SANBI
Biodiversity for Life
South African National Biodiversity Institute

African Biodiversity & Conservation
ISSN: (Online) 3078-8056, (Print) 3078-8048

An automated workflow to standardise taxon names for South African alien species lists

Author
Katelyn T. Faulkner[1,2]

Affiliations
[1]South African National Biodiversity Institute, Kirstenbosch Research Centre, Cape Town, 7735, South Africa.
[2]Department of Zoology and Entomology, University of Pretoria, Pretoria, 0028, South Africa.

Corresponding Author
Katelyn T. Faulkner; e-mail: katelynfaulkner@gmail.com

Dates
Submitted: 2 April 2025
Accepted: 1 December 2025
Published: 9 March 2026

How to cite this article:
Faulkner, K.T., 2026, 'An automated workflow to standardise taxon names for South African alien species lists', *African Biodiversity & Conservation* 56(2), a2. http://dx.doi.org/10.38201/abc.v56.2.a2

Accurate, up-to-date and interoperable alien species lists are required if South Africa is to estimate the status of biological invasions, effectively manage the problem, and meet its international obligations under the Global Biodiversity Framework. A plethora of South African alien species lists have been developed; but these have been standardised using different taxonomic backbones that change over time, and what has been done taxonomically is often not transparent or repeatable. Consequently, the lists are not interoperable, and the names of the listed taxa are often out-of-date and contain errors. Here a workflow to standardise taxon names for South African alien species lists is introduced. The workflow is open, has been automated and uses taxonomic backbones that are relevant to the South African context. The final output includes accepted scientific names, higher taxonomic information and details on taxonomic sources; and issues and errors are detected, corrected and flagged, including synonyms and duplicated taxa. Three lists of alien taxa, with error rates of 13–20%, and which differ in their taxonomic focus and backbones, serve as case studies. The case studies illustrate that the workflow can automatically standardise hundreds of taxon names in minutes, leaving a few (0–14%) for which manual work would be required. The workflow is a step-by-step guide of the processes required, will reduce error rates and facilitate the transparent and repeatable taxonomic standardisation of the names of alien taxa. In so doing, it will increase the interoperability and utility of South African alien species lists.

Keywords: alien species databases, alien species lists, biological invasions, R software environment, taxonomic backbones

Introduction

Biological invasions are a costly and serious threat to South Africa's biodiversity, ecosystem services, human health and livelihoods (Van Wilgen et al. 2022; Zengeya & Wilson 2023; McCulloch-Jones et al. 2024). As a signatory to the Convention on Biological Diversity, the country has committed to work towards meeting Target 6 of the Kunming-Montreal Global Biodiversity Framework for 2030 (Convention on Biological Diversity 2022) and thus needs to reduce the introduction and impact of invasive alien species (for the full text see https://www.cbd.int/gbf/targets/6). South Africa has also invested significantly in the management of biological invasions (McCulloch-Jones et al. 2024) and to better understand the challenge and the effectiveness of management interventions, a triennial national-level report on the status of biological invasions is produced (Van Wilgen & Wilson 2018; Zengeya & Wilson 2020; Zengeya & Wilson 2023). Lists of alien taxa (hereafter alien species lists) are key to all of these initiatives – they are used in research that improves our knowledge of biological invasions, are used to produce estimates of the status and trends in invasions, are an important tool for invasive species management and policy, and are used to measure and report on progress towards conservation targets, including Target 6 (McGeoch et al. 2012; Faulkner et al. 2015).

In South Africa, a plethora of alien species lists have been produced. These lists serve a wide range of purposes and have been developed following different methods (Faulkner et al. 2015). They are also often limited in scope and focus on specific taxonomic groups, spatial units (e.g., habitats, subnational political entities), time periods and introduction pathways (Faulkner et al. 2015). To give a few specific examples, lists of alien terrestrial molluscs (Herbert 2010) and marine organisms (Mead et al. 2011; Robinson et al. 2016, 2020) have been developed using surveys and reviews of historical records and museum specimens; roadside surveys were conducted to develop a list of alien plants along the Garden Route (Baard & Kraaij 2019); and a list of planted trees was recently developed using data from citizen science platforms (Richardson & Potgieter 2024). Surveys and inspections have been used to produce lists of alien taxa associated with the medicinal plant trade (Williams et al. 2021), pet trade (Nelufule et al. 2020; Shivambu et al. 2022) and agricultural imports (Saccaggi et al. 2021); while those associated with the aquarium trade were identified using DNA barcoding (Niemann et al. 2022). Lists are also available on the biological control agents that have been tested and released (Zachariades 2021); on high-risk alien taxa whose introduction should be prevented (Faulkner et al. 2014; Swart & Robinson 2019); and for taxa that are regulated under the Alien and Invasive Species Regulations of the National Environmental Management: Biodiversity Act (hereafter the NEM:BA A&IS Lists [for a list of acronyms see Table 1]; Wilson & Kumschick [2024]).

To fulfil their various purposes, these alien species lists need to be accurate, up-to-date and interoperable (McGeoch et al. 2012; Faulkner et al. 2015). Unfortunately, compiling such lists is a challenge, and they are often plagued by errors and uncertainties (McGeoch et al. 2012). Taxonomic complexities, ongoing taxonomic change and the decline in the number of taxonomists to handle taxonomic queries make it difficult to ensure that the listed taxon names are accurate (Soberón & Peterson 2004; Murray et al. 2017). Consequently, the taxon names in South African alien species lists are often out of date and contain errors, such as synonyms, misspellings, lexical variants (different ways of writing the same name), misapplications (where an incorrect name has mistakenly been given to a taxon) and duplicates (the same taxon listed multiple times either under the same or different name) (Soberón & Peterson 2004; Boyle et al. 2013; Murray et al. 2017). To give a specific example, almost half of the taxon names included in the NEM:BA A&IS Lists of 2020 do not correspond exactly to the accepted scientific name found in the relevant taxonomic backbone (i.e., collections of taxonomic information for many taxa; for a discussion on terms used see Grenié et al. [2022]), with this misalignment often due to slight formatting differences or an update in the nomenclature (Wilson & Kumschick 2024). South African alien species lists are also not interoperable, partly because taxon names are inconsistent across lists due to formatting differences (e.g., taxon names are provided with or without authorship information) and as different taxonomic backbones have been implemented. For example, lists of alien plants in South Africa have been taxonomically standardised using taxonomic backbones that include those of The Plant List (e.g., Baard & Kraaij 2019; Canavan et al. 2021), Plants of the World Online (e.g., Keet et al. 2022), and the Global Biodiversity Information Facility (GBIF; e.g., Niemann et al. 2022). In addition, what was done taxonomically is not always documented, or is not documented in enough detail to be transparent and repeatable. These issues make it difficult to consolidate and compare lists (Murray et al. 2017), compromise our ability to get accurate, comprehensive and up-to-date estimates of introductions and

Table 1. List of acronyms

Acronym	Expansion
BODATSA	Botanical Database of Southern Africa
FAIR	Findable, Accessible, Interoperable and Reuseable
GBIF	Global Biodiversity Information Facility
GISD	Global Invasive Species Database
ICTV	International Committee on Taxonomy of Viruses
ITIS	Integrated Taxonomic Information System
IUCN	International Union for Conservation of Nature
NEM:BA A&IS Lists	Lists of alien species regulated under the Alien and Invasive Species Regulations of the National Environmental Management: Biodiversity Act
POWO	Plants of the World Online
WCVP	World Checklist of Vascular Plants

invasions, and ultimately lead to misinformed policy and management (McGeoch et al. 2012).

A workflow that could serve as a step-by-step guide of the procedure required to source and process taxonomic data to standardise alien taxon names could assist to overcome these issues. Of course, such a workflow could be manually implemented. For example, the taxon names for the 560 taxa on the NEM:BA A&IS lists of 2020 have been manually corrected and standardised (Wilson & Kumschick 2024). But this would be time consuming, could result in duplicated efforts, has the potential to introduce human errors, and would be difficult in many cases due to the sheer amount of data involved (Boyle et al. 2013; Murray et al. 2017; Grenié et al. 2022). This is also no longer necessary, as large, online infrastructures, such as GBIF (https://www.gbif.org/), have made vast amounts of taxonomic information open-access, and because automated computational tools to extract, evaluate, manipulate and visualise this information have been developed (Grenié et al. 2022). These tools, although linked to various taxonomic backbones, are not linked to the regional taxonomic backbones that are relevant to the South African context (Grenié et al. 2022). Thus, the appropriateness of the existing automated tools for South African alien species lists is limited, and their implementation would result in alien species lists that are not interoperable with other lists of taxa for the region.

Here, a workflow to standardise taxon names for South African alien species lists is introduced. The workflow is open and has been automated. It implements taxonomic backbones that are relevant to the South African context; obtains accepted scientific names and higher taxonomic information; and detects, corrects and flags issues and errors. The script and detailed guidance documents are freely available, and so while the workflow was developed for South Africa, it can easily be adapted for use in other contexts. The development of this workflow should greatly assist the researchers, managers, and decision-makers that compile and use alien species lists. In the following sections, the workflow is described and its implementation is demonstrated using three South African alien species lists as case studies.

Description of the workflow

The automated workflow can be used to standardise digitised lists of taxon names that are in one of two formats: (1) full scientific name with authorship and date information (hereafter ‘scientific name’); and (2) scientific name without authorship and date information (hereafter ‘canonical name’). The workflow cannot be used on lists of common or vernacular names, and pre-processing of the names will likely be required (e.g., to ensure that the list is in the correct format [see Table 2]; Grenié et al. [2022]). While the workflow can be used to taxonomically standardise canonical names, it is recommended that information on authorship and date is retained as this decreases errors and makes it easier to resolve issues such as taxonomic homonyms (taxon names that are identical in spelling but belong to different taxa) (Jansen & Dengler 2010; Grenié et al. 2022). The workflow uses existing online taxonomic backbones and, as has

Table 2. A summary of the operations performed during each step and substep (in *italics*) of the workflow. A full description of the workflow is provided in Supplementary Material 1, and descriptions of the variables in the outputs are provided in Supplementary Material 2

Steps and substeps	Summary of operations
1. Preparation of R environment	• Required R packages are installed from the R CRAN and GitHub. • Required R packages are loaded.
2. Preparation of data	
2a. *Preparation of BODATSA dataset*	• BODATSA dataset, provided as a csv, is loaded (file name: BODATSA.csv). • The date the BODATSA dataset was downloaded is entered manually in day-month-year format. • Taxon name columns are merged to create a column for canonical names, and a column for scientific names. • Inconsistently formatted scientific names are flagged and manipulated to address this issue. • Misapplied taxon names are removed, unused autonyms are linked to their accepted names. • The prepared BODATSA dataset is written to file.
2b. *Preparation of list of taxon names*	• A csv file containing a single column (column name: verbatimScientificName) of taxon names is loaded (file name: OriginalNames.csv). • Information on the format of the names (whether canonical [without authorship and date information] or scientific [with authorship and date information]) is entered manually. • Formatting of spaces in taxon names is standardised.

Table 2. A summary of the operations performed during each step and substep (in *italics*) of the workflow. A full description of the workflow is provided in Supplementary Material 1, and descriptions of the variables in the outputs are provided in Supplementary Material 2 (continued)

3.	Match all taxon names to GBIF backbone	• Taxon names are passed to the GBIF application programming interface (API), accessed through the rgbif package (Chamberlain et al. 2025). • Taxonomic information for all taxa is returned. • Output is written to file.
4.	Extract plants	• Vascular plants are extracted from the output from GBIF.
5.	Assess and flag issues for non-plant taxa	• The output from GBIF for taxa that are not vascular plants is evaluated by determining the prevalence of various issues (e.g., names that are not accepted, matches that were not exact). • Subspecific entities, names with open nomenclature, duplicated names and names with other issues or errors are flagged.
6.	Preparation of output for non-plant taxa matched to GBIF	• A column is added with the source of the accepted scientific name, the date that source was consulted, and a stable link to GBIF that is unique to the taxon. • Unnecessary columns are removed from the dataset.
7.	Match plant names to BODATSA backbone	• Vascular plant names (see Step 4) are standardised using BODATSA (prepared in Step 2a). • Exact matches are identified based on the canonical name or scientific name (see Step 2b). • Taxon names with multiple matches are flagged and, if the issue cannot be resolved, all matches are returned. • The output is written to file.
8.	Assess and flag issues for plants	• The output from BODATSA for vascular plants is evaluated by determining the prevalence of various issues (e.g., names that are not accepted, names that had multiple matches). • Subspecific entities, names with open nomenclature, duplicated names and names with other issues or errors are flagged.
9.	Extract unmatched plants	• Vascular plants that could not be matched are extracted from the output from BODATSA.
10.	Preparation of output for plants matched to BODATSA	• A column is added with the source of the accepted scientific name, the date that source was downloaded (see Step 2a) and the general link to the BODATSA data. • Unnecessary columns are removed from the dataset.
11.	Match unmatched plant names to WCVP backbone	• Vascular plants for which a match could not be found in BODATSA are passed to the WCVP taxonomic backbone, accessed through the rWCVP package (Brown et al. 2023). • Either the taxon name provided by the user or the canonical name from GBIF is used (see Step 2b). • Taxon names with multiple matches are flagged and, if the issue cannot be resolved, all matches are returned. • The output is written to file.
12.	Assess and flag issues for plants matched to WCVP	• The output from WCVP is evaluated by determining the prevalence of various issues (e.g., names that are not accepted scientific names, matches that were not exact, names that had multiple matches). • Subspecific entities, names with open nomenclature, duplicated names and names with other issues or errors are flagged.
13.	Preparation of output for plants matched to WCVP	• A column is added with the source of the accepted scientific name, the date that source was consulted, and a stable link to POWO that is unique to the taxon. • Unnecessary columns are removed from the dataset.
14.	Merge	• The formats of the three interim outputs are standardised and aligned. • Interim outputs are merged. • The data are reordered so that the taxon names appear in the same order as provided. • Final output is written to file.

been recommended, uses and builds on existing tools where possible (Grenié et al. 2022).

There are a wide range of taxonomic backbones that vary in their taxonomic and spatial scope and selecting which to use is not trivial (Grenié et al. 2022). It has been recommended that regularly updated, regional, taxon-specific backbones should generally be preferred (Murray et al. 2017; Grenié et al. 2022), though if the list is of alien taxa not present in regional backbones, then global backbones are more appropriate (Grenié et al. 2022). Following these recommendations, this workflow uses three taxonomic backbones to standardise the taxon names in South African alien species lists: (1) the Botanical Database of Southern Africa (BODATSA) – a regional, regularly updated and publicly available (https://posa.sanbi.org/) taxonomic backbone for plants (Klopper et al. 2021) – is used to standardise the taxon names of alien vascular plants present outside of captivity and cultivation in South Africa (those only in captivity and cultivation are not included in BODATSA); (2) the World Checklist of Vascular Plants (WCVP) – a global plant taxonomic backbone that is updated weekly, and accessed through Plants of the World Online (POWO; https://powo.science.kew.org/) (Govaerts et al. 2021) – is used to standardise the taxon names of alien vascular plants that are not present outside of captivity and cultivation in South Africa (i.e., those not included in BODATSA); and, as a regional, open-access taxonomic backbone that is regularly updated does not exist for other groups, (3) the GBIF taxonomic backbone – a multi-taxa, global taxonomic backbone that is regularly updated – is used to standardise the taxon names of alien taxa that are not vascular plants (https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c).

The automated workflow is executed in R, and requires four inputs (Figure 1): (A) the BODATSA dataset, downloaded from https://posa.sanbi.org/sanbi/Explore; (B) the date that the BODATSA dataset was downloaded, which is manually entered into the script under Step 2a of the workflow; (C) the list of taxon names that need to be standardised; and (D) information on the format of the taxon names (i.e., scientific or canonical), which is manually entered into the script under Step 2b of the workflow. The workflow follows a sequence of 14 main steps, with each involving several substeps or operations (Figure 1). Step 2 requires the inputs detailed above, but all subsequent steps require as inputs the interim outputs of the previous steps, with some of these interim outputs being written to file (Figure 1).

When executed, the automated workflow passes all the provided taxon names to GBIF, and taxonomic information including canonical names and higher taxonomic information (e.g., kingdom, family, class) is returned for all taxa. Higher taxonomic information is required to divide the taxon names into taxonomic groups, and global databases such as GBIF have been recommended for this process (Grenié et al. 2022). Each taxon name is matched to the relevant taxonomic backbone and, if the provided name is a synonym, the accepted scientific name according to the backbone is returned. For vascular plants, if an exact match is not found in BODATSA, the name is passed to WCVP. If an exact match is not found for vascular plants in WCVP or for other taxa in GBIF, then fuzzy matching is performed. Fuzzy matching can correct misspellings and lexical variants but can result in errors; although these should be reduced here as higher taxonomic information is used to ensure that the returned fuzzy matched names correspond to the appropriate taxonomic group (Grenié et al. 2022). Duplicated taxon names, and taxon names with other issues and errors are detected and flagged, as are subspecific entities, and taxon names with open nomenclature (abbreviations that indicate uncertainty or the provisional status of the taxon's identification).

The final output is a single file with the taxon names as provided by the user; and for each taxon: the canonical name and higher taxonomic information obtained from GBIF; the full accepted scientific name with authorship and date information; and details on the source of the accepted scientific name. Taxa for which issues or errors were detected are flagged in the final output, whether these have been corrected or not. These include taxon names that are synonyms or are doubtful (e.g., taxonomic homonyms, misapplied names), duplicated taxon names, if the accepted scientific names were not identified through an exact match (e.g., fuzzy matching performed), and if an accepted scientific name could not be identified. Subspecific entities and taxon names with open nomenclature are also flagged in the final output. Global biodiversity data standards as set by the Darwin Core are followed as far as is possible (descriptions of the Darwin Core fields are available at: https://dwc.tdwg.org/terms/).

A summary of each step of the workflow is provided in Table 2, with a full description in Supplementary Material 1.

Case studies

Three lists were used as case studies to test and illustrate the workflow: a watchlist of invasive species whose introduction to South Africa should be prevented (Faulkner et al. 2014); a list of taxa intercepted at South Africa's ports of entry during agricultural inspections (Saccaggi et al. 2021); and a list of macrophytes traded in South Africa (Niemann et al. 2022). These lists were selected as they were developed for different purposes; focus on different taxonomic groups and habitats; the formatting of the taxon names differs; and as they were developed using different taxonomic backbones (Figure 2).

Of the taxon names included on the three lists, between 13 and 20% had errors (Figure 2). When the automated workflow was executed, the taxon names in the three lists were rapidly standardised (although note that the throughput will vary based on computer hardware and internet connection speed). Nonetheless, the watchlist, which lists 400 taxa and was both the longest and most taxonomically complex of the three lists (Figure 2), was

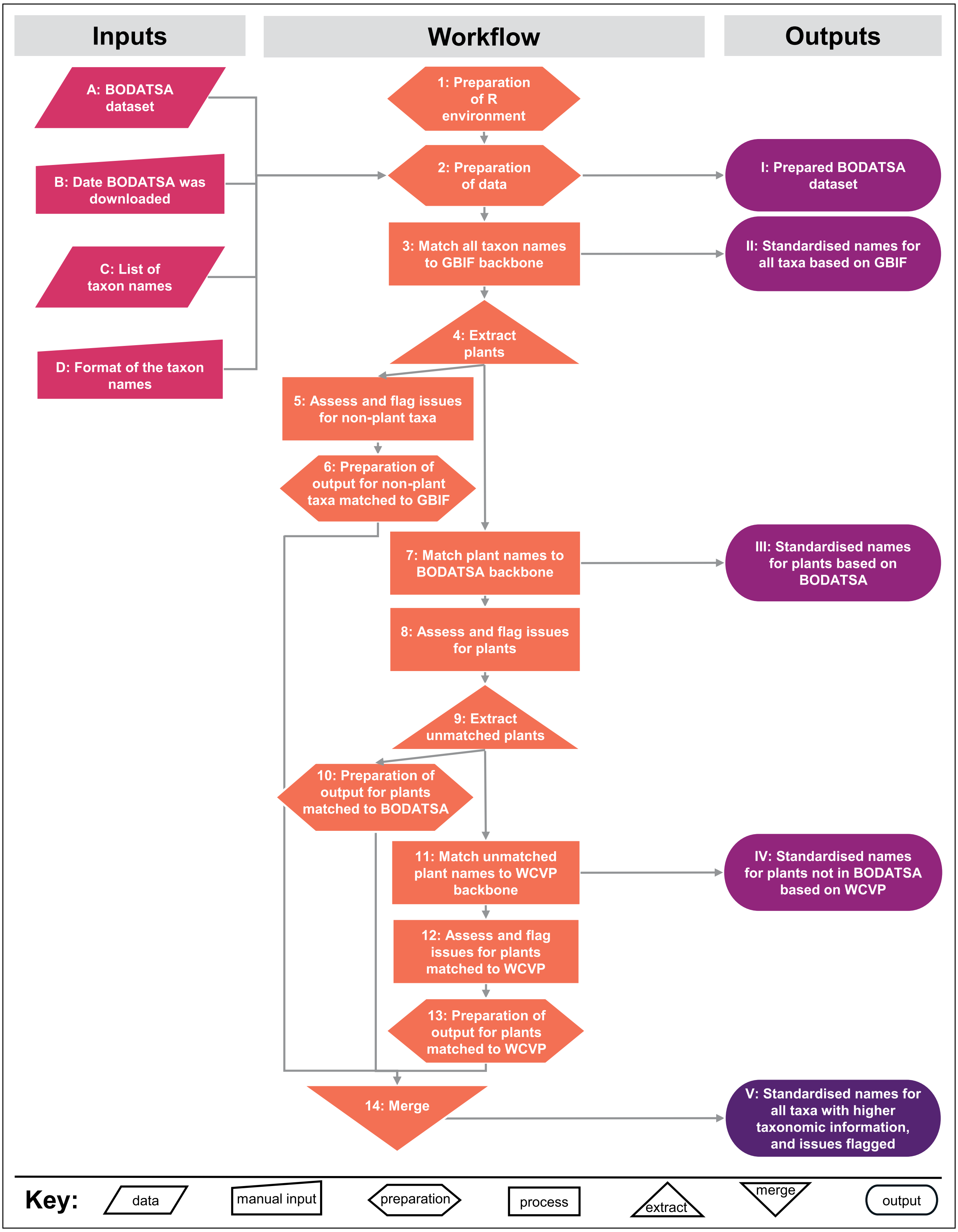
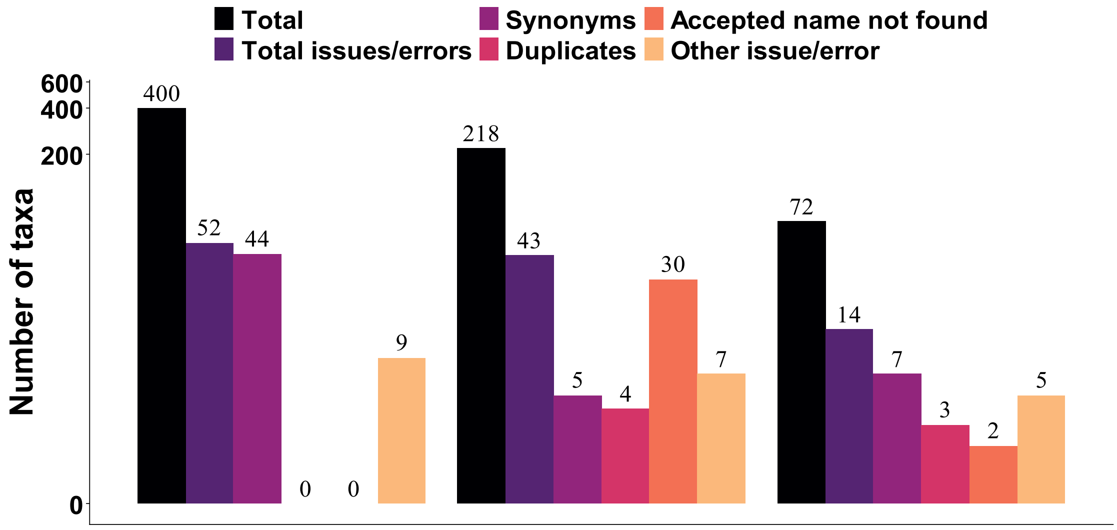


Figure 1. A workflow to standardise taxon names for South African alien species lists using taxonomic backbones relevant to South Africa. During execution, the accepted scientific name with authorship and date information and higher-level taxonomic information are obtained, and issues and errors are detected, corrected and flagged. Inputs are shown in pink (A–D), the workflow in orange (1–14), and the outputs are in purple (I–V), with interim outputs in light purple (I–IV) and the final output in dark purple (V). For a summary of each step see Table 2 and for details see Supplementary Material 1. Note that 'plants' refers to vascular plants, and 'non-plant taxa' refers to all taxa that are not vascular plants.

taxonomically standardised within four minutes. The workflow performed well for all case studies, with accepted scientific names automatically identified for most of the alien taxa (100% of the watchlist, 86.2% of the intercepted taxa and 97% of the traded aquatic macrophytes). Most of the taxa (0–14%) for which an accepted scientific name could not be automatically identified (Figure 2) were mites and viruses on the list of intercepted taxa. Synonyms were detected in all three lists, but were a particular issue for the watchlist, with 11% of the listed taxon names being synonyms (Figure 2). It is not surprising that the taxon names on the watchlist are out-of-date, as this is the oldest of the three lists, and as it was developed based on a data



	Watch list	Intercepted taxa	Traded aquatic macrophytes
Reference	Faulkner et al. (2014)	Saccaggi et al. (2021)	Niemann et al. (2022)
Taxonomic group (number of taxa)	Chromista (5) Fungi (4) Animals (207) Vascular plants (183) Algae (1)	Animals (200) Bacteria (3) Fungi (10) Viruses (5)	Vascular plants (72)
Habitats	Freshwater Marine Terrestrial	Terrestrial	Freshwater
Format of taxon names	Canonical	Canonical	Scientific
Taxonomic backbone	ITIS (early 2000s) with some updates based on IUCN (~2009)	GBIF for most taxa, ICTV for viruses, various for mites (29 July 2020)	GBIF (November 2017)
Taxonomic backbones (this workflow)	GBIF BODATSA WCVP	GBIF	BODATSA WCVP

Figure 2. Details on the lists used as case studies to test and illustrate the workflow. The bar chart shows the total number of taxa included on the three lists, as well as the number of taxa on the lists that were flagged due to the detection of various issues or errors. Note that the y-axis has been $\log_{10}$ transformed. The lower portion of the figure provides details on the taxonomic groups (with the number of taxa shown in brackets) and habitats of the listed taxa; the format of the taxon names (canonical or scientific); and the taxonomic backbones implemented during the development of the lists (with dates) and during the execution of this workflow. The watchlist was compiled using the Global Invasive Species Database (GISD), and so followed the taxonomy used in that database. Much of the content of the GISD was developed in the early 2000s (Pagad et al. 2015), and at that time the taxonomic backbone of the Integrated Taxonomic Information System (ITIS) was implemented, however, in ~2009, some revisions were made based on the International Union for Conservation of Nature (IUCN) (S. Pagad, pers. comm., 21 January 2025). GBIF – Global Biodiversity Information Facility; BODATSA – Botanical Database of Southern Africa; WCVP – World Checklist of Vascular Plants; ICTV – International Committee on Taxonomy of Viruses.

source that was last taxonomically standardised more than 15 years ago (Figure 2).

A manual evaluation of the outputs (see Supplementary Materials 3, 4 and 5) from all case studies indicated that synonyms had been correctly identified as such, and that the correct accepted scientific names were returned. A few of the returned accepted scientific names were duplicated in two of the lists (Figure 2). Two of the intercepted taxa – *Loria formosa* Cooreman, 1958 and *Sancassania oudemansi* (Zachvatkin, 1937) – were listed twice, once under the accepted scientific name and once under a synonym, while the list of traded aquatic macrophytes included three records for *Anubias* species that had not been identified to the species level. Between 2 and 7% of the taxon names were flagged as having other potential issues or errors (Figure 2). Several taxa (four on the watchlist, six on the list of intercepted taxa and one on the list of traded aquatic macrophytes) were flagged as they were matched to accepted scientific names through fuzzy matching. For example, *Didymosphenia geminata* was matched to *Didymosphenia geminatum* (Lyngbye) M.Schmidt, 1899. Five plants on the watchlist were flagged as their names matched to multiple names in the WCVP taxonomic backbone. For example, depending on its authority, *Polygonum cuspidatum* is either a synonym of *Reynoutria japonica* Houtt. or of *Persicaria acuminata* (Kunth) M.Gómez.

Discussion

The workflow presented here was developed to assist with keeping the taxon names in South African alien species lists up-to-date and can be used to map lists of taxon names standardised using other taxonomic backbones onto those that are relevant to the South African context. While the workflow can serve as a step-by-step guide of the required processes, its automation means that these processes can be easily and rapidly executed; and although it has some limitations, it can be adapted to meet the needs of users. Below the workflow's utility, limitations and adaptability are discussed.

Utility

The three alien species lists used as case studies were developed for different purposes, using different methods, and focus on alien taxa from different taxonomic groups and habitats. They, therefore, serve as an indication of the wide range of alien species lists that have been developed for South Africa. They also, however, demonstrate the various taxonomic issues that impact alien species lists – they are not interoperable as they present taxon names in different formats and were developed using different taxonomic backbones; contain a range of taxonomic errors (e.g., synonyms, misspellings, lexical variants); and are out-of-date. Indeed, even the taxon names in the most recently developed list, that of taxa intercepted during agricultural inspections at South African ports of entry (Saccaggi et al. 2021), were out-of-date, despite this list being taxonomically standardised in 2020 using the GBIF taxonomic backbone. The case studies also demonstrated the utility of the workflow – it can be used to reduce error rates; standardise taxon names in various formats and from a wide range of taxonomic groups and habitats in a transparent and repeatable manner; and its automation means that the time spent on taxonomic standardisation can be greatly reduced. The workflow can, therefore, assist with keeping South African alien species lists up-to-date, ensure that they meet international data standards (FAIR data principles [Findable, Accessible, Interoperable and Reuseable; Wilkinson et al. 2016]), and, in so doing, will increase their value (McGeoch et al. 2012; Faulkner et al. 2015). For example, a comprehensive, up-to-date list of alien taxa for the whole country is required for national-level reports on the status of biological invasions, and to monitor and report on progress towards Target 6 of the Kunming Montreal Global Biodiversity Framework (Zengeya & Wilson 2023); but to develop this list, the various lists that exist need to be compared and combined. This workflow can greatly assist with this currently difficult and time-consuming task (Zengeya & Wilson 2023).

Of course, not all alien species lists can be continuously updated to keep up with taxonomic changes (Murray et al. 2017; Garnett et al. 2020; but see Govaerts et al. 2021), but the workflow can also assist with keeping static, periodically published lists transparent; and facilitate tracking of taxonomic changes in these lists (Wilson & Kumschick 2024). For example, while the NEM:BA A&IS Lists are updated periodically (e.g., in 2016 and 2020) (Wilson & Kumschick 2024), the regulations, and therefore the lists, cannot be revised every time there is a taxonomic change. Such lists should therefore be published as versioned, dated, permanently accessible archives (Garnett et al. 2020). The workflow can facilitate this as both when the taxon names were last updated and the version of the backbone implemented are explicit in the final output (Grenié et al. 2022).

This workflow is not the only automated tool that can be used to standardise lists of taxon names according to specific taxonomic backbones, and indeed a range of such tools have been developed (Grenié et al. 2022). For example, GBIF offers a tool to standardise a list of taxon names based on its backbone (https://www.gbif.org/tools/species-lookup); and several R packages linked to different taxonomic backbones have been developed for this purpose (for a review see Grenié et al. 2022). However, the workflow presented here offers some advantages for South African alien species lists.

The workflow standardises alien plant names using BODATSA, which is the most current, scientifically accurate taxonomic backbone for southern African plants (Klopper et al. 2021), and its implementation means that South African alien species lists will be interoperable with other plant lists for the region.

Limitations

Automated tools for taxonomic standardisation can greatly assist with the detection and correction of various issues that are prevalent in lists of taxon names that are time consuming to manually address (Murray et al. 2017); such as synonyms, misspellings, lexical variants, and inconsistencies in authorship and date information. However, these tools, including the one presented here, have a common set of limitations.

Not all of the procedures required to taxonomically standardise lists of taxon names can be automated (Boyle et al. 2013; Murray et al. 2017) and thus some manual work is required after executing such tools. For example, unless taxon names are provided with authorship information, taxonomic homonyms cannot be resolved through automated processes, and so this relatively uncommon issue needs to be addressed manually on a case-by-case basis (Cayuela et al. 2012). In addition, these automated tools are linked to specific taxonomic backbones, and these will not be sufficient in all cases. Examples from the case studies presented here include viruses, for which the GBIF taxonomic backbone is not appropriate (Saccaggi et al. 2021). In such cases, the user will need to manually standardise the taxon names using a relevant and up-to-date taxonomic backbone (for some other suggestions see http://dx.doi.org/10.5281/zenodo.8217211); deviations from the taxonomic backbones used in the workflow will need to be reported; and the taxonomic backbones selected will need to be made explicit (Murray et al. 2017; Seebens et al. 2020). As an example, Saccaggi et al. (2021) implemented the GBIF taxonomic backbone for their list of intercepted taxa, but for viruses the taxonomic backbone of the ICTV was implemented. Nonetheless the case studies demonstrated that the workflow presented here can automatically standardise most taxon names in South African alien species lists, and thus manual work would be required for only a small proportion (9 taxa on the watchlist, 37 of the intercepted taxa, and 7 of the traded aquatic macrophytes).

When compiling, comparing and consolidating alien species lists, there are various non-taxonomic issues to contend with and scope-related, list-specific decisions are required. Automated taxonomic tools cannot address all these issues. For example, alien species lists are often not simple lists of alien taxa but provide a lot of additional taxon-specific data (McGeoch et al. 2012; Faulkner et al. 2015; Murray et al. 2017). These additional data need to be processed appropriately by the user after taxonomic standardisation. To assist the user with this process, the taxon names in the final output of the workflow presented here are in the same order as provided by the user. This ensures that any additional data are compatible with the outputs of the workflow, however, as a consequence duplicated taxon names are not removed, but are rather flagged to assist the user. Alien species lists also differ in how subspecific or supraspecific entities are treated, and, for example, subspecific entities are in some cases synonymised under the species to which they belong, although this can lead to a loss of important information (e.g., the three subspecies of *Trachemys scripta* have varying introduction histories, and only one, *Trachemys scripta* subsp. *elegans* has become widely invasive [Reshetnikov et al. 2023]) (Jansen & Dengler 2010; Murray et al. 2017). As this is a list-specific decision, the workflow does not standardise the taxonomic level at which the taxon names are presented. Similarly, decisions on what to do when taxon names are provided with open nomenclature are likely to be made on a case-by-case basis, thus these taxa are also flagged.

Importantly the workflow errs on the side of caution and flags all potential issues and errors whether or not they have been addressed. This increases transparency and allows for changes to be tracked, but it also means that in most cases the outputs for taxa with flagged issues will only need to be quickly examined to confirm that they have been dealt with appropriately (e.g., misspellings and lexical variants addressed through fuzzy matching). To further assist the user, interim outputs are also written to file, which will allow the sources of flagged issues to be identified.

The workflow presented here only speaks to taxonomy, but it is worth noting that guidance on data and processes for alien species list compilation and consolidation are available (Faulkner et al. 2015; Murray et al. 2017), that global standards for alien species data have been developed (Groom et al. 2019), and that an automated workflow exists to consolidate alien species lists (Seebens et al. 2020).

Adaptability

Software development standards (Huybrechts et al. 2024) and guidance on how to develop meaningful workflows for taxonomic standardisation (Grenié et al. 2022) were followed when developing this workflow; and the workflow is based on open-source tools, and itself follows open-source principles. As it stands, the workflow can be implemented for other southern African countries included in BODATSA; and to meet the needs of other countries, or to be interoperable with international alien species lists (Seebens et al. 2020), all taxon names standardised based on the GBIF backbone are available in an interim output of the workflow. However, it could be useful in some cases to alter the workflow, for example, so that all vascular plant names are standardised based on the WCVP taxonomic backbone (i.e., the steps that involve BODATSA [steps 7–10] could be omitted); or so that multiple lists can be iteratively standardised (Fernández Winzer et al. 2025). As the script for the workflow and detailed guidance documents are freely available, users can easily access and adapt the script to suit their purpose.

Conclusion

The automated, open workflow presented here can be used to taxonomically standardise alien taxon names in various formats, and from a wide range of taxonomic groups and habitats. The workflow was specifically developed to assist with keeping the taxon names in South African alien species lists up-to-date and can be used to map lists of taxon names standardised using other taxonomic backbones onto those that are relevant to the South African context. The workflow does not deal with all taxonomic issues, but these issues are flagged; it will reduce time spent on taxonomic standardisation; and it should greatly assist the researchers, managers and decision-makers that compile and use alien species lists. Importantly, the workflow is transparent and repeatable, will help to ensure that South African alien species lists meet the FAIR data principles (Wilkinson et al. 2016) and, in so doing, will improve the value of the lists for research, management and policy (McGeoch et al. 2012; McGeoch & Jetz 2019).

Data and code availability

The full workflow description and script, as well as detailed guidance documents are freely available at a dedicated GitHub repository (https://github.com/KatelynFaulkner/rsa-ans-workflow). The full description and other guidance documents are also provided as supplementary materials with this publication.

Acknowledgements

This automated workflow is based on open, online sources of taxonomic information. Thank you to the active and enthusiastic community of taxonomists that ensures that these sources are maintained. The South African Department of Forestry, Fisheries and the Environment (DFFE) are thanked for funding, noting that this publication does not necessarily represent the views or opinions of DFFE or its employees. K.T.F also received funding from the B-cubed project (Biodiversity Building Blocks for policy), which is funded by the European Union's Horizon Europe Research and Innovation Programme (ID No. 101059592). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them. Thank you to Tsungai Zengeya, John Wilson, Laura Fernández Winzer and Emily McCulloch-Jones for comments on an early draft of the workflow; to John Wilson, Promise Mtileni and two anonymous reviewers for comments on the manuscript; to Shyama Pagad for information on the GISD; and to Ronell Klopper for information on BODATSA.

Competing interests

The author declares that they have no financial or personal relationship(s) that may have inappropriately influenced them in writing this article.

Authors' contributions

K.T.F. conceptualised the workflow, developed the script and supporting documentation, set up the GitHub repository, applied and tested the workflow, and wrote the manuscript.

References

Baard, J.A. & Kraaij, T., 2019, 'Use of a rapid roadside survey to detect potentially invasive plant species along the garden route, South Africa', *Koedoe* 61(1), 1–10, https://doi.org/10.4102/koedoe.v61i1.1515.

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J.A., Mozzherin, D., Rees, T., Matasci, N., Narro, M.L., Piel, W.H., Mckay, S.J., Lowry, S., Freeland, C., Peet, R.K. & Enquist, B.J., 2013, 'The taxonomic name resolution service: an online tool for automated standardization of plant names', *BMC Bioinformatics* 14, 16, https://doi.org/10.1186/1471-2105-14-16.

Brown, M.J.M., Walker, B.E., Black, N., Govaerts, R.H.A., Ondo, I., Turner, R. & Nic Lughadha, E., 2023, 'rWCVP: a companion R package for the World Checklist of Vascular

Plants', *New Phytologist* 240(4), 1355–1365, https://doi.org/10.1111/nph.18919.

Canavan, S., Richardson, D.M., Le Roux, J.J., Kelchner, S.A. & Wilson, J.R.U., 2021, 'The status of alien bamboos in South Africa', *South African Journal of Botany* 138, 33–40, https://doi.org/10.1016/j.sajb.2020.11.027.

Cayuela, L., Granzow-de la Cerda, Í., Albuquerque, F.S. & Golicher, D.J., 2012, 'Taxonstand: an r package for species names standardisation in vegetation databases', *Methods in Ecology and Evolution* 3(6), 1078–1083, https://doi.org/10.1111/j.2041-210X.2012.00232.x.

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L. & Ram, K., 2025, *rgbif: interface to the Global Biodiversity Information Facility API*, *R package version 3.8.0*, computer software, viewed 25 March 2025 from https://cran.r-project.org/package=rgbif.

Convention on Biological Diversity, 2022, *Kunming-Montreal Global Biodiversity Framework: draft decision submitted by the President*, Secretariat of the Convention on Biological Diversity, Montreal.

Faulkner, K.T., Robertson, M.P., Rouget, M. & Wilson, J.R.U., 2014, 'A simple, rapid methodology for developing invasive species watchlists', *Biological Conservation* 179, 25–32, https://doi.org/10.1016/j.biocon.2014.08.014.

Faulkner, K.T., Spear, D., Robertson, M.P., Rouget, M. & Wilson, J.R.U., 2015, 'An assessment of the information content of South African alien species databases', *Bothalia: African Biodiversity and Conservation* 45(1), a1103, https://doi.org/10.4102/abc.v45i1.1103.

Fernández Winzer, L., Faulkner, K.T., Paap, T. & Wilson, J.R.U., 2025, 'From detection to action – a proposed workflow to ensure first reports of alien species from molecular analyses are acted upon', *NeoBiota* 104, 339–359, https://doi.org/10.3897/neobiota.104.162310.

Garnett, S.T., Christidis, L., Conix, S., Costello, M.J., Zachos, F.E., Bánki, O.S., Bao, Y., Barik, S.K., Buckeridge, J.S., Hobern, D., Lien, A., Montgomery, N., Nikolaeva, S., Pyle, R.L., Thomson, S.A., Van Dijk, P.P., Whalen, A., Zhang, Z.Q. & Thiele, K.R., 2020, 'Principles for creating a single authoritative list of the world's species', *PLoS Biology* 18(7), e3000736, https://doi.org/10.1371/journal.pbio.3000736.

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A., 2021, 'The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity', *Scientific Data* 8(1), 215, https://doi.org/10.1038/s41597-021-00997-6.

Grenié, M., Berti, E., Carvajal-Quintero, J., Dädlow, G.M.L., Sagouis, A. & Winter, M., 2022, 'Harmonizing taxon names in biodiversity data: a review of tools, databases and best practices', *Methods in Ecology and Evolution* 14(1), 12–25, https://doi.org/10.1111/2041-210X.13802.

Groom, Q., Desmet, P., Reyserhove, L., Adriaens, T., Oldoni, D., Vanderhoeven, S., Baskauf, S.J., Chapman, A., McGeoch, M., Walls, R., Wieczorek, J., Wilson, J.R.U., Zermoglio, P.F.F. & Simpson, A., 2019, 'Improving Darwin Core for research and management of alien species', *Biodiversity Information Science and Standards* 3, e38084, https://doi.org/10.3897/biss.3.38084.

Herbert, D.G., 2010, *The introduced terrestrial mollusca of South Africa*, South African National Biodiversity Institute, Pretoria.

Huybrechts, P., Trekels, M., Abraham, L. & Desmet, P., 2024, 'B-Cubed software development guide', viewed 21 May 2024, from https://docs.b-cubed.eu/dev-guide/.

Jansen, F. & Dengler, J., 2010, 'Plant names in vegetation databases - a neglected source of bias', *Journal of Vegetation Science* 21(6), 1179–1186, https://doi.org/10.1111/j.1654-1103.2010.01209.x.

Keet, J.-H., Arunava, D., Foxcroft, L.C., Kumschick, S., Nichols, G.R., Richardson, D.M. & Wilson, J.R.U., 2022, 'Assessing the level of compliance with alien plant regulations in a large African protected area', *Biological Invasions* 24, 3831–3844, https://doi.org/10.1007/s10530-022-02883-7.

Klopper, R., Winter, P. & Le Roux, M., 2021, 'The South African National Plant Checklist: maintaining the taxonomic backbone for a megadiverse country', *Biodiversity Information Science and Standards* 5, e73899, https://doi.org/10.3897/biss.5.73899.

McCulloch-Jones, E.J., Cuthbert, R.N., Van Wilgen, B.W. & Wilson, J.R.U., 2024, 'Estimating the monetary cost of biological invasions to South Africa', *Biological Invasions* 26(9), 3191–3203, https://doi.org/10.1007/s10530-024-03369-4.

McGeoch, M. & Jetz, W., 2019, 'Measure and reduce the harm caused by biological invasions', *One Earth* 1(2), 171–174, https://doi.org/10.1016/j.oneear.2019.10.003.

McGeoch, M.A., Spear, D., Kleynhans, E.J. & Marais, E., 2012, 'Uncertainty in invasive alien species listing', *Ecological Applications* 22(3), 959–971, https://doi.org/10.1890/11-1252.1.

Mead, A., Carlton, J.T., Griffiths, C.L. & Rius, M., 2011, 'Introduced and cryptogenic marine and estuarine species of South Africa', *Journal of Natural History* 45, 2463–2524, https://doi.org/10.1080/00222933.2011.595836.

Murray, B.R., Martin, L.J., Phillips, M.L. & Pyšek, P., 2017, 'Taxonomic perils and pitfalls of dataset assembly in ecology: a case study of the naturalized Asteraceae in Australia', *NeoBiota* 34, 1–20, https://doi.org/10.3897/neobiota.34.11139.

Nelufule, T., Robertson, M.P., Wilson, J.R.U., Faulkner, K.T., Sole, C. & Kumschick, S., 2020, 'The threats posed by the pet trade in alien terrestrial invertebrates in South Africa', *Journal for Nature Conservation* 55, 125831, https://doi.org/10.1016/j.jnc.2020.125831.

Niemann, H.J., Bezeng, B.S., Orton, R.D., Kabongo, R.M., Pilusa, M. & Van der Bank, M., 2022, 'Using a DNA barcoding approach to facilitate biosecurity: identifying invasive alien macrophytes traded within the South African aquarium and pond plant industry', *South African Journal of Botany* 144, 364–376, https://doi.org/10.1016/j.sajb.2021.08.041.

Pagad, S., Genovesi, P., Carnevali, L., Scalera, R. & Clout, M., 2015, 'IUCN SSC invasive species specialist group: invasive alien species information management supporting practitioners, policy makers and decision takers', *Management of Biological Invasions* 6(2), 127–135, https://doi.org/10.3391/mbi.2015.6.2.03.

Reshetnikov, A.N., Zibrova, M.G., Ayaz, D., Bhattarai, S., Borodin, O.V., Borzée, A., Brejcha, J., Çiçek, K., Dimaki, M., Doronin, I.V., Drobenkov, S.M., Gichikhanova, U.A., Gladkova, A.Y., Gordeev, D.A., Ioannidis, Y., Ilyukh, M.P., Interesova, E.A., Jadhav, T.D., Karabanov, D.P., Khabibullin, V.F., Khabilov, T.K., Khan, M.M.H., Kidov, A.A., Klimov, A.S., Kochetkov, D.N., Kolbintsev, V.G., Kuzmin, S.L., Lotiev, K.Y., Louppova, N.E., Lvov, V.D., Lyapkov, S.M., Martynenko, I.M., Maslova, I.V., Masroor, R., Mazanaeva, L.F., Milko, D.A., Milto, K.D., Mozaffari, O., Nguyen, T.Q., Novitsky, R.V., Petrovskiy, A.B., Prelovskiy, V.A., Serbin, V.V., Shi, H.-t., Skalon, N.V., Struijk, R.P.J.H., Taniguchi, M., Tarkhnishvili, D., Tsurkan, V.F., Tyutenkov, O.Y., Ushakov, M.V., Vekhov, D.A., Xiao, F., Yakimov, A.V., Yakovleva, T.I., Yang, P., Zeleev, D.F. & Petrosyan, V.G., 2023, 'Rarely naturalized, but widespread and even invasive: the paradox of a popular pet terrapin expansion in Eurasia', *NeoBiota* 81, 91–127, https://doi.org/10.3897/neobiota.81.90473.

Richardson, D.M. & Potgieter, L.J., 2024, 'A living inventory of planted trees in South Africa derived from iNaturalist', *South African Journal of Botany* 173, 365–379, https://doi.org/10.1016/j.sajb.2024.08.006.

Robinson, T.B., Alexander, M.E., Simon, C.A., Griffiths, C.L., Peters, K., Sibanda, S., Miza, S., Groenewald, B., Majiedt, P. & Sink, K.J., 2016, 'Lost in translation? Standardising the terminology used in marine invasion biology and updating South African alien species lists', *African Journal of Marine Science* 38(1), 129–140, https://doi.org/10.2989/1814232X.2016.1163292.

Robinson, T.B., Peters, K. & Brooker, B., 2020, 'Coastal invasions: the South African context', in B.W. van Wilgen, J. Measey, D.M. Richardson, J.R. Wilson & T.A. Zengeya (eds), *Biological invasions in South Africa*, pp. 229–247, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-030-32394-3_9.

Saccaggi, D.L., Arendse, M., Wilson, J.R.U. & Terblanche, J.S., 2021, 'Contaminant organisms recorded on plant product imports to South Africa 1994-2019', *Scientific Data* 8, 83, https://doi.org/10.1038/s41597-021-00869-z.

Seebens, H., Clarke, D.A., Groom, Q., Wilson, J.R.U., García-Berthou, E., Kühn, I., Roigé, M., Pagad, S., Essl, F., Vicente, J., Winter, M. & McGeoch, M., 2020, 'A workflow for standardising and integrating alien species distribution data', *NeoBiota* 59, 39–59, https://doi.org/10.3897/neobiota.59.53578.

Shivambu, T.C., Shivambu, N. & Downs, C.T., 2022, 'An assessment of avian species sold in the South African pet trade', *African Journal of Ecology* 60(4), 980–995, https://doi.org/10.1111/aje.13029.

Soberón, J. & Peterson, A.T., 2004, 'Biodiversity informatics: managing and applying primary biodiversity data', *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, 689–698, https://doi.org/10.1098/rstb.2003.1439.

Swart, C. & Robinson, T.B., 2019, 'Horizon scanning for alien predatory crabs: insights from South Africa', *African Journal of Marine Science* 41(2), 125–135, https://doi.org/10.2989/1814232X.2019.1588782.

Van Wilgen, B.W. & Wilson, J.R., 2018, *The status of biological invasions and their management in South Africa in 2017*, South African National Biodiversity Institute, Kirstenbosch and DST-NRF Centre of Excellence for Invasion Biology, Stellenbosch.

Van Wilgen, B.W., Zengeya, T.A. & Richardson, D.M., 2022, 'A review of the impacts of biological invasions in South Africa', *Biological Invasions* 24(1), 27–50, https://doi.org/10.1007/s10530-021-02623-3.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., Van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., Van der Lei, J., Van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B., 2016, 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* 3, 1–9, https://doi.org/10.1038/sdata.2016.18.

Williams, V.L., Wojtasik, E.M. & Byrne, M.J., 2021, 'A chronicle of alien medicinal plants used as traditional medicine in South Africa, and their status as invasive species', *South African Journal of Botany* 142, 63–72, https://doi.org/10.1016/j.sajb.2021.05.027.

Wilson, J.R.U. & Kumschick, S., 2024, 'The regulation of alien species in South Africa', *South African Journal of Science* 120(5/6), 17002, https://doi.org/10.17159/sajs.2024/17002.

Zachariades, C., 2021, 'A catalogue of natural enemies of invasive alien plants in South Africa: classical biological control agents considered, released and established, exotic natural enemies present in the field, and bioherbicides', *African Entomology* 29(3), 1077–1142, https://doi.org/10.4001/003.029.1077.

Zengeya, T.A. & Wilson, J.R., 2020, *The status of biological invasions and their management in South Africa in 2019*, South African National Biodiversity Institute, Kirstenbosch and DSI-NRF Centre of Excellence for Invasion Biology, Stellenbosch, http://dx.doi.org/10.5281/zenodo.3947817.

Zengeya, T.A. & Wilson, J.R., 2023, *The status of biological invasions and their management in South Africa in 2022*, South African National Biodiversity Institute, Kirstenbosch and DSI-NRF Centre of Excellence for Invasion Biology, Stellenbosch, https://dx.doi.org/10.5281/zenodo.8217182.

Supplementary material

Available online: http://dx.doi.org/10.38201/abc.v56.2.a2

Supplementary Material 1:	Description: an automated workflow to standardise taxon names for South African alien species lists.
Supplementary Material 2:	Description: outputs of an automated workflow to standardise taxon names for South African alien species lists.
Supplementary Material 3:	Final output for watchlist.
Supplementary Material 4:	Final output for list of taxa intercepted during agricultural inspections.
Supplementary Material 5:	Final output for list of traded aquatic macrophytes.