The role of digital herbarium collections in the taxonomic revision of complex plant families: Salicaceae *sensu stricto*

Irina Belyaeva[1,2*], and *Nataliya* Kovtonyuk[3]

[1]Collections, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK
[2]Ecology of Woody Plants Department, Botanical Garden-Institute Ural Branch RAS, 620144 Yekaterinburg, Russia
[3] Digital Herbarium Group of Plant Systematics Laboratory, Central Siberian Botanical Garden, SB RAS, 630090 Novosibirsk, Russia

Abstract. The importance of digitisation and digital resources in botanical research is underlined. Advantages and disadvantages of the use of digital herbaria in taxonomic research in Salicaceae *sensu stricto* are discussed and some suggestions on possible improvements are made.

1 Introduction

Digitisation has become a necessary element of modern life, and science is no exception. Botanical research benefits most from digital Internet resources which provide not only fast access to scanned old and rare literature where botanical names were published (Biodiversity Heritage Library, BHL [1]) but also nomenclatural (International Plant Names Index, IPNI [2]) and taxonomic databases (The World Checklist of Vascular Plants, WCVP [3], Plants of the World Online, POWO [4]), Catalogue of Life, CoL [5], Tropicos [6], and also digital herbarium collections [7, 8]. Some of these online resources have hyperlinks to each other as for example at Royal Botanic Gardens, Kew (IPNI, WCVP, POWO) and to other valuable external resources, such as BHL [1], and internal Kew resources, such as the Kew Herbarium's Digital Collection [9] and the Royal Botanic Gardens Kew DNA Bank [10], all of which is a good example for other institutions in the world to follow.

The most important target of botanical research is an inventory of worldwide and regional floras [11, 12, 13, 14, 15, 16] and it cannot be achieved without the study of herbarium collections. Therefore, providing access to herbarium collections for the scientific community and wider public is a major current target of all botanical institutions which possess herbarium collections. Such use of collections justifies the existence and financial support of these organisations to some extent. It is out of the question that digitising of existing collections is the modern way for institutions to share them with the rest of the world.

There are ca 383,671 accepted vascular plant species [13] and ca. 400 million herbarium specimens of them that are stored in approximately 3000 herbaria around the world [8,17].

* Corresponding author: i.belyaeva@kew.org

Bearing this in mind, the work of a monographer of one complex family such as Salicaceae *sensu stricto*, which includes the two genera, *Populus* L. and *Salix* L., comprising ca. 500 recognised species and nearly 5000 recorded names [18] looks like a drop in the ocean. However, for one scientist it could be a work for their entire life, especially, if this work involves visiting different Herbaria around the world in person.

The purpose of this publication is to highlight the advantages and disadvantages of using digital resources, including virtual herbarium collections, in the everyday scientific work of a taxonomist-monographer, i.e. Salicaceae *s. str*. The authors of this paper follow the taxonomic concept of including the genera *Chosenia* Nakai and *Toisusu* Kimura in the genus *Salix* which is based on modern molecular-genetic research listed in the recent publication [19].

It has been pointed out that herbarium specimens are the only reliable evidence about the existence and distribution of plants [20, 21]. Information about the locality of collected plants in the past and present helps to prognose the situation with the distribution of plants in the future which is very important for the conservation of endangered taxa [22, 23, 24], monitoring plant health and their distribution in the conditions of climate change [25]. According to the comprehensive review by James and colleagues [26] and other publications [17, 27, 28] herbarium collections have been used in ecological research, bioengineering, forestry, food and medical security, ethnobotany and in training a new generation of botanists.

2 Digital herbarium collections and their current use in taxonomic research

The mass digitisation of herbarium collections is becoming a major topic of discussion [7, 17, 20, 27, 28, 29, 30, 31, 32, 33, 34, 35]. As underlined by Takano [35], China [36] and the United States [37] are taking specimen digitisation as seriously as national projects.

Seregin [28] named the leading Institutions by the number of digitised herbarium specimens as the Muséum National d'Histoire Naturelle, Paris, France (P, Herbarium codes follow [8]; the numbers for Salicaceae *s. str.* cited here are provided by the authors of this paper) – 5.4 million scanned specimens (incl. 31442 of Salicaceae *s. str.*); Naturalis, Leiden, Netherlands (AMD, L, U, WAG) – 4.5 million (incl. 23280 of Salicaceae *s.str*.); the New York Botanical Garden, New York, U.S.A. (NY) – 2 million (incl. 21551 of Salicaceae *s. str*.); the Smithsonian Institute, Washington, U.S.A. (US, USNS) – 2 million (incl. 12633 of Salicaceae *s.str*.); and the Institute of Botany, Chinese Academy of Sciencees, Beijing, China (PE) – 1.6 million scanned specimens (incl. 74682 of Salicaceae *s.str*.).

The list of the major virtual herbaria resources in the world was provided by Takano and colleagues [35] which contains URLs for collections of eight digital herbarium sources in North America (A, BRIT, F, FLAS, GH, MO, NY, US, USNS), two sources in South America (VALLE and Centro de Referência em Informação Ambiental, CRIA), eleven in Europe (AAU, B, BM, BR, E, H, K, P, WU, Z, ZAGR), two sources in Oceania (The Australasian Virtual Herbarium and New Zealand national herbarium network) and six resources in Asia (Chinese Virtual herbarium, Makino Herbarium Type Specimen Image Database, Shimane University Digital Herbarium, Thai Forest Herbarium Type Specimen Database, PE, TH, TI, TNS). The URLs for the sources mentioned above and other digital herbarium collections can be found also on the Virtual Herbaria website [38].

There is another important digital herbarium resource, the Global Biodiversity Information Facility [39], where one can find not only digitised specimens but also useful information on nomenclature, taxonomy and distribution. However, when working with this

resource one should understand that most of the herbarium specimens included in GBIF were not verified by specialists.

Digitisation of herbarium collections in Russia began relatively recently, and 14 Russian virtual botanical collections are listed on the website GBIG.ru (http://gbif.ru/digitized). The Moscow Digital Herbarium was launched in 2016 by Moscow University with the support of the Russian Science Foundation [29, 40]. Today the Moscow Digital Herbarium (https://plant.depo.msu.ru/) is a consortium of six Russian university and academic herbaria: MW (931396 specimens), MHA (78193), IRKU (14784), TUL (9000), KUZ (7131) and MAG (896), that contribute to the documentation of the Russian flora. To our knowledge there are 1115973 digitised online available herbarium specimens in Russia and for 465565 of them the data from herbarium labels are given. For Salicaceae *s.str*. there are 27970 scanned specimens (24974 – *Salix* and 2996 – *Populus*). Part of this data is also available in GBIF [39].

There are other virtual collections of herbarium specimens in the following Russian Institutions:

1. Komarov Botanical Institute RAS, LE (http://herbariumle.ru) with 76619 scanned specimens, including 595 of *Salix*. Data are being published via GBIF after verification [41].
2. Botanical Garden-Institute FEB RAS, VLA (http://botsad.ru/herbarium) with 55781 scanned specimens, including 4740 of Salicaceae. Their data are not published in GBIF.
3. Institute of Ecology of Plants and Animals UB RAS, SVER (https://herbarium.ipae.uran.ru) with 2521 scanned specimens of Salicaceae 2521 specimens. Data are being published via GBIF after verification [42].
4. Voronezh State Kozo-Polyansky University, VOR (http://herbarium.bio.vsu.ru) with ca 10000 scanned specimens. Family Salicaceae *s. str*. – no scanned specimens.
5. South Siberian Botanical Garden, ALTB (http://altb.asu.ru/index.php). The information about the number of scanned specimens is not available at the moment.
6. Central Siberian Botanical Garden SB RAS, NSK, NS (http://herb.csbg.nsc.ru:8081) with 60400 scanned specimens of vascular plants including 8800 of Salicaceae. Data are being published via GBIF after verification. The digitisation of herbarium collections began in 2014 and undertaken according to the international standards, at 600 dpi, with a barcode, 24-color scale and spatial scale bar [17]. High quality images with included herbarium label information were published as a dataset via the GBIF portal [43] and in the CSBG SB RAS Digital Herbarium (http://herb.csbg.nsc.ru:8081). Combining biodiversity data using GBIF is a very valuable tool for sharing biological information worldwide.

The study of verified digital specimens for a whole family allows the analysis of the taxonomy and variability within existing species in the entire area of their distribution and the production of real maps of the distribution and this is one of the massive advantages of virtual herbarium collections. Using special digital herbarium collections such as type collections, voucher collections and collections of cultivated plants makes taxonomic work much faster and easier. The use of virtual herbaria helps to save herbarium specimens from unnecessary damage and even loss by sending them on loan to other institutions. Some institutions do not send herbarium specimens on loan for financial reasons and the complexity of custom procedures. Especially difficult is nowadays to get herbarium specimens on loan from Russia to other countries and vice versa.

The challenges in identification and naming in Salicaceae *s. str.* were discussed recently [19] and access to the virtual herbarium collections worldwide was not mentioned at that time as a challenge. Thus, here are some disadvantages in the use of virtual herbarium collections in taxonomic work: on digital specimens of low resolution the tiny morphological details of flowers and leaves, such as nectary glands, type of hairiness,

stomata etc. cannot be seen, and if it is important for identification, verification could be very difficult. The colours of upper and lower surfaces of leaves of representatives of Salicaceae *s. str*. are very important for identification but they can be distorted by scanning, and it could also upset the identification and verification process.

3 Special digital herbarium collections

3.1. Type collections

All research that involves plants usually starts with an investigation of their distribution and the history of where, when and by whom they were originally described and asking two main questions: (1) To which taxon does a plant belong and (2) What is the correct name of this taxon [19]. To resolve these questions taxonomic and nomenclature study should be undertaken. Taxonomic opinions are changing over time as new information from different research fields becomes available [19, 44, 45]. In contrast, nomenclature is determined by the concept of nomenclatural type which is defined as an element to which the name of the taxon is permanently attached. Thus, if necessary, the plant in question could be compared to the type, which is, in the case of plants, the herbarium specimen or illustration (Art. 8.1. of the International Code of Nomenclature for Algae, Fungi, and Plants, ICN, [46]). The concept of type requires for the scientific names of plants to be validly published (Art. 7.2 of ICN [46]). Before the decision about the taxonomic identity of a plant can be made, a study of authentic specimens that were in possession of the publishing author of the name should be undertaken. The next important step, typification of the name, is recommended to be done only by specialists who understand the subject, i.e. a taxonomic group, very well (Rec. 9.A.2 of ICN [46]) and also understand the method of working of the author of the name (Rec. 9.A.1 of ICN [46]). All names in Salicaceae *s. str*. that were published before January 1st, 1958 are needed typification, and ca. 87 % of the names in this family are not yet typified. Existing digital collections of types and authentic material makes this work possible.

The usual practice in Herbaria is to prevent collections of type specimens from damage. Therefore, they have special covers with red marks on them to draw the attention of researchers for careful use. In herbarium collections of Russian Botanical Institutions type specimens do not only have special covers but are also stored separately from other specimens, listed and taken under special curation. This prevents misplacing and damaging of type collections. Digitising of authentic specimens that could belong to the original material is a way of sharing information about them without sending herbarium material on loan. Valuable information on authentic specimens can be found online via JSTOR Global Plants [47]. However high-resolution pictures in this resource can be accessed only by subscribing institutions. Other researchers should ask for every high-resolution scan they need.

3.2. Voucher collections

For cultivated plants, there are different rules for validly published cultivar names and such an element as type is not required by the International Code of Nomenclature for Cultivated Plants, ICNCP [48]. Therefore, most cultivar names are difficult to apply without an herbarium specimen attached to the name. In this case, verified herbarium voucher specimens are valuable to document plants used in research [45, 49, 50, 51, 52, 53, 54, 55]. There are also voucher specimens of non-cultivated plants that are used for different types of biological research (i.e. molecular research, chromosome counts, palynological,

anatomical, physiological, ecological and other types of research). Usually, voucher specimens are marked on the herbarium label but are not separated from general herbarium collections. Disadvantageous is a slow searching process as it involves searching through the whole collection, carefully reading every label. It would be very useful to have the option of a quick search for voucher specimens in electronic resources in the same way as it is now possible for type specimens by cataloguing them.

3.3. Collections of cultivated plants

Many plant names were described from living cultivated plants without citing any specimens, so that old specimens of cultivated plants became a very important source for identity investigation and neotypification of these names. Any authentic material collected and annotated by or affiliated to the author of the name are not separated from the rest of the specimens. These collections are neglected in the Botanical Institutions and the search for necessary specimens could be very time consuming. By mass digitisation such specimens could be marked and catalogued.

References

1. Biodiversity Heritage Library, BHL, https://www.biodiversitylibrary.org/Default.aspx (Accessed 29 June 2021)
2. International Plant Names Index (IPNI), https://beta.ipni.org/ (Accessed 29 June 2021)
3. Govaerts, R.H.A. (Ed.) The World Checklist of Vascular Plants (WCVP), https://wcvp.science.kew.org/ (Accessed 25 June 2021).
4. Plants of the World Online (POWO), http://plantsoftheworldonline.org/ (Accessed 29 June 2021)
5. Catalogue of Life (CoL), https://www.catalogueoflife.org/col/ (Accessed 29 June 2021)
6. Tropicos. https://www.tropicos.org/home (Accessed 29 June 2021)
7. B.M. Thiers, M.C. Tulig, K.A. Watson, Brittonia **68**, 3 (2016)
8. B. Thiers (Ed.) Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium, http://sweetgum.nybg.org/science/ih/ (Accessed 29 June 2021)
9. Kew Herbarium's Digital Collection, http://apps.kew.org/herbcat/gotoHomePage.do (Accessed 29 June 2021)
10. Royal Botanic Gardens Kew DNA Bank, https://dnabank.science.kew.org/homepage.html (Accessed 29 June 2021)
11. A. J. Paton, N. Brummitt, R. Govaerts, K. Harman, S. Hinchcliffe, B. Alkin, E. Nic Lughanda, Taxon **57** (2008)
12. M. J. Christenhusz, J. W. Byng, Phytotaxa **261**, 3 (2016)
13. E. Nic Lughandha, R. Govaerts, I. Belyaeva, N. Black, H. Lindon, R. Alkin, R. E. Magill and N. Nicolson, Phytotaxa, **272**, 1 (2016)
14. A. Paton, R. Allkin. R., I. Belyaeva, E. Dauncey, R. Govaerts, S. Edwards, J. Irving, C. Leon, E. Nic Lughadha, Plant name resources: building bridges with users. Botanists of the twenty-first century. Roles, challenges and opportunities. UNESCO (2016).

15. T. Borsch, W. Berendson, E. Dalcin, M. Delmas, S. Demissew, A. Elliot, P. Fritsch, A. Fuchs, D. Geltman, A. Güner, T. Haevermans, S. Knapp, M.M. Roux, P. A. Loizeau, C. Miller, J. Miller, J. T. Miller, R. Palese, A. Paton, J. Parnell, C. Pendry, H. N. Qin, V. Sosa, M. Sosef, E. Raab-Straube, F. Ranwashe, L. Raz, R. Salimov, E. Smets, B. Thiers, W. Thomas, M. Tulig, W. Ulate, V. Ung, M. Watson, P. Wyse Jackson, N. Zamora, Taxon **69**, 6 (2020)
16. S. T. Garnett, L. Christidis, S. Conix, M.J. Costello, F. E. Zachos, O. S. Bánki, Y. Bao, S. K. Barik, J. S. Buckeridge, D. Hobern, A. Lien, N. Montgomery, S. Nikolaeva, R. L. Pyle, S. A. Thompson, P. P. Van Dijk, A. Whalen, Z.Q. Zhang, K.R. Thiele, PLOS Biol., **18**, 7 (2020)
17. N. K. Kovtonyuk, I. V. Han, E. A. Gatilova, The flora of Asian Russiai **4**, 32 (2018)
18. I. V. Belyaeva, R. H. A. Govaerts, *Populu*s L., *Salix* L., The World Checklist of Vascular Plants (WCVP) https://wcvp.science.kew.org/search?q=Populus, https://wcvp.science.kew.org/search?q=Salix (Accessed 29 June 2021)
19. I. V. Belyaeva, Skvortsovia, **5**, 3 (2020)
20. N. K. Kovtonyuk, The flora of Asian Russia **1**, 25 (2017)
21. N. Kovtonyuk, I. Han, E. Gatilova, Informations Technologies in the Research of Biodiversity, SPEES (2019)
22. N. Brummitt, S. P. Bachmann, E. Aletrari, H. Chardburn, J. Griffiths-Lee, M. Lutz, J. Moat, M.C. Rivers, M. M. Syfert, E. M. Nic Lughandha, Philos. Trans. Ser. B, **370** (2015)
23. E. Nic Lughadha, B. E. Walker, C. Canteiro, H. Chadburn, A. P. Davis, S. Hargreaves, E. J. Lucas, A. Schuiteman, E. Williams, S. P. Bachman, D. Baines, A. Barker, A. P. Budden, J. Carretero, J. J. Clarkson, A. Roberts, M. C. Rivers, Philos. Trans., Ser. B, **374**, 1763 (2018)
24. M. M. Syfert, N. A. Brummitt, D. A. Coomes, N. Bystriakova, M. J. Smith, Global Ecol. Conserv. **16**, e004330 (2018)
25. N. Bystiakova, P. H. Alves De Melo, J. Moat, E. Nic Lughandha, A. K. Monro, Sci. Rep., **9**, 17037 (2019)
26. S. A. James, P. S. Soltis, L. Belbin, A. D. Chapman, G. Nelson, D. L. Paul, M. Collins, Appl. Pl. Sci. **6**, 2, e1024 (2018)
27. A. P. Seregin, Taxon, **65**, 1 (2016)
28. A. P. Seregin, Bot. j. **102**, 3 (2017)
29. A. P. Seregin, Taxon, **67**, 2 (2018)
30. G. Le Bras, M. Pignal, M. L. Jeanson, S. Muller, C. Aupic, B. Carré, G. Flament, M. Gaudeul, C. Gonçalves, V. R. Invernón, F. Jabbour, E. Lerat, P. P. Lowry, B. Offroy, E. P. Pimparé, O. Poncy, G. Rouhan, T. Haevermans, Sci. Data, **4**, 170016 (2017)
31. D. E. Kislov, V. A. Bakalin, E. A. Pimenova, V. P. Verkholat, P. V. Krestov, Botanica Pacifica **2** (2017)
32. A. Barber, C. Skema, M. Mancini, T. Block, Biodivers. Inform. Sci. Standards, **2**, e26181 (2018)
33. S. De Smedt, A. Bogaerts, Q. Groom, H. Engledow, Biodivers. Inform. Sci. Standards, **2**, e26140 (2018)
34. P. W. Sweeney, B. Starly, P. J. Morris, Y. Xu, A. Jones, S. Radhakrishnan, C. J. Grassa, C. C. Davis, Taxon, **67**, 1 (2018)

35. A. Takano, Y. Horiuchi, Y. Fujimoto, K. Aoki, H. Mitsuhashi, A. Takahashi, PhytoKeys **118** (2019)
36. Chinese Virtual herbarium, http://www.cvh.ac.cn/ (Accessed 29 June 2021)
37. Integrated Digitized Biocollections, iDigBio, https://www.idigbio.org/ (Accessed 29 June 2021)
38. Virtual Herbaria, https://herbarium.univie.ac.at/database/ (Accessed 29 June 2021)
39. Global Biodiversity Information Facility (GBIF), www.gbif.org (Accessed 29 June 2021)
40. A. P. Seregin (Ed.), Moscow Digital Herbarium: Electronic resource, https://plant.depo.msu.ru/ (Accessed 29 June 2021).
41. D. Melnikov, G. Tyusov, I. Belyaeva, I. Han, N. Kovtonyuk Salix pseudopentandra (Flod.) Flod. in the Herbarium (LE) of the Komarov Botanical Institute RAS. Version 1.10 (2020) https://doi.org/10.15468/udzn9d (Accessed via GBIF.org on 2021-06-29).
42. N. Erokhin, N. Kovtonyuk, I. Han, A. Vorobiev, D. Veselkin, E. Zvezdina, I. Belyaeva. Genus *Salix* L. in SVER Herbarium Institute of Plant and Animal Ecology (IPAE), Occurrence dataset https://doi.org/10.15468/5npjcc (2021) (Accessed via GBIF.org on 29 June 2021)
43. N. Kovtonyuk, I. Han, E. Banaev, I. Belyaeva, Genus *Salix* L. at the CSBG SB RAS Digital Herbarium. Version 1.5. Central Siberian Botanical Garden SB RAS (2020) Occurrence dataset https://doi.org/10.15468/6f3ybc (Accessed via GBIF.org on 29 June 2021)
44. A. M. Humphreys, H. P. Linder, Taxon **58**, 4 (2009)
45. T. M. Culley, Appl. Pl. Sci. **1**, 11 (2013)
46. N. J. Turland, J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W.-H. Kusber, D.-Z. Li, K. Marhold, T. W. May, J. McNeill, A. M. Monro, J. Prado, M. J. Price, G. F. Smith (eds.), International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159 (2018)
47. JSTOR Global Plants https://plants.jstor.org/ (Accessed 25 June 2021)
48. C. D. Brickell, C. Alexander, J. J. Cubey, J. C. David, M. H. A. Hoffman, A. C. Leslie, V. Malécot, V., X. Jin (eds.) International Code of Nomenclature for Cultivated Plants (ICNCP or Cultivated Plant Code), Ninth Edition (2016)
49. R. V. Goldblatt, P. C. Hoch, L. M. McCook, Ann. Missouri Bot. Gard. **79** (1992)
50. V. Funk, 100 Uses for an Herbarium: Well at least (2004) http://botany.si.edu/documents/2003_Funk_100Uses.pdf
51. V. A. Funk, P. C. Hoch, L. A. Prather, W. L. Wagner 2005. Taxon **54** (2005)
52. V. Funk, Pl. Press **17** (2014)
53. Z. A. Goodwin, D. J. Harris, D. Filler, J. R. Wood, R. W. Scotland Widespread mistaken identity in tropical plant collections. Curr. Biol. **25**, R10066–P1067 (2015)
54. J. Carranza-Rojas, H. Goeau, P. Bonnet, E. Mata-Montero, A. Joby, BMC Evol. Biol. **17**(1) (2017)
55. I. V. Belyaeva, C. Dutton, R. H. A. Govaerts, H. Liesebach, R. McGinn, M. Steenackers, G. Taylor, J. Pickett, Skvortsovia **6**, 3 (2020)