Perspectives in Plant Ecology, Evolution and Systematics 17 (2015) 284–300

Contents lists available at ScienceDirect

Perspectives in Plant Ecology, Evolution and Systematics

journal homepage: www.elsevier.com/locate/ppees

Forum

An integrative and dynamic approach for monographing species-rich plant groups – Building the global synthesis of the angiosperm order Caryophyllales

Thomas Borsch[a,b,*], Patricia Hernández-Ledesma[a], Walter G. Berendsohn[a,b], Hilda Flores-Olvera[c], Helga Ochoterena[c], Fernando O. Zuloaga[d], Sabine von Mering[a], Norbert Kilian[a]

[a] *Botanischer Garten und Botanisches Museum Berlin-Dahlem, Dahlem Centre of Plant Sciences, Freie Universität Berlin, Königin-Luise Str. 6-8, 14195 Berlin, Germany*
[b] *Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Altensteinstr. 6, 14195 Berlin, Germany*
[c] *Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Apartado Postal 70-233, C.P. 04510 México, D.F., Mexico*
[d] *Instituto de Botánica, Darwinion, Argentina*

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 17 April 2015
Accepted 3 May 2015
Available online 5 June 2015

Keywords:
Taxonomic workflow
Biodiversity informatics
Species concept
Botanical monography
Digitasation
Evolutionary biology
Classification

ABSTRACT

One of the major goals of systematics is to provide a synthesis of knowledge on the diversity of a group of organisms, such as flowering plants. Biodiversity conservation and management call for rapid and accurate global assessments at the species level. At the same time the rapid development of evolutionary biology with a spectrum of approaches to test species relationships and species limits, has revolutionised and is still revolutionizing the science of plant systematics including taxonomy. We explore the relevant scientific and technological developments with the aim to suggest a conceptual framework for an integrated monographic synthesis which can reach global coverage. Our exemplar group are the Caryophyllales, which are a lineage of worldwide distribution, comprising approx. 5% of flowering plant species diversity. The current situation of classification is marked by a transition from pre-phylogenetic treatments to taxonomic treatments increasingly evaluated in an evolutionary context. Structured data (both molecular and morphological), linked to well-documented specimens will be important as fundamental entities of information that can be subjected to evolutionary analysis. As a result, taxon concepts are established as hypotheses which then can be used as basis for a classification system in a second step. Global syntheses need to provide information and use a classification system that reflects the current state of knowledge. In order to accommodate the constantly improved understanding of the organisms, eventually also resulting in the change of taxon concepts, the treatments need to be dynamic. The workflow for a global monographic synthesis as outlined here is supported by currently available biodiversity informatics tools such as the EDIT Platform for Cybertaxonomy. The availability of electronic sources (names, protologues, type images, literature) greatly facilitates the access to information, but as our case shows, considerable efforts for data curation and research are still needed. The implementation of a global monographic synthesis such as the Caryophyllales requires the involvement of the global scientific community.

© 2015 The Authors. Published by Elsevier GmbH. on behalf of Geobotanisches Institut ETH, Stiftung Ruebel. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Synthesising the current knowledge on lineages of organisms is still a major goal of systematic biology. This is normally achieved through monographs, which usually focus on the genus level and provide access to the known diversity of species. Monographs make explicit statements on taxonomic concepts and summarise the history of classification of lineages. In this way they deliver correct nomenclature including accepted names and synonyms and guide

* Corresponding author at: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Dahlem Centre of Plant Sciences, Königin Luise-Str. 6-8, 14195 Berlin, Germany. Tel.: +49 30 838 50132; fax: +49 30 838 4 50133.
E-mail addresses: t.borsch@bgbm.org (T. Borsch), phl@st.ib.unam.mx (P. Hernández-Ledesma), w.berendsohn@bgbm.org (W.G. Berendsohn), mahilda@ib.unam.mx (H. Flores-Olvera), helga@ib.unam.mx (H. Ochoterena), fzuloaga@darwin.edu.ar (F.O. Zuloaga), s.vonmering@bgbm.org (S. von Mering).

http://dx.doi.org/10.1016/j.ppees.2015.05.003
1433-8319/© 2015 The Authors. Published by Elsevier GmbH. on behalf of Geobotanisches Institut ETH, Stiftung Ruebel. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

us through the previous literature where often differing taxon concepts were applied. Monographs serve us with comparative data on the phenotype, increasingly also on the genotype (molecular data), distribution, ecology, and provide means to identify species and infraspecific taxa. Identification keys combined with thorough descriptions and illustrations not only enable the user to decide from a selection of characters and character states to which species an individual belongs but also help to recognise hitherto undescribed taxa.

Studies of species diversity have been revolutionised by phylogenetics. Technological developments in DNA sequencing yield the data sets that allow reconstructing trees and networks of evolutionary relationships. Still, in most studies the coverage is limited for practical reasons to a selection of individuals, constituting a sample of species, to infer a representative phylogeny of the group. But increasingly high coverage of species is being achieved due to methodological advances and decreasing costs per base sequenced (see, e.g. Mansion et al., 2012).

It is evident, however, that the establishment of molecular methods in systematic biology significantly affects monographing. Factual integration of molecular (genetic) and morphological (phenotypic) characters for an individual or taxon will be an essential task of any species diagnosis (González Gutiérrez et al., 2013) and any monograph. Also other kinds of data in the "omics" age will need to be assigned to well-documented individual specimens representing species. Therefore, efficient approaches to synthesise this knowledge in truly integrative monographs are needed. It is needless to say that unambiguous names for clearly and transparently described biological entities (species) are fundamental to all fields of biology as well as biodiversity conservation and use (Patterson et al., 2010; Hardisty and Roberts, 2013). On the other hand, the question of which species concept would apply to a group of organisms has been intensely discussed (De Queiroz and Donoghue, 2007; Goldstein and DeSalle, 2010; Mallet, 2007; McDade, 1995), and some evolutionary biologists even demurred from producing classifications because they would have had to make a compromise. We argue that the approach can be more pragmatic without loosening scientific precision.

In flowering plants, for example, species may either be monophyletic (in line with a species concept that proposes to only recognise monophyletic entities as species; e.g. Donoghue, 1985) or at least represent lineages with common origin (in line with a phylogenetic species concept that proposes to recognise entities that are the result of phylogenetic history; Nixon and Wheeler, 2008). There may be cases of incomplete lineage sorting (e.g. Gurushidze et al., 2010; Flores-Rentería et al., 2013), resulting in paraphyletic assemblages of morphologically very similar individuals while single populations out of the common gene pool may have adapted and phenotypically diversified with new traits (consistent with the phylogenetic species concept, but not with the monophyletic species concept). In any case (almost) all species concepts can be evaluated with phylogenetic methods (as it is the case also for the biological or typological species concepts, Mayr, 1969), but in the case of plants, non-hierarchical speciation has also been documented. For example, hybrid and allopolyploid speciation is in fact frequent (e.g. Kim et al., 2008; Rieseberg et al., 1990; Soltis and Soltis, 2009), even among more distantly related genera. Therefore the biological species concept is not always applicable, but thanks to the application of molecular evolutionary methods, this type of species origin can now be inferred with high levels of certainty. Our approach to describe species is to take the best available knowledge on the evolutionary history of a species, including continuities and discontinuities in the distribution of characters and character states among individuals and populations, as a basis to formulate a taxonomic concept. In this sense, biologically different mechanisms that have led to speciation are reflected in the taxonomic concept, and the purpose of species classification is to make this transparent. However, we do not argue that the criterion of monophyly for classifying higher level taxa (Hennig, 1966) should be questioned. We rather want to stress that a consistency of concepts of classification that spans from above to below the species level may not be appropriate or even pragmatic. Thus, non-dichotomous speciation histories related to complex patterns of gene flow can also be formulated in respective taxon concepts, which then will receive a name.

A further challenge for monographing is the constant accumulation of new knowledge. Reconstructing the tree of life has successfully illuminated the overall relationships of major groups such as flowering plants, providing the base for classifying monophyletic entities at the level of orders and families (APG, 2009). Considering the genus level, the coming years and even decades will yield a wealth of new insights. Like in other large groups such as the grasses (Vorontsova and Simon, 2012) this is also the case in Caryophyllales (e.g. Nyffeler and Eggli, 2010; Fuentes-Bazan et al., 2012; Dillenberger and Kadereit, 2014; Hernández-Ledesma et al., unpubl. data). And the same will happen at the species level, but probably much slower in terms of total coverage. We are therefore in a transition-phase from pre-phylogenetic to evolution-based taxon concepts and classifications. Apart from clear and stable rules in nomenclature, dynamic approaches are needed that can efficiently handle changing views on the concepts for taxa and which easily allow the incorporation of new knowledge into existing treatment frameworks. Traditional hard copy monographs by mostly individual authors, which take years or even decades to complete but become quickly outdated, are no longer an adequate means of synthesising current knowledge on lineages of organisms.

And then there is the enormous task of comparing organisms from different geographical parts of the world in order to unravel their evolutionary history, for which a representative set of characters needs to be compared in order to understand what species are and to evaluate alpha-taxonomic concepts. As a basis for this task, international collaboration has to be established. Piles of literature, often in a variety of languages, need to be evaluated. Collections, built upon centuries of exploration, have to be worked through and supplemented by new collecting activities and fieldwork to answer specific questions. And laboratory work has to be organised that will efficiently include representative sets of samples and will keep up with methodological advances. Often it takes years before the necessary material and a team of workers to analyse different types of characters are brought together. Nevertheless, the individual specialist is comparable to a historian who has to be specialised on an epoch, a country or a social group in order to connect the knowledge into a profound synthetic treatment.

Under these circumstances, the question arises how a monographic project can be efficiently organised? How can the data, gathered with considerable investment of resources, be collected, organised and stored in a dynamic, highly effective as well as sustainable way, which optimises their future use in updated synthesising approaches, as well as making them available for practical purposes? While the work is becoming increasingly facilitated by electronic access to type collections (e.g. JSTOR Global Plants, JSTOR, 2013), historical literature (e.g. Biodiversity Heritage Library, BHL, 2005) and specimens and species occurrence observations in general (e.g. Global Biodiversity Information Facility, GBIF, 2014) the complexity of the task of preparing an integrative biological monograph still remains.

Despite this enormous complexity, the products of this research are urgently needed. Societies demand reliable and accessible knowledge about organisms in order to efficiently deal with challenges in conservation and sustainable use of biodiversity.

This makes a need for efficient approaches to cover as much species diversity as possible in modern monographs. The need for taxonomic research as a basis has been repeatedly expressed (e.g. the Global Taxonomy Initiative, CBD, 1998). It is the only source of data for assessing the extinction risk of plants (Von Staden et al., 2013) and other organisms. It is of utmost importance to consider the discovery, collection, storing in collections, study and description of specimens and tissues of the still unknown species of the biosphere, before they turn extinct, as an absolute priority for biology in our century of extinctions (Dubois, 2010). As a consequence there are several proposals to revolutionise and streamline taxonomic research, using new technologies (Godfray, 2002; Mallet and Willmott, 2003; Agnarsson and Kutner, 2007; Godfray et al., 2007; Mayo et al., 2008).

Both a lack of knowledge or confusion about species limits can pose problems for conservation (Mace, 2004). A case study on Juncaceae and Potamogetonaceae indicated, for example, that accurate monographing revealed up to 25% of taxa categorised in the IUCN red lists representing synonyms or otherwise doubtful entities (Kirschner Kaplan, 2002), a percentage that poses serious challenges to the efficiency of conservation measures and investments. More recently, the Global Strategy of Plant Conservation (GSPC) as a programme of the United Nations Convention on Biological Diversity (CBD, 2006) with 16 practical targets, manifested the need for an "online flora of all known plants by 2020" (CBD-SBSTTA, 2012). While all these facts, needs, and arguments are well appreciated, precise conceptual work to develop the implementation of larger taxonomic treatments and monographic syntheses is still lacking.

A global treatment of plant species diversity is essentially not a new call – there have been several initiatives such as the Species Plantarum project (Brummitt et al., 2001) aiming at monographic coverage, Encyclopaedia of Life (EOL, 2014) using an aggregator approach with existing online resources, and attempts to create global checklists, as the result of database merging (The Plant List, TPL, 2013) or in a federated approach (Roskov et al., 2014). However, neither of these initiatives has progressed to monographically treat a significant amount of global plant diversity. In parallel, however, several network initiatives became quite successfully established in their respective communities, providing global up-to-date treatments for all species of those groups through electronic permanently curated portals (e.g. Brassicaceae – Koch et al., 2012; Ciardelli et al., 2009; Arecaceae – Palmweb, 2010; Solanaceae – Solanaceae Source, 2011; *Campanula* – Campanula Portal, 2013). Also, there are project websites on certain groups of plants that aim to organise taxonomic information and coordinate research (e.g. Euphorbiaceae – Rina and Berry, 2014, Sileneae – Oxelman et al., 2013).

Most of the current technical reports and opinion papers focus on either the development of information standards, data base and portal systems (Berendsohn et al., 2011; Holetschek et al., 2012) or the information needs for conservation (Paton, 2009; Nic Lughadha and Miller, 2009; Hardisty and Roberts, 2013). Only few studies address the taxonomic or monographic workflow as such (e.g. Stuessy and Lack, 2011; Marhold et al., 2013). Our motivation is to bridge between these issues in light of the recent, enormous advances in the study of phylogeny and speciation of plants that is fuelled by methodological advances in the genomics age. And we feel the importance to also consider developments in the scientific community that govern the generation of knowledge that then will be synthesised in monographic endeavours. In this paper we aim at (1) reviewing the current perspective for doing large scale monographs in an interdisciplinary and international environment, and (2) developing a work-flow for implementing a global synthesis of the angiosperm order Caryophyllales as a model.

Rationale for large scale syntheses: setting the base for the Caryophyllales

The step-wise approximation in classifying and naming biological entities

In an ideal world, evolutionary relationships and species limits will have been investigated including all types of data (morphology, anatomy, embryology, palynology, cytology, secondary compounds, DNA sequences; see Stuessy, 1990), using a comprehensive sampling of all taxa so far identified and of populations representing the whole range of putative species, and using a suite of inference methods ranging from tree and network reconstruction to the population-level modelling of shifts in allele frequencies and gene flow. However, in the real world this is not the case. Data sets and taxon sampling will be complemented over time by usually different workers, leading to a step-wise gain in knowledge on the evolutionary history and diversity of organisms. As a consequence, their classification is also improving in a step-wise manner.

In the real world including the Caryophyllales we find that there are different concepts of species and genera readily available, that concepts are differently used by different user groups, and that some of them are affected by rather rapid knowledge turnover. Scientific names need to be applied to such recognised entities even at preliminary stages where our knowledge is far from complete. This is necessary because taxonomy/systematics does not operate in a vacuum – concrete uses and other traits (e.g. conservation status, invasiveness, role in landscape and ecosystem) are linked to these entities and are widely applied.

The need of managing (different) taxonomic concepts has been pointed out (Berendsohn, 1995; Franz and Peet, 2009 and Franz and Cardona-Duque, 2013). The rank-issue ("mandatory categories", De Queiroz and Gauthier, 1992) advocated by phylocodists is only a small part of this problem [e.g. Moore, 2003; but see the entire special issue of The Botanical Review 69(1)]. In most cases where new evidence implies changes, these have to be either in the concept (content) or in the name. There is no benefit in maintaining a name stable if its content is dramatically changed. On the contrary, names have the purpose of communication and should therefore be as stable as possible with respect to their content.

It is evident that there have to be clear rules on how to translate evolutionary insights, and as a consequence taxonomic concepts into a formal system of nomenclature (e.g. McNeill et al., 2012). The objective of this paper is not to discuss the philosophy behind differing approaches such as a "Linnaean", hierarchical, rank-based nomenclature on the one hand (see Turland, 2013) and a phylogenetic nomenclature (sensu PhyloCode, Cantino and Queiroz, 2010). We do believe, though, that names of organisms have to facilitate unambiguous communication of taxonomic concepts which will not be achieved through applying the PhyloCode. Rules for describing these concepts should enable us to simplify and categorise the detailed hypothesis on the evolutionary history of a group of organisms that led to the formulation of the concept. For our synthesis of the Caryophyllales, we agree with the view that monophyly is an important criterion to make higher-level taxon concepts (i.e. above the level of species such as genera and families) as predictive as possible. Using current phylogenetic techniques, monophyly can be established with high degrees of confidence in a macroevolutionary context. At the level of species and below, where evolutionary histories are much more complex (microevolution) and cannot be necessarily described by dichotomous patterns, we should explicitly state what the concept for a species as classified includes: (a) an assemblage of populations that all go back to a common ancestor; (b) an assemblage of populations that represents a partial

spectrum of ancestral genotypes so that the species is paraphyletic (it is often very difficult to distinguish between incomplete lineage sorting or hybrid origin); (c) populations that go back to one or several polyploidization events (inferred from molecular data); (d) an assemblage of populations that go back to one or several hybridisation events without genome duplication. There is no theoretical justification to apply the criterion of monophyly as universal base for species-level and higher rank classification (Rieppel, 2009) because the distinctness of a species as a biological entity can be explicitly formulated as a hypothesis to entail a defined group of individuals. As Rieppel (2009) put it: "Monophyletic groups (clades) stand in phylogenetic relations to each other, constituents of species stand in tokogenetic relations to each other". Biologically meaningful assessments of species relationships and species limits therefore require a comprehensive approach (see e.g. Naciri and Linder, 2015, for a recent review). This also means that in order to be reproducible, the recognised entities and the respective taxon concepts have to be based on explicit sets of character data (both molecular and morphological) linked to individual specimens.

Considering that in a stepwise process there will be discrepancies in circumscriptions of species envisaged by different authors, a responsible taxonomist should procure the least number possible of taxonomic changes in taxon concepts in light of new evidence. Therefore, an important issue is the economy of change in classification systems procuring nomenclatural stability (Nixon et al., 2003). For instance, different philosophical criteria, such as reflecting monophyly in a classification, have been the cause of profound supraspecific changes in classifications. Only accepting monophyletic taxa could still result in contrasting opinions on supraspecific taxa. Therefore, changes should also be based on an expert view on clade stability (in light of the reliability of existing phylogenetic hypotheses to try to depict what it is considered "the true organismal phylogeny" or at least the best supported hypothesis) and of phenotypic diagnosability (e.g. Vences et al., 2013).

In spite of this, many of the current changes in classification do not consider economy of nomenclatural changes. Conservation of names is a currently seldom-used resource to promote name stability in this context. An analysis of stability in both concepts (species circumscription) and names of vascular plants and mosses in Germany (Berendsohn and Geoffroy, 2007) concluded that there is a considerable instability in a substantial proportion of plant taxa, even among works in current use.

Standards and tools for managing integrative concept-based classification systems

Classifications and species circumscriptions evolve with the growth of knowledge. In taxonomy and elsewhere in science, there is a fundamental difference between information provision by information systems and information provision directly mediated by human beings (normally specialists). The practical value of being able to name and identify a biological entity lies in the provision of an indexing system for knowledge about these entities, which allows integrating information from different sources on different topics over large periods of time. In traditional printed publications, the integration of information has been mediated by specialists who are able to qualify the relationships between the named and circumscribed taxonomic concepts. In contrast, in information managed by computer systems and published on the Internet, the possibility to automatically link information on taxa from various sources using the scientific name is increasingly used. Given the uncertainties stated above, without expert validation this poses a high risk of misinformation.

In order to come to terms with this problem public information systems need integrated explicit knowledge to reliably transmit information linked to taxonomic concepts (in this case: named biological entities; Beach et al., 1993; Berendsohn, 1995). Information models were developed, providing the theoretical base for handling taxonomic concepts (e.g. Zhong et al., 1996; Berendsohn, 1997; Pullan et al., 2000), followed by software development that demonstrated the practicability of such models for taxonomic data (e.g. Pullan et al., 2000; Gradstein et al., 2001; Berendsohn et al., 2003; Berendsohn and Geoffroy, 2007). In parallel, a number of projects drove standardisation of data items, especially in the context of the organisation for Biodiversity Information Standards (TDWG, formerly the Taxonomic Databases Working Group). Examples include the Taxonomic Concept Transfer Schema (TCS; Kennedy et al., 2006), Structured Descriptive Data (SDD; Hagedorn et al., 2005), Access to Biological Collection Data (ABCD; Berendsohn, 2010a,b; Holetschek et al., 2012) and Darwin Core (DwC; Wieczorek et al., 2009). This development culminated in the definition of an object-oriented Common Data Model for taxonomic information (CDM) that was devised for and forms the basis of the EDIT Platform for Cybertaxonomy (Berendsohn, 2010a,b; Berendsohn et al., 2011).

The EDIT Platform is a collection of more or less tightly coupled tools especially devised for taxonomists, incorporating the entire scope of nomenclatural, taxonomic, descriptive and geographic information contained in the products of taxonomic work (monographs, treatments for faunas and floras, and checklists; Berendsohn, 2010a,b; Venin et al., 2010). The EDIT Platform is natively concept-oriented. Therefore, the system allows applying a name to deviating taxonomic entities, all of which include the type specimen but are more or less inclusive. Concept relationships can be made explicit, stating that the circumscriptions according to the first concept is congruent to, overlapping, included in or includes the second one. Alternative classification systems can be shown (e.g. the alternative classifications for *Hieracium* and *Pilosella* in the *Cichorieae* portal; ICN, 2009). It is easy to adapt the classification to a new concept once there are new research results, without having to laboriously re-enter information (protologues, images, etc.) used previously. This will also be of considerable importance in the synthesis of Caryophyllales. Available Platform tools include those for local and on-line data input/editing and on-line and print data output (Ciardelli et al., 2009). Import and export interfaces handle files that adhere to community data standards such as ABCD, DwC, SDD, and TCS. The Xper2 software (Ung et al., 2010a), which can also be used as a stand-alone tool, handles descriptive information, i.e. character/character state lists that can be augmented by illustrations and used for keys and species descriptions as well as exported as a matrix into NEXUS (Maddison et al., 1997) for phylogenetic analysis. Descriptive information may refer to taxa or to specimens, or to taxa synthesising information from taxa of a lower rank or from specimens (Kilian et al., submitted for publication). Nomenclatural details are covered in full for botany and zoology, taking into account the often very specific requirements of taxonomists and taxonomic publications. On the technical side, the EDIT Platform is independent of the operating system and database management system used, and a full range of machine-accessible web services (Booth et al., 2004) provide the base for full interoperability e.g. in workflow environments, as well as connectivity with the global biodiversity informatics infrastructures. This is exemplified in the Biodiversity Virtual e-Laboratory (BioVeL) project using the taxonomic data of the Catalogue of Life (Mathew et al., 2014).

We posit that for a truly integrative approach to taxonomic information a system like this is necessary, combining the existing rigorous traditions in handling data from taxonomic research and nomenclature with the possibility of concept handling while remaining open for the advances that are currently being made in Internet technologies.

Continuous actualisation of knowledge and classification through involvement of the scientific community and the use of modern information technology

The beginning of the 21st century has seen important steps towards realising teams working on integrative monographs, often facilitated or event incentivized by the technical advances in information sharing, networking and new forms of publishing (Marhold et al., 2013). A more detailed look at those initiatives should help to consolidate a sensitive approach for large scale syntheses of diverse plant groups such as the Caryophyllales.

As one of the first projects using internet technologies, the Solanaceae Source (2011) began in 2004 to work towards a species-level treatment for *Solanum*, relying on existing modern monographs and revisions and coordinating work on the remaining groups using existing international taxonomic expertise. The continuous assembly of information on names, types, and specimens as well as bibliographic sources, and combining this with field work and phylogenetic research (e.g. Särkinen et al., 2013) resulted in an increasingly complete coverage of the information on the family. The network is coordinated by experts in the family, and a sustainably supported software platform (Scratchpads, Smith et al., 2011) provides the means for information storage, display and communication among the experts. The Lecythidaceae Pages (Mory and Prance, 2006; Mory et al., 2010) similarly attempt to provide a continuously amended information source on this Neotropical family, based on a widely distributed and institutionally supported network. The information system is now based on the commercial KE-EMU software (KEsoftware and About, 2014). The site developed for Melastomataceae (MELNet, 2014), though not yet going much beyond a species checklist and providing distribution data, is based on the Diversity Workbench Platform (Triebel et al., 1999). The Cichorieae Network (ICN, 2009) and Palmweb (2010) served as exemplar implementations for the above described EDIT Platform for Cybertaxonomy, a suite of software tools aimed at comprehensively supporting the entire taxonomic workflow. The EDIT platform supports alternative classifications, which are not found in other online syntheses.

For the monocots, eMonocot (2012) provides a portal integrating several information sources, among those Palmweb (2010), GrassBase (Clayton et al., 2006), and several Scratchpad-based systems (e.g. eMonocot Team, 2012a,b). For Brassicaceae, the more recently created BrassiBase (Koch et al., 2012) aims at "providing an online-accessible knowledge and database system of cross-referenced information and resources on Brassicaceae" including full taxonomic coverage of the family as well as character and trait studies. The software was developed specifically for the project.

Apart from sites with truly monographic aims there are checklist sites that cover plant groups, either covering certain taxa (e.g. WCSP, 2014) or geographic areas (see Crouch et al., 2013 for examples from southern Africa). A pioneer for collaborative organisation of such sites is the ILDIS network for Legumes, which started in the beginning of the 1990ies using the ALICE software (Bisby, 1993) and continues to be available on the Internet (ILDIS, 1996). Other initiatives merely came together to coordinate sampling and phylogenetic analysis of diverse families and usually include the effort of revised classification above the species level based on the criterion of monophyly (e.g. Poaceae, Grass Phylogeny Working Group, 2001; Fabaceae, Legume Phylogeny Working Group, 2013).

We can conclude that these initiatives have been quite successful for a number of plant groups to collaboratively achieve the collection, merging and electronic publishing of existing information sources (published and unpublished) and thus can serve as proof of concept for a de-centralised approach. As exemplified for ILDIS, however, institutional collaboration and commitment will be needed to ensure long-term sustainability (Crouch et al., 2013; Costello et al., 2014). While portals supported by existing initiatives present taxonomic information (treatments), they hardly make taxon concepts transparent with links between existing names, literature (incl. protologues), specimens (incl. types) and character data (from the phenotype and genotype). We therefore also conclude that a synthesis of Caryophyllales that builds upon an organised and electronically supported taxonomic workflow has to still pioneer in several ways.

Global taxonomic slicing and regional floras will mutually benefit

Global syntheses like the one envisaged for Caryophyllales have the advantage that they can study biological entities independent of political borders or regionally defined project settings. Hence, clade or taxon-wide monographs that are delimited only by scientific criteria, are essential for the understanding of evolutionary history. Very few clades exclusively occur within project regions such as countries: on the genus level such taxa are mostly restricted to islands (e.g. *Podonephelium* of the Sapindaceae in New Caledonia [Munzinger et al., 2013]). Nevertheless, a geographically much wider sampling had to be used to detect their monophyly. Considering further that such taxa constitute only a fraction of the respective floras, the knowledge obtained through large scale taxonomic syntheses will generally be beneficial to more regional flora projects. In fact, there are many cases of groups of closely allied species, for which strikingly different taxon concepts have been published. A Caryophyllales-example is *Alternanthera halimifolia* (Amaranthaceae) that may be an endemic from the lomas of Peru or a widespread Neotropical species (TROPICOS lists 48 synonyms incl. infraspecific taxa). It is obvious that a regional flora project will neither have the resources nor the mandate to clarify species concepts in such cases. Global syntheses will therefore be essential to provide the knowledge on species concepts that can then also be used for more regional treatments.

On the other hand, there are also clear advantages of regional studies such as good specimen coverage, the availability of a local/regional network that allows representative field sampling and knowledge about relevant specimens in regional herbaria. Floristic research often requires specific knowledge of localities, and thus involvement of local people and institutions. The question of whether a species occurs in certain geographic area can often only be clarified at that level. The same applies to the assessment of data which cannot be derived from a specimen in a collection, such as on habitat, phenology, pollinators, etc. In many cases botanists acquire enormous expertise on the plants of their region but do not study plants beyond this region.

Furthermore, regional treatments offer obvious benefits for the user. A selection of species in a particular area will lead to more usable, smaller keys, and in fact may be the only feasible way to allow plant identification by non-specialists of a group. Therefore, there are good reasons to produce both regional floras and worldwide treatments. Some major flora projects that cover very large and well delimited biogeographic regions may go beyond regional syntheses and may include the provision of more large scale original taxonomic research (e.g. Flora Malesiana, Roos et al., 2011). But even here, the overlap at genus and species level between the area covered by Flora Malesiana and the Flora of Nepal was recently estimated in the range of 30–40% (Pendry and Watson, 2009). The authors therefore suggested that coordinated efforts on the respective taxonomic groups could be a way to facilitate the production of treatments in both flora projects. Along the same line, Funk (2006) nicely pointed out that Floras are part of a continuum between revisions and monographs on one hand, and field work and data bases on the other.

We conclude that the authors who prepare treatments of specific taxonomic groups for a Flora project, should ideally be involved

in the respective taxonomic network. However, the interaction between taxonomic networks (global slicing) and regional flora projects has to be actively developed by the respective scientific communities and the communication has to be intense. We believe that certain mechanisms and work-flows need to be developed to support this interaction, which should also be an objective of the Caryophyllales synthesis.

User needs versus needs of the scientific community and the individual researcher – the example of the conservation community

During our preparations for a global synthesis of Caryophyllales it became evident that there is quite a gap between the interests of users of monographs and taxonomic information and the needs of the actual scientists generating this knowledge. We believe that large scale syntheses need to take that into account in order to be successful. The Global Taxonomy Initiative, established by the conference of the parties of the Convention on Biological Diversity in (CBD, 1998, 2006) highlighted the importance of taxonomy at the political level. User needs have then been more clearly recognised through the Global Strategy for Plant Conservation (GSPC) that as a programme of the CBD called upon the provision of a widely accessible working list of known plant species (Target 1 for 2011) as a step towards a complete world flora (adopted by the COP in 2011 as Target 1 in the updated version of the GSPC for 2020). However, the underlying discussions were largely conducted outside the academic community, and without much regard to issues such as scientific methodology and workflows, research funding, good scientific practice, academic curricula and career development. This may partly be the result of distributed responsibilities between ministries of the environment or natural resources (with responsibility for the CBD) on the one hand, and ministries for education and research (with responsibility for science) on the other. The result is not only a huge gap in funding for taxonomic research but also a decrease in quality. As a consequence, user needs are often only superficially met. Trying to fulfil such targets without responsible involvement of the scientific community, i.e. without the standard mechanisms of science-funding and science evaluation in effect, even increases the taxonomic impediment by cutting off many scientists and institutions from a positively evaluated contribution to solving global challenges. We hope that the facts and thoughts presented in this paper will not only stimulate a solution-oriented discussion on the example of the Caryophyllales but also a wider stakeholder dialogue.

User needs appear to be very pragmatic in the conservation community, and in other communities using taxonomic information: a stable reference system for their factual data that refer to organisms. A user may feel satisfied once a name is attached to a plant, regardless if a natural species concept is reflected by the obtained name. User needs may contradict the actual research needs and are on short term not necessarily correlated with the quality of the science. It is thus most important to convey beyond the taxonomic community that the provision of reliable information is more than a mere mobilisation of existing knowledge. In addition to the successful development of a biodiversity informatics infrastructure during the past two decades (Hobern et al., 2013), aspects of data curation and synthesis as well as the actual research processes need to be supported. The research community can and will contribute, provided that appropriate incentives exist i.e. that their contributions further careers of individuals or positive evaluation and funding of institutions. Synergies can be generated if research projects that often span three to four years (e.g. PhD thesis) are embedded in a sustainable structure that supports the management and synthesis of primary data. We believe that such synergies will be very important for global initiatives of practical relevance for conservation such as the World Flora Online (WFO; CBD-SBSTTA, 2012) to be successful.

There are first positive developments (e.g. the German Federation for the Curation of Biological Data financed by the German research council, DFG; Diepenbroek et al., 2014). Nevertheless, at the same time, there needs to be sufficient long term institutional funding to support data curation, large scale synthesis and result/data dissemination. It should be noted that in spite of the lack of adequate funding, much knowledge is generated by numerous smaller research projects worldwide. However, there so far has been no conceptual framework for integrating results into a global synthesis. Offering this together with the respective infrastructure, as proposed in this article, could lead to benefits for users and the scientific community at large.

We believe that the approach of a global network like the one laid out here for the Caryophyllales helps to increase the productivity of the individual researcher, by offering access to infrastructure, to a structured knowledge base, as well as to methodological background knowledge that directly furthers the research effort. Therefore, a positive effect on research quality can be expected. Benefits are also likely in the sense of a social network and by increasing the visibility and impact of publications.

The angiosperm order Caryophyllales as a model

Function of the model group

On a small scale, prototypic treatments exist that largely fulfil the criteria for an integrative monographic approach as outlined above. However, this has not yet been attempted on a large scale, involving thousands of species and representing a major clade of the angiosperms. Simply upscaling existing methods will not be sufficient. Rather a new, much more collaborative approach is needed that focuses on the knowledge generation process in biological systematics and also builds upon the existing data and information resources.

The approach outlined here needs to emphasise short and long-term goals. Taxonomists are challenged to provide as soon as possible a treatment which globally covers the species diversity in order to provide a base for decision making, e.g. in conservation. On the other hand there is the goal of building a scientifically comprehensive information base and synthesis on the group that is fully based on documented evidence and falsifiable in its postulated taxonomy. Nevertheless, several information components (e.g. database of names, citations, protologues, type specimen images) are essential in both short and long term perspective. An electronic information system will therefore have to structure and manage available published knowledge, evidence gathered in formulating new knowledge (including primary data linked to specimens), and provide the means to gather and re-integrate new data and knowledge (e.g. altered taxon concepts) and thus actively support the development towards a truly integrated scientific monograph of the entire group.

Evolution, diversity and importance of the Caryophyllales

The Caryophyllales were chosen as a group to serve as an exemplar for this new approach because the order is of significant size to have an impact in advancing our knowledge on flowering plants. Caryophyllales comprise about 5% of flowering plant species diversity, and are classified into 37 families. Currently, 732 genera (Hernández-Ledesma et al., unpubl. data) and an estimated 12,500 species are recognised. Caryophyllaceae, Aizoaceae and Cactaceae are the most species-rich families, while ten families comprise only one to three species. The Caryophyllales are an ecologically diverse

lineage with traits such as C_4 photosynthesis, salt and/or gypsum tolerance, succulence and carnivory. Centres of diversity are temperate and subtropical regions, especially in Mediterranean climates, deciduous vegetation with seasonal distribution of rainfall, semi-deserts and deserts, and montane vegetation. Due to the high number of endangered plants recognised in the order, prominently in Cactaceae, Droseraceae and Nepenthaceae, the group is also relevant in the context of the Global Strategy for Plant Conservation and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Furthermore, Caryophyllales comprise many species of economic importance such as ornamentals (e.g. Cactaceae, Caryophyllaceae, carnivorous groups), cereals and green vegetables (e.g. spinach, amaranth, quinoa, and sugar beet), but also as noxious weeds (e.g. *Alternanthera philoxeroides* (Mart.) Griseb., *Amaranthus spinosus* L., *Opuntia* spp., *Mirabilis* spp.) and allergens (e.g. *Amaranthus retroflexus* L., *Atriplex* spp., *Salsola kali* L.).

Several studies have shown that Caryophyllales form a monophyletic group (e.g. Cuénoud et al., 2002; Hilu et al., 2003; Schäferhoff et al., 2009; Brockington et al., 2009). The concept of the order has changed from the pre-phylogenetic Centrospermae (Eichler, 1878) to the expanded Caryophyllales that is largely based on DNA sequence data. A summary of the taxonomic history will be presented in Hernández-Ledesma et al. (unpubl. data) along with a more detailed discussion on the current understanding of phylogenetic relationships.

Existing data and information sources for Caryophyllales

The Caryophyllales are typical for most major groups of flowering plants, in that their family-level classification is beginning to stabilise while generic realignments based on phylogenetic insights are well underway (see section on establishing a taxonomic backbone below). Nevertheless, the big work ahead relates to the circumscription and treatment of species and infraspecific taxa, which will therefore be discussed here. Since integration of information from existing electronic sources is fundamental to streamline the work of the taxonomist, we focus on publicly available electronic data sources for Caryophyllales.

The International Plant Names Index (IPNI, 2004) provides a starting point for names and their original publication. Having been produced from three datasets (Index Kewensis, Grey Index and Australian Plant Name Index), there is considerable duplication of names. Tropicos (2014) similarly provides nomenclatural data including the original publication but also indications of the nomenclatural status of names and the usage of names in the form of being accepted or a synonym according to specific bibliographic references. It is especially this feature that renders Tropicos one of the important sources for Caryophyllales information, albeit it is somewhat geographically biased. A total of 36,835 species names are registered in Tropicos for the Caryophyllales (as of October 22, 2014). IPNI and Tropicos were used to complement the so-far existing global family checklists to form The Plant List (TPL, 2013). TPL is a checklist indicating a status (accepted, synonym or unresolved) for most species-level plant names that have been published. TPL lists a total of 50,192 names, about 37% of which are unresolved as to their status as correct names or synonyms (see Fig. 1 for details in four families, and Appendix 1). It is the only comprehensive species-level taxonomic checklist that covers the entire Caryophylles on a global scale.

We carried out a review to assess the species numbers for the different families in Caryophyllales comparing TPL and expert views taken from the literature (Appendix 1). While species number estimates match in some medium sized or small families (e.g. Didiereaceae, Frankeniaceae), numbers differ by a quarter and more for families such as Cactaceae, Caryophyllaceae, Droseraceae or

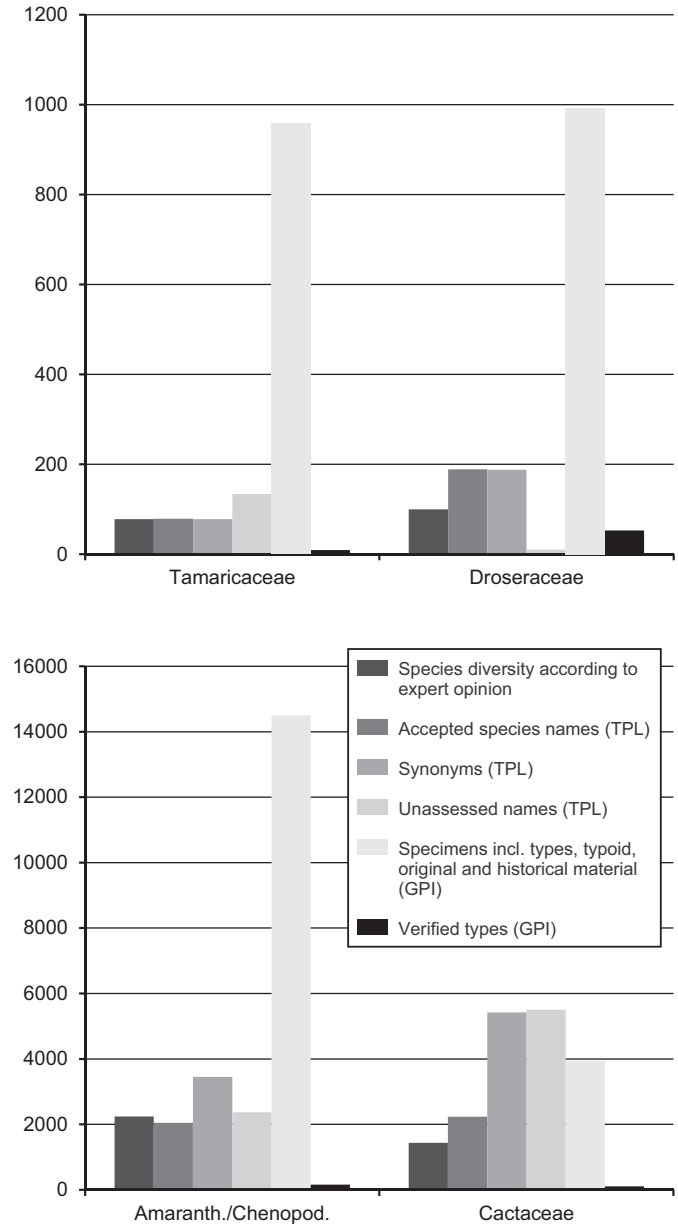

Fig. 1. Comparison of the estimate of species diversity in four Caryophyllales families based current consensus in literature by family experts and the numbers given in The Plant List. In all cases, there are significant numbers of unassessed names in TPL. From the digitised specimens made available by the Global Plant Initiative (GBI) all families only a small percentage has so far been verified as to their accurate type status.

Polygonaceae. This indicates the amount of work that has to be done in order to clarify species limits and arrive at a stable classification. In addition, the number of unresolved names in TPL points to the task of clarifying existing usage of species names. In order to tackle these problems, easy access to literature and specimen data is instrumental.

Fortunately, more and more of the necessary information becomes accessible on-line such as publications (e.g. BHL, 2005; JSTOR, 2000), digital type specimens from individual herbaria or from aggregators (e.g. JSTOR Global Plants, JSTOR, 2013), general specimen and species observation data from GBIF (2014), or entire checklists available for download or as a machine-accessible webservice (e.g. PESI, 2014). No reliable estimate can be given at this time of the degree of on-line availability of taxonomic literature relevant for the Global Synthesis of Caryophyllales

but an indication may be the number of names with digitised and accessible protologues. For example, for the genus *Anabasis* (Basellaceae) about 30% of the names in Tropicos offer a link to a protologue openly accessible in the Biodiversity Heritage Library; for *Nepenthes* (Nepenthaceae) this number reaches 45%. However, with these numbers it has to be considered that there is a big amount of digitised literature available on-line from other, non-specialised sources, e.g. from JSTOR (2000) and from the many commercially published journals with taxonomic content, albeit these sources are often behind a paywall.

Examples of electronic sources for Caryophyllales that specifically treat a defined geographic region are presented in Appendix 2. These include taxonomic checklists and Flora treatments in a wide variety of formats, from scanned documents that are not immediately usable for structured data access in an information system, to highly structured databases that can produce output which can be reused almost instantly. Although the list given in the appendix is by no means fully comprehensive, it becomes clear that still many treatments are not openly accessible (in spite of calls to the contrary, e.g. by Brach and Bouffort, 2011). Looking at the over-all available literature it also becomes clear that quality treatments and monographs on the species level are far from fully covering both the taxonomic diversity and the geographic spread of taxa in the Caryophyllales.

For the revision process, access to type specimens and other original material that has been used to name taxa is essential. JSTOR Global Plants (GP, JSTOR, 2013) is a prime electronic information source aggregating images of type specimens and particularly important historical specimen collections from more than 300 herbaria worldwide. A dataset for Caryophyllales families provided by JSTOR (8 December 2014) to the authors included records representing a total of close to 75,000 specimen images that were accessible through GP (JSTOR, 2013), of which more than 52,000 carried some indication by the data provider that the specimen was considered to be a type or original material. More than 12,000 were marked as holotypes, lectotypes or neotypes, 4500 as syntypes. Nearly 20,500 specimens were designated as isotypes, isolectotypes, isoneotypes and isosyntypes. Around 1000 records each refer to paratypes and original material. The remaining more than 13,000 records were just called Type, with more than 3000 explicitly marked as questionable. Because we lack a complete checklist of Caryophyllales species (and infraspecific) names and because the names given in GP are not fully standardised, we cannot be sure how many different names are represented in this dataset. To understand the situation in more detail, we looked at examples such as the seven genera of the tribe Pisonieae (Nyctaginaceae; Hernández-Ledesma, unpublished data), with an estimated 260 species. Global Plants provides images of the holotypes, lectotypes or neotypes for 89 of these names, i.e. for about one third. At least one type specimen (this then also includes isotypes and isolectotypes) is available for 167 names, i.e. for about two thirds, while for one third no type is available. This level of coverage is in rough agreement with the findings of preliminary gap analyses on other plant groups as cited above under Information Resources. However, it may also be noted that the coverage is better for Africa and the Americas, where the digitisation initiative funded by the Andrew W. Mellon Foundation was particularly active. As a consequence, this may explain why only 16 holotype or lectotype specimens are available for the names (species and infraspecific) in the largely Asian genus *Nepenthes*, which counts 259 largely unresolved species names in TPL.

Further work is therefore necessary to localise and digitise the remaining type specimens, the degree of which will vary from genus to genus, and may be substantial in several genera. Moreover, when the number of accepted species names and the overall total of names at species level is compared to the number of type images available (Fig. 1), another important issue becomes evident: in many cases a relatively large amount of specimens was actually digitised including types, specimens that are probably types or just historical material without scrutinising their correct status. The most important task will therefore be to verify the type status of the material available. Global Plants is therefore a source that facilitates taxonomic work but not yet a ready-to-use taxonomic product. This confirms the results of Smith and Figueiredo (2013) on the African species of *Polygala*, which clearly showed that the quality of the textual data (e.g. the category of types, collector information, etc.) in GP is not yet satisfactory. The conclusion is that JSTOR Global Plants is becoming the unique global resource for botanical type specimens, but that specialist input is needed for quality control of the type specimen data. The amount of work needed is difficult to assess and will not only depend on the respective genus or family but also on how Global Plant data curation interacts with the taxonomic work flow maintained by the teams working on these taxa. Within the Caryophyllales synthesis, this work should take place as one of the first steps in the context of verifying protologues, citations and type images as an activity of the scientific community. The results can be incorporated into commented global checklists which can also be published, at least as data papers. However, JSTOR also needs to provide resources to manage the information backflow and to implement curation standards for Global Plants.

With regard to specimens in general, herbaria have made much progress in providing digital images. Some major herbaria have been completely digitised, e.g. those in Paris and Leiden, others provide varying degrees of mobilisation of their holdings. Access is still not satisfactorily organised, although the Global Biodiversity Information Facility (GBIF, 2014) provides the technical infrastructure for a global, single point of entry. Currently (October 2014), GBIF offers access to 43.5 million records that have been marked as plant specimens, of which close to 2.4 million are classified as Caryophyllales. However, the number of already digitised specimens exceeds this by far. For example, the large herbarium in Paris, although fully digitised, is not included in this number, neither are large already digitised specimen holdings, e.g. in China, Brazil, and India. The conclusion is that, similar to the view on the literature presented above, incorporation of specimen information will be eased by the electronic resources available, and that this will make the results more comprehensive and reliable, but that this will in no way replace taxonomic scrutiny and synthesis.

As for DNA electronic data, the INSDC (International Nucleotide Sequence Database Collaboration; Karsch-Mizrachi et al., 2011) for example retrieves 1,623,294 records for Caryophyllales (http://www.ncbi.nlm.nih.gov/nuccore/?term=caryophyllales: accessed on December 22 2014). This amount of information would be paramount for the understanding of the phylogeny and hence naturalness of many taxa, but it is difficult to assess how many of those sequences are effectively linked to well-documented and accessible voucher specimens (even fewer of them would be linked to specimen images). Also, the sequenced loci are not standardised across taxa. It will therefore be necessary to access the original source in order to include sequence data into a global synthesis. Actual research will then have to evaluate if the taxonomic identification can be reassessed and the data placed into the respective taxon circumscription, in particular at the level of species.

In conclusion, a formidable amount of information is accessible and provides a broad base to accelerate the step-wise research process towards the Global Synthesis of Caryophyllales. Further in-depths analysis of the existing information resources is part of the work in progress including the identification of priorities for further data mobilisation and digitisation.

Main goals and application needs

In the following paragraphs we will consider the organisational prerequisites and present a strategy towards a global Synthesis of Caryophyllales (Fig. 2). There are two major lines of activity:

(1). A global *Caryophyllales Network* needs to be established in a way that

- integrates research activities to organise a quality and first hand outlet of knowledge, building on effective communication within the scientific community and formal agreements among the participating institutions
- facilitates interaction between established (taxonomic) specialists and young researchers, thereby promoting postgraduate education and career development
- promotes scientific exchange by organising workshops or conferences.

(2). A scholarly *information system* has to be established on-line that

- provides the state of knowledge on species diversity
- allows to present the currently accepted concepts of species (and supraspecific as well as infraspecific taxa)
- promotes the identification and discussion of conflicting taxon circumscriptions
- integrates information on names, synonyms, protologues, descriptions, and keys
- integrates character data both phenotypic ("morphological") and DNA ("molecular")
- facilitates the development of a standardised terminology used in Caryophyllales (in the sense of an ontology e.g. for morphological characters)
- provides synthetic distribution data and specimen-based distribution records for accepted taxa
- credits all sources of information and thereby increases the visibility of research
- informs original sources about data corrections, enhancements, and annotations.
- references source objects and provides stable references for newly created data objects
- represents an exemplar high quality data set through linking of specimen information (including digitised images) with their meta-data (e.g. geographical coordinates) and character data
- to advance the recognition and classification of species on the basis of well- studied biological entities
- to assess and model species distributions including their future trends
- to reconstruct the evolution of lineages in an approach illuminating molecular and morphological diversification
- provides a dynamic and easy to update platform to identify areas of future research and to promote collaboration
- enhances visibility of the network, its partner institutions and contributing scientists.

Strengthening the Caryophyllales network

In the taxonomic community, many examples of collaborations of individual researchers exist. Forces are often joined in the context of projects with defined kick off and completion dates or to review the state of knowledge about particular taxa or topics. Some collaborations also provide an organisational framework including resources for coordination, data management and curation, apart from the actual research activities. But such networks are often grounded in project financing, and consequently, collaborative activities tend to dwindle once the project period runs out. For long-term sustainability, we therefore believe that institutional commitments are imperative. Institutions need to sustain coordination and information management tasks in addition to allocating the resources for the dedicated individuals forming an expert network.

For the Caryophyllales network, an initial institutional partnership consists of the Freie Universität Berlin's Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM, herbarium B), the Universidad Nacional Autónoma de México's Instituto de Biología (incl. herbarium MEXU) and the Instituto de Botánica Darwinion in Argentina (herbarium SI). All three are committed to contribute their existing expertise in Caryophyllales research into a larger scale initiative. The BGBM has defined Caryophyllales as one of the focus areas on their long-term research agenda with a structural backup of a full position dedicated to sustainably support the network. A memorandum of understanding (MoU) among these core partners will formalise the agreement on the project's main aims, structure and governance, in order to coordinate the elaboration of products and the execution (and further development) of the work plan (including methods and tools as well as milestones and time frame). Efforts are under way to extend the geographic coverage to include further institutions with Caryophyllales on their research agenda.

The actual network, however, consists of the individual scientists. There are several benefits arising from participation, and the intention is to develop the network as to create mutual benefits. Apart from easier communication about research questions, the information infrastructure (incl. linked resources of structured data on Caryophyllales) will facilitate the generation of a variety of publications, including scholarly papers, flora treatments, and data papers. The network will be instrumental in increasing the visibility and recognition of the participating scientists and institutions in the scientific community and for users, which in turn can have positive effects for evaluation and funding purposes.

A distinct advantage of a jointly curated repository of the information on Caryophyllales is the possibility of linking into other initiatives by means of information interfaces. For example, the World Flora Online (WFO) project (CBD-SBSTTA, 2012) has synergies with the Caryophyllales monographic synthesis. WFO aims for a compilation of current floristic knowledge across the entire world while the Caryophyllales synthesis aims for a quality treatment maintained for an important group of flowering plants. It is not the intention of WFO to replace current regional floristic or revisionary efforts. The interaction between the Caryophyllales synthesis data and the (presumably much less detailed) WFO can be streamlined using existing data interfaces (Fig. 3) that, on the Caryophyllales side, need only minimal adjustments to adapt to the WFO data ingestion standard (a variant of the Darwin Core standard, in Darwin Core Archive format). In the future, we think that WFO data ingestion will be achieved by using the Web Services of the EDIT Platform, which already can provide Darwin Core standard data.

In building the human Caryophyllales network, a detailed assessment of relevant publications, information sources and experts was the starting point. In order to organise efforts towards the understanding of Caryophyllales promoting communication and international collaboration, we have contacted specialists, who are currently working on the systematics and/or taxonomy of Caryophyllales or who have contributed to the knowledge of the group at least in the last decades. Of the 110 specialists contacted, more than 80 have responded positively to the initiative of a global synthesis. These botanists are studying most diverse families of the order and are based in universities and other research institutions in 26 different countries. Since further integration and participation of the community (established senior researchers and students are invited to participate in the network) will be essential for the success of the initiative the overall positive responses from these

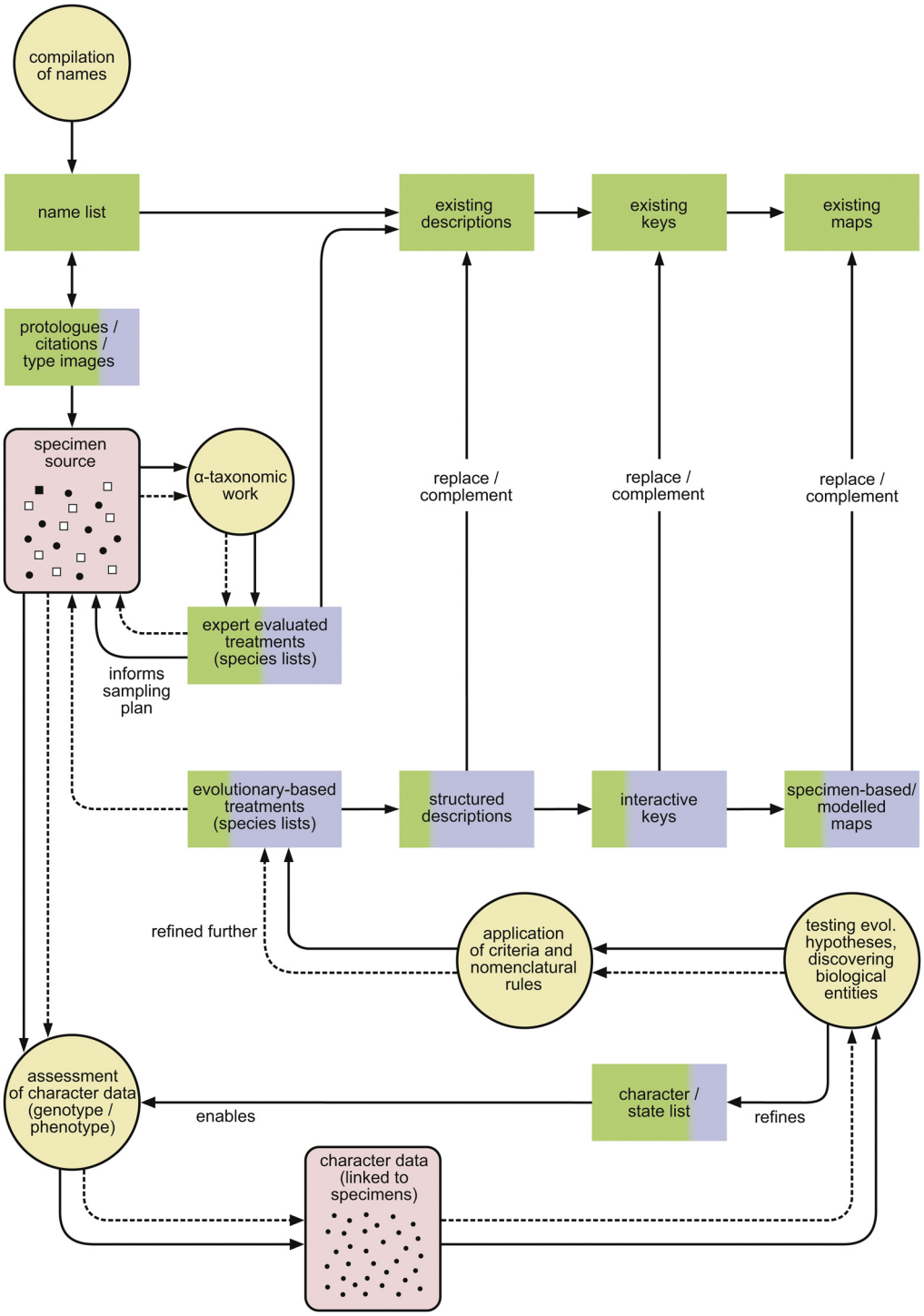

Fig. 2. Workflow for the global synthesis of a group of organisms such as the Caryophyllales that aims at complete species-level coverage. The compilation of names with the corresponding original descriptions (protologues) is the starting point. Specimens are a central information source. An alpha-taxonomic approach leads to proposing species concepts based on a series of criteria put forward by individual authors, along with descriptions, keys and maps (process shown in the area with white background). Because all species recognised have been treated using an alpha-taxonomic approach, this process can lead to full coverage relatively quickly. Nevertheless, only the assessment of a representative set of character data (both phenotype and genotype) from a representatively sampled number of individuals (specimens), and the subsequent analysis with an evolutionary methodology will allow the postulation of biological entities as reproducible and testable hypotheses. These entities can be named according to nomenclatural rules, leading to evolutionary-based treatments. At the same time, the process of testing hypotheses allows refining the character and state lists, which directly affects the assessment of phenotypic and genotypic character data. Since clear sets of individuals with their character data can be assigned to a species concept, the respective descriptions and keys can be generated in a structured way. This approximation process is shown in the area with grey background. While ongoing research will continuously improve the knowledge base and thus improve the treatments it will take considerable time to cover all species at this level. Alpha-taxonomic descriptions and keys will therefore be replaced in a step-by-step process as more knowledge becomes available. Circles represent actions, square boxes products. Green portions of the boxes reflect the status quo in Caryophyllales; blue corresponds to the estimated amount of work still needed. Different symbols at the specimen boxes reflect different quality of the specimens. Solid lines indicate direction of the actions in the workflow; dotted lines indicate options for iterative analyses.

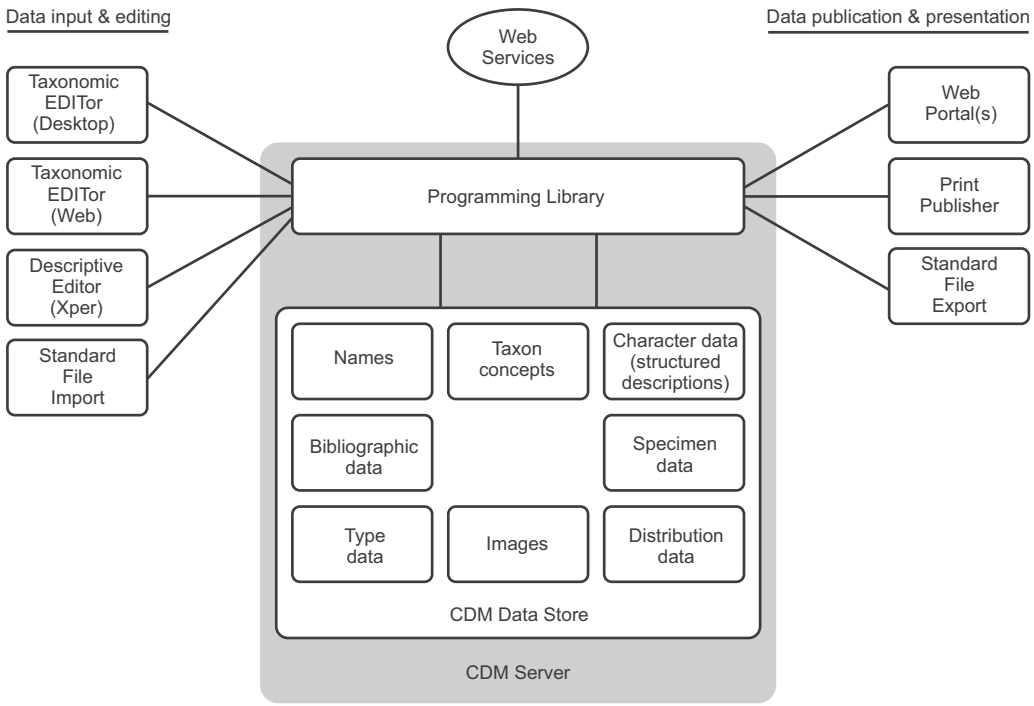

Fig. 3. Components of the EDIT Platform for Cybertaxonomy used in the synthesis workflow. Data input and editing can be effected by software clients (*Taxonomic EDITor*) over the internet or in a local area network; for descriptive data the Xper[2] software is used (*Descriptive Editor*). *Standard File Import* interfaces handle input in the form of community-standard data files. The *Data Store* holds the entire information used (names, character data, etc.) in highly standardised form (Common Data Model, CDM). The *Programming Library* is the part of the software that manages all operations including input, output, and complex processing of requests (such as putting a correct name into synonymy). Programmers use the Library to create application software for the users, such as the editors or the *Web Portals* and *Print Publisher* (for web and document-type output, including maps, descriptions and keys). *Standard File Export* interfaces allow to generate community standard files for import elsewhere (e.g. for the World Flora Online initiative) while *Web Services* allow machine-to-machine access to the Platform functionality in order to incorporate it into software elsewhere (e.g. specialised access portals with "life" access to the Platform, data quality control tools, research workflows, alternative in- and output tools, etc.).

specialists to our initiative are encouraging. An important step in further constituting the network community is the ongoing organisation of an international Caryophyllales conference in the autumn of 2015.

Implementing the work-flow and information system for the synthesis

An overview of the work-flow towards to synthesis of Caryophyllales is given in Fig. 2. In order to achieve dissemination of synthesis results as quickly as possible and to provide structured information sources for the community involved, the workflow generally follows two paths:

(1) Synthesis of existing information
   a. Compilation of species names (data ingestion)
   b. Integration of resources after expert review and verification (data ingestion: taxonomic treatments, literature, online resources, specimen information)
   c. Data publication and presentation (output products)
(2) Phylogenetic evaluation of taxon concepts based on structured data
   a. Assessment of molecular and morphological character data and ontology
   b. Testing hypotheses with phylogenetic methods
   c. Translating the results to species nomenclature and classification

The two main paths are strongly interrelated and will take place in parallel and with different speed in different taxonomic groups. Global coverage will therefore be reached independently for different sub-groups, and for the majority of these at an alpha-taxonomic level first. A fundamental principle is to streamline the replacement with more thoroughly evaluated treatments once they become available.

In the first step, existing resources such as taxonomic information already online as well as printed taxonomic treatments, other literature and specimens are being made technically available to be integrated. Most of the necessary interface specifications to access and import existing taxonomic information resources already exist, but the greater part of the electronic information available is insufficiently standardised to be directly used. Exceptions are nomenclatural and checklist information as well as specimen data, where many sources produce reasonably standardised output. Part of the expert review and verification will be to link these resources, e.g. the names from IPNI with the publication information present in JSTOR (2000) and BHL (2005) as well as with the type specimen information present in JSTOR Global Plants (JSTOR, 2013), the unresolved names in TPL (2013) with accepted names (where possible), and unambiguously circumscribed species from eFloras with expertly verified specimens that serve as vouchers (for information on sequences, pollen, morphological characters, and geographic occurrence, etc.). New data generated during that process include the references that were used in establishing these linkages and the information needed to give proper credit to the person who carried out the review and verification. In contrast to specimens and names, the available data from flora or monographic taxonomic treatments (i.e. descriptions, keys and supplemental information) are largely unstandardised. Thus, one of the first steps here will be to identify priority data sources, electronic or in print, in order to be able to mobilise that information and make it available for the Synthesis. These sources have of course the advantage to come with an authorship that can directly be used for proper credit at this stage. However, the process of transforming a taxonomic treatment from

print to database involves not only substantial technical but also a considerable amount of taxonomic expertise (Hamann et al., 2014), which both need to be properly credited.

In the following step, specialists for the different Caryophyllales groups will be using these existing resources as a basis for the revision and verification of information, and also for deeper evolutionary analysis. This step is crucial in terms of quality control and gap filling and highlights the necessity for the involvement of the respective specialist community. A literature database containing relevant literature on taxonomic, systematic, nomenclatural, phylogenetic, biogeographic and ecological aspects of Caryophyllales groups is currently under development. Several specialists have contributed reference libraries that are integrated and deduplicated to be made available online for the Caryophyllales community. The bibliography will be updated continuously to provide a useful tool for research.

Translating the results of phylogenetic and evolutionary analysis into a formal classification system as part of the work flow will thus help to overcome the "phylogeny-classification gap" (Franz, 2005). But this does not only hold true for the importance of classification systems that attach easily remembered names to organisms considering the vast majority of users (Franz, 2005). It also addresses another quality of a "phylogeny-classification gap" – which is that most phylogenetic studies still do not provide the needed data to support a species level classification, mostly due to limitations in taxon sampling. Our goal therefore is to more comprehensively integrate evolutionary analysis and classification work.

Establishing a dynamic taxonomic backbone as an integral part of the synthesis

As a first step the current understanding of family circumscriptions was evaluated with the aim of establishing well defined monophyletic groups. This resulted in the recognition of 37 families in the order, largely in congruence with the last classification of the Angiosperm Phylogeny Group (APG, 2009), but incorporating some recently published research results. Along the same line and aiming at full taxonomic coverage an annotated generic checklist of Caryophyllales is being created in an extensive collaborative effort (Hernández-Ledesma et al., unpubl. data). It will provide the first version of the genus level taxonomic backbone made available both as a dynamic electronic source and as a scientific paper. It builds upon the comprehensive but mostly pre-phylogenetic genus-level treatment for the entire order published by various family specialists in the "Families and genera of vascular plants" series; volumes two and five (Kubitzki and Bayer, 1993; Kubitzki et al., 2003). In order to reach a comprehensive cover of relevant genus names, the electronic version of Names in Current Use 3 (Greuter et al., 1997) was used as a starting point. More recent comprehensive family treatments including all genera were considered (e.g. for Aizoaceae – several authors in Hartmann, 2001; Cactaceae – Hunt, 2006; Basellaceae – Ericksson, 2007) as well as many specific papers that reflect the enormous progress in understanding the evolution of Caryophyllales in the last two decades. While the generic checklist provides a complete genus-level treatment of the order that represents current knowledge and identifies gaps, it must be emphasised that the individual quality of generic circumscriptions depends of the stage reached in the step-wise progress in research (see above). Ideally, genera represent well-defined monophyletic entities, but the treatment can also be provisional (phylogenetic trees lack resolution and support in respective nodes) or still at alpha-taxonomic level (no phylogenetic data available). Through explicitly stating the stage of knowledge, the generic synopsis also provides the framework for facilitating future knowledge generating activities, which will again lead to constant updates of the genus level treatment. The actual editorial procedures supporting and documenting this process are on the agenda for discussions in the Network.

To meet user needs, a full-coverage species-level taxonomic backbone will be built initially using available electronic sources (e.g. The Plant List) if there are no expert-reviewed resources (see below). Access will be name-based and conflicting classifications can be depicted in parallel (see the historical Provisional Global Plant Checklist, IOPPI, 2003). At the species level (including infraspecific taxa), the percentage of taxon concepts which have already been evaluated by truly evolutionary methods at this point cover only a small fraction of the estimated 12,500 entities. In case there are recent monographs (at least providing the character data in support of species concepts accepted) the species-level taxonomic backbone can reflect those monographs, easily giving credit to the respective author(s). This may extend to Flora treatments if they include whole genera or monophyletic subgeneric entities. For all Caryophyllales species names, the citations, protologues and type images will be included in or linked to the EDIT platform (Fig. 3) and presented on-line. This will be particularly important for those genera for which just an automatically compiled species list is available. The Caryophyllales network will help to identify colleagues who can more easily work out "expert validated species checklists" which are essentially alpha-taxonomic but progressed in quality over a simple names compilation and which will replace the initial species list. The resulting product will be credited to its author(s) and can be citable using formats such as data-papers (Chavan and Penev, 2011) or even a journal publication. It is estimated that a larger proportion of the total species diversity can be treated at least to this level until 2020.

The expert-validated species checklists will then also inform sampling plans for integrated morpho-molecular sampling of documented specimens in the course of evolutionary analyses. The hypotheses on biological entities identified in this research can then lead to an evaluation and eventually update of the species level treatment of the taxonomic backbone, then replacing the expert checklists. Since this kind of evolutionary research is more comprehensive, its publication as journal paper is likely, making it easier to credit authors. Developments in science evaluation (impact factors, etc.) will be reflected in the publication policies of the network. Print (or print-like) publications generated from the online resource will also be possible at appropriate times for significant taxonomic groups (e.g. families) or following specific stakeholder demands, also as part of journal articles dealing with new research results in evolutionary analysis.

Working towards a standardised ontology for characters, structured descriptions and linking specimens to character data

With regard to descriptive data, a forward-looking approach will turn to account for the plethora of knowledge on Caryophyllales accumulated in the literature and the research collections, while at the same time it will be oriented towards a more sustainable way of organising systematics than the common state of the art. This is a decisive part of the vision of the integrative on-line monographic approach.

Sustainability in systematics presupposes maximum data build-up. The corresponding requirements start with (1) unequivocal links of the gathered character data to the specimen from which they originate, continue with (2) standardised basic entities of information (i.e. terms and concepts of morphological characters and states) and (3) lead to iterative data aggregation procedures using congruent and interoperable character and state matrices from specimens to primary and higher taxa (see Kilian et al., submitted for publication). The first and last are basically information technological requirements. Standardised terms and concepts

for character data, in contrast, are core issues of systematics: they span from term ontologies to homology issues, they bridge between literature data and new data to be gathered, and they are fundamental to the entire process from data gathering to data aggregation and synthesis of information.

Playing such a central role, predefined standardised terms and concepts for character data receive a high priority in the timing and structure of the work process in the Caryophyllales Synthesis. Developing terms and concepts for a standard Caryophyllales character data matrix should thus be high on the agenda of the Network, starting out with an analysis of the characterisations ("descriptions") of the genera in the taxonomic backbone, taken from literature and unpublished own sources. In a first resulting work flow, contributing specialists extract terms and concepts of character data with references into a terminology wiki (Karam and Fichtmüller, 2014), forming the referenced elements of a growing illustrated standard term glossary with ontologies, which also builds on existing glossaries (Beentje, 2012; Eggli, 1993) and other activities regarding work flows and ontologies (Franz and Thau, 2010; Deans et al., 2012; Mathew and Güntsch, 2013). In a second resulting work flow, the unstructured textual genus descriptions are transformed into structured characterisations in a character data matrix (using the open source Xper2 software, see Ung et al., 2010a, 2010b) within the EDIT Platform (Fig. 3), with the support of the contributing specialists and employing the predefined standard entries of the Caryophyllales glossary. Working up all Caryophyllales genera in this way ensures that a vast majority of the actual range of characters and states in this order will be covered and compared. Of course, there will be a number of specialised characters that apply only for a specific group within the order, but we believe that also for these their descriptive value will be greatly enhanced with the inclusion into a joined semantic framework.

In parallel, three important tasks can be achieved in this way: (1) a standard Caryophyllales glossary and referenced ontology; (2) a standard Caryophyllales character data matrix; (3) a multi-access key to the genera of the Caryophyllales by using the functionalities of the Xper software on the basis of the Caryophyllales character data matrix. The structured way of data storage in both the standard glossary and the standard matrix enables their efficient editing, improving and supplementing, which is inevitable even in the long run. The standardisation of matrix and glossary ensures congruence and build up of the character data across the entire order.

In addition to this top-down approach, we envisage a complementary bottom-up approach, wherever appropriate, for specimen-based taxonomic work. This approach is based on the proposed solution to extent (a) the use of character data matrices from taxon characterisation to specimen data gathering, and (b) the data aggregation from lower to higher taxon ranks to the iterative aggregation of specimen-based data at species (or any other primary taxon) level (see Kilian et al., submitted for publication). Employing the Caryophyllales standard matrix for specimen data guarantees the data congruence. It brings the further advantage that subsequently revised taxon delimitations and specimen identifications can be easily brought into account by triggering a re-aggregation of the datasets.

Finally, since for the time being few structured data sets for taxon descriptions below generic rank will be available compared to the numerous unstructured textual descriptions in the literature (mostly of limited regional scope, as from floras), we envisage a mixed approach for taxon descriptions at lower taxon ranks: unstructured textual descriptions from the literature can be analysed in the same way as will be done for generic descriptions, thus adding both to the standard glossary and standard matrix. A literature item can then be transformed and incorporated as a separate item into an otherwise specimen-based application of the Caryophyllales standard character data matrix along with the specimen items. The mixed matrix data set can subsequently be aggregated into a taxon level matrix data set. Moreover, with growing numbers of specimen-based items, comparative evaluations of the specimen-based and literature-based items become possible. The data of the literature-based items are then either increasingly backed by the specimen-based data, or the discrepancies become apparent. In this way, our approach supplies a qualified assessment of the concept relationships (Berendsohn, 1995; Geoffroy and Berendsohn, 2003) between literature-based items and a preferred taxonomic concept.

Data publication and presentation

The synthesis of information will be carried out in the EDIT Platform for Cybertaxonomy (see above). Different interfaces are used to integrate information from all components, specialists (or editors) review and re-organise the content on-line, with support from the Network coordinators.

One key product of the global synthesis will be the Caryophyllales Portal, a scholarly information system that will serve as a single entry point for expert-reviewed quality information about the order. The Portal will be a product of the EDIT Platform (Fig. 3), web services and data exchange via standardised formats will be used to secure connectivity to international biodiversity informatics initiatives such as IPNI and GBIF, wherever considered appropriate by the Network and technically feasible. The platform web services are also the primary interface to advanced computational workflow environments. In terms of products, scholarly content and scientific publications are at the core of activities, but products for the interested general public (education purposes; updated quality material for teaching) should not be neglected. Illustrations (photographs, botanical illustrations) must be included.

The portal is not seen as a means to publish primary research results; it should provide the interface and synthesis of such publications and make the results available in standard formats for open data exchange. However, authors may wish to publish expertly revised alpha-taxonomic species catalogues as starting points for treatments through the portal. Using standard interfaces, quality data with respect to the taxonomic backbone as well as with respect to detailed species level treatments can also be made available to ongoing flora projects, the WFO initiative, and to other initiatives like the Catalogue of Life and the Encyclopaedia of Life.

The editorial processes and responsibilities involved in establishing the taxonomic backbone (and thus the baseline for the Synthesis) will be discussed and developed as a collaborative effort of the Caryophyllales Network. Basically, the principle is that the scientists or teams of scientists who work on a family or genus are also updating the backbone. The exact editorial process is being discussed. Clearly, this will be and for some time stay a highly iterative process, with new research results being constantly fed back into the information system. The resulting instability will represent a problem to users outside the taxonomic community. As in other systems (e.g. the Catalogue of Life, Roskov et al., 2014), regular stabilised and citable versions of the backbone and the Synthesis should be part of the output.

What is a realistic timeline for the synthesis?

Of course, this is a question that is intimately related to the availability of resources, monetary and intellectual, that can be dedicated to the initiative. Commitment of resources thus will be one of the principal discussion items within the network. The online bibliography is expected to be available by the end of 2015, and the genus-level backbone is already implemented in the EDIT Platform; its first edition will be available and published in 2015, which

will also be the time for the first launch of the Data Portal. With this available, the import of species names from online sources can commence and the EDIT Platform will have feedback mechanisms to these sources for corrections generated in the integration process. Where feedback interfacing cannot be automated, an annotation system based on the AnnoSys approach (Tschöpe et al., 2013) can be used. Specialist' review and updating of the species level backbone will start with a few exemplar groups, and first publication of comprehensive species checklists can be expected for 2016, including links to on-line protologues, type specimen images and (partial) distribution information (as soon as expertly identified specimens become available, point distribution maps can be generated by the portal). Work on the integration of descriptions and other character data from published works will be an ongoing accompanying process, in many cases seconded by negotiations with publishers on open access to such resources. Species and generic boundaries are still controversial in many groups and detailed studies at the species-level are often lacking. Once smaller monophyletic groups are identified in broader studies, these clades can be the focus of more detailed studies. During this step-wise process knowledge gaps will be identified and result in research providing evolutionary hypotheses for the generic and species limits in clades that are currently understudied. Information on the geographic distribution based (initially) on TDWG units (Brummitt, 2001) will considerably increase the utility of the portal over the usual checklist formats, as will the explicit stating of caveats with respect to taxonomic classification, distribution, nomenclature, etc., in order to specify potential problems for proper usage.

We do think that a minimum coverage of all species in a taxonomically consistent and comprehensive taxonomic backbone should be achievable by 2020, which incidentally is the deadline formulated by the CBD for a World Flora Online to become available for the conservation community – (CBD-SBSTTA, 2012). How far the comprehensive Synthesis including the species level sampling for phylogenetic analysis and the character matrix and ontology will have progressed by that year cannot be forecast at this stage, it is too dependent on the resources made available for research as well as on the degree of community integration achieved by the Network.

Conclusion

The current international initiatives to integrate and synthesise the actual taxonomic and systematic knowledge of flowering plants such as WFO or Catalogue of Life provide information standards for the compilation and accessibility of the available data using modern electronic tools. Nevertheless, their strategies neither launch international collaboration of experts nor the guidance of new generation of scientists focused on the study of biodiversity. They also do not actively promote more in-depth research, although only a rather small proportion of the species concepts currently used has hitherto been evaluated in an evolutionary context. We propose mechanisms to achieve these objectives using Caryophyllales as a model group. We aim to implement a system that allows not only the compilation of existing data, but also the generation of new information that can replace it in a stepwise process. This iterative process intends to shed light on the currently controversial taxon circumscriptions at different hierarchical levels. The basis of this progressive workflow is the integration of phylogenetic methodology to test current hypotheses of both taxa and characters based on specimens using structured and re-usable character data that are clearly linked to specimens. Our approach will also provide a perspective to the further development of the science of systematics (and taxonomy) in terms of improving the additivity and re-usability of its results.

Considering the methodological developments in systematics and the current stage of knowledge on the diversity of flowering plants, as exemplified by the order Caryophyllales, we conclude that users and the scientific community need to get into a closer dialogue. This will be essential in order to maintain the long-term quality of information for the users but also to promote application-oriented research in systematics, and thus highlight the relevance of this work. In this sense, the integrated global synthesis of the Caryophyllales can contribute to global initiatives such as the World Flora Online and can be a model how to sustainably organise this contribution through involving the scientific community.

Acknowledgements

We thank A. Güntsch and N. Turland for very helpful comments on a previous version of the manuscript. The Deutsche Forschungsgemeinschaft, the Bundesministerium für Bildung und Forschung (BMBF), the German Federal Agency for Nature Conservation and the European Commission are acknowledged for supporting a range of projects that contributed to the development of ideas towards this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ppees.2015.05.003

References

Agnarsson, I., Kutner, M., 2007. Taxonomy in a changing world: seeking solutions for a science in crisis. Syst. Biol. 56, 531–539, http://dx.doi.org/10.1080/10635150701424546

APG III, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161, 105–121, http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x

Beach, J.H., Pramanik, S., Beaman, J.H., 1993. Hierarchic taxonomic databases. In: Fortuner, R. (Ed.), Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision. John Hopkins University Press, Baltimore, pp. 241–256.

Beentje, H.J., 2012. The Kew Plant Glossary: An Illustrated Dictionary of Plant Terms. Kew Publishing, Richmond.

Berendsohn, W.G., 1995. The concept of "potential taxa" in databases. Taxon 44, 207–212.

Berendsohn, W.G., 1997. A taxonomic information model for botanical databases: the IOPI model. Taxon 46, 283–309.

Berendsohn, W.G. (Ed.), 2007–2010. ABCD – Access to Biological Collection Data. , http://wiki.tdwg.org/twiki/bin/view/ABCD/ (accessed 07.04.15).

Berendsohn, W.G., 2010b. Devising the EDIT platform for cybertaxonomy. In: Nimis, P.L., Vignes-Lebbe, R. (Eds.), Tools for Identifying Biodiversity: Progress and Problems. EUT Edizioni Università di Trieste, Trieste, pp. 1–6, http://hdl.handle.net/10077/3737 (accessed 07.01.15).

Berendsohn, W.G., Döring, M., Geoffroy, M., Glück, K., Güntsch, A., Hahn, A., Kusber, W.-H., Li, J.-J., Röpert, D., Specht, F., 2003. The Berlin taxonomic information model. Schr.reihe Veg.kd. 39, 15–42.

Berendsohn, W.G., Geoffroy, M., 2007. Networking taxonomic concepts – uniting without "unitaryism". In: Curry, G., Humphries, C. (Eds.), Biodiversity Databases – Techniques, Politics, and Applications. CRC Taylor & Francis, Baton Rouge, pp. 13–22.

Berendsohn, W.G., Güntsch, A., Hoffmann, N., Kohlbecker, A., Luther, K., Müller, A., 2011. Biodiversity information platforms: from standards to interoperability. Zookeys 150, 71–78, http://dx.doi.org/10.3897/zookeys.150.2166

BHL, 2005. Biodiversity Heritage Library. BHL Project, http://www.biodiversitylibrary.org/ (accessed 19.10.14).

Bisby, F.A., 1993. The ILDIS prototype for legumes. Ann. N. Y. Acad. Sci. 700, 159–164, http://dx.doi.org/10.1111/j.1749-6632.1993.tb26316.x

Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., Orchard, D. (Eds.), 2004. W3C Web Services Architecture. W3C Working Group Note 11 February 2004. , http://www.w3.org/TR/ws-arch/ (accessed 07.01.15).

Brach, A.R., Bouffort, D.E., 2011. Why are we still producing paper floras? Ann. Missouri Bot. Gard. 98, 297–300, http://dx.doi.org/10.3417/2010035

Brockington, S.F., Alexandre, R., Ramdial, J., Moore, M.J., Crawley, S., Dhingra, A., Hilu, K.W., Soltis, P.S., Soltis, D.E., 2009. Phylogeny of the Caryophyllales sensu lato: revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. Int. J. Plant Sci. 170, 627–643.

Brummitt, R.K., 2001. World geographical scheme for recording plant distributions. In: Plant Taxonomic Database Standards 2, 2nd ed. TDWG, Hunt Institute,

Pittsburg, http://www.nhm.ac.uk/hosted_sites/tdwg/TDWG_geo2.pdf (accessed 19.10.14).
Brummitt, R.K., Castrovieja, S., Chikuni, A.C., Orchard, A.E., Smith, G.F., Wagner, W.L., 2001. The Species Plantarum Project, an international collaborative initiative for higher plant taxonomy. Taxon 50, 1217–1230, http://www.jstor.org/stable/1224752
Campanula Portal, 2013. The Campanula Data Portal, Available at: http://dev.e-taxonomy.eu/dataportal/campanulaceae (accessed 16.01.15).
Cantino, P.D., de Queiroz, K., 2010. International Code of Phylogenetic Nomenclature, Version 4c, http://www.ohio.edu/phylocode/ (accessed 07.01.15).
CBD, 1998. Global Taxonomy Initiative, https://www.cbd.int/gti/ (accessed 19.10.14).
CBD, 2006. Convention on Biological Diversity. Guide to the Global Taxonomy Initiative. CBD Technical Series No. 30, http://www.cbd.int/doc/publications/cbd-ts-30.pdf (accessed 20.10.14).
CBD-SBSTTA, 2012. Global Strategy for Plant Conservation: World Flora Online by 2020. Note by the Executive Secretary, http://www.cbd.int/doc/meetings/sbstta/sbstta-16/information/sbstta-16-inf-38-en.pdf (accessed 07.01.15).
Chavan, V., Penev, L., 2011. The data paper – a mechanism to incentivize data publishing in biodiversity science. BMC Bioinform. 12 (Suppl. 15), S2, http://dx.doi.org/10.1186/1471-2105-12-S15-S2
Ciardelli, P., Kelbert, P., Kohlbecker, A., Hoffmann, N., Güntsch, A., Berendsohn, W.G., 2009. The EDIT platform for cybertaxonomy and the taxonomic workflow: selected components. Lect. Notes Inform. 154, 625–638.
Clayton, W.D., Vorontsova, M.S., Harman, K.T., Williamson, H., 2006. GrassBase – The Online World Grass Flora (accessed 20.10.14).
Costello, M.J., Appeltans, W., Bailly, N., Berendsohn, W.G., de Jong, Y., Edwards, M., Froese, R., Huettmann, F., Los, W., Mees, J., Segers, H., Bisby, F.A., 2014. Strategies for the sustainability of online open-access biodiversity databases. Biol. Conserv. 173, 155–165, http://dx.doi.org/10.1016/j.biocon.2013.07.042
Crouch, N., Smith, G.F., Figueiredo, E., 2013. From checklists to an e-flora for Southern Africa: past experiences and future prospects for meeting target 1 of the 2020 global strategy for plant conservation. Ann. Missouri Bot. Gard. 99, 153–160, http://dx.doi.org/10.3417/2011107
Cuénoud, P., Savolainen, V., Chatrou, L.W., Powell, M., Grayer, R.J., Chase, M.W., 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18SrDNA and plastid *rbcL*, *atpB*, and *matK* sequences. Am. J. Bot. 89, 132–144.
Deans, A.R., Yoder, M.J., Balhoff, J.P., 2012. Time to change how we describe biodiversity. Trends Ecol. Evol. 27, 78–84, http://dx.doi.org/10.1016/j.tree.2011.11.007
De Queiroz, K., Donoghue, M., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886, http://dx.doi.org/10.1080/10635150701701083
De Queiroz, K., Gauthier, J., 1992. Phylogenetic taxonomy. Ann. Rev. Ecol. Syst. 23, 449–480.
Diepenbroek, M., Glöckner, F., Grobe, P., Güntsch, A., Huber, R., König-Ries, B., Kostadinov, I., Nieschulze, J., Seeger, B., Tolksdorf, R., Triebel, D., 2014. Towards an integrated biodiversity and ecological research data management and archiving platform: the German Federation for the Curation of Biological Data (GFBio). In: Plödereder, E., Grunske, L., Schneider, E., Ull, D. (Eds.), Informatik 2014 – Big Data Komplexität meistern. GI-Edition: Lecture Notes in Informatics (LNI) – Proceedings 232. Köllen Verlag, Bonn, pp. 1711–1724.
Dillenberger, M.S., Kadereit, J.W., 2014. Maximum polyphyly: multiple origins and delimitation with plesiomorphic characters require a new circumscription of Minuartia (Caryophyllaceae). Taxon 63, 64–88.
Donoghue, M.J., 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist 88, 172–181, http://www.jstor.org/stable/3243026
Dubois, A., 2010. Taxonomy in the century of extinctions: taxonomic gap, taxonomic impediment, taxonomic urgency. Taprobanica 2, 1–5, http://dx.doi.org/10.4038/tapro.v2i1.2702
Eggli, U., 1993. Glossary of Botanical Terms with Special Reference to Succulent Plants. British Cactus and Succulent Society, Richmond.
Eichler, A.W., 1878. Blüthendiagramme, Zweiter Teil. W. Engelmann, Leipzig.
eMonocot, 2012. An Online Resource for Monocot Plants, http://e-monocot.org/ (accessed 20.10.14).
eMonocot Team, 2012a. eMonocot Cyperaceae, http://cyperaceae.e-monocot.org/ (accessed 20.10.14).
eMonocot Team, 2012b. CATE Araceae, http://araceae.e-monocot.org/ (accessed 20.10.14).
EOL, 2014. Encyclopedia of Life, http://www.eol.org (accessed 19.10.14).
Ericksson, R., 2007. A synopsis of Basellaceae. Kew Bull. 62, 297–320.
Flores-Rentería, L., Wegier, A., Ortega Del Vecchyo, D., Ortíz-Medrano, A., Piñero, D., Whipple, A.V., Molina-Freaner, F., Domínguez, C.A., 2013. Genetic, morphological, geographical and ecological approaches reveal phylogenetic relationships in complex groups, an example of recently diverged pinyon pine species (Subsection *Cembroides*). Mol. Phylogenet. Evol. 69, 940–949, http://dx.doi.org/10.1016/j.ympev.2013.06.010
Franz, N.M., 2005. On the lack of good scientific reasons for the growing phylogeny/classification gap. Cladistics 21, 495–500, http://dx.doi.org/10.1111/j.1096-0031.2005.00080.x
Franz, N.M., Peet, R.K., 2009. Towards a language for mapping relationships among taxonomic concepts. Syst. Biodivers. 7, 5–20, http://dx.doi.org/10.1017/S147720000800282X
Franz, N.M., Cardona-Duque, J., 2013. Description of two new species and phylogenetic reassessment of *Perelleschus*. O'Brien & Wibmer, 1986 (Coleoptera: Curculionidae), with a complete taxonomic concept history of Perelleschus sec Franz & Cardona-Duque, 2013. Syst. Biodivers. 11, 209–236, http://dx.doi.org/10.1080/14772000.2013.806371
Franz, N.M., Thau, D., 2010. Biological taxonomy and ontology development: scope and limitations. Biodivers. Inform. 7, 45–66, https://journals.ku.edu/index.php/jbi/article/view/3927/3852 (accessed 07.01.15).
Fuentes-Bazan, S., Uotila, P., Borsch, T., 2012. A novel phylogeny-based generic classification for *Chenopodium* sensu lato, and a tribal rearrangement of Chenopodioideae (Chenopodiaceae). Willdenowia 42, 5–24, http://dx.doi.org/10.3372/wi.42.42101
Funk, V.A., 2006. Floras: a model for biodiversity studies or a thing of the past? Taxon 55, 581–588, http://www.ingentaconnect.com/content/iapt/tax/2006/00000055/00000003/art00006 (accessed 07.01.15).
Geoffroy, M., Berendsohn, W., 2003. The concept problem in taxonomy: importance, components, approaches. Schr. Vegetationskd. 39, 5–14.
GBIF, 2014. Global Biodiversity Information Facility, http://www.gbif.org (accessed 20.10.14).
Godfray, H.C.J., 2002. Challenges for taxonomy. Nature 417, 17–19, http://dx.doi.org/10.1038/417017a
Godfray, H.C.J., Clark, B.R., Kitching, I.J., Mayo, S.J., Scoble, M.J., 2007. The web and the structure of taxonomy. Syst. Biol. 56, 943–955, http://dx.doi.org/10.1080/10635150701777521
Goldstein, P.Z., DeSalle, R., 2010. Phylogenetic species, nested hierarchies, and character fixation. Cladistics 16, 364–384, http://dx.doi.org/10.1006/clad.2000.0141
González Gutiérrez, P.A., Köhler, E., Borsch, T., 2013. New species of Buxus (Buxaceae) from northeastern Cuba based on morphological and molecular characters, including some comments on molecular diagnosis. Willdenowia 43, 125–137, http://dx.doi.org/10.3372/wi.43.43115
Gradstein, S.R., Sauer, M., Braun, M., Koperski, M., Ludwig, G., 2001. TaxLink, a program for computer-assisted documentation of different circumscriptions of biological taxa. Taxon 50, 1075–1084, http://dx.doi.org/10.2307/1224722
Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. Missouri Bot. Gard. 88, 373–457.
Greuter, W., Zimmer, B., Berendsohn, W.G. (Eds.), 1997. NCU-3e. Names in Current Use for Extant Plant Genera. Electronic version 1.0. , http://www.bgbm.org/iapt/ncu/genera/ (accessed 19.10.14).
Gurushidze, M., Fritsch, R.M., Blattner, F.R., 2010. Species-level phylogeny of *Allium* subgenus *Melanocrommyum*: incomplete lineage sorting, hybridization and trnF gene duplication. Taxon 59, 829–840, http://www.jstor.org/stable/25677671
Hagedorn, G., Thiele, K., Morris, R., Heidorn, P.B., 2005. The Structured Descriptive Data (SDD) w3c-xml-schema, Version 1.0, http://www.tdwg.org/standards/116/ (accessed 19.10.14).
Hamann, T.D., Müller, A., Roos, M.C., Sosef, M., Smets, E., 2014. Detailed mark-up of semi-monographic legacy taxonomic works using FlorML. Taxon 63, 377–393, http://dx.doi.org/10.12705/632.11
Hardisty, A., Roberts, D., 2013. A decadal view of biodiversity informatics: challenges and priorities. BMC Ecol. 13, e16, http://dx.doi.org/10.1186/1472-6785-13-16
Hartmann, H.E.K., Aizoaceae, A-E, Eggli, U., Hartmann, H.E.K. (Eds.), 2001. Illustrated Handbook of Succulent Plants. Springer, Heidelberg.
Hennig, W., 1966. Phylogenetic Systematics. Univ. Illinois Press, Urbana.
Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T.A.B., Rohwer, J.G., Campbell, C.S., Chatrou, L.W., 2003. Angiosperm phylogeny based on *matK* sequence information. Am. J. Bot. 90, 1758–1776.
Hobern, D., Apostolico, A., Arnaud, E., Bello, J.C., Canhos, D., Dubois, G., Field, D., García, E.A., Hardisty, A., Harrison, J., Heidorn, B., Krishtalka, L., Mata, E., Page, R., Parr, C., Price, J., Willoughby, S., 2013. Global Biodiversity Informatics Outlook – Delivering Biodiversity Knowledge in the Information Age. Copenhagen, http://www.biodiversityinformatics.org/ (accessed 28.12.14).
Hunt, D., 2006. The New Cactus Lexicon. DH Books, Milborne Port.
Holetschek, J., Dröge, G., Güntsch, A., Berendsohn, W., 2012. The ABCD of primary biodiversity data access. Plant Biosyst. 146, 771–779, http://dx.doi.org/10.1080/11263504.2012.740085
ICN, 2009. In: Hand, R., Kilian, N., von Raab-Straube, E. (Eds.), International Cichorieae Network: Cichorieae Portal. , http://wp6-cichorieae.e-taxonomy.eu/portal/ (accessed 17.10.14).
ILDIS, 1996. International Legume Database & Information Service, http://www.ildis.org/ (accessed 19.10.14).
IOPI, 1997–2003. Provisional Global Plant Checklist, http://www.bgbm.org/iopi/gpc/default.asp (accessed 02.01.15).
IPNI, 2004. The International Plant Names Index. The Plant Names Project, http://www.ipni.org (accessed 02.01.14).
JSTOR, 2000. JSTOR. ITHACA. http://www.jstor.org/ (accessed 19.10.14).
JSTOR, 2013. JSTOR Global Plants. New York. http://about.jstor.org/download/file/fid/22128 (accessed 19.10.14).
Karam, N., Fichtmüller, D., 2014. GFBio Terminology Server, http://terminologies.gfbio.org/ (accessed 08.01.15).
Karsch-Mizrachi, I., Nakamura, Y., Cochrane, G., 2011. The international nucleotide sequence database collaboration. Nucleic Acids Res. 40, D33–D37, http://dx.doi.org/10.1093/nar/gkr1006 (database issue).
KEsoftware, 2014. About KE Software's Electronic Museum Management System, EMu, http://emu.kesoftware.com/about-emu/ (accessed 19.10.14).
Kennedy, J., Hyam, R., Kukla, R., Paterson, T., 2006. Standard data model representation for taxonomic information. OMICS 10, 220–230, http://dx.doi.org/10.1089/omi.2006.10.220

Kilian, N., Henning, T., Plitzner, P., Müller, A., Güntsch, A., Stöver, B.C., Müller, K.F., Berendsohn, W.G., Borsch T., 2015. Sample data processing in an additive and reproducible taxonomic workflow by using character data persistently linked to preserved individual specimens. (submitted for publication).

Kim, S.-T., Sultan, S., Donoghue, M.J., 2008. Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. Proc. Natl. Acad. Sci. U. S. A. 105, 12370–12375, http://dx.doi.org/10.1073/pnas.0805141105

Kirschner, J., Kaplan, Ž., 2002. Taxonomic monographs in relation to global red lists. Taxon 51, 155–158, http://dx.doi.org/10.2307/1554973

Koch, M.A., Kiefer, M., German, D.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Schmickl, R., 2012. BrassiBase: tools and biological resources to study characters and traits in the Brassicaceae – version 1.1. Taxon 61, 1001–1009, http://www.ingentaconnect.com/content/iapt/tax/2012/00000061/00000005/art00007 (accessed 07.01.15).

Kubitzki, K., Bayer, C. (Eds.), 1993. The Families and Genera of Vascular Plants 5. Springer, Berlin.

Kubitzki, K., Rohwer, J.G., Bittrich, V. (Eds.), 2003. The Families and Genera of Vascular Plants 2. Springer, Berlin.

Legume Phylogeny Working Group, 2013. Legume phylogeny and classification in the 21st century: progress, prospects, and lessons for species-rich clades. Taxon 62, 217–248.

Mace, G.M., 2004. The role of taxonomy in species conservation. Philos. Trans. R. Soc. B 359, 711–719, http://dx.doi.org/10.1098/rstb.2003.1454

Maddison, D.R., Swofford, D.L., Maddison, W.P., 1997. NEXUS: an extensible file format for systematic information. Syst. Biol. 46, 590–621, http://dx.doi.org/10.1093/sysbio/46.4.590

Mallet, J., 2007. Species concepts. In: Levin, S., et al. (Eds.), Encyclopedia of Biodiversity 5. Academic Press (online update 1, pp. 1–15, Elsevier, Oxford).

Mallet, J., Willmott, K., 2003. Taxonomy: renaissance or Tower of Babel? Trends Ecol. Evol. 18, 57–59, http://dx.doi.org/10.1016/S0169-5347(02)00061-7

Mansion, G., Parolly, G., Crowl, A.A., Mavrodiev, E., Cellinese, N., Oganesian, M., Fraunhofer, K., Kamari, G., Phitos, D., Haberle, R., Akaydin, G., Ikinci, N., Raus, T., Borsch, T., 2012. How to handle speciose clades? Massive taxon sampling as a strategy towards illuminating the natural history of Campanula (Campanuloideae). PLoS ONE 7 (11), e50076, http://dx.doi.org/10.1371/journal.pone.0050076

Marhold, K., Stuessy, T. (Eds.) in collaboration with 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. http://www.ingentaconnect.com/content/iapt/tax/2013/00000062/00000001/art00003 (accessed 07.01.15).

Mathew, C., Güntsch, A., 2013. A Workflow for Data Refinement. BioVeL Newsletter No. 2, Autumn 2012: 3–5, https://www.biovel.eu/images/publications/newsno2autumn12.pdf (accessed 07.01.15).

Mathew, C., Güntsch, A., Obst, M., Vicario, S., Haines, R., Williams, A.R., de Jong, Y., Goble, C., 2014. A semi-automated workflow for biodiversity data retrieval, cleaning, and quality control. Biodivers. Data J. 2, e4221, http://dx.doi.org/10.3897/BDJ.2.e4221

Mayo, S.J., Allkin, R., Baker, W., Blagoderov, V., Brake, I., Clark, B., Govaerts, R., Godfray, C., Haigh, A., Hand, R., Harman, K., Jackson, M., Kilian, N., Kirkup, D.W., Kitching, I., Knapp, S., Lewis, G.P., Malcolm, P., von Raab-Straube, E., Roberts, D.M., Scoble, M., Simpson, D.A., Smith, C., Smith, V., Villalba, S., Walley, L., Wilkin, P., 2008. Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis. Kew Bull. 63, 1–16, http://dx.doi.org/10.1007/s12225-008-9014-1

Mayr, E., 1969. Principles of Systematic Zoology. McGraw-Hill, New York.

McDade, L.A., 1995. Species concepts and problems in practice: insight from botanical monographs. Syst. Bot. 20, 606–622, http://dx.doi.org/10.2307/2419813

McNeill, J., Barrie, F.R., Buck, W.R., Demoulin, V., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Marhold, K., Prado, J., Prud'homme van Reine, W.F., Smith, G.F., Wiersema, J.H., Turland, N.J. (Eds.), 2012. International Code of Nomenclature for Algae, Fungi, and Plants (Melbourne Code) Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011. , Gantner, Ruggell [Regnum Veg. 154].

MELNet, 2007–2014. Melastomataceae.Net. A Site with Information on the Biodiversity of Melastomataceae, http://www.melastomataceae.net (accessed 19.10.14).

Moore, G., 2003. Should taxon names be explicitly defined? Bot. Rev. 69, 2–21.

Mory, S.A., Prance, G.T., 2006–2010. The Lecythidaceae Pages. The New York Botanical Garden. Bronx, New York.

Mory, S.A., Smith, N.P., Cornejo, X., Prance, G.T., 2010. The Lecythidaceae Pages (http://sweetgum.nybg.org/lp/index.php). The New York Botanical Garden. Bronx, New York, http://sweetgum.nybg.org/lp/index.html (accessed 07.01.15).

Munzinger, J., Lowry, I.I., Callmander, P.P., Buerki, M.W.S., 2013. A taxonomic revision of the endemic New Caledonian genus *Podonephelium* Baill (Sapindaceae). Syst. Bot. 38, 1105–1124, http://dx.doi.org/10.1600/036364413X674814

Naciri, Y., Linder, P.L., 2015. Species delimitation and relationships: the dance of the seven veils. Taxon 64, 3–16, http://dx.doi.org/10.12705/641.24

Nic Lughadha, E., Miller, C., 2009. Accelerating global access to plant diversity information. Trends Plant Sci. 14, 622–628, http://dx.doi.org/10.1016/j.tplants.2009.08.014

Nixon, K.C., Carpenter, J.M., Stevenson, D.W., 2003. The PhyloCode is fatally flawed, and the "Linnaean system" can easily be fixed. Bot. Rev. 69, 111–120.

Nixon, K.C., Wheeler, Q.D., 2008. An amplification of the phylogenetic species concept. Cladistics 6, 211–223, http://dx.doi.org/10.1111/j.1096-0031.1990.tb00541.x

Nyffeler, R., Eggli, U., 2010. Disintegrating Portulacaceae: a new familial classification of the suborder Portulacinae (Caryophyllales) based on molecular and morphological data. Taxon 59, 227–240, http://www.jstor.org/stable/27757065

Oxelman, B., Rautenberg, A., Thollesson, M., Larsson, A., Frajman, B., Eggens, F., Petri, A., Aydin, Z., Töpel, M., Brandtberg-Falkman, A., 2013. Sileneae Taxonomy and Systematics, http://www.sileneae.info (accessed 05.01.15).

Palmweb, 2010. Palmweb: Palms of the World Online, http://www.palmweb.org/ (accessed 19.10.14).

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, http://dx.doi.org/10.1016/j.tree.2010.09.004

Paton, A., 2009. Biodiversity informatics and the plant conservation baseline. Trends Plant Sci. 14, 629–637, http://dx.doi.org/10.1016/j.tplants.2009.08.007

Pendry, C.A., Watson, M.F., 2009. Flora Malesiana and the Flora of Nepal: floristic links and the potential for collaboration. Blumea 54, 18–22, http://dx.doi.org/10.3767/000651909X474050

Pan-European Species directories Infrastructure (PESI), 2014. http://www.eu-nomen.eu/portal (accessed 19.10.14).

Pullan, M.R., Watson, M.F., Kennedy, J.B., Raguenaud, C., Hyam, R., 2000. The Prometheus Taxonomic Model: a practical approach to representing multiple classifications. Taxon 49, 55–75, http://dx.doi.org/10.2307/1223932

Rieppel, O., 2009. Species monophyly. J. Zool. Syst. Evol. Res. 48, 1–8, http://dx.doi.org/10.1111/j.1439-0469.2009.00545.x

Rieseberg, L.H., Carter, I.R., Zona, S., 1990. Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). Evolution 44, 1498–1511, http://dx.doi.org/10.2307/2409332

Rina, R., Berry, P.E., 2014. Euphorbia Plantetary Biodiversity Inventory Database, http://app.tolkin.org/projects/72/taxa

Roos, M.C., Berendsohn, W.G., Dessein, S., Hamann, T., Hoffmann, N., Hovenkamp, P., Janssen, T., Kirkup, D., de Kok, R., Sierra, S.E.C., Smets, E., Webb, C., van Welzen, P.C., 2011. e-Flora Malesiana: state of the art and perspectives. Gard. Bull. Singap. 63, 189–195, http://biostor.org/reference/140281

Roskov, Y., Kunze, T., Orrell, T., Abucay, L., Paglinawan, L., Culham, A., Bailly, N., Kirk, P., Bourgoin, T., Baillargeon, G., Decock, W., De Wever, A., Didžiulis.F V., 2014. Species 2000 & ITIS Catalogue of Life, 2014 Annual Checklist. DVD. Species 2000, Naturalis, Leiden, the Netherlands.

Särkinen, T., Olmstead, R.G., Bohs, L., Knapp, S., 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol. Biol. 13, 212, http://dx.doi.org/10.1186/1471-2148-13-214

Schäferhoff, B., Müller, K.F., Borsch, T., 2009. Caryophyllales phylogenetics: disentangling Phytolaccaceae and Molluginaceae and description of Microteaceae as a new isolated family. Willdenowia 39, 209–228, http://dx.doi.org/10.3372/wi.39.39201

Smith, G.F., Figueiredo, E., 2013. Type specimens online: what is available, what is not, and how to proceed. Reflections based on an analysis of the type specimens of southern African *Polygala* L. (Polygalaceae) accessible on the worldwide web. Taxon 62, 801–806, http://dx.doi.org/10.12705/624.5

Smith, V.S., Rycroft, S., Brake, I., Scott, B., Baker, E., Livermore, L., Blagoderov, V., Roberts, D., 2011. Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. Zookeys 150, 53–70, http://dx.doi.org/10.3897/zookeys.150.2193

Solanaceae Source, 2011. A Global Taxonomic Resource for the Nightshade Family, http://www.solanaceaesource.org/ (accessed 28.12.14).

Soltis, P.S., Soltis, D.E., 2009. The role of hybridization in plant speciation. Ann. Rev. Plant Biol. 60, 561–588, http://dx.doi.org/10.1146/annurev.arplant.043008.092039

Stuessy, T.F., 1990. Plant Taxonomy. The Systematic Evaluation of Comparative Data. Columbia University Press, New York, pp. 514 S.

Stuessy, T.F., Lack, H.W. (Eds.), 2011. Monographic Plant Systematics. Fundamental Assessment of Plant Biodiversity. Koeltz, Königstein. [Regnum Veg. 153].

TPL, 2013. The Plant List. Version 1.1, http://www.theplantlist.org/ (accessed 19.10.14).

Triebel, D., Hagedorn, G., Jablonski, S., Rambold, G. (Eds.), 1999. Diversity Workbench – A Virtual Research Environment for Building and Accessing Biodiversity and Environmental Data. , http://www.diversityworkbench.net (accessed 19.10.14).

Tropicos, 2014. Tropicos.org. Missouri Botanical Garden, http://www.tropicos.org (accessed 19.10.14).

Tschöpe, O., Macklin, J.A., Morris, R.A., Suhrbier, L., Berendsohn, W.G., 2013. Annotating biodiversity data via the internet. Taxon 62, 1248–1258, http://dx.doi.org/10.12705/626.4

Turland, N.J., 2013. The Code Decoded. A User's Guide to the International Code of Nomenclature for Algae, Fungi, and Plants. Koeltz, Königstein [Regnum Veg. 155].

Ung, V., Dubus, G., Zaragüeta-Bagils, R., Vignes-Lebbe, R., 2010a. Xper^2: introducing e-taxonomy. Bioinformatics 26, 703–704, http://dx.doi.org/10.1093/bioinformatics/btp715

Ung, V., Causse, F., Vignes Lebbe, R., 2010b. Xper2: managing descriptive data from their collection to e-monographs. In: Nimis, P.L., Vignes-Lebbe, R. (Eds.), Tools for identifying Biodiversity: Progress and Problems. EUT Edizioni Università di Trieste, Trieste, pp. 113–120, http://dbiodbs1.units.it/bioidentify/files/volume_bioidentify_low.pdf (accessed 07.01.15).

Vences, M., Guayasamin, J.M., Miralles, A., de la Riva, I., 2013. To name or not to name. Criteria to promote economy of change in Linnaean classification schemes. Zootaxa 3636, 201–244, http://dx.doi.org/10.11646/zootaxa.3636.2.1

Venin, M., Kirchhoff, A., Fradin, H., Güntsch, A., Hoffmann, N., Kohlbecker, A., Kuntzelmann, E., Maiocco, O., Müller, A., Vignes Lebbe, R., Berendsohn, W.G., 2010. Descriptive data in the EDIT Platform for Cybertaxonomy. In: Nimis, P.L., Vignes-Lebbe, R. (Eds.), Tools for identifying Biodiversity: Progress and Problems. EUT Edizioni Università di Trieste, Trieste, pp. 7–11, http://hdl.handle.net/10077/3738

Von Staden, L., Raimondo, D., Dayaram, A., 2013. Taxonomic research priorities for the conservation of the South African flora. S. Afr. J. Sci. 109, 1–10, http://dx.doi.org/10.1590/sajs.2013/1182

Vorontsova, M.S., Simon, B.K., 2012. Updating classifications to reflect monophyly: 10 to 20 percent of species names change in Poaceae. Taxon 61, 735–746, http://www.ingentaconnect.com/content/iapt/tax/2012/00000061/00000004/art00003 (accessed 07.01.15).

WCSP, 2014. World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew, http://apps.kew.org/wcsp/ (accessed 20.10.14).

Wieczorek, J., Döring, M., De Giovanni, R., Robertson, T., Vieglais, D., 2009. Darwin Core Terms: A Quick Reference Guide, http://rs.tdwg.org/dwc/terms/index.htm (accessed 20.10.14).

Zhong, Y., Jung, S., Pramanik, S., Beaman, J.H., 1996. Data model and comparison and query methods for interacting classifications in a taxonomic database. Taxon 45, 223–241, http://www.jstor.org/stable/1224663