

## **CLASS I: DATA SET DESCRIPTORS**

### **A. Data set identity**

**Title:** The Global Naturalized Alien Flora (GloNAF) database

### **B. Data set identification code**

#### **Suggested data set identity codes:**

Taxon\_x\_List\_GloNAF\_vanKleunenetal2018Ecology.csv,  
List\_GloNAF\_vanKleunenetal2018Ecology.csv,  
Region\_GloNAF\_vanKleunenetal2018Ecology.csv,  
Reference\_GloNAF\_vanKleunenetal2018Ecology.csv,  
Shapefile\_GloNAF\_vanKleunenetal2018Ecology.zip

### **C. Data set description**

**Authors:** Mark van Kleunen<sup>1,2</sup>, Petr Pyšek<sup>3,4</sup>, Wayne Dawson<sup>5</sup>, Franz Essl<sup>6</sup>, Holger Kreft<sup>7,8</sup>, Jan Pergl<sup>3</sup>, Patrick Weigelt<sup>7</sup>, Anke Stein<sup>2</sup>, Stefan Dullinger<sup>6</sup>, Christian König<sup>7</sup>, Bernd Lenzner<sup>6</sup>, Noëlie Maurel<sup>2</sup>, Dietmar Moser<sup>6</sup>, Hanno Seebens<sup>9</sup>, John Kartesz<sup>10</sup>, Misako Nishino<sup>10</sup>, Alla Aleksanyan<sup>11</sup>, Michael Ansong<sup>12</sup>, Liubov A. Antonova<sup>13</sup>, Julie F. Barcelona<sup>14</sup>, Siegmund W. Breckle<sup>15</sup>, Giuseppe Brundu<sup>16</sup>, Francisco J. Cabezas<sup>17</sup>, Dairon Cárdenas<sup>18</sup>, Juliana Cárdenas-Toro<sup>19</sup>, Nicolás Castaño<sup>18</sup>, Eduardo Chacón<sup>20</sup>, Cyrille Chatelain<sup>21</sup>, Barry Conn<sup>22,23</sup>, Michele de Sá Dechoum<sup>24</sup>, Jean-Marc Dufour-Dror<sup>25</sup>, Aleksandr L. Ebel<sup>26</sup>, Estrela Figueiredo<sup>27,28</sup>, Ori Fragman-Sapir<sup>29</sup>, Nicol Fuentes<sup>30</sup>, Quentin J. Groom<sup>31</sup>, Lesley Henderson<sup>32</sup>, Inderjit<sup>33</sup>, Nejc Jogan<sup>34</sup>, Pavel Krestov<sup>35</sup>, Andrey Kupriyanov<sup>36</sup>, Silvana Masciadri<sup>37</sup>, Jan Meerman<sup>38</sup>, Olga Morozova<sup>39</sup>, Daniel Nickrent<sup>40</sup>, Arkadiusz Nowak<sup>41,42</sup>, Annette Patzelt<sup>43</sup>, Pieter B. Pelsler<sup>14</sup>, Wen-sheng Shu<sup>44</sup>, Jacob Thomas<sup>45</sup>, Ahmet Uludag<sup>46</sup>, Mauricio Velayos<sup>47</sup>, Alla Verkhosina<sup>48</sup>,

José L. Villaseñor<sup>49</sup>, Ewald Weber<sup>50</sup>, Jan J. Wieringa<sup>51</sup>, Ayşe Yazlık<sup>52</sup>, Abida Zeddami<sup>53</sup>, Elena Zykova<sup>54</sup>, Marten Winter<sup>55</sup>

**Institutions:** <sup>1</sup>Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, China; <sup>2</sup>Ecology, Department of Biology, University of Konstanz, Universitätsstrasse 10, D-78464 Konstanz, Germany; <sup>3</sup>The Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, CZ-252 43 Půhonice, Czech Republic; <sup>4</sup>Department of Ecology, Faculty of Science, Charles University, Viničná 7, CZ-128 44 Prague, Czech Republic; <sup>5</sup>Conservation Ecology Group, Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, United Kingdom; <sup>6</sup>Division of Conservation Biology, Vegetation and Landscape Ecology, University Vienna, Rennweg 14, 1030 Vienna, Austria; <sup>7</sup>Biodiversity, Macroecology & Biogeography, University of Goettingen, Büsgenweg 1, 37077 Göttingen, Germany; <sup>8</sup>Centre of Biodiversity and Sustainable Land Use (CBL), University of Goettingen, Büsgenweg 1, 37077 Göttingen, Germany; <sup>9</sup>Senckenberg Biodiversity and Climate Research Centre (SBIK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany; <sup>10</sup>Biota of North America Program (BONAP), Chapel Hill, NC, USA; <sup>11</sup>Institute of Botany aft. A.L. Takhtajyan NAS RA / Department of Geobotany and Plant Eco-Physiology, Acharyan 1, 0040, Yerevan, Armenia; <sup>12</sup>Department of Silviculture and Forest Management, Faculty of Renewable Natural Resources, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; <sup>13</sup>Institute for Aquatic and Ecological Problems, Far East Branch, Russian Academy of Sciences, 680000 Khabarovsk, Russia; <sup>14</sup>School of Biological Sciences, University of Canterbury, Private bag 4800, Christchurch 8140, New Zealand; <sup>15</sup>Department of Ecology, University of Bielefeld, Wasserfuhr 24-26, D-33619 Bielefeld, Germany; <sup>16</sup>University of Sassari, Department of Agriculture, Viale Italia 39, 07100 Sassari, Italy; <sup>17</sup>Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Jose Antonio Novais 12, Ciudad Universitaria, 28040-Madrid, Spain; <sup>18</sup>Herbario Amazónico Colombiano, Instituto Amazónico de Investigaciones Científicas Sinchi, 110311 Bogotá, Colombia; <sup>19</sup>Instituto de Investigacion de Recursos Biológicos Alexander von Humboldt, 111311 Bogotá, Colombia; <sup>20</sup>Escuela de Biología, Universidad de Costa Rica, 11501-2060, San José, Costa Rica; <sup>21</sup>Conservatoire et jardin botaniques (CJB), Département de la culture et du sport, chemin de l'Impératrice 1, 1292 Chambésy, Switzerland; <sup>22</sup>National Herbarium of New South Wales, Royal Botanic Gardens, Sydney, NSW 2000, Australia; <sup>23</sup>School of Life and Environmental Sciences, The University of Sydney, NSW 2006, Australia;

<sup>24</sup>The Horus Institute for Environmental Conservation and Development. AC Baía Sul, Caixa Postal 4543, Florianopolis, SC, 88047-971, Brazil; <sup>25</sup>Independent Ecologist, Research & Consultancy in Applied Ecology, Shachar street 1, 9626301 Jerusalem, Israel; <sup>26</sup>Department of Botany, Research Laboratory "Herbarium", Tomsk State University, Lenin Prospect 36, 634050 Tomsk, Russia; <sup>27</sup>Department of Botany, P.O. Box 77000, Nelson Mandela University, Port Elizabeth, 6031 South Africa; <sup>28</sup>Centre for Functional Ecology, Departamento de Ciências da Vida, Universidade de Coimbra, 3001-455 Coimbra, Portugal; <sup>29</sup>The Jerusalem Botanical Gardens and the Hebrew University Herbarium, Giv'at-Ram, 9190401, Jerusalem, Israel; <sup>30</sup>Departamento de Botánica, Facultad de Ciencias Naturales y Oceanograficas, Universidad de Concepción, Casilla 160-C, Concepción, Chile; <sup>31</sup>Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium; <sup>32</sup>ARC-Plant Protection Research Institute, Pretoria 0001, South Africa; <sup>33</sup>Department of Environmental Studies, Centre for Environmental Management of Degraded Ecosystems (CEMDE), University of Delhi, Delhi 110007, India; <sup>34</sup>Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; <sup>35</sup>Botanical Garden-Institute FEB RAS, Vladivostok 690024, Russia; <sup>36</sup>Institute of Human Ecology, Siberian Branch of Russian Academy of Sciences, Prospect Leningradsky 10, 650065 Kemerovo, Russia; <sup>37</sup>Consejo de Educación Técnico Profesional (CETP-UTU), Uruguay; <sup>38</sup>Belize Tropical Forest Studies, PO Box 208, Belmopan, Belize; <sup>39</sup>Institute of Geography RAS, Staromonetny 29, 119017 Moscow, Russia; <sup>40</sup>Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901-6509, USA; <sup>41</sup>Polish Academy of Sciences, Botanical Garden, Prawdziwka St. 2, 00-952 Warszawa, Poland; <sup>42</sup>University of Opole, Department of Biosystematics, 45-052 Opole, Poland; <sup>43</sup>Oman Botanic Garden, Diwan of Royal Court, P.O. Box 808, P.C. 122, Muscat, Sultanate of Oman; <sup>44</sup>State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, College of Ecology and Evolution, Sun Yat-sen University, Guangzhou 510275, China; <sup>45</sup>Department of Botany & Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; <sup>46</sup>Faculty of Agriculture, Canakkale Onsekiz Mart University, Canakkale, Turkey; <sup>47</sup>Departamento de Biodiversidad y Conservación, Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain; <sup>48</sup>Siberian Institute of Plant Physiology & Biochemistry of the Siberian Branch of the Russian Academy of Sciences (SIPPB SB RAS), Irkutsk, Russia; <sup>49</sup>Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México. Apartado Postal 70-233, 04510, Ciudad de México, México; <sup>50</sup>Biodiversity Research, University of Potsdam, Maulbeerallee 2, 14469 Potsdam, Germany; <sup>51</sup>Naturalis Biodiversity Centre, Darwinweg 2, Leiden, The Netherlands; <sup>52</sup>Department of Plant Protection, Faculty of Agriculture and Nature

Sciences, Düzce University, Düzce, Turkey; <sup>53</sup>Ingenieur en Ecologie vegetale, 17 rue Aissa Akif, Hussein-Dey. Algiers. Algeria; <sup>54</sup>Central Siberian Botanical Garden, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia; <sup>55</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

**Abstract:** This dataset provides the Global Naturalized Alien Flora (GloNAF) database, version 1.2. GloNAF represents a data compendium on the occurrence and identity of naturalized alien vascular plant taxa across geographic regions (e.g. countries, states, provinces, districts, islands) around the globe. The dataset includes 13,939 taxa and covers 1,029 regions (including 381 islands). The dataset is based on 210 data sources. For each taxon-by-region combination, we provide information on whether the taxon is considered to be naturalized in the specific region (i.e. has established self-sustaining populations in the wild). Non-native taxa are marked as ‘alien’, when it is not clear whether they are naturalized. To facilitate alignment with other plant databases, we provide for each taxon the name as given in the original data source and the standardized taxon and family names used by The Plant List Version 1.1 (<http://www.theplantlist.org/>). We provide an ESRI shapefile including polygons for each region and information on whether it is an island or a mainland region, the country and the Taxonomic Databases Working Group (TDWG) regions it is part of (TDWG levels 1-4). We also provide several variables that can be used to filter the data according to quality and completeness of alien taxon lists, which vary among the combinations of regions and data sources. A previous version of the GloNAF dataset (version 1.1) has already been used in several studies on e.g. historical spatial flows of taxa between continents, and geographical patterns and determinants of naturalization across different taxonomic groups. We intend the updated and expanded GloNAF version presented here to be a global resource useful for studying plant invasions and changes in biodiversity from regional to global scales. We release these data into the public domain under a Creative Commons Zero license waiver (<https://creativecommons.org/share-your-work/public-domain/cc0/>). When you use the data in your publication, we request that you cite this data paper. If GloNAF is a major part of the data analyzed in your study, you should consider inviting the GloNAF core team (see Originators in the Overall project description) as collaborators. If you plan to use the GloNAF dataset, we encourage you to contact the GloNAF core team to check whether there have been recent updates of the dataset, and whether similar analyses are already ongoing.

**D. Key words:** *alien plants, exotic plants, global distribution, naturalized plants, neophytes, non-native plants, species invasions, vascular plants.*

## **CLASS II: RESEARCH ORIGIN DESCRIPTORS**

### **A. Overall project description**

**Title:** GLONAF – a global database on naturalized alien vascular plants

**Originators:** Wayne Dawson, Franz Essl, Holger Kreft, Jan Pergl, Petr Pyšek, Mark van Kleunen, Patrick Weigelt, Marten Winter (in alphabetical order; all contributed equally to the building of the database)

**Period of study:** The database covers neophytes, i.e. alien plants that were introduced in the period 1492–2018. Note that the earliest and latest references used are from 1918 and 2018, respectively.

**Objectives:** The purpose of the GloNAF database is to provide a standardized compendium for the global distribution of naturalized alien vascular plant taxa.

**Abstract:** Same as above.

**Sources of funding:** Deutsche Forschungsgemeinschaft DFG, Kennedyallee 40, 53175 Bonn, Germany (project numbers KL 1866/9-1 and 264740629); Wissenschaftsfonds FWF, Sensengasse 1, 1090 Vienna, Austria (project number I2096-B16); The Czech Academy of

Sciences, Narodni 3, 117 20 Prague, Czech Republic (long-term research development project RVO 67985939 and Praemium Academiae award to PP).

## **B. Specific subproject description**

**Research methods:** From 2011 to 2018, the originators (see above) thoroughly searched the literature and internet, primarily using the English language, but also several other languages (e.g. Dutch, English, German, Russian and Spanish), to find inventories of naturalized alien vascular plants for geographic regions worldwide. This search also yielded contacts with taxonomists and invasion biologists who contributed to the GloNAF database with data from specific regions. The major data contributors are included as co-authors on this paper. Most of the regions correspond to countries or subnational administrative units, islands and island groups. In total, the database includes lists of alien plants for 1029 regions (Fig. 1), and 13,939 taxa. A region is defined here as the smallest geographic area for which we had a list of alien vascular plant species. We used 210 different data sources (see `References_GloNAF_vanKleunenetal2018Ecology.csv`), including naturalized alien plant compendia, national and subnational lists of naturalized alien plant taxa published in scientific journals, in reports, in books or on the internet. In addition to sources restricted to alien organisms, we also used books and online compendia of national or subnational floras that besides native taxa also include information on taxa that are alien and occur in the wild. Some of the inventories of naturalized alien taxa were specifically compiled for the GloNAF database (version 1.2) and have not been published (e.g. for the provinces of China) or the publication has been completed in the meantime (e.g. Inderjit et al. 2018, Ansong et al. in revision). The non-published data sources are marked in `References_GloNAF_vanKleunenetal2018Ecology.csv` by having ‘pers. communication’ in the reference.

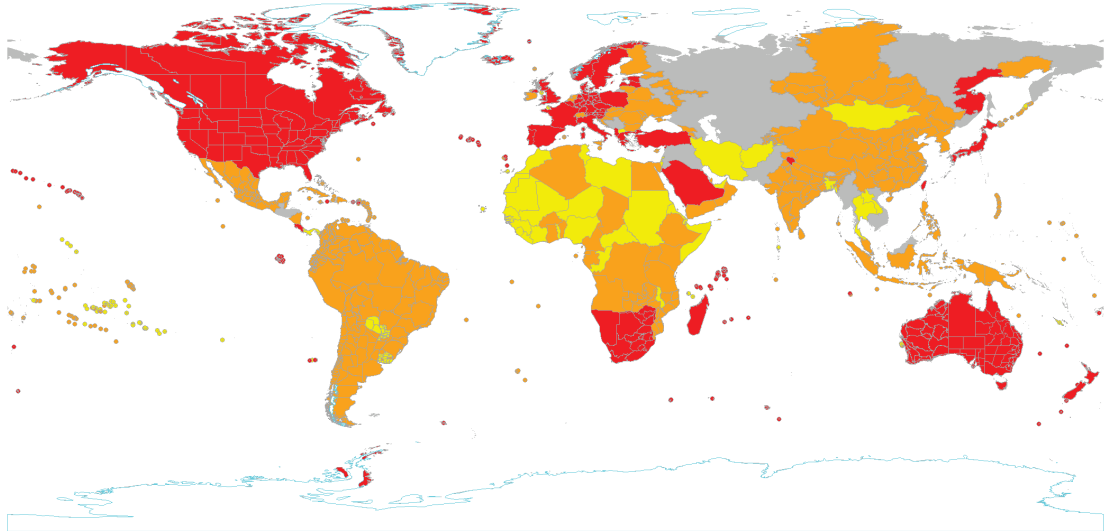


FIG. 1. Map of the world showing the 1029 regions included in the GloNAF database, version 1.2. Small islands and island groups are represented by dots to increase visibility. Regions in yellow likely have *very incomplete* naturalized plant inventories, regions in orange likely have *incomplete* inventories, and regions in red likely have *nearly complete* inventories (see Table 2). For the areas in grey, GloNAF does not have naturalized plant inventories. The permanent ice sheets are shown in white with a cyan border, and are from ESRI basemaps. The polygons for the GloNAF regions are provided in `Shapefile_GloNAF_vanKleunenetal2018Ecology.zip`.

Our primary focus was on naturalized vascular plant taxa, i.e. alien taxa that were introduced by humans and have established self-sustaining populations (Richardson et al. 2000, Pyšek et al. 2004). For regions whose lists differentiated between archaeophytes (e.g. alien taxa that were introduced in Europe before the year 1492) and neophytes (taxa that were introduced after the year 1492), we included only the latter, for reasons of comparability (most non-European countries only include neophytes), and because the alien status of some taxa classified as archaeophytes in some European regions is disputed. For some data sources, it was not clear whether all alien taxa in the inventory are naturalized or whether the inventory might also include alien taxa that are sometimes found in the wild but do not have established self-sustaining populations yet (so-called casuals; Richardson et al. 2000) or alien taxa that are only cultivated. In these cases, all alien taxa (in the given inventory) were assigned the status ‘alien’ instead of ‘naturalized’ for the specific region and data source. Taxa that were marked in a data source as being invasive were included as naturalized,

because invasive taxa are a subset of naturalized ones (Richardson et al. 2000). We did not include those taxa as a separate “invasive” category, because most data sources are not clear about the criteria used to define a naturalized taxon as being invasive. Of the 13,939 alien taxa in the database, 13,083 are naturalized in at least one region. For 947 of the 1029 regions in the database, we have lists of naturalized taxa, and for the remaining 82 regions, we only have lists of alien species (i.e. lists of non-native species for which it is not clear whether all species are naturalized).

Inventories of alien taxa that were not available in digital format were first converted into digital format. We copied the taxon names as they were given in the data source, including, if reported, infraspecific names and author names. To standardize scientific names, each alien plant inventory was compared to The Plant List (TPL) version 1.1 (<http://www.theplantlist.org>). This taxonomic standardization was done with the help of the *Taxonstand* package (Cayuela et al. 2017) in the statistical programming language R (version 3.4.2; R Core Team 2017), and also by manual corrections and checking of other websites and databases (in particular Tropicos [<http://tropicos.org/>] and IPNI [<http://www.ipni.org/ipni/plantnamesearchpage.do>]). Taxon names identified as synonyms were supplemented by the names accepted by TPL. Other taxon names were identified as unresolved in TPL or could not be found in TPL, in which case original names were kept as standardized names. From TPL, we also extracted for each taxon the name of the plant family it belongs to. For angiosperms, TPL follows the family circumscription of the Angiosperm Phylogeny Group (Chase and Reveal 2009).

We aimed to include the most comprehensive and most recent regional inventories, and for some regions multiple lists were included. Nevertheless, the availability and completeness of data varies geographically (Meyer et al. 2016), and it is therefore unavoidable that data quality differs among regional inventories of alien taxa. Consequently, for each inventory (i.e. list), we included several data-quality parameters as described subsequently (also see Table 2). If a single region had several lists, the quality of each of them was scored separately. Although we aimed to include inventories that considered all vascular plants (Tracheophyta), some were restricted to a taxonomic subset, for example seed plants (Spermatophyta) or angiosperms. Therefore, we included data on the taxonomic group covered by an inventory (variable ‘taxon\_group’). We also noted whether a data source explicitly refers to taxa as being naturalized (= established) alien plants (‘naturalized\_explicit’), whether it may have included casuals (‘casuals\_included’) or cultivated aliens (‘cultivated\_included’), and whether or not the inventory was restricted to



the subset of naturalized taxa that are considered invasive ('invasives\_only'). The inventories also likely vary in completeness. It is impossible to know with certainty how complete any list of taxa provided in an inventory really is. However, some lists, particularly very short ones and those restricted to invasive taxa, were judged to be likely incomplete by us or by the data-source authors (indicated by 'completeness\_judge'). Other lists, particularly very long ones with detailed information on the taxa, were judged to be near complete, and many were somewhere in-between. Therefore, we added a completeness score ('completeness'), which can be used to filter out likely less complete lists (see Fig. 1). However, since the scoring was subjective, the completeness scoring should be used with caution.

To facilitate mapping and spatial analyses in a geographical information system, we provide a zipped shapefile (Shapefile\_GloNAF\_vanKleunenetal2018Ecology.zip) with ESRI polygons (using projection WGS 84; EPSG: 4326) for each of the 1029 regions (see Fig. 1). We also provide for each region information on whether it is an island (also including island groups) or a mainland region. We also linked all regions to the four levels of the World Geographical Scheme for Recording Plant Distributions (Taxonomic Databases Working Group, TDWG; Brummit 2001), and include the name of the country the region belongs to. As a few GloNAF regions overlapped with multiple TDWG regions (e.g. Indonesia, Turkey), we created a few dummy TDWG regions at the different TDWG levels. The dummy TDWG IDs (tdwg1, tdwg2, tdwg3 and tdwg4 in Region\_GloNAF\_vanKleunenetal2018Ecology.csv) can be distinguished by their format: three digit-two digit combination at TDWG level 4 (instead of a three letter-two letter combination), three digits at level 3 (instead of three letters), three digits at level 2 (instead of two digits), 0 ('mixed') at level 1.

Our taxonomic harmonization and the provisioning of polygons of the different regions facilitates the alignment of GloNAF data with other databases. For example, for some analyses (e.g. van Kleunen et al. 2015) information on the native distribution might be useful. For such data, we refer to the Germplasm Resources Information Network (<https://npgsweb.ars-grin.gov/gringlobal/taxonomybrowse.aspx>) and the World Checklist of Selected Plant Families ([https://wcp.science.kew.org/prepareChecklist.do?checklist=selected\\_families%40%40254110920180947194](https://wcp.science.kew.org/prepareChecklist.do?checklist=selected_families%40%40254110920180947194)) databases. Some of the inventories that we used also reported information on the year of first record of an alien species in a particular region. These data are freely available from <http://dataportal-senckenberg.de/database/metacat/bikf.10029/bikf>.

**Project personnel:** For DFG KL 1866/9-1, the project postdoc was Anke Stein, and the principal and associated investigators were Mark van Kleunen and Wayne Dawson, respectively. For FWF I2086B16, the project PhD was Bernd Lenzner, and the principal and associated investigators were Franz Essl, Dietmar Moser, and Stefan Dullinger. For the Praemium Academiae project (The Czech Academy of Sciences), Petr Pyšek was the lead investigator, Jan Pergl and Martin Hejda were collaborators and Zuzana Sixtová a technician.

### **CLASS III: DATA SET STATUS AND ACCESSIBILITY**

#### **A. Status**

**Latest update:** 13 September 2018

**Data verification:** Data quality assurance checking has finished. We now call for experts to check the data and contribute new data. If someone identifies a potential mistake, has naturalized plant lists for regions that are not yet included in GloNAF version 1.2, or has more recent plant lists with information on naturalization status for regions that are already included, please, notify Mark van Kleunen or any of the other contact persons listed below.

#### **B. Accessibility**

**Storage location and medium:** The primary data are stored on a server at the University of Konstanz, Germany. The published data files are stored by the journal, and have been archived in the iDiv data repository (<https://idata.idiv.de/DDM/Data/ShowData/257>).

**Contact persons:** In alphabetical order:

Wayne Dawson, Conservation Ecology Group, Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, United Kingdom, email: [wayne.dawson@durham.ac.uk](mailto:wayne.dawson@durham.ac.uk)

Franz Essl, University Vienna, Rennweg 14, 1030 Vienna, Austria, email: [franz.essl@univie.ac.at](mailto:franz.essl@univie.ac.at)

Holger Kreft, Biodiversity, Macroecology & Biogeography, University of Goettingen, Büsgenweg 1, 37077 Göttingen, Germany, email: [hkreft@uni-goettingen.de](mailto:hkreft@uni-goettingen.de)

Jan Pergl, Institute of Botany, The Czech Academy of Sciences, Průhonice, Czech Republic, email: [jan.pergl@ibot.cas.cz](mailto:jan.pergl@ibot.cas.cz)

Petr Pyšek, Institute of Botany, The Czech Academy of Sciences, Průhonice, Czech Republic, email: [petr.pysek@ibot.cas.cz](mailto:petr.pysek@ibot.cas.cz)

Mark van Kleunen, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany, tel. +49 7531 88 2997, email: [mark.vankleunen@uni-konstanz.de](mailto:mark.vankleunen@uni-konstanz.de)

Patrick Weigelt, Biodiversity, Macroecology & Biogeography, University of Goettingen, Büsgenweg 1, 37077 Göttingen, Germany, email: [pweigel@uni-goettingen.de](mailto:pweigel@uni-goettingen.de)

Marten Winter, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; Deutscher Platz 5e; 04103 Leipzig; Germany, email: [marten.winter@idiv.de](mailto:marten.winter@idiv.de)

**Proprietary restrictions:** We release these data into the public domain under a Creative Commons Zero license waiver (<https://creativecommons.org/share-your-work/public-domain/cc0/>). When you use the data in your publication, we request that you cite this data paper. If GloNAF is a major part of the data analyzed in your study, you should consider inviting the GloNAF core team (see Originators in the Overall project description) as collaborators. If you plan to use the GloNAF dataset, we encourage you to contact the GloNAF core team to check whether there have been recent updates of the dataset, and whether similar analyses are already ongoing.

## **CLASS IV: DATA STRUCTURAL DESCRIPTORS**

### **A. Dataset files**

The GloNAF dataset has four csv files and a zipped shapefile that are linked as illustrated in Fig. 2.

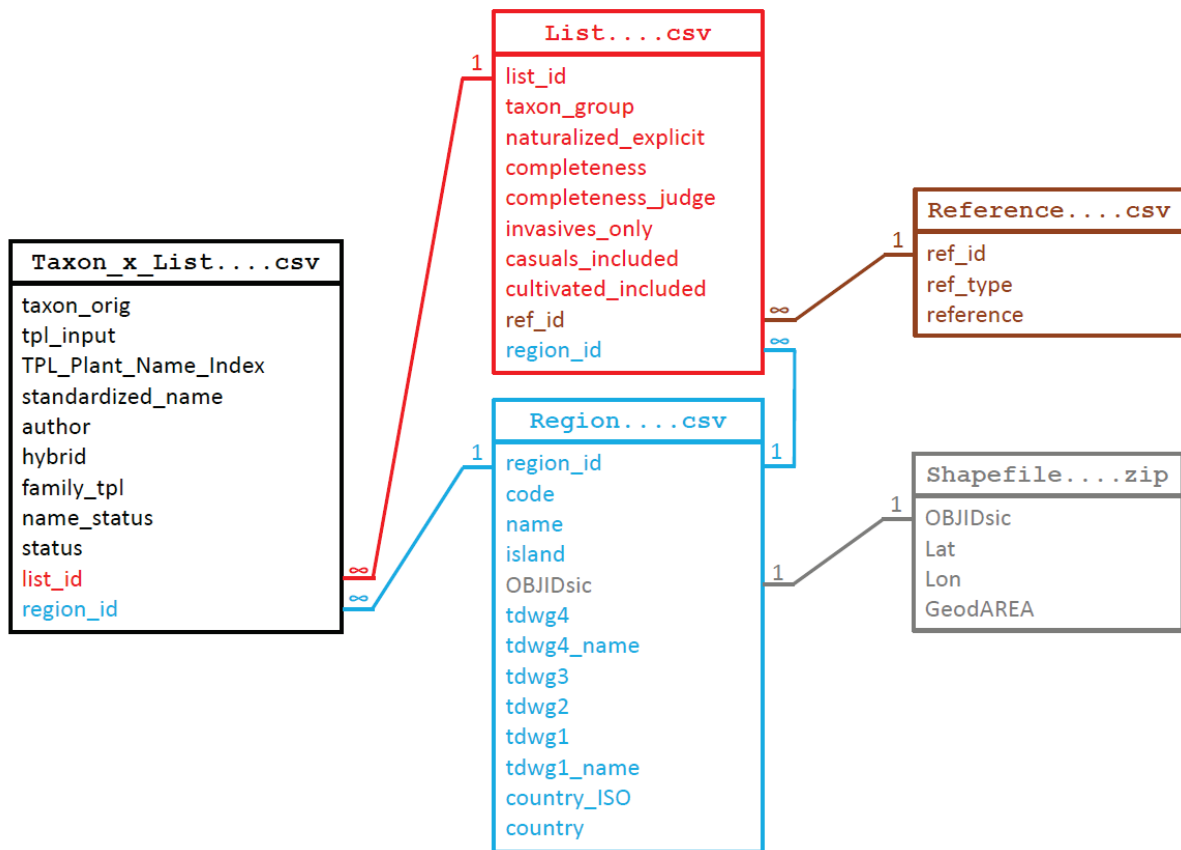


FIG. 2. Diagram illustrating how the four csv files and the zipped shapefile (for full names see below) are linked. The colors of the id variables (list\_id, region\_id, ref\_id, OBJIDsic), which link the different files, correspond to the colors used for the primary files they are part of. For a description of the variables in each file, see Tables 1-5.

**Identity:** Taxon\_x\_List\_GloNAF\_vanKleunenetal2018Ecology.csv

**Size:** 232,042 rows (excluding the header), 11 columns.

**Format and storage mode:** csv.

**Header information:** see Table 1

**Identity:** List\_GloNAF\_vanKleunenetal2018Ecology.csv

**Size:** 1463 rows (excluding the header), 10 columns.

**Format and storage mode:** csv.

**Header information:** see Table 2

**Identity:** Region\_GloNAF\_vanKleunenetal2018Ecology.csv

**Size:** 1029 rows (excluding the header), 13 columns.

**Format and storage mode:** csv.

**Header information:** see Table 3

**Identity:** Reference\_GloNAF\_vanKleunenetal2018Ecology.csv

**Size:** 210 rows (excluding the header), 3 columns.

**Format and storage mode:** csv.

**Header information:** see Table 4

**Identity:** Shapefile\_GloNAF\_vanKleunenetal2018Ecology.zip

**Size:** 1029 rows (excluding the header), 4 columns.

**Format and storage mode:** zip.

**Header information:** see Table 5

## B. Variable information

TABLE 1. Header information of

Taxon\_x\_List\_GloNAF\_vanKleunenetal2018Ecology.csv. Table indicates in which other table(s) the variable is used.

Variable name	Variable definition	Table
taxon_orig	The name of the taxon as provided in the data source	-
tpl_input	The name of the taxon as used as input for standardization of the taxon names according to The Plant List. This name may deviate from taxon_origin due to typos in the latter, cleaning of names (e.g. dropping superfluous text), modifying author information and changing of taxon names based on other taxonomic databases (e.g. infrataxa, basionyms).	

TPL_Plant_Name_Index	Variable indicating whether the tpl_input name is (1) or is not (0) included in The Plant List.	-
standardized_name	The taxon name (consisting of genus name, specific epithet, and in the case of infraspecific taxa also the infraspecific rank and epithet) after standardization according to The Plant List	-
author	Author of the standardized name according to The Plant List	-
hybrid	Variable indicating whether the taxon is (1) or is not (0) a hybrid according to The Plant List	-
family_tpl	The family of the taxon according to The Plant List	-
name_status	Variable indicating whether standardized_name is an 'accepted' or 'unresolved' name in The Plant List, or 'is not in The Plant List'	-
list_id	The id number of a naturalized/alien taxon inventory. Each list is a combination of a data source and a region. So, when there are two data sources for one region, we have two lists, and when one data source provides inventories for two regions, we also have two lists.	2
region_id	The id number of a GloNAF region, which is the smallest geographic area for which we had a list of alien vascular plant species	2, 3
status	Invasion status (alien, naturalized) of the alien taxon in the respective region according to the data source. Naturalized taxa are marked 'naturalized', and alien taxa for which it is not clear from the data source whether they are naturalized, casuals or only cultivated are marked 'alien'.	-

TABLE 2. Header information of

List\_GloNAF\_vanKleunenetal2018Ecology.csv. Table indicates in which other table(s) the variable is used.

<b>Variable name</b>	<b>Variable definition</b>	<b>Table</b>
list_id	The id number of a naturalized/alien taxon inventory. Each list is a combination of a data source and a region. So, when there are two data sources for one region, we have two lists, and when one data source provides inventories for two regions, we also have two lists.	1
taxon_group	Variable indicating the taxonomic group considered by a list: (1) Tracheophyta, (2) Spermatophyta, (3) Angiospermae, (4) Gymnospermae, (5) Pteridophyta (6) Other, (7) not clear, (8) Embryophyta	-
naturalized_explicit	Data-quality variable indicating whether the data source explicitly mentions that it includes naturalized taxa (for example by using the words 'naturalized',	-

	‘naturalization’ or ‘established’; note that use of these words does not necessarily mean that all taxa in the source are naturalized): (0) data source does not explicitly refer to naturalized taxa (but uses e.g. alien, exotic, introduced), (1) source explicitly refers to naturalized taxa, (2) source does not explicitly refer to naturalized taxa, but from the context it seems reasonable to consider the taxa to be naturalized (e.g. when it is mentioned that all species are widespread, or when the region is an uninhabited atoll were plants are not likely to be cultivated).	
completeness	Data-quality variable indicating the completeness of the taxon list: (1) likely very incomplete (most likely less than 50% of naturalized taxa included), (2) likely incomplete (most likely 50–90% of naturalized taxa included), (3) likely nearly complete (most likely >90% of naturalized taxa included). These scores are very crude and subjectively done by the GloNAF originators or respective source authors, and should be used cautiously.	-
completeness_judge	Variable indicating whether the completeness score was judged (0) explicitly by the author(s) of the data source, (1) implicitly by the author(s) of the data source, (2) by someone of the GloNAF originators	-
invasives_only	Data-quality variable indicating whether the taxon list provided by the data source is restricted to invasive taxa (1) or not (0)	-
casuals_included	Data-quality variable indicating whether the taxon list is likely to include casuals (i.e. alien taxa that sometimes occur in the wild but have no persistent populations): (0) casuals were unlikely to be included in the list because the data source focused on naturalized species or because casuals were marked and excluded, (1) casuals may be included in the list because information in the data source was not clear, (2) casuals were likely to be included because the data source focused on both naturalized and casual species and did not distinguish between these categories.	-
cultivated_included	Data-quality variable indicating whether the taxon list is likely to include cultivated alien plants that are not naturalized: (0) non-naturalized cultivated taxa were unlikely to be included in the list because the data source focussed on naturalized species (and casuals) or because non-naturalized cultivated alien plants were marked and excluded, (1) non-naturalized cultivated taxa may be included in the list because information in the data source was not clear, (2) non-naturalized cultivated taxa were likely to be included because the data source focussed on all kinds of alien plants and did not distinguish between different categories, (3)	-

	non-naturalized cultivated taxa were unlikely to be included in the list because they were excluded using ancillary data sources (this applies to the data for Mozambique, where cultivated aliens could be identified using <a href="http://www.mozambiqueflora.com/cult/index.php">http://www.mozambiqueflora.com/cult/index.php</a> , accessed on 27 July 2016).	
ref_id	The id number of the reference of the data source	4
region_id	The id number of a GloNAF region, which is the smallest geographic area for which we had a list of alien vascular plant species	1, 3

TABLE 3. Header information of

Region\_GloNAF\_vanKleunenetal2018Ecology.csv. Table indicates in which other table(s) the variable is used.

Variable name	Variable definition	Table
region_id	The id number of a GloNAF region, which is the smallest geographic area for which we had a list of alien vascular plant species	1, 2
code	The short code of a GloNAF region. This is usually based on the 3-letter ISO national code, which in the case of intra-national regions was extended by a dot and one or more 2- or 3-letter codes (e.g. USA.CA for California, United States of America).	-
name	The full name of the region, mostly based on the <i>wrld-sipl</i> dataset of the R package <i>maptools</i> (Bivand and Lewin-Koh 2017) for country names and on the names used in the data sources for subcountry level names	-
island	Variable indicating whether the region is an island (1) or a mainland region (0)	-
OBJIDSic	Polygon id linking to the shape file of the region	5
tdwg4	The TDWG-level 4 code to which the region belongs	-
tdwg4_name	Name of the TDWG-level 4 region	-
tdwg3	The TDWG-level 3 code to which the region belongs	-
tdwg2	The TDWG-level 2 code to which the region belongs	-
tdwg1	The TDWG-level 1 code to which the region belongs	-
tdwg1_name	Name of the TDWG-level 1 region	-
country_ISO	ISO 3166-1 alpha-3 country codes	-
country	Name of country, according to ISO	-



TABLE 4. Header information of

Reference\_GloNAF\_vanKleunenetal2018Ecology.csv. Table indicates in which other table the variable is used.

Variable name	Variable definition	Table
ref_id	The id number of the reference of the data source	2
ref_type	A variable indicating the type of data source: (1) peer-reviewed research paper (also including PhD theses), (2) reports or checklists on alien taxa by governmental or non-governmental organizations, (3) printed general flora or taxa checklist, which marked the alien taxa, (4) online general flora or checklist, which marked the alien taxa, (5) online alien taxa database, (6) compiled for GloNAF from various sources (including personal communications)	-
reference	The full reference to the data source	-

TABLE 5. Header information of

Shapefile\_GloNAF\_vanKleunenetal2018Ecology.zip. Table indicates in which other table the variable is used.

Variable name	Variable definition	Table
OBJIDsic	Polygon id linking to the shape file of the region	3
LAT	Latitude of the central point of the polygon of a region	-
LON	Longitude of the central point of the polygon of a region	-
GeodAREA	Area of the polygon of a region in km <sup>2</sup>	-

## CLASS V: SUPPLEMENTAL DESCRIPTORS

### A. Data acquisition

See the description in the section “Specific subproject description” of CLASS II.

### B. Quality assurance/quality control procedures

See the description in the section “Specific subproject description” of CLASS II.

### C. Related materials

The GIS polygons of the GloNAF regions, as shown in Fig. 1, are provided in the zipped shape file Shapefile\_GloNAF\_vanKleunenetal2018Ecology.zip.

### D. Computer programs and data-processing algorithms

Not applicable

### **E. Archiving**

The primary data are stored on a server at the University of Konstanz, Germany, and have been archived in the iDiv data repository (<https://idata.idiv.de/DDM/Data/ShowData/257>).

### **F. Publications and results**

GloNAF version 1.2 may be used to test fundamental questions and hypotheses in invasion biology and macroecology at global and regional scales. GloNAF may also be used to assess the naturalization risk of species in specific regions. A previous version of GloNAF (version 1.1) has already been used to test several hypotheses in invasion biology (for an overview, see <https://glonaf.org/index.php/publications/>). GloNAF made it possible to quantify, for the first time, the accumulation and exchange of naturalized alien plant species among continents (van Kleunen et al. 2015). In addition, GloNAF has been used in combination with species distribution models to forecast the risks of future invasions (Dullinger et al. 2017, Klonner et al. 2017, Mayer et al. 2017). These studies have revealed that climate change will increase the naturalization risk from alien ornamental garden plants in Europe (Dullinger et al. 2017). Furthermore, a modelling of invasion probability showed that countries with emerging economies, which are mostly countries with a high biodiversity, are the regions that are most vulnerable to future plant invasions because of the interaction of global trade and climate change (Seebens et al. 2015).

GloNAF has also been used in studies showing that niche dynamics of alien species do not differ between apomictic and sexually reproducing flowering plants (Dellinger et al. 2016), and that selfing ability is directly as well as indirectly associated with global naturalization success (Razanajatovo et al. 2016). Very recently, GloNAF helped to identify anthropogenic habitats as the major donors of European natives that have become naturalized elsewhere (Kalusová et al. 2017), and that stress-tolerators (S in the CSR scheme of Grime 1979) are less successful than ruderals (R) and competitors (C) in becoming naturalized (Guo et al. 2018). One should note, however, that inferring the probability of naturalization of individual species from its presence in GloNAF is limited due to missing data on whether species have actually been introduced to a region (when they failed to naturalized). Finally, the patterns in regional diversity and geographic distribution of naturalized plant species,

taxonomic, phylogenetic and life-history structure of the global naturalized flora as well as levels of naturalization and their determinants have been summarized in Pyšek et al. (2017a).

GloNAF has also inspired the creation of a new manuscript category “Alien Floras and Faunas” in the journal *Biological Invasions*, to accommodate the publication of regional alien species inventories (Pyšek et al. 2017b). Beyond vascular plants, GloNAF has a strong potential to be used in cross-taxon studies, when combined with similar data for other taxonomic groups. For example, GloNAF was already used to compare hot- and coldspots of alien species and their drivers across taxonomic groups at the global level (Dawson et al. 2017).

### **G. History of data set usage**

See the section “Publications and results”.

### **Acknowledgements**

We thank Zuzana Sixtová, Beate Rüter, Katya Mamonova, Vanessa Pasqualetto, Nina Gunselmann and Peter Kunze for technical support. The study was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, via iDiv [FZT 118] to MW, grant SE 1891/2-1 to HS, grant KL 1866/9-1 to MvK and WD, grant 264740629 to MvK), and Fondecyt Regular (grant N°1181688) to NF. PP and JP were supported by project no. 14-36079G Centre of Excellence PLADIAS (Czech Science Foundation) and long-term research development project RVO 67985939 (The Czech Academy of Sciences). PP acknowledges the support from Praemium Academiae award from The Czech Academy of Sciences. We thank the national football teams of France and Uruguay as the outcome of their match in the quarter final of the 2018 World Cup decided on the title of this data paper.

### **Literature cited**

- Bivand, R., and N. Lewin-Koh. 2017. *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.9-2. <https://CRAN.R-project.org/package=maptools>.
- Brummit, R. K. 2001. *World geographical scheme for recording plant distributions*. Ed. 2. Hunt Institute for Botanical Documentation, Pittsburgh.

- Cayuela, L., A. Stein, and J. Oksanen. 2017. Taxonstand: Taxonomic Standardization of Plant Species Names. R package version 2.1. <https://CRAN.R-project.org/package=Taxonstand>.
- Chase, M. W., and J. L. Reveal. 2009. A phylogenetic classification of the land plants to accompany APG III. *Botanical Journal of the Linnean Society* 161:122–127.
- Dawson, W., D. Moser, M. van Kleunen, H. Kreft, J. Pergl, P. Pyšek, P. Weigelt, M. Winter, B. Lenzner, T. Blackburn, E. Dyer, P. Cassey, S. Scrivens, E. Economo, B. Guénard, C. Capinha, H. Seebens, P. García-Díaz, W. Nentwig, E. García-Berthou, C. Casal, N. Mandrak, P. Fuller, C. Meyer, and F. Essel (2017) Global hotspots and correlates of alien species richness across taxonomic groups. *Nature Ecology & Evolution* 1: ~~Article No.~~ 0186.
- Dellinger, A., F. Essl, D. Hojsgaard, B. Kirchheimer, S. Klatt, W. Dawson, J. Pergl, P. Pyšek, M. van Kleunen, E. Weber, M. Winter, E. Hoerandl, and S. Dullinger. 2016. Niche dynamics of alien species do not differ among sexual and apomictic flowering plants. *New Phytologist* 209:1313–1323.
- Dullinger, I., J. Wessely, O. Bossdorf, W. Dawson, F. Essl, A. Gattringer, G. Klonner, M. Kuttner, D. Moser, J. Pergl, P. Pyšek, W. Thuiller, M. van Kleunen, P. Weigelt, M. Winter, and S. Dullinger. 2017. Climate change will increase naturalization risk from garden plants in Europe. *Global Ecology and Biogeography* 26:43–53.
- Grime, J. P. 1979. *Plant strategies and vegetation processes*. Wiley, Chichester.
- Guo, W.-Y., M. van Kleunen, M. Winter, P. Weigelt, A. Stein, S. Pierce, J. Pergl, D. Moser, N. Maurel, B. Lenzner, H. Kreft, F. Essl, W. Dawson, and P. Pyšek. 2018. The role of adaptive strategies in plant naturalization. *Ecology Letters* ~~doi: 10.1111/ele.13104~~.
- Inderjit, J. Pergl, M. van Kleunen, M. Hejda, C. R. Babu, S. Majumdar, P. Singh, S. P. Singh, S. Salamma, B. R. P. Rao & Pyšek. 2018. Naturalized alien flora of the Indian states: biogeographic patterns, taxonomic structure and drivers of species richness. *Biological Invasions* 20:1625–1638.
- Kalusová, V., M. Chytrý, M. van Kleunen, L. Mucina, W. Dawson, F. Essl, H. Kreft, J. Pergl, P. Weigelt, M. Winter, and P. Pyšek. 2017. Naturalization of European plants on other continents: the role of donor habitats. *Proceedings of the National Academy of Sciences*, 114:13756–13761.
- Klonner, G., I. Dullinger, J. Wessely, O. Bossdorf, M. Carboni, W. Dawson, F. Essl, A. Gattringer, E. Haeuser, M. van Kleunen, H. Kreft, D. Moser, J. Pergl, P. Pyšek, W. Thuiller, P. Weigelt, M. Winter, and S. Dullinger. 2017. Will climate change increase

- hybridization risk between potential plant invaders and their congeners in Europe? *Diversity and Distributions* 23:934–943.
- Mayer, K., E. Haeuser, W. Dawson, F. Essl, H. Kreft, J. Pergl, P. Pyšek, P. Weigelt, M. Winter, B. Lenzner, and M. van Kleunen. 2017. Naturalization of ornamental plant species in public green spaces and private gardens. *Biological Invasions* 19:3613–3627.
- Meyer, C., P. Weigelt, and H. Kreft. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters* 19:992–1006.
- Pyšek, P., D. M. Richardson, M. Rejmánek, G. Webster, M. Williamson, and J. Kirschner. 2004. Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon* 53:131–143.
- Pyšek, P., L. Meyerson, and D. Simberloff. 2017. Introducing “Alien Floras and Faunas”, a new series in *Biological Invasions*. *Biological Invasions* 6:1375–1376.
- Pyšek, P., J. Pergl, F. Essl, B. Lenzner, W. Dawson, H. Kreft, P. Weigelt, M. Winter, J. Kartesz, M. Nishino, L. A. Antonova, J. F. Barcelona, F. J. Cabezas, D. Cárdenas, J. Cárdenas-Toro, N. Castaño, E. Chacón, C. Chatelain, S. Dullinger, A. L. Ebel, E. Figueiredo, N. Fuentes, P. Genovesi, Q. J. Groom, L. Henderson, Inderjit, A. Kupriyanov, S. Masciadri, N. Maurel, J. Meerman, O. Morozova, D. Moser, D. Nickrent, P. M. Nowak, S. Pagad, A. Patzelt, P. B. Pelsler, H. Seebens, W. Shu, J. Thomas, M. Velayos, E. Weber, J. J. Wieringa, M. P. Baptiste, and M. van Kleunen. 2017a. Naturalized alien flora of the world: species diversity, taxonomic and phylogenetic patterns, geographic distribution and global hotspots of plant invasion. *Preslia* 89: 203–274.
- Razanajatovo, M., N. Maurel, W. Dawson, F. Essl, H. Kreft, J. Pergl, P. Pyšek, P. Weigelt, M. Winter, and M. van Kleunen. 2016. Plants capable of selfing are more likely to become naturalized. *Nature Communications* 7:13313.
- R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Richardson D. M., P. Pyšek, M. Rejmánek, M. G. Barbour, F. D. Panetta, and C. J. West. 2000. Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions* 6:93–107.
- Seebens, H., F. Essl, W. Dawson, N. Fuentes, D. Moser, J. Pergl, P. Pyšek, M. van Kleunen, E. Weber, M. Winter, and B. Blasius. 2015. Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology* 21:4128–4140.
- van Kleunen, M., W. Dawson, F. Essl, J. Pergl, M. Winter, E. Weber, H. Kreft, P. Weigelt, J. Kartesz, M. Nishino, L. A. Antonova, J. F. Barcelona, F. J. Cabezas, D. Cárdenas, J.

Cárdenas-Toro, N. Castaño, E. Chacón, C. Chatelain, A. L. Ebel, E. Figueiredo, N. Fuentes, Q. J. Groom, L. Henderson, Inderjit, A. Kupriyanov, S. Masciadri, J. Meerman, O. Morozova, D. Moser, D. Nickrent, A. Patzelt, P. B. Pelsler, M. P. Baptiste, M. Poopath, M. Schulze, H. Seebens, W. Shu, J. Thomas, M. Velayos, J. J. Wieringa, and P. Pyšek. 2015. Global exchange and accumulation of non-native plants. *Nature* 525:100–103.