Cladistics

Cladistics (2014) 1–9

10.1111/cla.12095

Taxonomy in the electronic age and an e-monograph of the papaya family (Caricaceae) as an example

Fernanda Antunes Carvalho[a,*], Denis Filer[b] and Susanne S. Renner[a]

[a]*Systematic Botany and Mycology, University of Munich (LMU), Menzinger Strasse 67, 80638, Munich, Germany;* [b]*Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK*

Accepted 5 July 2014

Abstract

The need for taxonomists to take full advantage of biodiversity informatics has been clear for at least 10 years. Significant progress has been made in providing access to taxonomic resources online, including images of specimens (especially types), original species descriptions, and georeferenced collection data. However, in spite of persuasive calls for e-monography, there are few, if any, completed project, even though monographic research is the only mechanism for reducing synonymous names, which are estimated to comprise 50% of all published names. Caricaceae is an economically important family of flowering plants from Africa and the Neotropics, best known for the fruit crop papaya. There is a large amount of information on the family, especially on chemistry, crop improvement, genomics, and the sex chromosomes of papaya, but up-to-date information on the 230 names and which species they might belong to was not available. A dynamically updated e-monograph of the Caricaceae now brings together all information on this family, including keys, species descriptions, and specimen data relating the 230 names to 34 species and one hybrid. This may be the first taxonomic monograph of a plant family completely published online. The curated information will be continuously updated to improve the monograph's comprehensiveness and utility.

Introduction

The Plant List (2010) shows 1 040 426 published names for plants, of which 29% are accepted, 25% of unclear status, and 46% considered synonymous with other species names. The problem of synonymous names arises because taxonomists inadvertently name the same species several times, usually because it is widespread and has been collected in far-apart regions and/or because widespread species often are morphologically variable, sometimes in correlation with their environment, making it difficult to assess species status until a dense collection series can be studied. In the flowering plants, there may be 3–4 synonyms for every accepted name (Scotland and Wortley, 2003; Wortley and Scotland, 2004; Paton et al., 2008; The Plant List, 2010). The problem of synonymous names is by no means restricted to plants, although reliable estimates for all eukaryotes are difficult to obtain (Alroy, 2002; Mora et al., 2011). Synonymous names are not a harmless nuisance, and their rate seems to be increasing apace with the rate at which new species are described (Fig. 1). When it comes to conserving species or using species for medical or any other kind of purpose, synonymous names will result in two kinds of errors: they result in wrong, usually narrower, species range estimates than warranted because each name will be associated with its own "species" range; and they make it difficult to find material of, or published information on, a particular biological entity because users cannot know which names refer to which good species.

The assessment, and reassessment, of the status of a name as either a synonym or a good species is done during monographic research. Monographic research is based on bringing together the information pertaining to all names that have ever been published for some group, typically a genus or a family. This will

*Corresponding author:
E-mail address: antunesfc@gmail.com

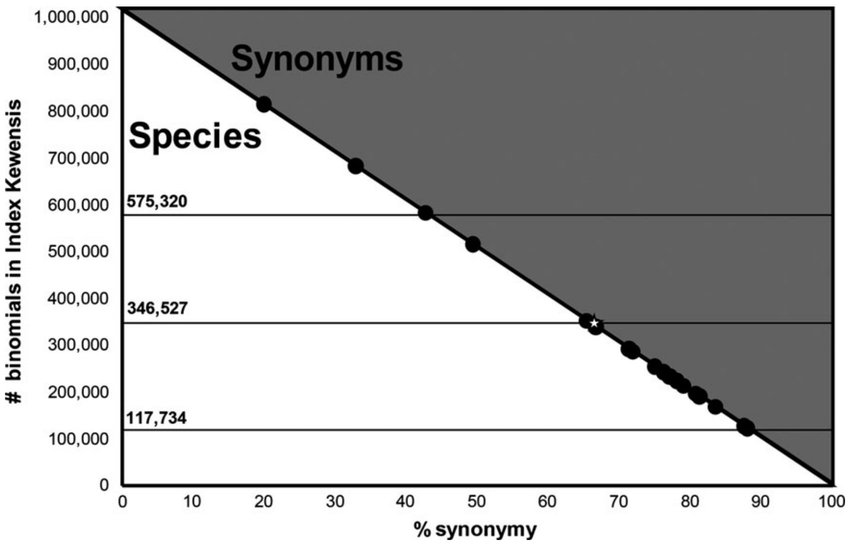

Fig. 1. Relationship between synonymy percentage and number of species from Wortley and Scotland (2004), reproduced with permission of the authors.

include the publication in which a name was first proposed (the so-called protologue), all specimens to which the name has been applied (rightly or wrongly), the accepted names and their synonyms, morphological descriptions for each species, geographical coordinates of relevant collections, chromosome numbers, chemical traits, flowering or fruiting times, and DNA sequences from specimens given one or several of the names in question. A monographer will study specimens, often do some phylogenetic work based on DNA sequences of a representative subset, and reach a conclusion about which names refer to which species. He/she next constructs a key to identify the accepted species and prepares an authoritative list of the accepted and synonymized names. Monography is the only known mechanism for achieving quality control in taxonomy and for reducing the number of synonymous names that clutter up databases and hinder progress in our knowledge of the World's biodiversity and its conservation status.

Because taxonomy is the portal to the entire information available about species, the need for taxonomic research to "move into the electronic age" has long been clear (Bisby et al., 2002; Godfray, 2002; Wilson, 2003; Kress, 2004; Wheeler, 2004; Scotland and Wood, 2012). Indeed, species descriptions of animals and plants are now increasingly being published online (Blagoderov et al., 2010; Knapp, 2010; Penev et al., 2010; Knapp et al., 2011). Monography, however, has not followed suit, in spite of the availability of massive online databases of literature and digitized specimens, wikis, ever cheaper digital photography and microscopy (essential to the study of herbarium specimens), and dedicated platforms for taxonomy, such as the Botanical Research and Herbarium Management System (BRAHMS, http://herbaria.plants.ox.ac.uk/bol/) and Scratchpads (http://scratchpads.eu/). The new "cyber-taxonomy" or "e-taxonomy" (Zauner, 2009; Wheeler and Valdecasas, 2010) is reality only for species descriptions and lists of names but not yet for monographic research (Scotland and Wood, 2012). Although there are several ongoing taxon-centred initiatives (Appendix 1), to our knowledge no revision or monograph of any large group has been completed. The advantages of online monography, such as the possibility of including near-unlimited colour images and the option of updating information, have thus not been realized.

Overview of the electronic monograph of Caricaceae and its underlying database

Here we present a recently completed electronic monograph of a plant family (Caricaceae), the result of research that brought together the available collections with digital libraries, digitized specimen data, and other taxonomic and methodological tools available, including DNA sequencing for barcoding the recognized species (Carvalho and Renner, 2012, 2013).

Caricaceae is a small family of flowering plants from Africa and the Neotropics, best known for the fruit crop *Carica papaya*. The family's economic importance lies not only in the papaya fruit, but also in the production of papain, a cysteine proteinase widely used in food and pharmaceutical industries. A search for the topics "papaya" and "papain" in Web of Knowledge retrieves approximately 20 823 and 42 100 citations, respectively (ReutersISI, 2013). Several Caricaceae are considered as unexploited crops because of their nutritive fruits, high concentration of papain-like enzymes, and resistance to pathogens (Kyndt et al., 2007; Ramos-Martínez et al., 2012). Among these are the so-called highland papayas, species of *Vasconcellea*, a genus thought to be synonymous with *Carica* until Badillo (2000) cleared up their morphological distinctness. Molecular data have revealed that the closest relative of papaya is a clade of four species in Mexico and Guatemala entirely neglected by ecologists and breeders (Carvalho and Renner, 2012). The lack of knowledge before 2012 on the true closest relatives of papaya resulted in the assumption that the highland papayas (*Vasconcellea* species) were the best group to use in papaya improvement (Scheldeman et al., 2011; Coppens d'Eeckenbrugge et al., 2014).

As required in a taxonomic monograph, the e-monograph of Caricaceae (http://herbaria.plants.ox.ac.uk/bol/caricaceae) allocates all names (here 230) to recognized species (here 34 and one hybrid), providing a comprehensive data infrastructure for scientists and non-scientists alike. The database is being developed, managed and published online using BRAHMS (http://herbaria.plants.ox.ac.uk/bol) developed at the University of Oxford. In carrying out this research on the Caricaceae, we added a range of new features to BRAHMS that facilitate cyber-monography, thus emphasizing the importance of close collaborations among taxonomists and bioinformaticians (Stein, 2008).

The e-monograph of Caricaceae and its underlying database store (and make available) data and images on collections, herbarium specimens, literature, and the revised nomenclature (including accepted names, synonyms, *nomina nuda*, illegitimate names, and excluded names). The monographic research resulted in updated circumscriptions of the recognized species, including detailed plates (Fig. 2), and precise geographical distribution of all relevant collections. Links to supportive literature and high-resolution images of type specimens are provided for each species as are cross-references to databases, such as The Plant List, TROPICOS, IPNI, and GBIF. General information on the family, including its ecology, sex chromosomes, and molecular phylogeny, is provided, along with identification keys to all genera and species.

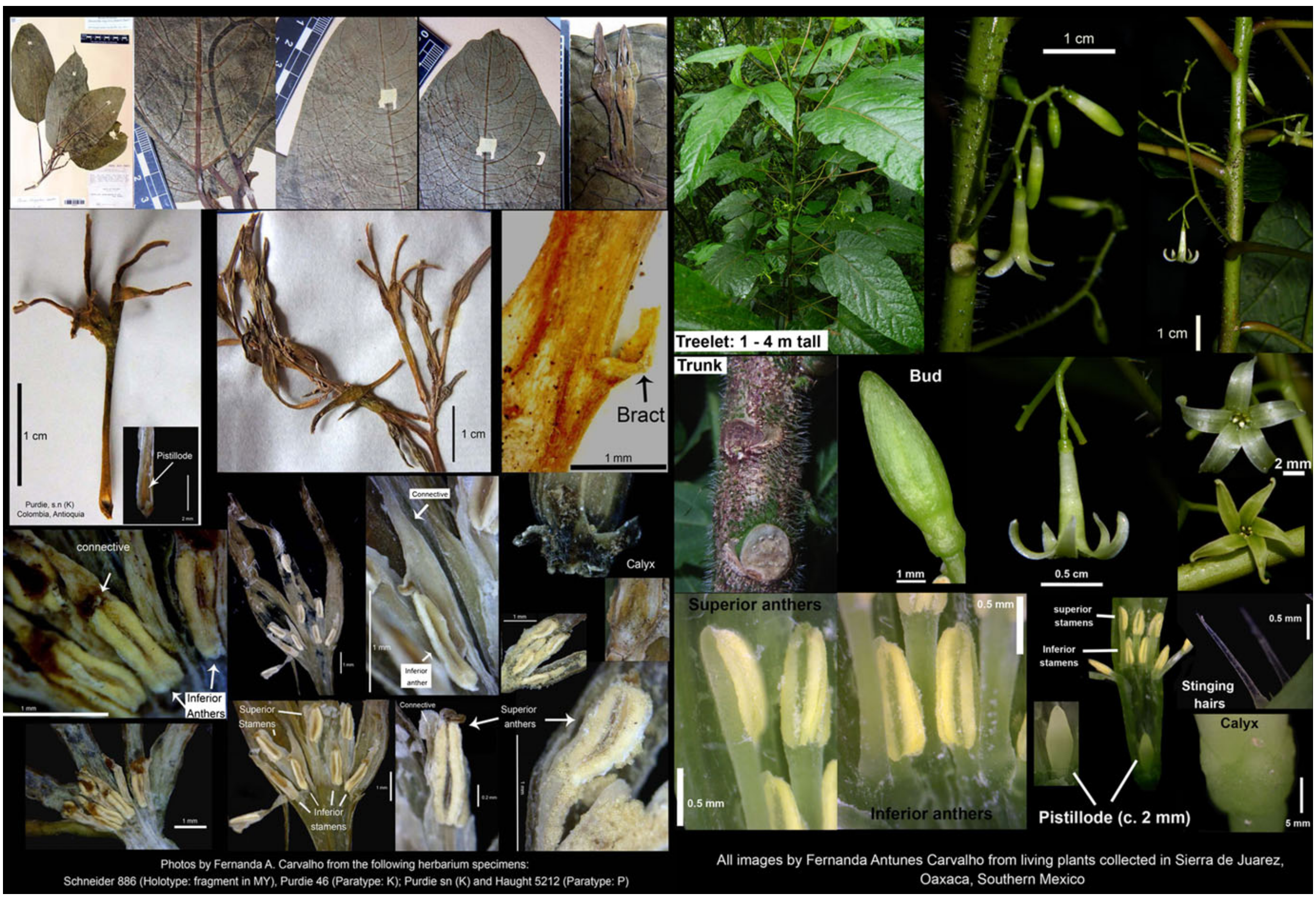

Fig. 2. Examples of species plates used to describe species in the website. To the left are images of details of male flowers and inflorescences based on herbarium specimens of *Vasconcellea longiflora*; to the right, images of living material of *Horovitzia cnidoscoloides*, one of the four little known closest relatives of papaya.

All these data are accessible through BRAHMS online and summarized in Table 1. Searches by taxon, collector, geographical place name, and map area (Fig. 3) generate tables that can also be shown in text format. Images can be grouped and filtered, and viewed at different resolutions. Maps are available using clustered Google Maps or Google Earth, both configurable with zoom features. A detailed description of the methods used to build the e-monograph is given in Appendix 2.

Discussion

Among the challenges for taxonomy today are to incorporate results and insights from molecular phylogenetic work and to tackle the problem of the 46–50% synonymous names already published (Scotland and Wortley, 2003; Wortley and Scotland, 2004; The Plant List, 2010). Both challenges can only be addressed through monographic work in which species and genus circumscriptions are vetted and updated, based on the study of specimens and consideration of relevant phylogenetic results on relationships.

Reliably circumscribed and named species are also required to fulfil the promise of DNA barcodes, at least if that promise is finding names for unidentified specimens via matching of short DNA sequences (obviously, one can also match unnamed material to unnamed sequences). Simply increasing the rate of species discovery, while important, does not address either of these challenges because naming a newly discovered species does not require a complete assessment of all existing names that might apply (which would often

Table 1
Summary of the Caricaceae e-monograph data available online as of 13 February 2014. Invalid, Illegitimate, Excluded and Uncertain names are not included in this table

Genera (6)	Species (34 + 1 hybrid)	Synonyms (160)	Collections examined (2950)	Specimens examined (4337)	Georeferenced collections (2204)	Images (10 988)
Cylicomorpha	*C. parviflora* Urb.	1	36	57	28	246
	C. solmsii Urb.	1	18	27	12	158
Carica	*C. papaya* L.	21	590	773	30	1911
Horovitzia	*Horovitzia cnidoscoloides*	1	97	26	19	68
Jarilla	*Jarilla chocola* Standl.	1	37	52	36	136
	Jarilla caudata (Brandegee) Standl.	4	50	62	48	159
	Jarilla heterophylla (Cerv. ex La Llave) Rusby	4	71	85	69	219
Jacaratia (7 species)	*J. digitata* (Poepp. & Endl.) Solms-Laub.	3	178	251	167	512
	J. spinosa (Aubl.) A.DC.	8	209	329	190	849
	J. chocoensis A.H.Gentry & Forero	0	15	21	15	31
	J. corumbensis Kuntze	3	41	86	34	281
	J. dolichaula (Donn.Sm.) Woodson	1	128	172	120	450
	J. mexicana A. DC.	7	142	158	132	500
	J. heptaphylla (Vell.) A.DC.	1	30	37	26	134
Vasconcellea (21 species and 1 hybrid)	*V. candicans* (A.Gray) A.DC.	3	26	38	19	141
	V. cauliflora (Jacq.) A.DC.	8	111	159	87	452
	V. crassipetala (V.M.Badillo) V.M.Badillo	1	6	16	5	55
	V. glandulosa A.DC	9	80	153	71	419
	V. goudotiana Triana & Planch.	4	20	35	11	107
	V. horovitziana (V.M.Badillo) V.M.Badillo	1	13	34	3	125
	V. longiflora (V.M.Badillo) V.M.Badillo	1	6	10	2	26
	V. microcarpa (Jacq.) A.DC.	22	401	774	336	1508
	V. monoica (Desf.) A.DC.	7	32	70	14	156
	V. omnilingua (V.M.Badillo) V.M.Badillo	1	2	3	1	16
Vasconcellea (21 species and 1 hybrid)	*V. palandensis* (V.M.Badillo, Van den Eynden & Van Damme) V.M.Badillo	1	3	6	3	27
	V. parviflora A.DC.	5	61	109	46	323
	V. pubescens A.DC.	10	102	230	68	501
	V. pulchra (V.M.Badillo) V.M.Badillo	1	13	36	10	98
	V. quercifolia A.St.-Hil.	14	157	253	114	675
	V. sphaerocarpa (García-Barr. & Hern.Cam.) V.M.Badillo	1	13	25	12	60
	V. sprucei (V.M.Badillo) V.M.Badillo	1	27	57	7	184
	V. stipulata (V.M.Badillo) V.M.Badillo	1	19	40	15	115
	V. weberbaueri (Harms) V.M.Badillo	1	8	31	7	82
	V. chilensis Planch. ex A.DC.	3	16	38	6	86
	V. × heilbornii (V.M.Badillo) V.M.Badillo	9	31	84	13	188

Values in parentheses are the total numbers available in the database.

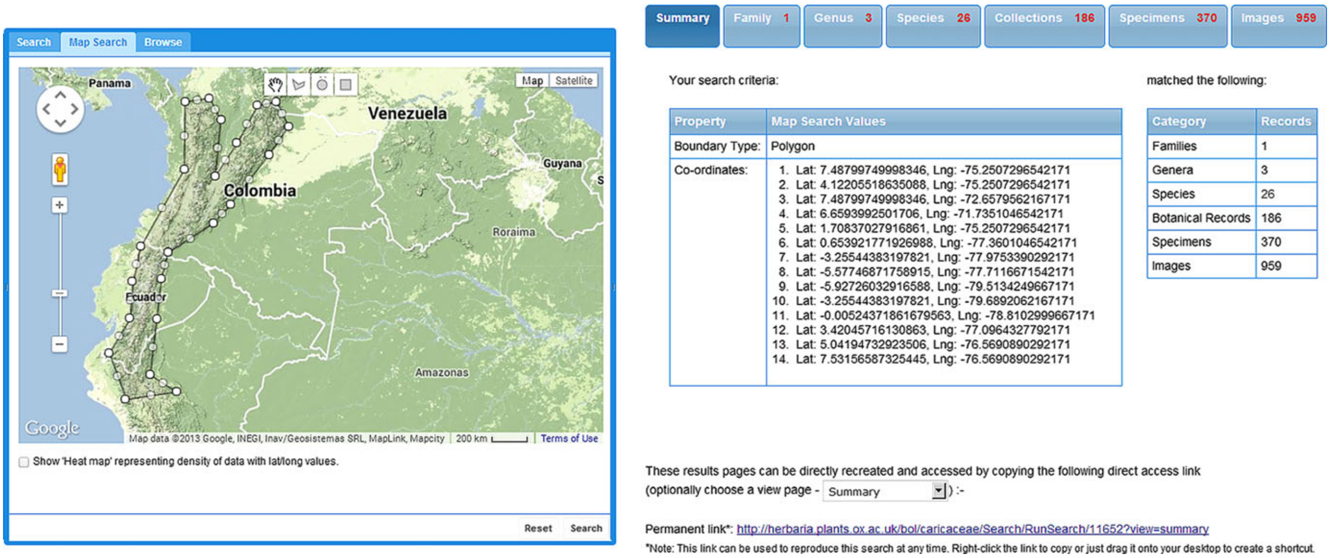

Fig. 3. Map search in BRAHMS. The left figure shows a polygon that can be drawn by the user to delimit the area of interest, in this case the Andes from northern Peru to northern Colombia. To the left is a summary of the results, which includes number of genera, collections, specimens, and images available in the database. It also provides the coordinates of the polygons, which can be used to create a shape file.

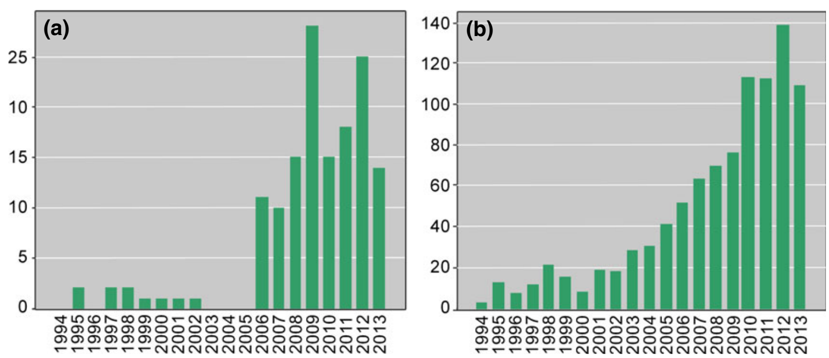

Fig. 4. (a) Number of published studies with the topic search fields "Caricaceae" and "genome"; a total of 168 records were found. (b) Number of published studies on bioclimatic modelling per year; in total 1002 records were found. (Web of Knowledge, accessed 18 November 2013.)

take too much time). It is therefore likely that as the number of species descriptions increases (Costello et al., 2013), so does the number of newly created synonyms (Fig. 1).

A well-resolved, expert-vetted nomenclature and detailed information on the distribution of species are of great importance for many fields of research (Yesson et al., 2007; Bortolus, 2008; Patterson et al., 2010; Lis and Lis, 2011; Santos and Branco, 2012). However, high-quality data produced by taxonomists in revisions and monographs are of little use unless they are widely accessible (Kress, 2004). This is especially important for economically important groups, which often are also groups with a high rate of nomenclatural changes (as is the case for Caricaceae). Open-access information to this highly organized set of online data and images for the Caricaceae benefits the scientific community broadly as well as those working on the food and medicinal aspects of the family. This includes the community of herbarium curators, researchers focusing on papaya genomics (Fig. 4a), breeders, and the non-scientific public. In addition, georeferenced specimens are the basis for the growing field of bioclimatic modelling (Fig. 4b) and for a reliable baseline to document the effects of ongoing climatic changes on plant ranges.

In the case of the papaya family, the most recent taxonomic accounts were by Victor Manuel Badillo (1920–2008; http://herbaria.plants.ox.ac.uk/bol/caricaceae#badillo) a Venezuelan taxonomist who dealt with *c.* 200 names described in the family, 64 of these basionyms (meaning that the remainder result from changing generic concepts). The work of Badillo (1971, 1993, 2000) is poorly accessible, and since his last publication 13 years ago (Badillo, 2001) no further taxonomic work on the Caricaceae has been published.

Meanwhile, molecular work on the family took off (Van Droogenbroeck et al., 2002; Kyndt et al., 2005a, b; Carvalho and Renner, 2012). The IUCN Red List of Threatened Species (IUCN, 2013) lists six endangered species of Caricaceae, none under the correct name; the new e-monograph available at http://herbaria.plants.ox.ac.uk/bol/caricaceae now includes updated information on the vulnerability of species that together with the geographical and ecological information should help in conservation efforts.

A major problem in building the Caricaceae database was to gather data from different herbaria that use different standards and field definitions. This occurs despite proposals to standardize biodiversity databases such as HISPID (Herbarium Information Standards and Protocols for Interchange of Data; Conn, 1995), Darwin Core (Wieczorek et al., 2012), and ABCD (TDWG, 2013). The last two are recommended by the Taxonomic Database Working Group (TDWG). New software and platforms are being developed each year by different institutions, but the communication among them is not being improved at the same pace. Database mapping can be used to integrate two distinct data models. However, if the same piece of information is digitized slightly differently among institutes, queries that address multiple databases may not be adequately solved (Willemse et al., 2008). Standardization in data entry would increase the value of freely available biodiversity data by facilitating the use and re-use, distribution, and integration of this information. Initiatives, like speciesLink (http://splink.cria.org.br/), which integrates primary data from biological collections deposited in different scientific collections using Darwin Core standards, are laudable and should be linked to worldwide programs, such as the Encyclopedia of Life (EOL, http://eol.org/) and the Global Biodiversity Information Facility (GBIF, http://www.gbif.org/).

With the development of digital photography technology, professional and amateur photographers are unknowingly discovering and informally documenting new species by placing images of plants and animals in online image databases (Winterton et al., 2012). Species identification via images is becoming more and more important, and freely available e-monographs that combine images (which will be picked up, for example, by "google images") with professionally curated names and descriptions can support such citizen science. Systematists, however, are not yet producing freely accessible taxonomic monographs (or floras) despite more than 10 years' worth of admonitions (Bisby et al., 2002; Godfray, 2002; Wilson, 2003; Kress, 2004; Wheeler et al., 2004; Scotland and Wood, 2012). This probably has two (related) reasons: the small number of people in a position to populate the existing cyber-infrastructure with data and the pressure for publishing in citable journals or monograph series. Overcoming the second problem will require citation of online publications as has long been standard in physics, mathematics, computer science, and chemistry.

The e-monograph of Caricaceae includes all features of a traditional monograph (Marhold et al., 2013), and is a single portal to access information on all taxon names, thus facilitating the communication among different groups of researchers. Different from any hard-copy work, however, it is rapidly searchable and links specimens and species to other kinds of data and knowledge; for example, specimens used in DNA isolation are linked to the respective GenBank entries. Another obvious advantage of online monography is the ease of updating. A newly discovered species, a range expansion, or a newly available set of images can easily be added to an online database, but not to a printed monograph. E-monographs will greatly improve access to knowledge about species, while at the same time feeding other databases with invaluable information for scientific research, society, and industry. As John Kress (2004, p. 2152 and 2127) envisaged, "With remote wireless communication the field botanist will be able to immediately compare newly collected plants with type specimens and reference collections archived and digitized in museums thousands of miles away. ...[The e-monographs] of the future, including web-based, computer-based, image-based, and even DNA-based products, are ... fulfilling new functions that paper-based and word-based floras of the past could never attain."

Acknowledgements

We thank all herbaria that are providing open access to images of specimens, indispensable for the development of e-taxonomy; all curators of the herbaria visited by F.A.C. during the development of the monograph; Andrew Liddell at Plant Sciences, Oxford, for his work on the BRAHMS online system; Carmen Benítez for useful information on Victor Badillo and for giving F.A.C. access to all his literature, including original hand-writings; Theodor C. H. Cole provided editorial support in reviewing the e-monograph.

References

Alroy, J., 2002. How many named species are valid? Proc. Natl Acad. Sci. USA 99, 3706–3711.

Badillo, V.M., 1971. Monografia de la familia Caricaceae. Asociación de profesores, Universidad Central de Venezuela, Maracay.

Badillo, V.M., 1993. Caricaceae. Segundo esquema. Rev. Fac. Agron. (UCV) 43, 1–111.

Badillo, V.M., 2000. *Vasconcella* St.-Hil. (Caricaceae) con la rehabilitación de este último. Ernstia 10, 74–79.

Badillo, V.M., 2001. Nota correctiva *Vasconcellea* St. Hill. y no Vasconcella (Caricaceae). Ernstia 11, 75–76.

Bisby, F.A., Shimura, J., Ruggiero, M., Edwards, J., Haeuser, C., 2002. Taxonomy, at the click of a mouse. Nature 418, 367.

Blagoderov, V., Brake, I., Georgiev, T., Penev, L., Roberts, D., Ryrcroft, S., Scott, B., Agosti, D., Catapano, T., Smith, V.S., 2010. Streamlining taxonomic publication: a working example with Scratchpads and ZooKeys. ZooKeys 28, 17–28.

Bortolus, A., 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37, 114–118.

Carvalho, F.A., Renner, S.S., 2012. A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. Mol. Phylogenet. Evol. 65, 46–53.

Carvalho, F.A., Renner, S., 2013. Correct names for some of the closest relatives of *Carica papaya*: a review of the Mexican/Guatemalan genera *Jarilla* and *Horovitzia*. PhytoKeys 29, 63–74.

Conn, B.J., 1995. HISPID – Herbarium Information Standards and Protocols for Interchange of Data. Version 3. Retrieved 13 December 2013 from http://plantnet.rbgsyd.nsw.gov.au/HISCOM/HISPID/HISPID3/H3.html.

Coppens d'Eeckenbrugge, G., Drew, R., Kyndt, T., Scheldeman, X., 2014. *Vasconcellea* for papaya improvement. In: Ming, R., Moore, P.H. (Eds.), Genetics and Genomics of Papaya, 1st edn. Springer, New York, pp. 47–79.

Costello, M.J., May, R.M., Stork, N.E., 2013. Can we name Earth's species before they go extinct? Science 339, 413–416.

Garcia-Milagros, E., Funk, V.A., 2010. Improving the use of information from museum specimens: using Google Earth to georeference Guiana Shield specimens in the US National Herbarium. Front. Biogeogr. 2.3, 71–77.

Godfray, H.C.J., 2002. Challenges for taxonomy. Nature 417, 17–19.

IUCN., 2013. The IUCN Red List of Threatened Species. Version 2013.2. www.iucnredlist.org. Retrieved 8 December 2013 from www.iucnredlist.org.

Knapp, S., 2010. Four new vining species of *Solanum* (Dulcamaroid Clade) from montane habitats in tropical America. PLoS One 5, 1–8, e10502.

Knapp, S., McNeill, J., Turland, N.J., 2011. Changes to publication requirements made at the XVIII International Botanical Congress in Melbourne – what does e-publication mean for you? Taxon 60, 250.

Kress, W.J., 2004. Paper floras: how long will they last? A review of flowering plants of the neotropics Am. J. Bot. 91, 2124–2127.

Kyndt, T., Romeijn-Peeters, E., Van Droogenbroeck, B., Romero-motochi, J.P., Gheysen, G., Goetghebeur, P., 2005a. Species relationships in the genus *Vasconcellea* (Caricaceae) based on molecular and morphological evidence. Am. J. Bot. 92, 1033–1044.

Kyndt, T., Van Droogenbroeck, B., Romeijn-Peeters, E., Romero-Motochi, J.P., Scheldeman, X., Goetghebeur, P., Van Damme, P., Gheysen, G., 2005b. Molecular phylogeny and evolution of Caricaceae based on rDNA internal transcribed spacers and chloroplast sequence data. Mol. Phylogenet. Evol. 37, 442–459.

Kyndt, T., Van Damme, E.J.M., Van Beeumen, J., Gheysen, G., 2007. Purification and characterization of the cysteine proteinases in the latex of *Vasconcellea* spp. FEBS J. 274, 451–462.

Lis, J.A., Lis, B., 2011. Is accurate taxon identification important for molecular studies? Several cases of faux pas in pentatomoid bugs (Hemiptera: Heteroptera: Pentatomoidea) Zootaxa 2932, 47–50.

Marhold, K., Stuessy, T., Agababian, M., Agosti, D., Alford, M.H., Crespo, A., Crisci, J.V, Dorr, L.J., Ferencová, Z., Frodin, D., Geltman, D.V., Kilian, N., Linder, H.P., Lohmann, L.G., Oberprieler, C., Penev, L., Smith, G.F., Thomas, W., Tulig, M., Turland, N., Zhang, X.-C., 2013. The Future of Botanical Monography: report from an international workshop, 12–16 March 2012, Smolenice, Slovak Republic. Taxon 62, 4–20.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., Worm, B., 2011. How many species are there on Earth and in the ocean? PLoS Biol. 9, e1001127.

Paton, A.J., Brummitt, N., Govaerts, R., Harman, K., Hinchcliffe, S., Allkin, B., Lughadha, E.N., 2008. Towards Target 1 of the Global strategy for plant conservation: a working list of all known plant species – progress and prospects. Taxon 57, 602–611.

Patterson, D.J., Cooper, J., Kirk, P.M., Pyle, R.L., Remsen, D.P., 2010. Names are key to the big new biology. Trends Ecol. Evol. 25, 686–691.

Penev, L., Kress, W.J., Knapp, S., Li, D.-Z., Renner, S., 2010. Fast, linked, and open – the future of taxonomic publishing for plants: launching the journal PhytoKeys. PhytoKeys 14, 1–14.

Ramos-Martínez, E.M., Herrera-Ramírez, A.C., Badillo-Corona, J.A., Garibay-Orijel, C., González-Rábade, N., Oliver-Salvador, M.D.C., 2012. Isolation of cDNA from *Jacaratia mexicana* encoding a mexicain-like cysteine protease gene. Gene 502, 60–68.

ReutersISI, T., 2013. Web of Knowledge. Retrieved 19 October 2013 from http://apps.webofknowledge.com/.

Santos, A.M., Branco, M., 2012. The quality of name-based species records in databases. Trends Ecol. Evol. 27, 6–7.

Scheldeman, X., Kyndt, T., Coppens d'Eeckenbrugge, G., Ming, R., Drew, R., Van Droogenbroeck, B.V., Van Damme, P., Moore, P.H., 2011. *Vasconcellea*. In: Kole, C. (Ed.), Wild Crop Relatives: Genomic and Breeding Resources. Tropical and Subtropical Fruits. Springer, Berlin, pp. 213–249.

Scotland, R.W., Wood, J.R.I., 2012. Accelerating the pace of taxonomy. Trends Ecol. Evol. 27, 415–416.

Scotland, R.W., Wortley, A.H., 2003. How many species of seed plants are there? Taxon 52, 101–104.

Stein, L.D., 2008. Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. Nat. Rev. Genet. 9, 678–688.

TDWG., 2013. Biodiversity information standards. Retrieved 16 November 2013 from http://rs.tdwg.org/dwc/.

The Plant List., 2010. Version 1. Retrieved 10 December 2013 from http://www.theplantlist.org/.

Ung, V., Dubus, G., Zaragüeta-Bagils, R., Vignes-Lebbe, R., 2010. Xper2: introducing e-taxonomy. Bioinformatics 26, 703–704.

Van Droogenbroeck, B., Breyne, P., Goetghebeur, P., Romeijn-Peeters, E., Kyndt, T., Gheysen, G., 2002. AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador. Theor. Appl. Genet. 105, 289–297.

Wheeler, Q.D., 2004. Taxonomic triage and the poverty of phylogeny. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 571–583.

Wheeler, Q.D., Valdecasas, A.G., 2010. Cybertaxonomy and ecology. Nat. Educ. Knowl. 1, 6.

Wheeler, Q.D., Raven, P.H., Wilson, E.O., 2004. Taxonomy: impediment or expedient? Science 303, 285.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin Core: an evolving community-developed biodiversity data standard. PLoS One 7, e29715.

Willemse, L.P.M., Welzen, P.C., Van Mols, J.B., Taxon, S., May, N., 2008. Standardisation in data-entry across databases: avoiding babylonian confusion. Taxon 57, 343–345.

Wilson, E.O., 2003. The encyclopedia of life. Trends Ecol. Evol. 18, 77–80.

Winterton, S.L., Guek, H.P., Brooks, S.J., 2012. A charismatic new species of green lacewing discovered in Malaysia (Neuroptera, Chrysopidae): the confluence of citizen scientist, online image database and cybertaxonomy. ZooKeys 11, 1–11.

Wortley, A.H., Scotland, R.W., 2004. Synonymy, sampling and seed plant numbers. Taxon 53, 478–480.

Yesson, C., Brewer, P.W., Sutton, T., Caithness, N., Pahwa, J.S., Burgess, M., Gray, W.A., White, R.J., Jones, A.C., Bisby, F.A., Culham, A., 2007. How global is the global biodiversity information facility? PLoS One 2, e1124.

Zauner, H., 2009. Evolving e-taxonomy. BMC Evol. Biol. 9, 141.

Appendix 1
Weblinks and short descriptions of e-taxonomy projects

Only websites aiming to be a taxonomic monograph and already containing considerable information are included. All weblinks were accessed on 6 April 2014.

Name of the website and weblink	Short description
Hymenopteran systematics http://hymenoptera.ucr.edu/mx-database	Detailed information on the main groups of Hymenoptera and a dichotomous key for identification of North American eulophid genera. Missing distribution data and images of specimens
EuphORBia http://www.euphorbiaceae.org/pages/data_portal.html	General information on the genus *Euphorbia*. Missing identification keys, distribution maps, and descriptions of the species. Also, there is no information or images of specimens and the project seems not to have advanced since 2010
Milichiidae online http://milichiidae.info/content/milichiella-online-revision	Revision of Milichiidae or “freeloader” flies including descriptions, dichotomous keys, list of specimens, and distribution maps for most of the taxa.
Bombus Bumblebees of the world http://www.nhm.ac.uk/research-curation/research/projects/bombus/	A taxonomic review of the group, including an overview of all bumblebees species with interactive keys, distribution maps, and descriptions for most of the taxa
Lacistemataceae http://tolweb.org/Lacistemataceae	A short overview of the plant family Lacistemataceae including nomenclature for all taxa. Detailed species descriptions with images, information on vouchers, and distribution maps are missing for most of the taxa
Solanaceae Source. A global taxonomic resource for the nightshade family http://solanaceaesource.org/	An introduction, including general morphology, distribution, and other data. Also information on taxonomic status of genera and species. Distribution maps and images of many taxa are missing, and the monograph so far covers a small percentage of the *c*. 4000 species
FishBase http://www.fishbase.org/	A database of the world’s fishes with information on common and scientific names, descriptions and environmental requirements of each taxon, images, and distribution maps
WoRMS World Register of Marine Species http://www.marinespecies.org/	Detailed information on nomenclature (95% of accepted names in the database were already checked). Missing descriptions and illustrations of many taxa
Miconieae http://sweetgum.nybg.org/melastomataceae/	Descriptions, phenology, distribution, and nomenclature of many taxa from the tribe Miconieae, which includes *c*. 1800 species. Identification keys and descriptive images of each accepted name are missing, although images of specimens are available
The genus *Leucaena* http://herbaria.plants.ox.ac.uk/bol/leucaena	Monograph of the genus *Leucaena* with two dichotomous keys (using vegetative, floral, or fruit characters), detailed species descriptions, nomenclature, images of each accepted species, and specimen list. Missing images of specimens
ILDIS International Legume Database & Information Service http://www.ildis.org/	Taxonomic database of Fabaceae with list of taxa and taxon status. Includes short descriptions for the accepted names but lacks images
$Xper^2$ http://lis-upmc.snv.jussieu.fr/xper2/infosXper2Bases/en/liste-bases.php	List of knowledge bases using $Xper^2$, i.e. a list of interactive keys for different taxonomic groups. The keys constitute morphological databases, many of them well illustrated. No information on the taxonomic status of the names is provided, and there are no images of specimens
Scratchpads—Biodiversity online http://scratchpads.eu/explore/sites-list	List of websites produced with scratchpads. Although the number of available websites is large, few are active projects with much information, while many are rather incomplete and appear stagnant
BRAHMS online websites http://herbaria.plants.ox.ac.uk/bol/brahms/Websites	List of websites published using BRAHMS. Most are databases with name list and sometimes information on taxonomic status

Appendix 2
Methods used to produce the e-monograph of Caricaceae

The database is developed, managed, and published online using the Botanical Research and Herbarium Management System (BRAHMS, http://herbaria.plants.ox.ac.uk/bol/) developed at the University of Oxford. The principal reasons for choosing this software were (i) its established use in more than 60 countries around the world, facilitating communication among databases and researchers at different institutions, (ii) a user-friendly interface with many tutorials, (iii) freely available resources, and (iv) presence of a powerful module for publishing data online.

We imported into BRAHMS draft lists of taxon names available in the International Plant Names Index (http://www.ipni.org/) and TROPICOS (http://www.tropicos.org/). Duplicated names were marked and selected for deletion using the BRAHMS editing function "Tag identical entries". Protologues were then located on the web and linked to each name in the database. Main providers of old relevant literature at this stage were Botanicus (http://www.botanicus.org/) and BHL (http://www.biodiversitylibrary.org/), both dynamically accessed using BRAHMS web toolbar links. Smaller online libraries, such as Internet Archive (http://archive.org/) and Gallica (http://gallica.bnf.fr/), were also important for texts not found elsewhere. Protologues not available online were acquired using the library facilities of the University of Munich. Following the entry of the protologues, details of type specimens were initially entered in the database following the most recent taxonomic work on Caricaceae (Badillo, 1971, 1993) and updated later. Data relevant to nomenclature and taxonomic decisions, such as synonymization, the taxonomic status of each name, and legitimacy, were edited using further formatting tools in BRAHMS. We also kept in the database not validly published names (*nomina nuda*) because some still populate other digital databases.

Web links to high-resolution images (as provided by some herbaria) were then added to the specimen records. Herbarium specimens form the base of this e-monograph, and the first author photographed all specimens she could find in relevant herbaria of Latin America (BHCB, GUADA, HUEFS, IBUG, MBM, MEXU, MY, R, RB, UPCB, VEN), North America (NY, F), and Europe (B, BM, GB, K, M, OXF, P, W, WU, GENT, S), either in loaned material or during personal visits. CGE, FI, INPA, IAN, and MG provided images of important Caricaceae specimens. At least two images of each specimen were taken: first, the label (to facilitate data entering) and, second, the complete specimen. Images of details, such as stipules, trichomes, and venation, were also taken for some specimens. All images were processed and renamed using tools provided in BRAHMS, following the tutorial available on http://herbaria.plants.ox.ac.uk/bol/caricaceae. Photography relied on the Macro function of a RICOH CX5 camera at resolutions of 3, 5, 7 or 10 megapixels and a Dino-Lite AM-413ZT digital microscope, a portable device hooked up to a laptop. Measurements of morphological characters were made with either ImageJ (http://rsbweb.nih.gov/ij/) or DinoCapture 2.0 (http://www.dino-lite.com/support.php), and a morphological database was built with $Xper^2$ (Ung et al., 2010), which allows the creation of interactive keys and can be integrated with the BRAHMS online system. Using one of the text reporting tools in BRAHMS, the $Xper^2$ database was exported to text format to generate standardized species descriptions.

For distribution maps, coordinates were taken from the specimen labels when available and then checked on Google Earth, easily accessible through a BRAHMS toolbar. Localities of collections without coordinates were found using available gazetteers and then corrected using Google Earth, following guidelines provided by Garcia-Milagros and Funk (2010). We also used other information present on specimen labels, such as elevation, distance from other locations (e.g. "10 km South of…"), and habitat (e.g. "across the river, up the slopes, in a dry area"). Where locality names were not in Google Earth, we checked historical maps, Wikipedia, and studies of botanical itineraries. The sources of all geographical coordinates were added to the field "llorig" (source of the latitude and longitude values) in BRAHMS.