall so that users can assess how much of the species diversity is covered by the collection. Sepsidae satisfy both conditions. Male foreleg and genital morphology is sufficient for identifying most species. In addition, given that the number of newly described species has dropped to <5 per year and even the tropical fauna has received considerable attention, the evidence indicates that most species have been described. The family Sepsidae currently comprises 347 valid species in 36 genera. Several landmark monographs from the 20th century created a solid foundation for understanding the species diversity and all the research culminated in the publication of a first world catalogue that listed 318 valid species (Ozerov, 2005). Since then approximately 30 new species have been described and only few genera such as *Dicranosepsis* Duda, 1926 and *Sepsis* Fallén, 1810 are likely to include many more undescribed species.

Despite the overall good state of sepsid taxonomy, identification tools suffer from all the problems mentioned earlier. For example, access to species descriptions is difficult because they are scattered throughout a number of journals with varying accessibility. The first sepsid was described >250 years ago [*Sepsis cynipsea* (Linnaeus, 1758)] and descriptions of almost one sixth of the existing valid species date from before 1900 (Ozerov, 2005). As such, a substantial proportion of the descriptions are in old and/or discontinued journals that are

available in few libraries. Some of the most important, old publications and the recently published literature are available online via commercial journal subscriptions or publication digitization projects that are an excellent example for the meaningful use of digitization for supporting taxonomic research (e.g. Biodiversity Heritage Library: <http://www.biodiversitylibrary.org/>). However, much of the 20th century literature is less accessible due to a toxic mix of lack of digitization and copyright restrictions. Access to primary type material is even more difficult. Despite having only about 350 described species, the types are distributed across *c.* 50 institutions on all continents. Borrowing types is getting increasingly difficult due to the risk of damage or loss given that the proportion of old and fragile types is high and bound to increase because most species were described in the past centuries; i.e. types are aging. Already many type specimens are so damaged that little morphological information can be obtained. This is particularly problematic in sepsids where many types lack legs or abdomina, and/or have claspers that are withdrawn into shrivelled bodies (e.g. [http://sepsidnet-rmbr.nus.edu.sg/Sepsis\\_kyandolirensis.html](http://sepsidnet-rmbr.nus.edu.sg/Sepsis_kyandolirensis.html)).

In addition to the access problems, there are numerous quality problems with existing identification tools. More than 70 authors have described sepsid species and the quality of the descriptions varies tremendously. Those from the 18th and 19th century generally are very brief and lack illustrations and information on genitalia. The descriptive literature of the 20th and 21st century is of higher quality, but some authors underestimated the intraspecific variability, leading to some lumping and many synonymy problems [see Ozerov (1995) for the large number of synonyms created by Vanschuytbroeck]. Potentially all descriptions suffer from a lack of information on character systems with relevance recognized only after the species were described. For example, the authors of the 18th and 19th century did not anticipate the need for information on genitalia, and microtrichosity ('pruinosity') patterns were rarely described in the literature of the early 20th century. Today's descriptions are also likely to suffer from yet unknown deficiencies. These quality problems can be partially overcome through digital photography (Cranston, 2005). High-quality photographs provide similar amounts of information for all species. In addition, good photographs capture not only the morphological features that are regarded currently as important, but they may serendipitously also capture information that will become relevant only in the future. For example, if such images had been used by taxonomists of the early 20th century, microtrichosity patterns would have been captured. Of course, digital photography also has shortcomings and drawings can be more effective in highlighting critical features and usually are better at summarizing intraspecific variability. Thus, digital reference collections should consist of a mixture of photographs and other types of illustrations.

An additional problem with the traditional literature used for specimen identification is that it becomes quickly outdated through the description of new species and new synonymy proposals. For example, in Sepsidae more than 650 species-level taxa have been described of which >300 have been

synonymized. Many synonyms were proposed recently so that identification literature predating these proposals is outdated. The synonymy problem alone is widespread and can be quite severe for some species. For example, as of 2005, 87 of the 318 valid species had junior synonyms (Ozerov, 2005) and variable species such as *Saltella sphondylii* (Schrank 1803) had attracted > 20 separate descriptions (Pont & Meier, 2002). Digital reference collections also can become outdated if not updated regularly. However, in contrast to the traditional identification tools, they can remain up-to-date and revision dates can be specified on the web page; that is, synonymy changes and descriptions of new species are accommodated more readily.

Another much lamented problem with the identification literature is that it is incomprehensible to the non-specialist because of ontology problems (Gotelli, 2004). For example, sepsid taxonomists routinely differ in the terms used to describe the same morphological structure. The distribution of microtrichia on certain pleural sclerites is diagnostic for some sepsid species. Sometimes these sclerites are referred to as anepisternum, katapisternum and anepimeron (e.g. Ozerov, 1993) whereas other authors use mesopleuron, sternopleuron and pteropleuron, respectively (e.g. Iwasa & Tewari, 1990). The most direct way of overcoming such ontology issues is to agree upon a standard terminology, and many Dipterists use the terminology of the Diptera manuals (McAlpine *et al.*, 1981; Courtney *et al.*, 2000; Merz & Haenni, 2000; Sinclair, 2000; Cumming & Wood, 2009). However, the published literature containing alternative ontologies still needs to be translated to the standard. With a digital reference collection, two solutions are apparent. First, the specimens can remain unlabelled. The user then directly compares the query specimen with the imaged specimen, thus potentially bypassing terminological problems. However, this could result in overlooking morphological features that are critical for species identification. Alternatively, the specimens can be labelled and an interactive glossary is provided [see Hamilton *et al.* (2006)]. In Sepsidnet we have opted for unlabelled images, but provide an interactive glossary that allows the user to match terms in the identification literature with structures of a model sepsid species.

## Building the digital reference collection for Sepsidae

### *Specimen acquisition*

We obtained material for 139 of the *c.* 350 described species. This included type material for 32 species from several European institutions [Muséum National d'Histoire Naturelle, Paris, France (MNHN), Musée Royal de l'Afrique Centrale, Tervuren, Belgium (MRAC) and Royal Belgian Institute of Natural Sciences, Brussels, Belgium (RBINS)]. However, the specimens for most species were taken from cultures that were started from field-collected females as described in Ang *et al.* (2008) and Ang & Meier (2010). Most sepsid specimens in collections are pinned, but often

it is preferable to preserve them in 80% ethanol so that the specimens are available for both morphological and molecular work (storage in a fridge/freezer is recommended). If cultures are needed, field-collected males usually can be killed immediately, while females are used to establish a culture. This ensures ample high-quality specimens with a definite male–female association and the option for additional work on behaviour, natural history and reproductive isolation.

### *Specimen identification*

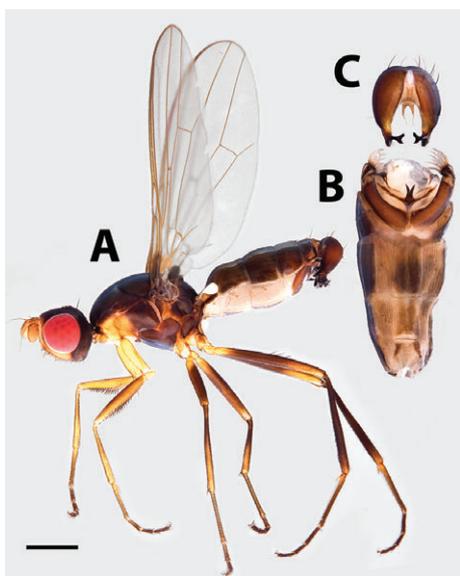
Sepsidnet is a community effort by sepsid workers from various parts of the world. This means that the species pages have been and continue to be checked by several experts in the group. Most specimens were identified initially based on the literature and then compared to the reference collection of the Evolutionary Biology Laboratory at the National University of Singapore (NUS). For most species we have generated a DNA barcode that can be accessed via a Genbank link.

### *Specimen imaging*

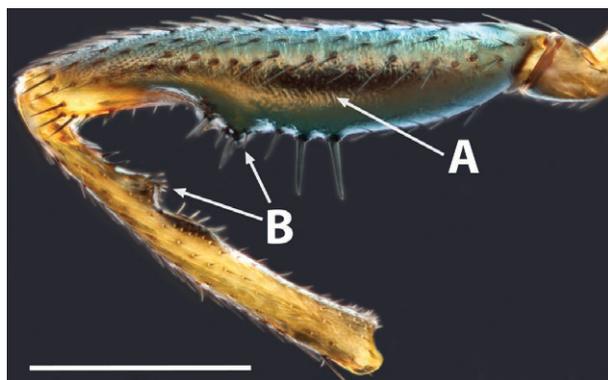
Sepsid specimens were imaged first before extracting DNA. Males were preferred because they tend to show species-specific diagnostic characters. Females were imaged if the association with males was confirmed via a culture or rearing. Where possible, we documented the male lateral habitus (Fig. 1A; to view the leg ornamentation and thoracic pleura), the ventral view of the abdomen (Fig. 1B; to view any modified sternites), and where available, the dissected claspers (Fig. 1C). These structures were combined into a single image (as Fig. 1). Specimens from the reference collection at NUS as well as type material from RBINS were imaged serially at different focal distances using a Visionary Digital™ BK Plus Lab System. The images were then focus-stacked using Helicon Focus™ Pro (v5.2.16). The image resolution generated with this system is sufficient enough to show cuticular sculpturation (Fig. 2). It is comparable in quality to observations by eye through a good quality stereomicroscope. Specimens from the European museums (MNHN and MRAC) were imaged and focus-stacked with the Auto-Montage® system using a Leica® DFC450 microscope. Image quality from this imaging system also is adequate for diagnostic features. Composite images were edited with Adobe® Photoshop® CS4 to remove background noise and optimize brightness, contrast and colour balance. In addition to providing digital photographs, we included line drawings from the literature for all species.

### *Sepsidnet design*

Edited photographs were exported in Zoomify™ format (as a Zoomifyer) and embedded into the Sepsidnet website. The Zoomifyer is a specialized Flash object that allows users to stream high-magnification images of structures that are



**Fig. 1.** Focus-stacked composite image of *Nemopoda speiseri* Duda, 1926. (A) habitus, lateral view; (B) abdomen, ventral view; (C) dissected claspers, dorsal view. Scale bar = 1 mm.



**Fig. 2.** Focus-stacked image of male *Sepsis cynipsea* (Linnaeus, 1758) fore femur and tibia, anterior view. The resolution of the image is high enough to show: (A) cuticular sculpturation detail on the femur; (B) sexually dimorphic ornamentation in detail on both femur and tibia. Scale bar = 0.5 mm.

critical for identification: it divides an image into a series of smaller-sized picture tiles at different resolutions and sizes that are presented onto a fixed frame. Because the viewer frame requires only few picture tiles to be loaded at any time, viewing is fast and smooth. As a Zoomifyer comprises a simple package of HTML code, small image files and a simple Flash movie code, it can be played readily on any browser with Flash support. One caveat to using Zoomify™ for viewing specimen images is the decline of Flash support in certain platforms (e.g. mobile Apple iOS requires plugins). However, should Flash become undesirable in the future, large-image viewing can easily be implemented using different software packages.

A Sepsidnet page consists of two columns. The main specimen image for the target species (usually a male) is shown in a 700-by-700 pixel Zoomifyer on the top of the left column, with a link to specimen label information directly below it. Below the Zoomifyer are diagnostic illustrations for the species. On the right column are several drop-down tabs that when selected give access to additional information. From top to bottom, the first gives access to taxonomic information that is drawn mostly from Pont & Meier (2002) and Ozerov (2005). Links to external sources (literature: PubMed and Biodiversity Heritage Library; DNA sequences: Genbank) are provided. The second tab provides information on species distribution, while the third opens up into an interactive morphoglossary that has separate sections for 'head', 'thorax', 'abdomen (male)' and 'abdomen (female)'. Below these three tabs are the species comparison and navigation tools. By clicking the 'Compare with other species' button, a new comparison frame appears with four separate Zoomifyers for quadruple species comparison. Additional frame options for different screen resolutions are also offered (from two to six species comparisons available), and the comparison frame is easily closed to revert back to the species page. Underneath the comparison tab are three buttons for navigation. The left 'previous species' and right 'next species' buttons direct to (alphabetically) antecedent and subsequent species pages, respectively, while the middle 'species list' drop-down tab opens a list of all species available on Sepsidnet. In some species there is one last additional tab that links to additional images. To ensure permanency, Sepsidnet is hosted by the server of the Faculty of Science at the National University of Singapore under an account that is associated with the Raffles Museum of Biodiversity Research: <http://sepsidnet-rmbr.nus.edu.sg/>.

## Conclusion

We present here a digital reference collection for 139 species of Sepsidae (Diptera). The collection is a work-in-progress in that it comprises currently *c.* 40% of the described species-level diversity but additional pages for *c.* 30 species are in preparation. Currently, many species are represented by only one specimen. This does no justice to the intraspecific variability and the morphological differences between males and females, nor are the immature stages shown. Future additions to the database will address these deficiencies. Nonetheless, we believe that Sepsidnet demonstrates that digital reference collections are powerful taxonomic tools because they can provide visual data on many species without the need for establishing physical reference collections. Such 'virtual specimens' are a good substitute for the actual specimens and can even be a better choice given that many physical specimens are in poor condition. Specialists and non-specialists alike can utilize these images for confirming identifications that were based initially on traditional tools; thus, Sepsidnet is not intended to replace existing identification keys. Instead it is a tool that should be used after a preliminary identification

has been obtained. As such Sepsidnet also can relieve sepsid specialists of the burden of routine identification and in many cases avoid the need for mailing specimens and/or consulting type specimens. We have used the online images on Sepsidnet to inform the public and other biologists about what kind of structures have to be imaged before sepsid species can be identified based on photographs, aiding in reducing identification requests based on inappropriate images.

As pointed out by many authors (Scoble, 2004; Cranston, 2005; Godfray, 2005; Knapp, 2008), taxonomy could benefit if data were available online. One way to organize these data is in the form of online databases. Credible, well-managed online databases undoubtedly can provide much information in a user-friendly interface and include many images. However, some proponents of cybertaxonomy have created a bewildering array of different digitization initiatives and online databases that are often cluttered with too many irrelevant features and empty fields (a prime example being EOL). Many also lack clearly defined goals and add to the taxonomic impediment by diverting time of taxonomists away from scientific research. We argue that only initiatives and tools that help with overcoming the taxonomic impediment should be subject to a digital upgrade. For example, digital reference collections make it easier to confirm species identifications after an initial identification. All subsequent work such as a species-level revision can also be online but this does not require re-inventing taxonomic procedures and it is not necessary or desirable to include all the conceivable bells and whistles. For example, until there is evidence for overwhelming demand, the label information for unidentified sepsids do not have to be captured and we will resist the temptation of adding 'Tweet' and 'like' buttons on Sepsidnet pages.

## Acknowledgements

We would like to acknowledge the help from numerous colleagues who collected specimens or helped with species identifications for this project. Special thanks go to A. L. Ozerov who confirmed many identifications. The construction of Sepsidnet was supported by MOE grant R-154-000-476-112. We thank Peter Cranston, who as editor of *Systematic Entomology* encouraged this contribution, as well as his coeditors for their comments on earlier drafts.

## References

Ang, Y.C. & Meier, R. (2010) Five additions to the list of Sepsidae (Diptera) for Vietnam: *Perochaeta cuirassa* sp. n., *Perochaeta lobo* sp. n., *Sepsis spura* sp. n., *Sepsis sepsi* Ozerov, 2003 and *Sepsis monostigma* Thompson, 1869. *Zookeys*, **70**, 41–56.

Ang, Y.C., Puniemoorthy, N. & Meier, R. (2008) Secondarily reduced foreleg armature in *Perochaeta dikowi* sp.n. (Diptera: Cyclorrhapha: Sepsidae) due to a novel mounting technique. *Systematic Entomology*, **33**, 552–559.

Berendsohn, W.G. & Güntsch, A. (2012) OpenUp! Creating a cross-domain pipeline for natural history data. *Zookeys*, **209**, 47–54.

Bertone, M.A., Blinn, R.L., Stanfield, T.M., Dew, K.J., Seltmann, K.C. & Deans, A.R. (2012) Results and insights from the NCSU Insect Museum GigaPan project. *Zookeys*, **209**, 115–132.

Blagoderov, V. & Smith, V. (2012) No specimen left behind: mass digitization of natural history collections. *Zookeys*, **209**, 1–267.

Blagoderov, V., Kitching, I.J., Livermore, L., Simonsen, T.J. & Smith, V.S. (2012) No specimen left behind: industrial scale digitization of natural history collections. *Zookeys*, **209**, 133–146.

Blanckenhorn, W.U., Reusch, T. & Muehlauser, C. (1998) Fluctuating asymmetry, body size and sexual selection in the dung fly *Sepsis cynipsea* – testing the good genes assumptions and predictions. *Journal of Evolutionary Biology*, **11**, 735–753.

Blanckenhorn, W.U., Morf, C., Mühlhauser, C. & Reusch, T. (1999) Spatiotemporal variation in selection on body size in the dung fly *Sepsis cynipsea*. *Journal of Evolutionary Biology*, **12**, 563–576.

Blanckenhorn, W.U., Mühlhauser, C., Morf, C., Reusch, T. & Reuter, M. (2000) Female choice, female reluctance to mate and sexual selection on body size in the dung fly *Sepsis cynipsea*. *Ethology*, **106**, 577–593.

Blanckenhorn, W.U., Hosken, D.J., Martin, O.Y., Reim, C., Teuschl, Y. & Ward, P.I. (2002) The costs of copulating in the dung fly *Sepsis cynipsea*. *Behavioral Ecology*, **13**, 353–358.

Blanckenhorn, W.U., Kraushaar, U., Teuschl, Y. & Reim, C. (2004) Sexual selection on morphological and physiological traits and fluctuating asymmetry in the black scavenger fly *Sepsis cynipsea*. *Journal of Evolutionary Biology*, **17**, 629–641.

Blanckenhorn, W.U., Dixon, A.F.G., Fairbairn, D.J. *et al.* (2007) Proximate causes of Rensch's rule: does sexual size dimorphism in arthropods result from sex differences in development time? *American Naturalist*, **169**, 245–257.

Bowsher, J.H. & Nijhout, H.F. (2007) Evolution of novel abdominal appendages in a sepsid fly from histoblasts, not imaginal discs. *Evolution & Development*, **9**, 347–354.

Bowsher, J.H. & Nijhout, H.F. (2009) Partial co-option of the appendage patterning pathway in the development of abdominal appendages in the sepsid fly *Themira biloba*. *Development Genes and Evolution*, **219**, 577–587.

Bowsher, J.H., Ang, Y., Ferderer, T. & Meier, R. (2013) Deciphering the evolutionary history and developmental mechanisms of a complex sexual ornament: the abdominal appendages of Sepsidae (Diptera). *Evolution*, DOI: 10.1111/evo.12006.

Courtney, G.W., Sinclair, B.J. & Meier, R. (2000) Morphology and terminology of Diptera larvae. *Contributions to a Manual of Palaearctic Diptera, Vol. 1* (ed. by L. Papp and B. Darvas), pp. 85–161. Science Herald, Budapest.

Cumming, J.M. & Wood, D.M. (2009) Adult morphology and terminology. *Manual of Central American Diptera, Vol. 1*. (ed. by B.V. Brown, A. Borkent, J.M. Cumming, D.M. Wood, N.E. Woodley and M.A. Zumbado), pp. 9–50. NRC Research Press, Ottawa, Canada.

Cranston, P.S. (2005) Ancient and modern – toolboxes for e-bugs. *Systematic Entomology*, **30**, 183–185.

Dietrich, C., Hart, J., Raila, D., *et al.* (2012) InvertNet: a new paradigm for digital access to invertebrate collections. *Zookeys*, **209**, 165–181.

Eberhard, W.G. (1999) Mating systems of sepsid flies and sexual behavior away from oviposition sites by *Sepsis neocynipsea* (Diptera: Sepsidae). *Journal of the Kansas Entomological Society*, **72**, 129–130.

Eberhard, W.G. (2000) Behavior and reproductive status of *Microsepsis armillata* (Diptera: Sepsidae) flies away from oviposition sites. *Annals of the Entomological Society of America*, **93**, 966–971.

Eberhard, W.G. (2001a) Courtship and multi-stage transfer of material to the female's wings during copulation in *Microsepsis armillata*

- (Diptera: Sepsidae). *Journal of the Kansas Entomological Society*, **74**, 70–78.
- Eberhard, W.G. (2001b) Multiple origins of a major novelty: moveable abdominal lobes in male sepsid flies (Diptera: Sepsidae), and the question of developmental constraints. *Evolution & Development*, **3**, 206–222.
- Eberhard, W.G. (2001c) Species-specific genitalic copulatory courtship in sepsid flies (Diptera, Sepsidae, *Microsepsis*) and theories of genitalic evolution. *Evolution*, **55**, 93–102.
- Eberhard, W.G. (2002a) The function of female resistance behavior: intromission by male coercion vs. female cooperation in sepsid flies (Diptera: Sepsidae). *Revista de Biología Tropical*, **50**, 485–505.
- Eberhard, W.G. (2002b) Physical restraint or stimulation? The function(s) of the modified front legs of male *Archisepsis diversiformis* (Diptera, Sepsidae). *Journal of Insect Behavior*, **15**, 831–850.
- Eberhard, W.G. (2002c) The relation between aggressive and sexual behavior and allometry in *Palaeosepsis dentatiformis* flies (Diptera: Sepsidae). *Journal of the Kansas Entomological Society*, **75**, 317–332.
- Eberhard, W.G. (2003) Sexual behavior and morphology of *Themira minor* (Diptera: Sepsidae) males and the evolution of male sternal lobes and genitalic surstyli. *Canadian Entomologist*, **135**, 569–581.
- Eberhard, W.G. (2005) Sexual morphology of male *Sepsis cynipsea* (Diptera: Sepsidae): lack of support for lock-and key and sexually antagonistic morphological coevolution hypotheses. *Canadian Entomologist*, **137**, 551–565.
- Eberhard, W.G. & Huber, B.A. (1998) Copulation and sperm transfer in *Archisepsis* flies (Diptera, Sepsidae) and the evolution of their intromittent genitalia. *Studia Dipterologica*, **5**, 217–248.
- Eberhard, W.G. & Pereira, F. (1996) Functional morphology of male genitalic surstyli in the dungflies *Achisepsis diversiformis* and *A. ecalcarata* (Diptera: Sepsidae). *Journal of the Kansas Entomological Society*, **69**, 43–60.
- Godfray, H.C.J. (2002) Challenges for taxonomy. *Nature*, **417**, 17–19.
- Godfray, H.C.J. (2005) Taxonomy as Information Science. *Proceedings of the California Academy of Sciences*, **56**, 170–181.
- Gotelli, N.J. (2004) A taxonomic wish-list for community ecology. *Philosophical Transactions of the Royal Society B*, **359**, 585–597.
- Hamilton, J.R., Yeates, D.K., Hastings, A., et al. (2006) *On the Fly: An Interactive Key, Guide and Anatomical Atlas for the Australian Fly Fauna*. ABRIS Identification Series, Canberra.
- Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R. & Eisen, M.B. (2008) Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genetics*, **4**, e1000106.
- Haston, E., Cubey, R., Pullan, M., Atkins, H. & Harris, D.J. (2012) Developing integrated workflows for the digitisation of herbarium specimens using a modular and scalable approach. *Zookeys*, **209**, 93–102.
- Hill, A., Guralnick, R., Smith, A. et al. (2012) The notes from nature tool for unlocking biodiversity records from museum records through citizen science. *Zookeys*, **209**, 219–233.
- Ingram, K.K., Laamanen, T., Puniamoorthy, N. & Meier, R. (2008) Lack of morphological coevolution between male forelegs and female wings in *Themira* (Sepsidae : Diptera : Insecta). *Biological Journal of the Linnean Society*, **93**, 227–238.
- Iwasa, M. & Tewari, R.R. (1990) Two new species of the genus *Dicranosepsis* Duda from India (Diptera, Sepsidae). *Japanese Journal of Entomology*, **58**, 794–798.
- Knapp, S. (2008) Taxonomy as a team sport. *The New Taxonomy* (ed. by Q. D. Wheeler), pp. 33–53. CRC Press, Boca Raton.
- Kwong, S., Srivathsan, A. & Meier, R. (2012) An update on DNA barcoding: low species coverage and numerous unidentified sequences. *Cladistics*, **28**, 639–644.
- Lim, G.S., Balke, M. & Meier, R. (2012) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology*, **61**, 195–169.
- Linnaeus, C. (1758) *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*, 10th edition, Vol. 1. Laurentii Salvii, Holmiae.
- Longino, J.T., Coddington, J. & Colwell, R.K. (2002) The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology*, **83**, 689–702.
- Mantle, B.L., La Salle, J. & Fisher, N. (2012) Whole-drawer imaging for digital management and curation of a large entomological collection. *Zookeys*, **209**, 147–163.
- McAlpine, J.F., Peterson, B.V., Shewell, G.E., Teskey, H.J., Vockeroth, J.R. & Wood, D.M. (1981) *Manual of Nearctic Diptera*, Vol. 1. Research Branch Agriculture Canada, Ottawa.
- Meier, R. (1997) A test and review of the empirical performance of the ontogenetic criterion. *Systematic Biology*, **46**, 699–721.
- Meier, R. & Dikow, T. (2004) Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Conservation Biology*, **18**, 478–488.
- Merz, B. & Haenni, J.P. (2000) Morphology and terminology of adult Diptera (other than terminalia). *Contributions to a Manual of Palaearctic Diptera, Vol. 1* (ed. by L. Papp and B. Darvas), pp. 21–52. Science Herald, Budapest.
- Mühlhauser, C. & Blanckenhorn, W.U. (2002) The costs of avoiding matings in the dung fly *Sepsis cynipsea*. *Behavioral Ecology*, **13**, 359–365.
- Mühlhauser, C. & Blanckenhorn, W.U. (2004) The quantitative genetics of sexual selection in the dung fly *Sepsis cynipsea*. *Behaviour*, **141**, 327–341.
- van Oever, J.P. & Gofferje, M. (2012) ‘From Pilot to production’: large scale digitisation project at Naturalis Biodiversity Center. *Zookeys*, **209**, 87–92.
- Ozerov, A.L. (1993) A new genus and three new species of Sepsidae (Diptera) from Vietnam. *Entomological Reviews*, **71**, 123–127.
- Ozerov, A.L. (1995) To the fauna and taxonomy of African Sepsidae (Diptera). *Russian Entomological Journal*, **4**, 127–144.
- Ozerov, A.L. (2005) World catalogue of the family Sepsidae (Insecta: Diptera). *Zoologicheskie Issledovania (Zoological Studies)*, **8**, 1–74.
- Pont, A.C. & Meier, R. (2002) The Sepsidae (Diptera) of Europe. *Fauna Entomologica Scandinavica*, **37**, 1–221.
- Puniamoorthy, N., Su, F.Y.K. & Meier, R. (2008) Bending for love: losses and gains of sexual dimorphisms are strictly correlated with changes in the mounting position of sepsid flies (Sepsidae: Diptera). *BMC Evolutionary Biology*, **8**, 155.
- Puniamoorthy, N., Ismail, M.R.B., Tan, D.S.H. & Meier, R. (2009) From kissing to belly stridulation: comparative analysis reveals surprising diversity, rapid evolution, and much homoplasy in the mating behaviour of 27 species of sepsid flies (Diptera: Sepsidae). *Journal of Evolutionary Biology*, **22**, 2146–2156.
- Puniamoorthy, N., Kotrba, M. & Meier, R. (2010) Unlocking the “Black box”: internal female genitalia in Sepsidae (Diptera) evolve fast and are species-specific. *BMC Evolutionary Biology*, **10**, 275.
- Puniamoorthy, N., Schäfer, M.A. & Blanckenhorn, W.U. (2012) Sexual selection accounts for the geographic reversal of sexual size dimorphism in the dung fly, *Sepsis punctum* (Diptera: Sepsidae). *Evolution*, **66**, 2117–2126.
- Reusch, T. & Blanckenhorn, W.U. (1998) Quantitative genetics of the dung fly *Sepsis cynipsea*: Cheverud’s conjecture revisited. *Heredity*, **81**, 111–119.
- Schmidt, S., Balke, M. & Lafogler, S. (2012) DScan—a high-performance digital scanning system for entomological collections. *Zookeys*, **209**, 183–191.

- Schuh, R.T. (2012) Integrating specimen databases and revisionary systematics. *Zookeys*, **209**, 255–267.
- Scoble, M.J. (2004) Unitary or unified taxonomy? *Philosophical Transactions of the Royal Society B*, **359**, 699–710.
- Sinclair, B.J. (2000) Morphology and terminology of Diptera male genitalia. *Contributions to a Manual of Palaearctic Diptera, Vol. 1* (ed. by L. Papp and B. Darvas), pp. 53–74. Science Herald, Budapest.
- Su, K.F.Y., Kutty, S.N. & Meier, R. (2008) Morphology versus molecules: the phylogenetic relationships of Sepsidae (Diptera: Cyclorrhapha) based on morphology and DNA sequence data from ten genes. *Cladistics*, **24**, 902–916.
- Tan, D.S.H., Ang, Y., Lim, G.S., Ismail, M.R.B. & Meier, R. (2010) From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). *Zoologica Scripta*, **39**, 51–61.
- Tan, D.S.H., Ng, S.R. & Meier, R. (2011) New information on the evolution of mating behaviour in Sepsidae (Diptera) and the cost of male copulations in *Saltella sphondylii*. *Organisms, Diversity and Evolution*, **11**, 253–261.
- Tegelberg, R., Haapala, J., Mononen, T., Pajari, M. & Saarenmaa, H. (2012) The development of a digitising service centre for natural history collections. *Zookeys*, **209**, 75–86.
- Teuschl, Y. & Blanckenhorn, W.U. (2007) The reluctant fly: what makes *Sepsis cynipsea* females willing to copulate? *Animal Behaviour*, **73**, 85–97.
- Teuschl, Y., Hosken, D.J. & Blanckenhorn, W.U. (2007) Is reduced female survival after mating a by-product of male-male competition in the dung fly *Sepsis cynipsea*? *BMC Evolutionary Biology*, **7**, 194.
- Teuschl, Y., Reim, C. & Blanckenhorn, W.U. (2010a) No size-dependent reproductive costs in male black scavenger flies (*Sepsis cynipsea*). *Behavioral Ecology*, **21**, 85–90.
- Teuschl, Y., Reim, C., Meister, B., Egger, J. & Blanckenhorn, W.U. (2010b) Strategic ejaculation in the black scavenger fly *Sepsis cynipsea* revisited: copula duration as a function of sperm depletion and body size. *Ethology*, **116**, 1118–1126.
- Thomer, A., Vaidya, G., Guralnick, R., Bloom, D. & Russell, L. (2012) From documents to datasets: a MediaWiki-based method of annotating and extracting species observations in century-old field notebooks. *Zookeys*, **209**, 235–253.
- Tulig, M., Tarnowsky, N., Bevans, M., Anthony, K. & Thiers, B.M. (2012) Increasing the efficiency of digitization workflows for herbarium specimens. *Zookeys*, **209**, 103–113.
- Wheeler, Q., Bourgoin, T., Coddington, J. *et al.* (2012) Nomenclatural benchmarking: the roles of digital typification and telemicroscopy. *Zookeys*, **209**, 193–202.
- Wild, G., Pizzari, T. & West, S.A. (2011) Sexual conflict in viscous populations: the effect of the timing of dispersal. *Theoretical Population Biology*, **80**, 398–316.

Accepted 31 January 2013  
 First published online 14 May 2013