Methods

The big four of plant taxonomy – a comparison of global checklists of vascular plant names

David Schellenberger Costa[1,2], Gerhard Boehnisch[3], Martin Freiberg[1], Rafaël Govaerts[4], Matthias Grenié[2], Michael Hassler[5], Jens Kattge[3], Alexandra N. Muellner-Riehl[2,6], Blanca M. Rojas Andrés[7,8], Marten Winter[2], Mark Watson[9], Alexander Zizka[2,10] and Christian Wirth[1,2]

[1]Department of Special Botany and Functional Biodiversity, Faculty of Life Sciences, University of Leipzig, Johannisallee 21-23, 04103 Leipzig, Germany; [2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr 4, 04103 Leipzig, Germany; [3]Research Group Functional Biogeography, Max Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany; [4]Jodrell Laboratory, Royal Botanic Gardens, Kew, Kew Road, Richmond, TW9 3DS, UK; [5]Weiherbergstr 77, 76646 Bruchsal, Germany; [6]Department of Molecular Evolution and Plant Systematics & Herbarium (LZ), Faculty of Life Sciences, University of Leipzig, Johannisallee 21-23, 04103 Leipzig, Germany; [7]Departamento de Botánica y Fisiología Vegetal, Universidad de Salamanca, Ave Licenciado Méndez Nieto s/n, 37007 Salamanca, Spain; [8]Biobanco de ADN Vegetal, Edificio Multiusos I+D+i, Universidad de Salamanca, Calle Espejo s/n, 37007 Salamanca, Spain; [9]Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh, EH3 5LR, UK; [10]Department of Biology, Philipps-University Marburg, 35043 Marburg, Germany

Author for correspondence:
David Schellenberger Costa
Email: david.schellenberger.costa@uni-leipzig.de

Received: *14 October 2022*
Accepted: *12 April 2023*

New Phytologist (2023) **240:** 1687–1702
doi: 10.1111/nph.18961

Key words: checklist, Leipzig Catalogue of Vascular Plants (LCVP), plant species names, taxonomic name resolution, The Plant List (TPL), World Checklist of Vascular Plants (WCVP), World Flora Online (WFO), WorldPlants (WP).

Summary

- Taxonomic checklists used to verify published plant names and identify synonyms are a cornerstone of biological research. Four global authoritative checklists for vascular plants exist: Leipzig Catalogue of Vascular Plants, World Checklist of Vascular Plants, World Flora Online (successor of The Plant List, TPL), and WorldPlants. We compared these four checklists in terms of size and differences across taxa.
- We matched taxon names of these checklists and TPL against each other, identified differences across checklists, and evaluated the consistency of accepted names linked to individual taxon names. We assessed geographic and phylogenetic patterns of variance.
- All checklists differed strongly compared with TPL and provided identical information on *c.* 60% of plant names. Geographically, differences in checklists increased from low to high latitudes. Phylogenetically, we detected strong variability across families. A comparison of name-matching performance on taxon names submitted to the functional trait database TRY, and a check of completeness of accepted names evaluated against an independent, expert-curated checklist of the family Meliaceae, showed a similar performance across checklists.
- This study raises awareness on the differences in data and approach across these checklists potentially impacting analyses. We propose ideas on the way forward exploring synergies and harmonizing the four global checklists.

Introduction

With the ongoing global crises of biodiversity loss and climate change, botanical and ecological research are of critical importance. Together with technical innovations and the aim to increase global collaboration, more and larger datasets on biodiversity are now available than ever before, including global datasets on plant traits and vegetation plots (e.g. Kattge *et al.*, 2011; Telenius, 2011; Enquist *et al.*, 2016; Bruelheide *et al.*, 2019; Weigelt *et al.*, 2020). To create these global databases and make use of the wealth of information across them, common identifiers are necessary. While environmental data can be integrated by location and time, ecological data are usually indexed by scientific names. When joining different datasets, it is of critical importance to link information belonging to the same taxonomic units. This can be challenging due to differences in synonyms, variation in the way names are cited, spelling mistakes, and differences in the taxonomic frameworks and taxon concepts used. The fast progress in the field of taxonomy with high numbers of molecular-based phylogenies published every year means these

This article is part of the Special Collection 'Global plant diversity and distribution'. See https://www.newphytologist.org/global-plant-diversity for more details.

problems are likely to persist. Therefore, before joining datasets, scientific names need to be desynonymized, standardized, and corrected for spelling mistakes (Grenié *et al.*, 2023). To do so, an authoritative reference is essential.

Traditionally, botanical nomenclatural databases have been compiled in botanical gardens, museums, and universities. Before the digital era, probably the largest collections of nomenclatural data were located at the Royal Botanical Gardens, Kew: Index Kewensis (Hooker, 1893) for seed plants and Index Filicum (Moore, 1857) for ferns and allies. These were among the nuclei from which the International Plant Names Index (IPNI, https://www.ipni.org/) and other global plant names checklists emerged, the latter adding taxonomic information, that is, the classification into genera, families, and higher taxonomic units, to the nomenclatural data. In 2010, The Plant List (TPL) was created as a response of the botanical community to Target 1 of the 2002–2010 Global Strategy for Plant Conservation (GPSC): to produce 'a working list of known plant species, as a step towards a complete world flora' (The Royal Botanic Gardens *et al.*, 2013). It was not the first resource on global plant names: Catalogue of Life (CoL) started in 2001 already (Cachuela-Palacio, 2006). However, CoL received its data from individual curators of Global Species Databases, who made taxonomic decisions on their own. The vision of TPL was to create a resource built by the taxonomic community in a consensual way. The Plant List was and continues to be used widely by the scientific community, with > 1900 occurrences in Google Scholar in 2022 alone. However, while botanical science has moved on, TPL remained static since 2013. Four other global checklists now exist, some led by dedicated individuals, and one by an international consortium (Fig. 1).

These four checklists are (in alphabetical order throughout the article): (1) the Leipzig Catalogue of Vascular Plants (LCVP), developed by Martin Freiberg, the curator of the Leipzig Botanical Garden, Germany, initially in an effort to build the basis for his LifeGate project, a biodiversity informatics portal based on a 2D-representation of the tree of life (https://lifegate.idiv.de). Since 2020, the vascular plants subset of the data can be downloaded as LCVP (Freiberg *et al.*, 2020). (2) The World Checklist of Vascular Plants (WCVP) can be traced back to when Rafaël Govaerts started to work on the World Checklist of Seed Plants in 1988, and from 1994 at the Royal Botanic Gardens, Kew.

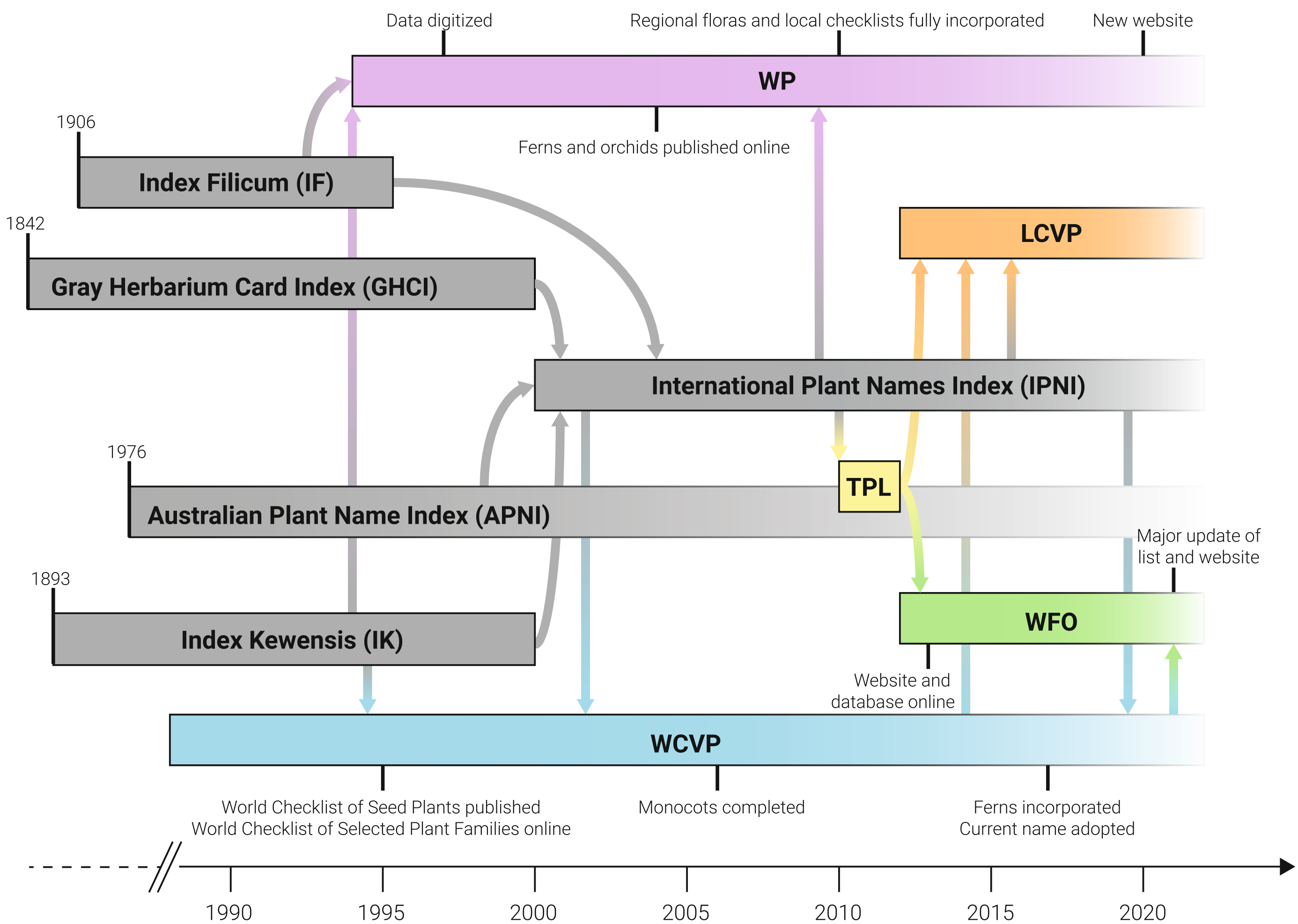

Fig. 1 History and time line of the development of the four global plant name checklists. Nomenclatural databases providing plant names only are filled in grey, while taxonomic databases informing about the relationships between taxa are filled in colours. Arrows indicate directions and approximate times of data flows between databases and checklists. TPL, The Plant List; LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online; WP, WorldPlants.

This dataset was published online in 1995 as the World Checklist of Selected Plant Families (WCSP). With increasing coverage of all plant families, including pteridophytes, WCSP became WCVP in 2016 (Govaerts *et al*., 2021). The data from WCVP are used to feed Plants Of The World Online (POWO, https://powo.science.kew.org), a website designed to retrieve information on individual plant species and their distributions (The Royal Botanic Gardens, Kew, 2022). (3) World Flora Online (WFO) can be seen as the successor to TPL, as TPL itself was imperfect due to a lack of time for revising many taxa and built with an architecture that made it unfeasible for further development. To continue the community-driven approach, WFO was founded in 2012 in response to the 2011–2020 GPSC with its revised Target 1: to produce 'an online flora of all known plants'. This included the creation of a sustainable global plant checklist as a taxonomic backbone. It is curated by a series of appointed Taxonomic Expert Networks (TENs) charged to provide a global consensus for their taxonomic groups (Borsch *et al*., 2020). It currently involves over 300 taxonomists from *c.* 50 botanical institutions world-wide. The WFO portal, including data downloads, was launched in 2019, and the first version of the WFO plant list, replacing TPL, was released in May 2021 (https://wfoplantlist.org/). (4) WorldPlants (WP) was started in the early 1990s and is developed by Michael Hassler, who based his dataset on Index Kewensis and Index Filicum. These datasets were critically cross-checked against as many floras, country checklists, and revisions as possible to create an authoritative, synonymic checklist including complete distribution data. Due to an initial lack of a suitable web resource, only parts were published online in 2002 (WorldFerns and WorldOrchids). In 2013, 60% of the dataset was included in The Catalogue of Life (https://www.catalogueoflife.org, Bánki *et al*., 2022) and since 2019, the whole WorldPlants dataset is accessible and fully searchable on its own website (https://www.worldplants.de, Hassler, 2022).

It is striking, how one of the foundations of data-driven plant ecology and biodiversity science – taxonomic name resolution – has largely relied on the efforts of a few dedicated individuals. Furthermore, there is no comprehensive information on potential differences, and therefore, on the consequences of choosing one list over the other for downstream data analyses, development of related tools or other databases (Grenié *et al*., 2023). Additionally, the question remains, if and how the work invested in these four different resources may be brought together, acknowledging the efforts of every individual involved, creating synergy in terms of cross-checking, replacing outdated or missing information, and sustaining the legacy of these checklists into the future.

As a first step toward this goal, our paper for the first time united authors and curators of the four global checklists of vascular plants. We compared general characteristics, the changes made since the outdated, but still widely used TPL, and assessed the degree to which these checklists provide consistent information, both in geographic and in phylogenetic contexts. We evaluated two use-cases that test two important properties of the checklists: their completeness, that is, their ability to match names from different sources; and their taxonomic resolution capability, that is, their ability to identify synonyms and keep taxonomic units apart. For the first test, representing one of the most common tasks for users of these global checklists, we used the taxon names submitted to the v.6 of TRY, the largest community-assembled global database for plant functional traits (Kattge *et al*., 2020). For the second test, we made use of an up-to-date checklist of the Meliaceae, independently curated by taxonomic experts of this family, and compared the accepted names of this list with those found in the global checklists.

We expected to find turnover in the taxon names of the four recent checklists compared with TPL, due to the imperfect nature TPL was composed, and the taxonomic advancements of the past decade. As current work on the individual checklists is largely independent from each other, the way taxonomic revisions and new names are incorporated may vary. Therefore, we assumed idiosyncrasy across the lists. This should manifest itself geographically, because research activity and plant diversity are highly variable across the globe, and phylogenetically, because some groups have received greater taxonomic attention and their taxonomies are more disputed than others. For the use cases, we expected the first to reflect similarity in taxon names across checklists, while the second is likely related to their update frequency.

In the short term, this study informs users about major differences and commonalities across checklists. In the long term, it is a direct contribution to the improvement and harmonization of these resources, highlighting problems and inconsistencies across checklists that need to be addressed, and preparing the ground for a collaborative framework uniting global checklists to the benefit of plant diversity-based research.

Materials and Methods

Acquisition and preprocessing of checklists

The following versions of the global checklists were used: LCVP v.1.05, released May 2022, downloaded from https://github.com/idiv-biodiversity/LCVP; WCVP February 2022 release, used by all authors of this special issue, provided directly to us; WFO v.2021.12, released December 2021, downloaded from http://www.worldfloraonline.org/downloadData; WP v.13.0, released June 2022, downloaded genus by genus from https://www.worldplants.de, using the R programming language (R Core Team, 2021) together with the R package RSELENIUM (v.1.7.7, Harrison, 2020) and converted into a tabular text format. To assess changes compared with TPL, v.1.1, released September 2013, we downloaded each genus from http://theplantlist.org and converted data into tabular text format. As the taxonomic scopes of WFO and TPL are wider than those of the other checklists, including mosses in addition to vascular plants, we removed mosses from WFO and TPL after download to make it comparable to the other checklists.

While the information given differs across checklists, their core structure is the same: They contain a set of taxon names including authors and their taxonomic status, which falls into the main categories 'accepted', 'synonym' (several sub-categories in WCVP) and 'other', the latter being used by us as an umbrella term for 'unresolved' (in LCVP), 'unplaced' (in WCVP), 'doubtful' or 'unchecked' (in WFO). WP does not list taxon names with a taxonomic status other than 'accepted' or 'synonym', the exception being a separate list of *c.* 1500 fern names. The 'other' category may contain names that – according to the International Code of Nomenclature – were not validly published and so do not even qualify as being called names, they are 'designations' (Turland, 2019). For the ease of writing, throughout the article, we refer to 'taxon names' as all entries irrespective of their taxonomic or nomenclatural status. For taxon names that fall into the 'synonym' category, the checklists give the accepted names or IDs of the accepted names. For the 'accepted' category, the taxon name and the accepted name are the same. For the 'other' category, no accepted name is known, but the databases sometimes return the taxon name instead, which could lead to a confusion with actually accepted names. For this reason, the analyses that involve accepted names exclusively use only those taxon names that have the taxonomic status 'accepted'.

To enable us to match taxon names against each other, we performed a set of column-joining and column-splitting operations on the checklists. This was necessary because the checklists differ in whether taxon and accepted names are given as one name string or separated by genus and epithet(s), and whether authorship citations are included in taxon names or not. In this process, we standardized rank abbreviations (i.e. subsp., var., and f.). During this step and all subsequent steps, we identified a multitude of minor and major errors within the checklists. We fixed errors that needed to be fixed for name matching, and reported back to the respective curators. The most obvious errors are listed in Table 1. They include missing author names, or infraspecies for which the species they belong to is missing. In any of the lists, errors were found in < 1% of the taxon names.

We created accepted names columns in the cases where accepted names were not directly given, but referred to by IDs. This was done to speed up computation and be able to process all checklists in a uniform way. We ended up with the following four columns: taxon name including, potentially, infraspecies name without author(s); author(s) of the taxon name; and the same two columns, but for the linked accepted names.

In the next step, we replaced special characters including all but standard whitespaces, Cyrillic letters, German sharp s, and standardized Greek letters used for infraspecies designation (replaced by their written English form, e.g. 'α' to 'alpha'). We also removed hybrid signs of various kinds that were not always consistently used within and across lists. This does not change the results of subsequent analyses, as a hybrid taxon must not only differ by the hybrid sign from another taxon. Hybrid signs included the multiplication sign, the plus sign, the letter x both lower- and uppercase, and an appended '_x'. We also removed diacritics (e.g. é → e, ç → c), leading, trailing, and repetitions of whitespace characters and hyphens ('-'), as taxon names must not differ by a hyphen only, and orthographic variants with and without hyphen often exist. Finally, we removed duplicate taxon name–author(s) combinations from the dataset. While often no duplicates in the strict sense, as an author may have described a taxon several times (e.g. *Acalypha mapirensis* Pax can be found twice in WCVP, with two different places and dates of publication), these entries slow down and complicate the matching process and are hardly of interest to users of the lists.

Table 1 Major problems found in the four global plant names checklists.

	LCVP	WCVP	WFO	WP
Missing taxon authors in accepted names (excluding autonyms)	183	227	77	36
Missing taxon authors in synonyms (excluding autonyms)	819	348	106	214
Missing species name for accepted infraspecies	92	14	6888	5
Missing species name for synonym infraspecies	781	361	1738	2082
Missing species name for other infraspecies	36	10	15	11
Missing accepted names in taxon names	910	6	77	7
Duplicate name–author combinations	372	1695	3267	1862

For the entry 'Missing accepted names in taxon names', depending on the list, a taxon name or ID is generally given as an accepted name for each entry. In some cases, the taxon name or ID given may not exist as an individual row in the dataset, meaning that in the case where an ID is given, the accepted name cannot be retrieved, and in the case of an accepted name given this particular name cannot be found in the taxon names column. LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online; WP, WorldPlants.

Checklist matching

Comparison of the taxon names across the lists requires some degree of fuzzy matching, that is, matching allowing for some differences in characters, as both genera and epithets as well as authors may be written differently. Taxon names and authors were treated differently, because authors show much higher variability in terms of whitespaces used and forenames mentioned, abbreviated, or omitted. For taxon names, the algorithm to calculate differences was designed to include three measures returning penalties for differences: the Levenshtein distance, the number of nonmutual characters, and the number of nonmutual two-character substrings. Additionally, differences in the first characters of words were weighted double compared with others. For author names, some preprocessing was undertaken, inspired by the TAXAMATCH algorithm (Rees, 2014). In short, author names were normalized, generic words removed (e.g. *sensu*, hort., *et al.*), and author names split into individual components representing single authors. Then, the overlap of single authors between pairs of author names was evaluated, allowing for a small degree of fuzzy matching. Overlap was calculated symmetrically as a relative measure ranging from zero to two, resulting in a total of zero when there was no overlap in single authors between author names pairs and two if all individual authors were reciprocally found. The performance of the matching algorithm was tested on *c.* 2000 author names pairs from our dataset with manually assigned true/false values for matches and nonmatches. See Supporting Information Notes S1 for details, including a comparison to the author matching function of the TAXAMATCH algorithm, and rates of matching success, false negatives, and false positives depending on the threshold used.

The actual matching of taxon names between checklists was done iteratively and consisted of the following steps: First, select a list to compare against, then, find all matches with other lists, then, store matched and unique taxa, and finally, remove all names matched from the other lists. Then, select the next list and compare against all but the one list that was already compared against. This process continued until the last list, which was already reduced to its unique names because of the removal of matches with the others. R scripts of the difference calculations and the name-matching can be found in Notes S2 and S3, respectively.

Taxonomic overlap, and geographic and phylogenetic patterns

The superlist constructed from all lists had as many rows as there are distinct taxon names across the lists, and several taxon and accepted names columns, one of each for every of the recent four global checklists and for TPL. These columns allowed for the comparisons of taxon names and accepted names which were fundamental to our analyses.

If a taxon name was unique to one checklist, the respective row of the superlist must have empty cells for all taxon name columns but one. On the contrary, if several taxon name cells are filled, it occurs in several checklists. We did not have to check the authors, because this had already been done by the matching algorithm: Thus, two identical taxon names with different authors would appear in two different rows of the superlist.

If a taxon name occurred in several checklists, there were also several accepted names. We distinguished two cases: all accepted name–author combinations were the same; or they were not. Throughout the article, we will refer to these two possibilities as consistent and inconsistent, respectively. As we had to consider spelling differences in accepted names and authors, we defined thresholds based on the penalties ('diff') calculated by our fuzzy matching algorithms (Notes S2) when comparing genera, species epithets, rank abbreviations, infraspecific epithets ('infra'), and authors. Two accepted name–author combinations were deemed consistent, if the following condition was fulfilled:

$$\mathrm{diff}_{\mathrm{genera}} \leq 4 \text{ AND } \mathrm{diff}_{\mathrm{epithets}} \leq 4 \text{ AND } \mathrm{diff}_{\mathrm{infra}} \leq 4 \text{ AND } \mathrm{diff}_{\mathrm{genera}} + \mathrm{diff}_{\mathrm{genera}} + \mathrm{diff}_{\mathrm{genera}} \leq 6 \text{ AND (authors not available OR } \mathrm{diff}_{\mathrm{authors}} < 1.6).$$

On the contrary, two names were deemed inconsistent if:

$$\mathrm{diff}_{\mathrm{genera}} > 4 \text{ OR } \mathrm{diff}_{\mathrm{epithets}} > 4 \text{ OR } \mathrm{diff}_{\mathrm{infra}} > 4 \text{ OR } \mathrm{diff}_{\mathrm{genera}} + \mathrm{diff}_{\mathrm{genera}} + \mathrm{diff}_{\mathrm{genera}} > 6 \text{ OR (authors available AND } \mathrm{diff}_{\mathrm{authors}} \geq 1.6).$$

The individual values of these thresholds were identified empirically, as there is a trade-off between judging two different name–author combinations to be the same or judging two orthographic variants as to be different (see Notes S1 for the derivation of the author matching threshold). All accepted names belonging to a taxon name could be classified as consistent or inconsistent following our definition.

While WCVP and WP had species distributions attached to each accepted name, mostly corresponding to TDWG level 3 regions (Brummitt *et al.*, 2001), there was no such information in LCVP and WFO. To compare checklists geographically, information needed to be linked to LCVP and WFO. We used the following approach: (1) each row of the superlist corresponds to a taxon name. If the taxon name is present in WCVP or WP, there is distribution information to the corresponding accepted names attached. No matter whether the accepted names across checklists for the particular taxon name are consistent or inconsistent, they refer to the same taxonomic unit and therefore distribution information can be transferred to the corresponding accepted names in LCVP and WFO. (2) If other taxon names within LCVP and WFO have the same accepted names, they also receive this distribution information. (3) For all remaining taxon names with missing distribution information, including few from WCVP and WP, we queried POWO, the online database building on the WCVP taxonomy, using the R package TAXIZE (v.0.9.98, Chamberlain & Szöcs, 2013). We found POWO sometimes had data that were not included in our downloaded version of WCVP. In total, distribution information was completed to 99%, 99%, 98%, and 100% for LCVP, WCVP, WFO, and WP, respectively.

To show phylogenetic family-wise statistics across lists, we displayed them on a phylogenetic tree of vascular plants. As the taxonomic frameworks used and thus the families are partly different across checklists, we harmonized families before the analysis to the Angiosperm Phylogeny Group classification for the orders and families of flowering plants (APG4), if possible. A phylogenetic supertree was taken from Carta *et al.*'s (2022) study and read into R and printed using the V.PHYLOMAKER package (Jin & Qian, 2019). Not all families (18 out of 491) were present in the supertree. They were displayed at the position of the most closely related taxon.

To understand to what degree checklist choice influences analyses, we calculated a measure of list dependency. We define list dependency as the ratio between list-dependent and all taxon names. A taxon name is list-dependent, if it does not occur in all four checklists, or occurs in all checklists, but with inconsistent accepted names. A taxon name is list-independent, if it occurs in all four checklists and has one consistent accepted name in all checklists. List dependency ranges from 0 for identical lists to 1 for lists that do either not share any taxon names or only taxon names with inconsistent accepted names.

For our first use case of matching the names submitted to v.6 of TRY against the four global checklists, we undertook some preprocessing of the TRY names to get more easily interpretable results. We removed special characters, numbers, many vernacular names, fungi and lichen names, location attributes, etc. We only kept taxon names with at least one whitespace, that is, such that potentially represent a genus name with a species epithet. We then used a modified version of the across-list comparison algorithm (Notes S3). Specifically, we additionally allowed for the replacement of rank abbreviations against each other (e.g. replace subsp. by var.) and for the recovery of truncated epithets (some of the names in TRY have epithets truncated after several characters). As all matching was done using the same algorithm, differences in matching results depended solely on the checklists.

To generate an up-to-date list of the plant family Meliaceae, taxonomic experts for this group (Alexandra Muellner-Riehl and Blanca Rojas-Andrés) compiled and critically evaluated information from the literature published up to October 2021, resulting in a final count of 749 accepted species names for this pantropical family. First, the most updated taxonomic and phylogenetic studies conducted for each genus were reviewed, when available. Then, floras, monographs, and other taxonomic literature were reviewed. The following 64 literature sources were used: de Wilde (1968, 2007), Pennington & Styles (1975), Mabberley (1979, 2003, 2011), Pennington *et al.* (1981, 2021), Smith (1985), White & Styles (1986), Styles & White (1990, 1991), Pannell (1992, 1997, 2004, 2020), Hajra *et al.* (1993), Palacios (1994, 2012, 2016), Dassanayake *et al.* (1995), Mabberley *et al.* (1995), Scott (1997), Takeuchi (2000, 2009a,b), Anil Kumar *et al.* (2001), Grierson & Long (2001), Bosser (2002), Kress *et al.* (2003), Lê (2003), Rogers *et al.* (2006), Hua *et al.* (2008), Muellner & Mabberley (2008), Pandey & Dilwakar (2008), Pennington & Muellner-Riehl (2010), Figueiredo *et al.* (2011), Friedmann (2011), Kenfack (2011), Kenfack & Peréz (2011), Lebrun & Stork (2011), Wongprasert *et al.* (2011), Callmander *et al.* (2012), Koenen & de Wilde (2012), Mabberley & Pannell (2013), Pennington & Clarkson (2013), Steinmann & Ramírez-Amezcua (2013), Cuong *et al.* (2014), Velayos *et al.* (2014), Rueangruea *et al.* (2015), Baider & Vincent Florens (2016), Pennington (2016), Bouka Dipelet *et al.* (2017), Randrianarivony *et al.* (2017), García-Gómez *et al.* (2018), Kasongo-Yakusu *et al.* (2018), Sivaraj *et al.* (2018), Flores *et al.* (2019), Palacios *et al.* (2019), Crouch & Styles (2020), Fischer *et al.* (2021), Madagascar Catalogue (2021), Holzmeyer *et al.* (2021).

For the results of the comparison with the Meliaceae expert list, we pruned the same supertree as used above for the family-wise comparison with only include Meliaceae (Carta *et al.*, 2022).

Results

The number of taxon names at differing ranks and numbers of accepted, synonym, and other names was mostly similar in the four checklists (Table 2). The largest differences were WP having *c.* 100 000 taxon names less than the other checklists, and the number of names whose taxonomic status was 'other': WFO had *c.* 141 000, LCVP 62 000, WCVP 46 000, and WP only 1491, respectively. On top of taxonomic information, WCVP and WP provided distributions, which LCVP and WFO did not, and WCVP and WFO included links to IPNI, which were missing in LCVP and WP.

Comparing the occurrences of taxon names across checklists, we found marked differences (Fig. 2). A number of taxon names from TPL were not present in the four recent checklists, a result mainly of the exclusion of some names, which are not validly published. The magnitude of these exclusions differed substantially: It ranged between 20 000 in LCVP and WFO to nearly 200 000 in WP (Fig. 2a). The taxonomic status of these excluded or missing names, according to TPL, was mostly 'unresolved' or 'synonym', but some were also listed as 'accepted' (Fig. 2b). All four recent lists included taxon names, which were not in TPL. Roughly 100 000 of these new or additional names were shared between LCVP, WCVP, and WP (Fig. 2c). Half of these were also shared with WFO. A number of taxon names have been added in several, but not all checklists.

Each of the four checklists had unique taxon names absent in the other lists (Fig. 2d). In WCVP, which had most unique taxon names, they made up *c.* 5% of the total number. As can be seen by subtracting from the additions to TPL, these taxon names were mainly absent from TPL, but in the cases of LCVP and WFO, of the 25 000 and 21 000 unique taxon names, 6000 and 7000 were shared with TPL, respectively. In WCVP and WP, there were < 800 such names. Geographically, the unique taxon names in the different checklists had accepted names with the following distribution centers: Southwestern Europe in LCVP; Europe in WCVP; no center in WFO; and Southwestern Europe, parts of China and India in WP (Fig. S1, upper row).

Accepted names linked to a particular taxon name were not always consistent across checklists. Instead, there was a considerable degree of inconsistency, where one or several opinions on the accepted name of a given taxon name existed (Fig. 2e vs 2f). In total, we identified *c.* 720 000 taxon names with consistent accepted names occurring within all four lists. Depending on the checklist considered, there were between 100 000 and 180 000 additional taxon names with consistent accepted names, which were not present in all the lists. On the contrary, *c.* 300 000 taxon names had inconsistent accepted names across the four checklists, although we did not quantify whether two, three, or even four different opinions on the respective accepted names existed. Again, we found an additional 50 000–125 000 taxon names with inconsistent accepted names only present in two or three of the lists, depending on which checklist was considered. The geographic distribution of these inconsistent taxa was centered in Southwestern Europe, China, and parts of Northern and Central South America (Fig. S1, lower row).

When considering only accepted names, we found *c.* 300 000 shared across all lists, with another 70 000 shared across several, but not all lists, and between 16 000 and 28 000 accepted names found in each of the lists exclusively (Fig. 2g).

Homonyms, that is identical taxon names with differing author names, were shared to *c.* 50%, equalling *c.* 20 000 names, across all lists. The remainder was either shared across only several lists or only occurring in one list (Fig. 2h). The exception was LCVP, which contained 15 000 more homonyms than any other checklist.

For the entire checklists, list dependency, the proportion of taxon names not occurring in all checklists or with differing accepted names across lists, was: 0.46 (LCVP), 0.46 (WCVP), 0.44 (WFO), and 0.4 (WP). When calculated for a specific geographic region or plant family, we report the average across the four checklists throughout.

Using the distributions retrieved from WCVP and WP and transferred to the other lists, we found that list dependency

Table 2 General comparison of the four global plant names checklists.

	LCVP	WCVP	WFO	WP
Version	1.05	2022	2021.12	13.0
Release date	May 2022	February 2022	December 2021	June 2022
Scheduled update frequency	Every 2 yr	Every 3 months	Every 6 months (more on demand)	Every 2 months
Taxonomic curation	The author	The author with advice from expert reviewers	Community-based with taxonomic experts for groups	The author with advice from expert reviewers
Taxonomic scope	Vascular plants (tracheophyta)	Vascular plants (tracheophyta)	Land plants (embryophyta = tracheophyta + mosses)	Vascular plants (tracheophyta)
Taxonomic levels	Species and infraspecies	Genus, species, and infraspecies	Phylum, order, family, genus, species, and infraspecies	Phylum, order, family, genus, species, and infraspecies
Number of vascular plant taxon names	1339 275	1354 108	1315 336	1210 720
Number of species taxon names	1079 040	1019 804	1034 677	970 612
Number of subspecies taxon names	55 178	67 357	60 779	70 396
Number of variety taxon names	175 903	216 494	186 598	149 321
Number of form taxon names	28 974	40 018	32 391	18 345
Number of other infraspecific taxon names	180	10 435	891	2046
Number of accepted names	416 099	407 961	388 342	415 170
Number of synonyms	861 282	900 455	785 732	794 059
Number of other names (unresolved, unplaced, etc.)	61 894	45 692	141 262	1491
ID provided	Yes	Yes	Yes	No
Accepted name provided directly	Yes	No	No	Yes
Family provided	Yes	Yes	Yes	Yes
Order provided	Yes	No	No	Yes
Distribution data provided	No	Yes (through POWO)	No	Yes
IPNI link provided	No	Yes (mostly)	Yes (mostly)	No
Original publication reference	No	Yes	Yes	Yes
Recent publication reference	Yes (but mostly missing)	No (but available through POWO)	Yes (but mostly links to TPL)	No
Number of autonyms	36	20 110	34 724	11 157
Several rank abbreviations in one name	Rarely (40)	No	Rarely (7)	Rarely (124)

Accepted name provided directly refers to all lists listing an accepted name for each taxon name. In some cases (WCVP and WFO), the ID of the accepted species name, but not the actual name is provided, which means users have to search for the accepted name using this ID. Note that WFO includes mosses in addition to vascular plants. They were removed here and in all subsequent analyses to ensure comparability. LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online; WP, WorldPlants.

increased from the tropics and subtropics toward the temperate regions by a factor of 1.5–2 (0.4 vs 0.6 and 0.2 vs 0.4, respectively) for taxon and accepted names, respectively (Fig. 3). While in the tropics, two out of five taxon names randomly taken from a checklist would not be present in all lists or have inconsistent accepted names, in the temperate zones, this ratio was three out of five. For accepted names, starting with one out of five in the tropics, it reached two out of five in the temperate zones.

On the family level, we found pronounced differences, some having low list dependency, while others showing a high value (Fig. 4). As can be seen comparing the inner and outer rings representing accepted and taxon names, respectively, list dependency of those two was highly correlated ($r = 0.79$). Two hotspots could be identified: One was within pteridophytes, that is, Desmophlebiaceae and related taxa, and one in the Santalales, including Aptandraceae and Nanodeaceae. On the other side of the spectrum, low list dependency could be found scattered throughout vascular plant families. In absolute terms, numbers of unique taxon names and taxon names with inconsistent accepted names across checklists were highest in the largest plant families, that is, Asteraceae, Orchidaceae, Fabaceae, and Poaceae (Fig. S2).

Our first use case, the name-matching of the names of the TRY database, resulted in 76% consistent matches within the four checklists (Table 3). A further 21% of the matches were found in all four lists, but with inconsistent accepted names. Depending on the checklist considered, the remaining 3% included a fraction not found and taxon names that were not found by all checklists with consistent or inconsistent accepted names. As a consequence of the high overall agreement, the number of resolved names returned by the checklists was similar and accounts for *c.* 56% of the number of taxon names, showing that about half of the taxon names fell into the categories of 'accepted' or 'other', while the others were synonyms.

For our second use case, we counted the number of accepted species names in the four checklists using an independent expert-derived list of the family Meliaceae (Table 4; Fig. 5; refer to Fig. S3 for a figure including species names). We differentiated three outcomes: accepted names found as accepted names;

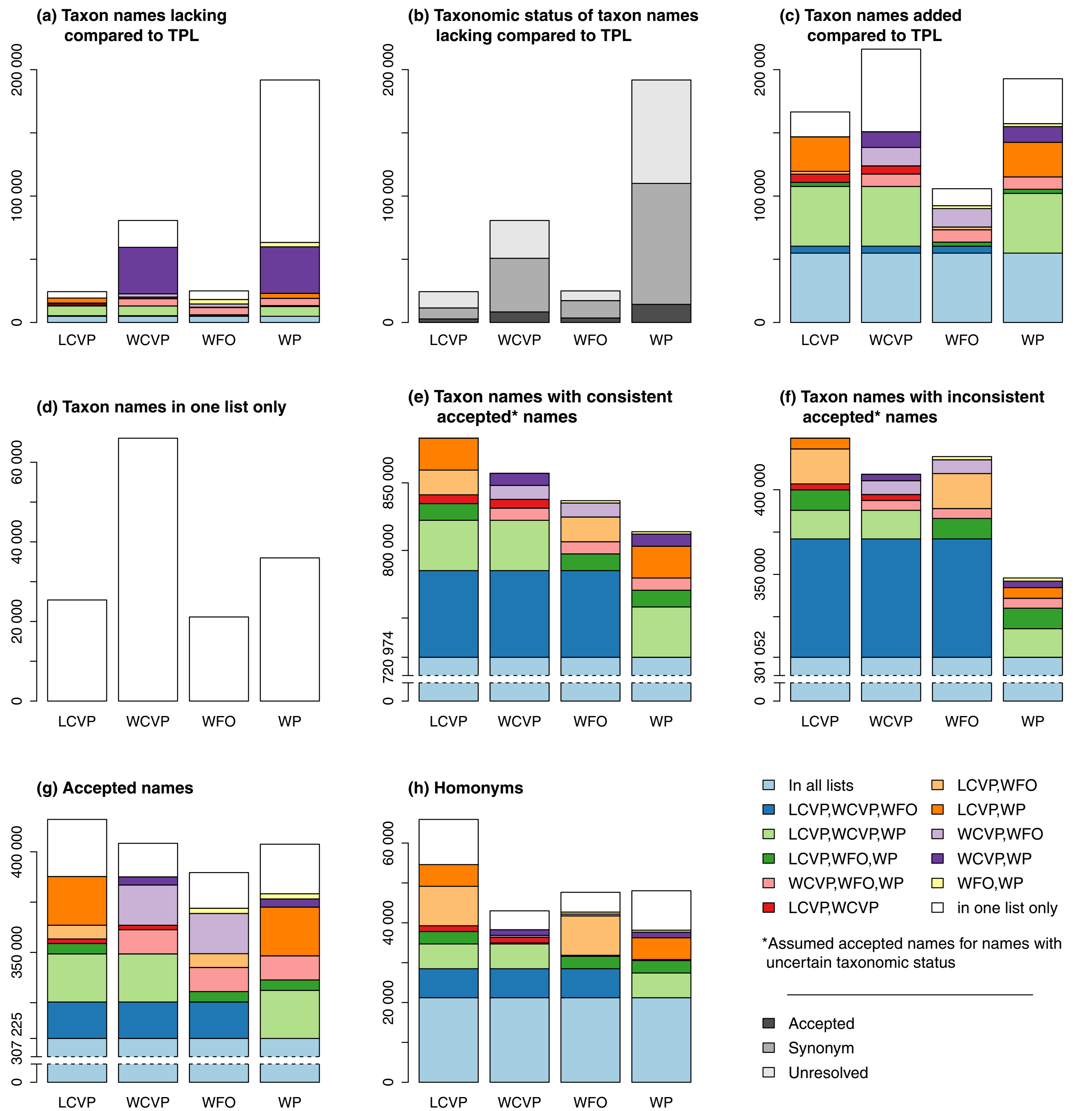

Fig. 2 Comparison of taxa numbers across the four global checklists and to The Plant List (TPL). (a) Taxon names lacking compared to TPL. These names are present in TPL, but not in the respective lists. (b) Taxonomic status of taxon names lacking compared to TPL. The taxonomic status of the names from (a) falls into the categories accepted, synonym, and unresolved. (c) Taxon names added compared to TPL. Added names may arise from new descriptions or names not included in TPL for different reasons. (d) Taxon names in one list only. These names occur in one checklists exclusively. (e) Taxon names with consistent accepted names. The taxonomic opinion on the referenced taxa is shared across checklists. (f) Taxon names with inconsistent accepted names. The taxonomic opinion on the referenced taxa differs between checklists. (g) Accepted names. These names are consistently classified as accepted by the checklists. (h) Homonyms. Homonyms but arise from several authors using the same taxon name in descriptions. Same colours indicate the same taxon names across lists. White filling shows taxon names found in one list only. Note the different scales and the cut bars in the plots (d–f). The exact size of the lowest stacks of those is always given on the respective scale. LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online; WP, WorldPlants.

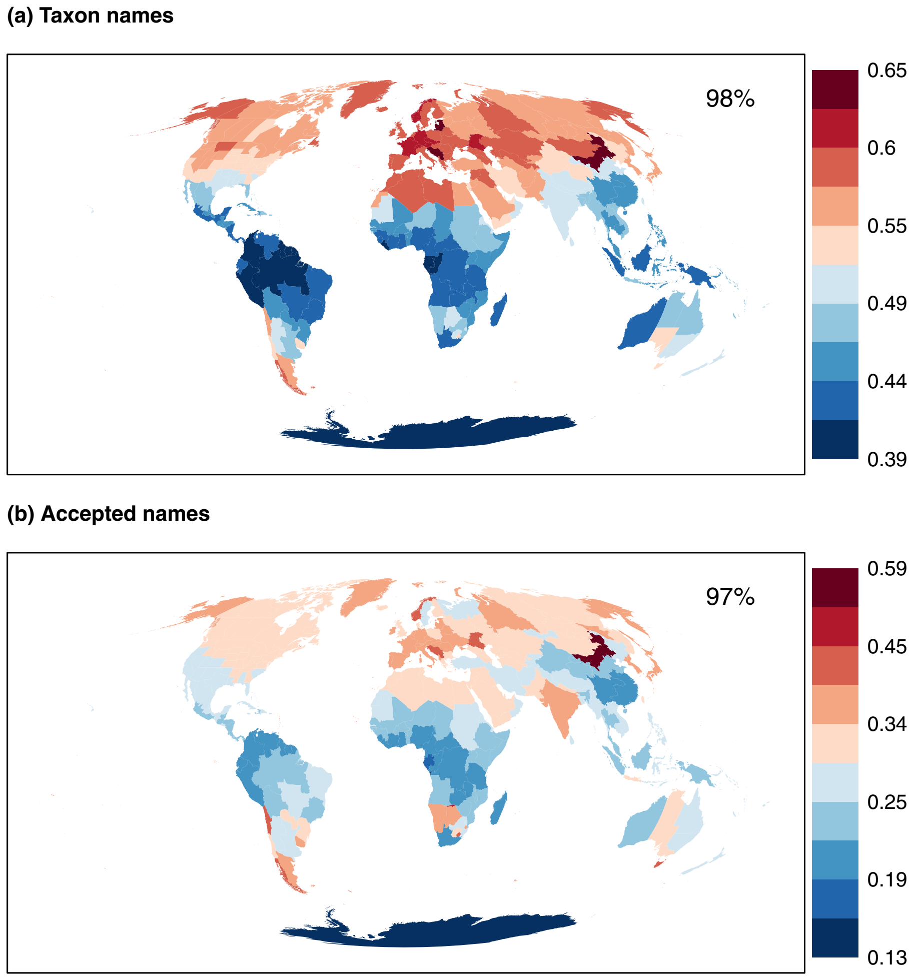

Fig. 3 Geographic distribution of list dependency of taxon names and accepted names. Maps show the averaged probability to select a taxon name (a) or accepted name (b) from one of the four lists that is not found in all lists, or, in case of taxon names, refers to different accepted names across lists. Numbers in the upper right indicate average completeness of taxa displayed given that for some taxa, especially from Leipzig Catalogue of Vascular Plants (LCVP) and World Flora Online (WFO), no distribution information was available.

accepted names found as synonyms; and accepted names not found. Nearly all accepted names were found in LCVP and WP, and a large proportion also in WCVP and WFO. Most names were correctly classified as accepted, although all checklists showed disagreements with the expert-compiled list. We also tested whether other taxon names not appearing in the expert list were deemed to be accepted, as this may result in incomplete name resolution. All checklists included such taxon names, the least were in WP and the most in WFO. The phylogenetic tree shows that the main differences between the four checklists were with a number of species in the genera *Didymocheton* Blume, *Goniocheton* Blume, and *Prasoxylon* M.Roem. (Fig. 5).

Discussion

Our main expectations, the turnover in taxon names compared with TPL, and the expressed differences across checklists, were corroborated by our results. Yet, we find the degree to which the checklists differed is large, given that they, to greater or lesser extents, are all used in scientific works without the perception that their individual choice may affect analyses.

Compared with the outdated and never thoroughly revised TPL, many taxon names were removed or lack in the current versions of the recent checklists. Errors in harmonizing algorithms, a large number of taxon names that were never validly published and unresolved names, were cross-checked and sorted out by the other checklists. This process cannot be automatized. It has to be done through comparisons with local floras and checklists, and new taxonomic revisions. The flagging of such names in TPL was unreliable in this respect: The taxonomic status of taxon names from TPL not present in the four checklists included 'unresolved', an equal amount of 'synonym', and even a smaller number of 'accepted' names (Fig. 2b). As decisions on removing, or not including, taxon names depend on the rules on which the lists are based, it is no surprise that there were differences. Concerning 'names' which are not validly published (termed designations by the Nomenclatural Code), there is an ongoing debate on whether to and how to include them in taxonomic databases. For informed users, global lists that include designations may help to identify and deal with them in their datasets. However, other users may ignore the 'invalid' flagging and continue to misuse these names. WP and WCVP take a purist nomenclatural approach and limit the inclusion of designations, whereas LCVP and WFO take a more expansive approach and seek to include all 'names' which are effectively published. WP additionally excluded unresolved names and synonyms from Index Kewensis that could not be verified with local checklists or floras, although in its most

New Phytologist

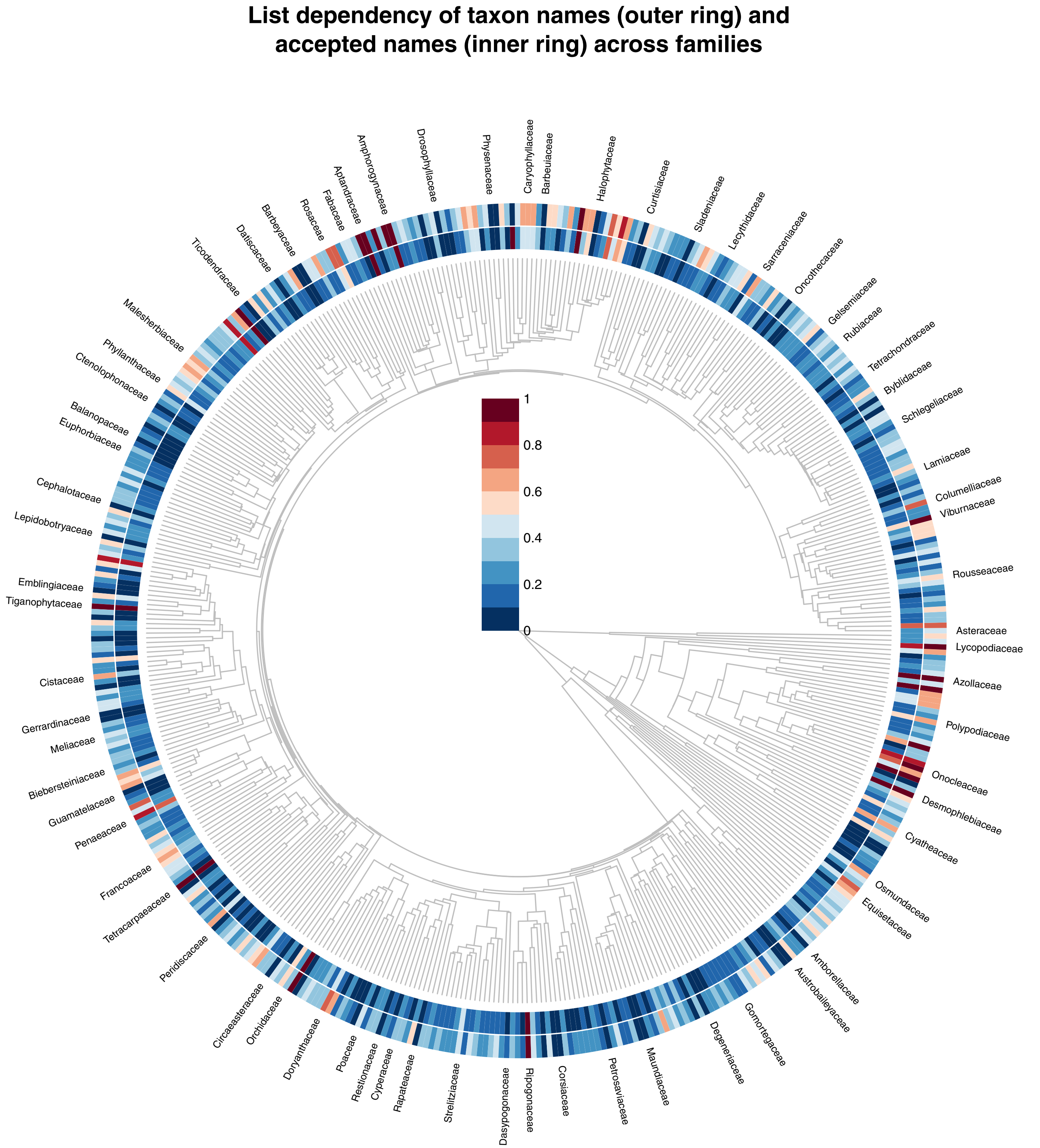

Fig. 4 Phylogenetic tree of vascular plants showing list dependency for taxon names (outer circle) and accepted names (inner circle). The colours symbolize the averaged probability to select a taxon name or accepted name from the respective family that is not found in any other list, or, in the case of taxon names, that has different accepted names across lists.

recent version, the WP dataset includes these synonyms. Unresolved names are still not included in WP, with the exception of a separate list of ferns. In the age of increasing use of data aggregators, such as Catalogue of Life or the Global Biodiversity Information Facility (GBIF), individual decisions by global lists on what names/designations they include presents a challenge,

Table 3 Results of the name resolution of the TRY species list.

	LCVP	WCVP	WFO	WP
In one list only	288 (0.1%)	303 (0.1%)	301 (0.1%)	188 (< 0.1%)
Consistent in less than four lists	4676 (1%)	4741 (1%)	3745 (0.8%)	2382 (0.5%)
Inconsistent in less than four lists	5594 (1.2%)	4304 (0.9%)	4873 (1%)	1671 (0.3%)
Consistent in four lists	367 078 (75.5%)	367 078 (75.5%)	367 078 (75.5%)	367 078 (75.5%)
Inconsistent in four lists	100 203 (20.6%)	100 203 (20.6%)	100 203 (20.6%)	100 203 (20.6%)
Not found	8453 (1.7%)	9663 (2%)	10 092 (2.1%)	14 770 (3%)
Number of resolved names	274 202 (*c.* 56.4%)	276 079 (*c.* 56.8%)	283 166 (*c.* 58.2%)	271 420 (*c.* 55.8%)

Names delivered to TRY were preprocessed removing special characters and numbers, among others. Only names consisting of at least two words, that is genus and epithet, were used. All names were matched using the same matching algorithm. Results are therefore entirely dependent on the respective list data. Percentages refer to the 486 292 names processed in total. The number of resolved names refers to the number of taxon names obtained per list when all synonyms are resolved. Percentages of resolved names also refer to the total number of processed names, indicating that roughly every second species name was found to be a synonym on average. LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online, WP WorldPlants.

Table 4 Comparison of accepted species found in an expert-compiled list of Meliaceae and the four global checklists.

	LCVP	WCVP	WFO	WP
Accepted names found as accepted	724 (96.7%)	659 (88%)	643 (85.8%)	709 (94.7%)
Accepted names found as synonyms	25 (3.3%)	42 (5.6%)	45 (6%)	35 (4.7%)
Accepted names not found	0 (0%)	48 (6.4%)	61 (8.1%)	5 (0.7%)
Synonyms found as accepted names	90 (12%)	175 (23.4%)	323 (43.1%)	56 (7.5%)
Total number of taxon names (including infraspecies)	3864	3669	3592	3572

'Synonyms found as accepted names' refers to taxon names that are classified as accepted names by the respective list. The total number of taxon names includes synonyms and accepted species names. LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online; WP, WorldPlants.

as if data are aggregated from several sources, taxon names excluded by some but not all global lists may be present.

Considering our second expectation of an idiosyncratic development of the checklists, our data confirm that as of the December 2022, WFO had not caught up with the other lists, considering it only shared half of the taxon names the other three lists have included as taxonomic additions since TPL. This is not surprising as, at that point, WFO was in development and had only incorporated updated names and classifications from a few Taxonomic Expert Networks.

It is notable that 50 000–100 000 taxon names found in just two of the checklists or exclusive to a single checklist either were not known to the other checklists, or there exist different opinions on whether these should be listed. While all new taxon names should be registered in IPNI, it happens that work published in regional journals, particularly on the re-classification of taxa, is overlooked. This is why we strongly advocate for mandatory taxon name registration, as is already the case for fungi (Aime *et al*., 2021). WCVP and WFO include links to IPNI for most taxon names. Although difficulties arise for a small percentage of names (Table S1), this has the potential to facilitate comparisons and offers a straightforward way to link additional data.

Apart from the exclusion or inclusion of certain taxon names, the larger part of the differences across checklists was due to the accepted names linked to them. Some of these differences will be mitigated in the near future, for example the high number of unresolved names in WFO compared with the other checklists, a legacy of TPL which is being addressed. But the sheer scale of differences in accepted names linked to taxon names across all lists means that it is unlikely they will be checked in the short term, leaving users with no choice but to acknowledge the idiosyncrasies and potentially test for their effect on name resolution. Such tests, however, should include botanical author names, but unfortunately, at a time where large global databases as TRY, GBIF, or GIFT make huge amounts of data available to ecologists, there is a tendency to omit author names for simplicity. This is not only the fault of the users themselves – the majority (76%) of the taxon names submitted to the v.6 of TRY had no author names. Problems arising through this are well known among taxonomists, but mostly ignored or unknown in other research communities. Examples include wrong distribution models or animal names ending up in botanical databases. While only 3–5% of taxon names in the different checklists are homonyms (Fig. 2h), their author names being necessary for correct assignment, we argue increased attention should be paid when submitting data involving taxonomic information to any database. Especially for homonyms, which have a list dependence of > 0.5, inconsistencies in databases may have a substantial impact on inferences drawn.

In a geographic context, the global maps presented in Fig. S1 offer some explanations for the differences across checklists: The strong centering in Europe is due to intensive taxonomic work in this region, with several apomictic genera, for example *Hieracium*

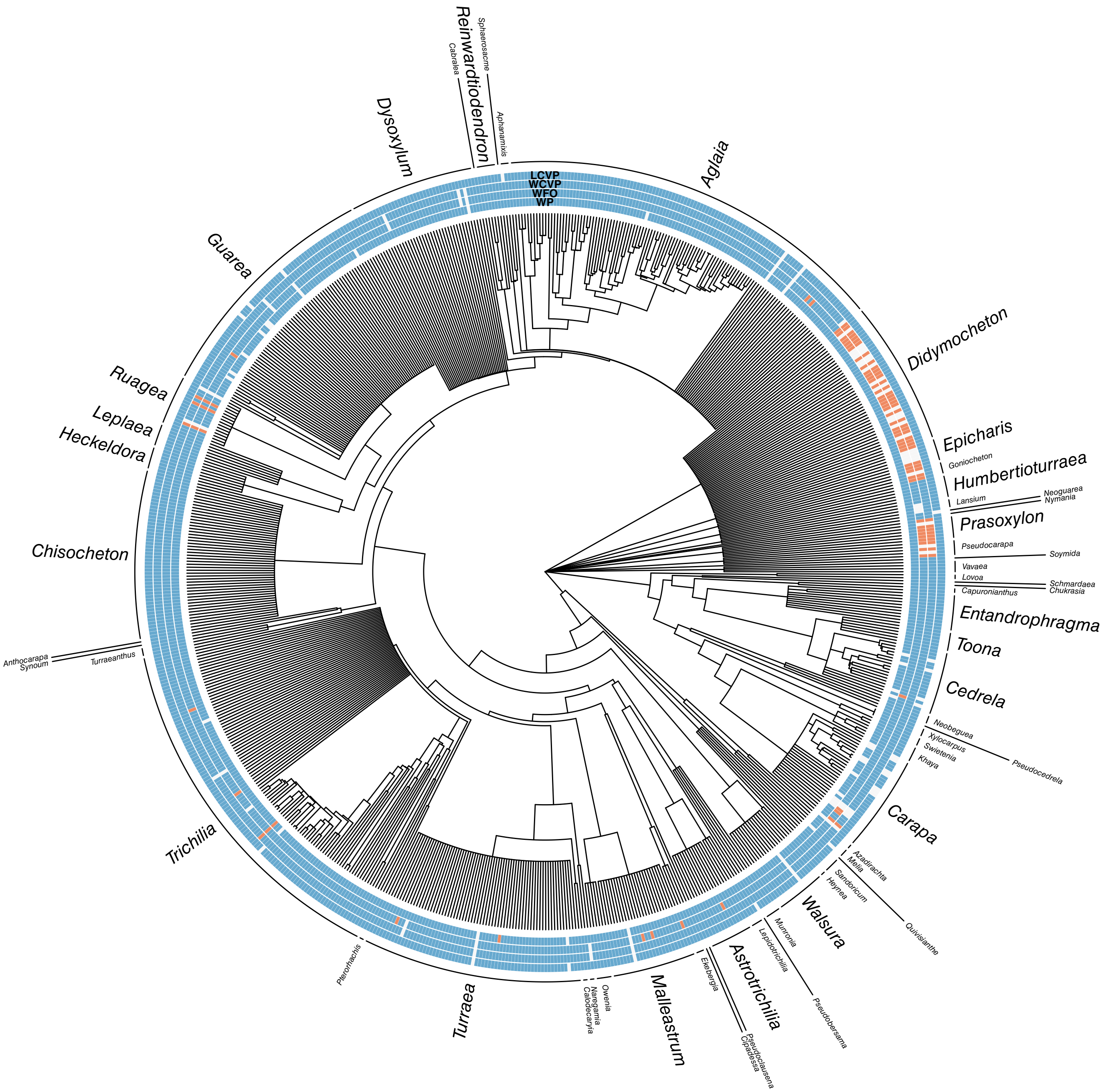

Fig. 5 Phylogenetic tree of the Meliaceae showing whether accepted species names can be found in the four global checklists. The four circles represent Leipzig Catalogue of Vascular Plants (LCVP), World Checklist of Vascular Plants (WCVP), World Flora Online (WFO), and WorldPlants (WP) from outermost to innermost, respectively. Colours indicate the following: blue, found; white, found as a synonym, not an accepted name; red, not found. Refer to Supporting Information Fig. S3 for individual species names.

L. or *Rubus* L., which have a large number of species or subspecies defined, so-called microspecies. The differences across checklists in the number of species and infraspecies of any taxonomic status within these genera are large (Table S2, 'genus statistics'), and this results in the centering in Europe.

Absolute numbers are important when trying to understand the reasons for differences across the checklists, but for practitioners, measures such as list dependency, which show the influence of these differences, are more relevant. Due to variation in total taxon numbers, the geographic distribution of list dependency does not mirror the absolute numbers. The gradient from the equator to the poles is likely explained by a larger number of scientists with more financial resources in temperate and economically more developed countries working on relatively few taxa compared with the tropical and subtropical countries (Meyer *et al.*, 2016).

From a phylogenetic perspective, the pattern of list dependency reflects taxonomic activity in the different clades. The hotspot in list dependency in the pteridophytes can be traced to WCVP using a classification different from the other lists, expanding Polypodiaceae and Aspleniaceae to include many other fern families, but also to a couple of recent taxonomic treatments in Thelypteridaceae and Polypodiaceae (Fawcett *et al*., 2021; Wei *et al*., 2021). The other hotspot in Santalales is due to vastly different family circumscriptions in the recent literature, the group being split in *c*. 20 families by Nickrent *et al*. (2019), which has not been accepted by all lists.

For the evaluation of the first use case matching the TRY data, it is necessary to compare the results to the overall checklist list dependency, which was *c*. 0.44 on average, meaning that, out of 10 taxon names, four to five are list-dependent. The TRY data, however, with three out of four taxon names being consistent across checklists, differ clearly from what we expected. It shows that – assuming the TRY taxon names are representative – differences across checklists may be relatively low in commonly investigated taxa and higher in others. Given the relatively small differences in overall matching success, the results do not warrant the recommendation of one checklist over the other.

In our second use case that tested name resolution capability, we found a high overall retrieval rate of accepted names in all checklists. The large difference in the number of Meliaceae taxon names missing or being listed as synonyms in WCVP and WFO can be traced to the work of Holzmeyer *et al*. (2021), published in October 2021. It reinstates five genera of Meliaceae based on phylogenetic evidence, leading to a considerable number of changes in accepted species names. The versions of WCVP and WFO were *c*. 3–6 months older than those of LCVP and WP, pointing at the pace with which plant taxonomy is changing and the effort needed to incorporate recent changes into the checklists. More importantly, all checklists included more accepted names within Meliaceae than the expert list. Depending on their number, this affects name resolution negatively, especially in WFO, where they accounted for over 40% of the accepted names listed. The details would need to be thoroughly investigated, but this example emphasizes that the corrections in taxonomy are of equal, if not larger importance, than new descriptions in keeping checklists up-to-date. However, no premature conclusions should be drawn from this use case. In all databases, new taxonomic work is constantly incorporated. This is, however, taxon-dependent, especially in WFO, where Taxonomic Expert Networks working on specific groups, but not on all, exist. An informed assessment of the taxonomic quality would therefore need a larger number of samples.

While the results presented here can objectively be evaluated, it is important to remember the following two aspects: (1) from neither of our analyses (the Meliaceae use case cannot be assumed representative for all vascular plants), we can judge the taxonomic quality of the checklists. Apart from being subjective to some extent, this would need taxonomic experts to review the differences found across lists, an endeavour beyond the scope of this study. (2) There are different approaches, or even philosophies, behind each list's assembly and development. In terms of up-to-dateness, the checklists mainly curated by single individuals have an advantage compared with the community approach involving groups of taxonomic experts for specific groups. Also, decisions on changes in functionality and design are easier and faster with fewer individuals involved. The community approach, however, ensures best available knowledge is used in the revision of individual groups, and the results are likely accepted as authoritative by the scientific community. In the long run, the community approach may prove more sustainable, as it does not depend on the persistence of individuals and is more consensus-based. The community approach is followed by WFO, and by WCVP to a much lesser extent, being curated by Rafaël Govaerts liaising with taxonomists from around the world. LCVP and WP receive help for specific tasks, but are to a large extent dependent on the individual work of Martin Freiberg and Michael Hassler, respectively. Together, the community vs individual approach, and the decisions made during list development, account for the majority of the differences seen across checklists throughout this work: whether taxonomic research has been integrated or not, different conclusions drawn from taxonomic revisions, and different decisions regarding the inclusion or exclusion of problematic taxon names. While the other lists have arrived at a state of 'maturity', WFO is still explicitly work in progress. Limitations of the underlying software used to manage the WFO database restricted progress in updating the outdated classification derived from TPL and the incorporation of new data in the past. A new online

Table 5 Selected R packages available for name resolution using the four checklists.

Checklist	Recommended R package	Checklist version used	URL	Reference
LCVP	LCVPLANTS	2.0	https://github.com/idiv-biodiversity/lcvplants	–
WCVP	WCVP package under development	2022	https://github.com/barnabywalker/kewr	–
WFO	WORLDFLORA	v.2022.07	https://cran.r-project.org/web/packages/WorldFlora/index.html	Kindt (2020)
WP	U.TAXONSTAND	User-dependent	https://github.com/ecoinfor/U.Taxonstand	Zhang & Qian (2022)

Currently, there is no R package available specifically for WorldPlants, but R package U.TAXONSTAND may be used. However, the user needs to provide an own offline copy of the WP dataset. Checklist versions are those available at the time of writing. LCVP, Leipzig Catalogue of Vascular Plants; WCVP, World Checklist of Vascular Plants; WFO, World Flora Online; WP, WorldPlants.

software system is now being used, easing incorporation of data updates on a regular basis.

Considering the long outdated TPL, we show its use should be abandoned. The use of any one of the four checklists should be done with care and understanding. Although our name matching use case showed a high agreement across lists, when working in particular regions or with specific taxonomic groups, differences across checklists may be substantial. When working with specific genera or families, this can be checked in the overview of the number of taxon names in each checklist (Table S2). While convenience of use may have been the reason for using a particular checklist (as was the case for TPL), several packages exist now in the R programming language (one specifically designed for WCVP is presented in this issue) to ease the use of the four checklists by providing matching algorithms and prepared functions for automated data processing (Table 5). We acknowledge that the existence of the four global checklists may be confusing to users, but many of the possibilities for improvement shown in this work were only revealed because these several checklists exist. The group of authors of this study, comprising taxonomic experts including authors and curators of the four reference lists, but also of representatives from user groups in the field of functional biodiversity research, biogeography and macroecology, are working on establishing a workflow to facilitate cross-talking between the four checklists. We see the current activity as a first explorative step toward a future integrated system.

Experiences and lessons learned from this study are the following: exploring and harmonizing the datasets for comparative analysis leads to the detection of errors (e.g. wrong encoding, special characters in names, missing authors, and duplicate name–author combinations); matching names across datasets allows for the incorporation of new information that was previously not included in some checklists (e.g. adding geographical information to LCVP and WFO, adding IPNI links to LCVP and WP); and it creates self-awareness of the strengths and weaknesses of the different approaches and products that help us to improve and to engage in targeted collaborations, harmonizing existing or importing missing content. At the same time, the diversity of approaches, reflecting individual interests, expertise and working styles, is embraced as a precondition to create both complementarity and synergy. In more general terms, scientific progress and quality control profits from a certain degree of redundancy and replication as a basis for testing reproducibility. This said, the outcome of the presented study is a major motivation for us to discuss roles and workflows for future collaboration.

Acknowledgements

We want to thank the *New Phytologist* for the possibility to submit our manuscript for the forthcoming special issue. Special thanks also to the high-performance cluster computing system of the University of Leipzig, where time-consuming list-matching calculations were performed. We thank an anonymous reviewer for valuable suggestions for improvements of a previous version of this manuscript. Open Access funding enabled and organized by Projekt DEAL.

Competing interests

None declared.

Author contributions

AZ, CW, DSC, MG and M Winter conceived the first ideas for this manuscript. MF, RG, M Watson and MH provided inside knowledge on LCVP, WCVP, WFO and WP, respectively. JK and GB supplied the taxon names of TRY v.6. ANM-R and BMRA compiled the Meliaceae checklist. DSC conceived and ran the analyses. DSC wrote the first version of the manuscript. All authors contributed substantially to the interpretation of the results and the final version of the manuscript.

ORCID

Martin Freiberg https://orcid.org/0000-0002-6036-5293
Rafaël Govaerts https://orcid.org/0000-0003-2991-5282
Matthias Grenié https://orcid.org/0000-0002-4659-7522
Jens Kattge https://orcid.org/0000-0002-1022-8469
Alexandra N. Muellner-Riehl https://orcid.org/0000-0002-2710-469X
Blanca M. Rojas Andrés https://orcid.org/0000-0001-7164-1313
David Schellenberger Costa https://orcid.org/0000-0003-1747-1506
Mark Watson https://orcid.org/0000-0002-4989-3976
Marten Winter https://orcid.org/0000-0002-9593-7300
Christian Wirth https://orcid.org/0000-0003-2604-8056
Alexander Zizka https://orcid.org/0000-0002-1680-9192

Data availability

All data used in this study are openly available through the sources mentioned in the Materials and Methods section, with the exception of the version of the World Checklist of Vascular Plants (WCVP) used, which was made available in November 2022 at http://sftp.kew.org/pub/data-repositories/WCVP/.

References

Aime MC, Miller AN, Aoki T, Bensch K, Cai L, Crous PW, Hawksworth DL, Hyde KD, Kirk PM, Lücking R *et al.* 2021. How to publish a new fungal species, or name, v.3.0. *IMA Fungus* **12**: 11.

Anil Kumar N, Sivadasan M, Ratheesh Narayanan MK. 2001. A new species of *Dysoxylum* Blume (Meliaceae) from India. *Rheedea* **11**: 115–118.

Baider C, Vincent Florens FB. 2016. A new and critically endangered species of *Turraea* (Meliaceae) endemic to the Island of Mauritius. *Phytotaxa* **247**: 219–228.

Bánki O, Roskov Y, Döring M, Ower G, Vandepitte L, Hobern D, Remsen D, Schalk PP, DeWalt RE, Keping M *et al.* 2022. *Catalogue of Life Checklist.* doi: 10.48580/dfqc.

Borsch T, Berendsohn W, Dalcin E, Delmas M, Demissew S, Elliott A, Fritsch P, Fuchs A, Geltman D, Güner A *et al.* 2020. World Flora Online: placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants. *Taxon* **69**: 1311–1341.

Bosser J. 2002. Une nouvelle espèce de *Turraea* (Meliaceae) des Mascareignes. Localisation de *T. thouarsiana* et identité de *T. casimiriana*. *Adansonia* **24**: 113–116.

Bouka Dipelet UG, Florence J, Doumenge C, Loumeto JJ, McKey D. 2017. Khayae (Meliaceae) specierum Nomenclator. *Adansonia* **39**: 15–30.

Bruelheide H, Dengler J, Jiménez-Alfaro B, Purschke O, Hennekens SM, Chytrý M, Pillar VD, Jansen F, Kattge J, Sandel B *et al.* 2019. sPLOT – a new tool for global vegetation analyses. *Journal of Vegetation Science* **30**: 161–186.

Brummitt RK, Pando F, Hollis S, Brummitt N. 2001. *World geographical scheme for recording plant distributions.* Pittsburg, PA, USA: International Working Group on Taxonomic Databases for Plant Sciences (TDWG).

Cachuela-Palacio M. 2006. Towards an index of all known species: the Catalogue of Life, its rationale, design and use. *Integrative Zoology* **1**: 18–21.

Callmander MW, Phillipson PB, Lowry PP. 2012. Novelties from the northern mountains complex of Madagascar. III. Two new species of *Turraea* L. (Meliaceae). *Adansonia* **34**: 93–102.

Carta A, Peruzzi L, Ramírez-Barahona S. 2022. A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytologist* **233**: 1494–1504.

Chamberlain SA, Szöcs E. 2013. TAXIZE: taxonomic search and retrieval in R. *F1000Research* **2**: 191.

Crouch NR, Styles DGA. 2020. *Turraea tugelensis*, a new species of *Turraea* sect. Nurmonia (Meliaceae) from South Africa. *Phytotaxa* **460**: 137–142.

Cuong NT, Hoan DT, Mabberley DJ. 2014. *Munronia petiolata* (Meliaceae), a new species from Vietnam. *Blumea – Biodiversity, Evolution and Biogeography of Plants* **59**: 139–141.

Dassanayake MD, Fosberg FR, Clayton WD, eds. 1995. *A revised handbook to the flora of Ceylon, vol. 9.* New Delhi, India; Rotterdam, the Netherlands: Amerind Publishing; Balkema.

Enquist BJ, Condit R, Peet RK, Schildhauer M, Thiers BM. 2016. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprints* doi: 10.7287/peerj.preprints.2615v2.

Fawcett S, Smith AR, Sundue M, Burleigh JG, Sessa EB, Kuo L-Y, Chen C-W, Testo WL, Kessler M, GoFlag Consortium *et al.* 2021. A global phylogenomic study of the Thelypteridaceae. *Systematic Botany* **46**: 891–915.

Figueiredo E, Paiva J, Stévart T, Oliveira F, Smith GF. 2011. Annotated catalogue of the flowering plants of São Tomé and Príncipe. *Bothalia* **41**: 41–82.

Fischer E, Killmann D, Leh B, Janssens SB. 2021. *Carapa wohllebenii* (Meliaceae), a new tree species from montane forests in the Democratic Republic of Congo, Rwanda, and Burundi. *Phytotaxa* **511**: 20–36.

Flores TB, Souza VC, Coelho RLG. 2019. A new species of *Trichilia* (Meliaceae) from the Atlantic Forest of Brazil. *Rodriguesia* **70**: e01312018.

Freiberg M, Winter M, Gentile A, Zizka A, Muellner-Riehl AN, Weigelt A, Wirth C. 2020. LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Scientific Data* **7**: 416.

Friedmann F. 2011. *Flore des Seychelles: dicotylédones.* Marseille, France: Publications Scientifiques du Muséum.

García-Gómez A, Figueroa-Brito R, García Serrano LA, Jiménez-Pérez A. 2018. *Trichilia* (Meliaceae) plants: an important source of biomolecules with insecticidal properties. *Florida Entomologist* **101**: 470–479.

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**: 215.

Grenié M, Berti E, Carvajal-Quintero J, Dädlow GML, Sagouis A, Winter M. 2023. Harmonizing taxon names in biodiversity data: a review of tools, databases and best practices. *Methods in Ecology and Evolution* **14**: 12–25.

Grierson AJC, Long DG. 2001. *Flora of Bhutan, vol. 2, part 3.* Huddersfield, UK: Royal Botanic Garden Edinburgh and Royal Government of Bhutan.

Hajra PK, Nair VJ, Daniel P, eds. 1993. *Flora of India, vol. 4, Malpiguiaceae – Dichapetalaceae.* Calcutta, India: Botanical Survey of India.

Harrison J. 2020. *RSELENIUM: R bindings for 'Selenium WebDriver'.* [WWW document] URL https://CRAN.R-project.org/package=RSelenium [accessed 19 January 2022].

Hassler M. 2022. *World Plants 12.3. Synonymic checklist and distribution of the World flora.* [WWW document] URL www.worldplants.de [accessed 17 May 2022].

Holzmeyer L, Hauenschild F, Mabberley DJ, Muellner-Riehl AN. 2021. Confirmed polyphyly, generic recircumscription and typification of *Dysoxylum* (Meliaceae), with revised disposition of currently accepted species. *Taxon* **70**: 1248–1272.

Hooker JD. 1893. *Index Kewensis.* Kew, Australia: Clarendon Press.

Hua P, Mabberley DJ, Pannell CM, Edmonds J, Bartholomew B. 2008. Meliaceae. In: Shukun C, ed. *Flora of China, vol. 11.* Beijing, China; St Louis, MO, USA: Science Press; Missouri Botanical Garden Press, 111–131.

Jin Y, Qian H. 2019. V.PHYLOMAKER: an R package that can generate very large phylogenies for vascular plants. *Ecography* **42**: 1353–1359.

Kasongo-Yakusu E, Monthe F, Bourland N, Hardy OJ, Louppe D, Bola F, Hubau W, Kahindo J-M, Bulcke J, Acker J *et al.* 2018. Le genre *Entandrophragma* (Meliaceae): taxonomie et écologie d'arbres africains d'intérêt économique (synthèse bibliographique). *Biotechnologie, Agronomie, Société et Environnement* **22**: 1–15.

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M *et al.* 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* **26**: 119–188.

Kattge J, Diaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ. 2011. TRY – a global database of plant traits. *Global Change Biology* **17**: 2905–2935.

Kenfack D. 2011. Resurrection in *Carapa* (Meliaceae): a reassessment of morphological variation and species boundaries using multivariate methods in a phylogenetic context. *Botanical Journal of the Linnean Society* **165**: 186–221.

Kenfack D, Peréz AJ. 2011. Two new species of *Carapa* (Meliaceae) from Western Ecuador. *Systematic Botany* **36**: 124–128.

Kindt R. 2020. WORLDFLORA: an R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. *Applications in Plant Sciences* **8**: e11388.

Koenen EJM, de Wilde JJFE. 2012. A taxonomic revision of the reinstated genus *Leplaea* and the newly recognized genus *Neoguarea* (Meliaceae, Sapindales): the exclusion of *Guarea* from Africa. *Plant Ecology and Evolution* **145**: 209–241.

Kress WJ, DeFilipps RA, Farr E, Kyi DYY. 2003. A checklist of the trees, shrubs, herbs, and climbers of Myanmar. *Contributions from the United States National Herbarium* **45**: 1–590.

Lê TC. 2003. *Danh lục các loài thực vật Việt Nam, 2 (Checklist of plant species of Vietnam, vol. 2).* Hanoi, Vietnam: Nhà xuất bản Nông nghiệp.

Lebrun J-P, Stork AL. 2011. *Tropical African flowering plants, vol. 6: Burseraceae – Apiaceae.* Geneva, Switzerland: Editions des Conservatoire et Jardin Botaniques.

Mabberley DJ. 1979. The species of Chisocheton (Meliaceae). *Bulletin of the British Museum (Natural History): Botany Series* **6**: 301–386.

Mabberley DJ. 2003. New species of, and other notes on, Chisocheton and Walsura (Meliaceae). *The Gardens' Bulletin, Singapore* **55**: 189–200.

Mabberley DJ. 2011. Meliaceae. In: Kubitzki K, ed. *The families and genera of vascular plants.* Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 185–211.

Mabberley DJ, Pannell CM. 2013. Meliaceae. In: Wilson A, ed. *Flora of Australia, vol. 26, Meliaceae, Rutaceae, Zygophyllaceae.* Canberra, ACT, Australia: Australian Biological Resoruces Study, 1–42.

Mabberley DJ, Pannell CM, Sing AM. 1995. Meliaceae. In: *Flora Malesiana series I.* Leiden, the Netherlands: Rijksherbarium/Hortus Botanicus, Leiden University.

Madagascar Catalogue. 2021. *Catalogue of the vascular plants of Madagascar.* St Louis, MO, USA; Antananarivo, Madagascar: Missouri Botanical Garden.

Meyer C, Weigelt P, Kreft H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters* **19**: 992–1006.

Moore T. 1857. *Index Filicum: a synopsis, with characters, of the genera, and an enumeration of the species of ferns, with synonymes, references.* Kew, Australia: Pamplin.

Muellner AN, Mabberley DJ. 2008. Phylogenetic position and taxonomic disposition of *Turraea breviflora* (Meliaceae), a hitherto enigmatic species. *Blumea – Biodiversity, Evolution and Biogeography of Plants* **53**: 607–616.

Nickrent DL, Anderson F, Kuijt J. 2019. Inflorescence evolution in Santalales: integrating morphological characters and molecular phylogenetics. *American Journal of Botany* **106**: 402–414.

Palacios WA. 1994. Especies nuevas de Meliaceae del Ecuador y áreas adyacentes. *Novon* **4**: 155–164.

Palacios WA. 2012. Four new tree species from Ecuador. *Caldasia* **34**: 75–85.
Palacios WA. 2016. Una especies nueva de *Guarea* (Meliaceae) del suroriente de Ecuador. *Neotropical Biodiversity* **2**: 69–71.
Palacios WA, Santiana J, Iglesias J. 2019. A new species of *Cedrela* (Meliaceae) from the eastern flanks of Ecuador. *Phytotaxa* **393**: 84.
Pandey RP, Dilwakar PG. 2008. An integrated check-list flora of Andaman and Nicobar islands, India. *Journal of Economic and Taxonomic Botany* **32**: 403–500.
Pannell CM. 1992. *A taxonomic monograph of the genus* Aglaia. *Kew bulletin additional series, XVI.* London, UK: HMSO.
Pannell CM. 1997. A new, cassowary-dispersed, species of *Aglaia* (Meliaceae, Section Amoora) from Papua New Guinea. *Kew Bulletin* **52**: 715–717.
Pannell CM. 2004. Three new species, two new subspecies and five new combinations at the subspecific level in *Aglaia* Lour. (Meliaceae). *Kew Bulletin* **59**: 87–94.
Pannell CM. 2020. Aglaia. *Flora of Australia.* Canberra, ACT, Australia: Australian Biological Resources Study, Department of Agriculture, Water and the Environment.
Pennington TD. 2016. Systematic treatment of American *Trichilia* (Meliaceae). *Phytotaxa* **259**: 18–162.
Pennington TD, Barker A, Rojas-Andrés BM. 2021. A revision of the genus *Ruagea* (Meliaceae: Melioideae). *Kew Bulletin* **76**: 565–604.
Pennington TD, Clarkson JJ. 2013. A revision of *Guarea* (Meliaceae). *Edinburgh Journal of Botany* **70**: 179–362.
Pennington TD, Muellner-Riehl AN. 2010. *A monograph of* Cedrela *(Meliaceae).* Milborne Port, UK: Dh Books.
Pennington TD, Styles BT. 1975. A generic monograph of the Meliaceae. *Blumea – Biodiversity, Evolution and Biogeography of Plants* **22**: 419–540.
Pennington TD, Styles BT, Taylor DAH. 1981. Meliaceae. *Flora Neotropica* **28**: 1–470.
R Core Team. 2021. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.
Randrianarivony TN, Randriatsivery MF, Andriamihajarivo TH, Callmander MW. 2017. Une nouvelle espèce de *Turraea* (Meliaceae) du Sud-Ouest de Madagascar. *Candollea* **72**: 283–288.
Rees T. 2014. TAXAMATCH, an algorithm for near ('fuzzy') matching of scientific names in taxonomic databases. *PLoS ONE* **9**: e107510.
Rogers ZS, Malécot V, Sikes KG. 2006. A synoptic revision of *Olax* L. (Olacaceae) in Madagascar and the Comoro Islands. *Adansonia* **28**: 71–100.
Rueangruea S, Tagane S, Suddee S, Tetsana N, Poopath M, Nagamasu H, Naiki A. 2015. *Toona calcicola*, a new species and *Reinwardtiodendron humile*, a new record to Thailand. *Thai Forest Bulletin (Botany)* **43**: 79–86.
Scott AJ. 1997. 69. Méliacées. In: Bosser J, Cadet TH, Guého J, Marais W, eds. *Flore des Mascareignes: La Réunion, Maurice, Rodrigues. Méliacées a Connaracées, vol. 69–79.* La Réunion, France: Mauritius: The Sugar Industry Research Institute, 1–17.
Sivaraj I, Nithaniyal S, Bhooma V, Senthilkumar U, Parani M. 2018. Species delimitation of *Melia dubia* Cav. from *Melia azedarach* L. complex based on DNA barcoding. *Botany* **96**: 329–336.
Smith AC. 1985. *Flora Vitiensis nova.* Lawaii, HI, USA: Pacific Tropical Botanical Garden.
Steinmann VW, Ramírez-Amezcua Y. 2013. *Trichilia pugana* (Meliaceae), una especie nueva de la tierra caliente de Michoacán, México. *Ibugana* **4**: 59–62.
Styles BT, White F. 1990. Two new species of *Turraea* (Meliaceae) from Northeast Africa. *Bulletin du Jardin Botanique National de Belgique* **60**: 415–416.
Styles BT, White F. 1991. Meliaceae. In: Turrill WB, Polhill RM, eds. *Flora of tropical east Africa.* Rotterdam, the Netherlands: A. A. Balkema.
Takeuchi W. 2000. A floristic and ethnobotanical account of the Josephstaal forest management agreement area, Papua New Guinea. *SIDA, Contributions to Botany* **19**: 1–63.
Takeuchi W. 2009a. Occurrence records in Papuasian *Aglaia* (Meliaceae): *A. pannelliana* and *A. puberulanthera* from the southern karst of Papua New Guinea. *Harvard Papers in Botany* **14**: 31–38.
Takeuchi W. 2009b. *Dysoxylum middletonianum* (Meliaceae), a distinctive new species from the southern old mountains of Papua New Guinea. *Edinburgh Journal of Botany* **66**: 347–353.
Telenius A. 2011. Biodiversity information goes public: GBIF at your service. *Nordic Journal of Botany* **29**: 378–381.
The Royal Botanic Gardens, Kew. 2022. *Plants of the World Online (POWO).* [WWW document] URL www.plantsoftheworldonline.org [accessed 17 September 2022].
The Royal Botanic Gardens, KewGrassBase, The Global Compositae Checklist, The International Legume Database and Information Service, Royal Botanic Gardens Edinburgh, The iPlants project, The International Organization for Plant Information, Missouri Botanical Gardens. 2013. *The Plant List 1.1.* [WWW document] URL www.theplantlist.org [accessed 17 May 2022].
Turland N. 2019. The code decoded. *Advanced Books* **1**: e38075.
Velayos M, Barberá P, Cabezas FJ, de la Estrella M, Fero M, Aedo C. 2014. Checklist of the vascular plants of Annobón (Equatorial Guinea). *Phytotaxa* **171**: 1–78.
Wei R, Yang J, He L-J, Liu H-M, Hu J-Y, Liang S-Q, Wei X-P, Zhao C-F, Zhang X-C. 2021. Plastid phylogenomics provides novel insights into the infrafamilial relationship of Polypodiaceae. *Cladistics* **37**: 717–727.
Weigelt P, König C, Kreft H. 2020. GIFT – a Global Inventory of Floras and Traits for macroecology and biogeography. *Journal of Biogeography* **47**: 16–43.
White F, Styles BT. 1986. Meliaceae. In: Leistner OA, ed. *Flora of southern Africa, vol. 18, part 3: Simaroubaceae-Malpighiaceae.* Pretoria, South Africa: Govt. Printer, 39–61.
de Wilde JJFE. 1968. *A revision of the species of* Trichilia *P. Browne (Meliaceae) on the African continent.* PhD thesis, Wageningen University, Wageningen, the Netherlands, 207.
de Wilde JJFE. 2007. Revision of the African genus *Heckeldora* (Meliaceae). *Blumea – Biodiversity, Evolution and Biogeography of Plants* **52**: 179–199.
Wongprasert T, Phengklai C, Boonthavikoon T. 2011. A synoptic account of the Meliaceae of Thailand. *Thai Forest Bulletin (Botany)* **39**: 210–266.
Zhang J, Qian H. 2022. U.TAXONSTAND: an R package for standardizing scientific names of plants and animals. *Plant Diversity* **45**: 1–5.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Geographic distribution of unique and inconsistent taxon names.

Fig. S2 Phylogeny showing absolute numbers of unique and inconsistent taxon names.

Fig. S3 Meliaceae accepted species occurrence in checklists with full species names.

Notes S1 Benchmark test of the authorMatch function.

Notes S2 Difference calculation for taxon names and authors.

Notes S3 Global checklist name-matching algorithm.

Table S1 Links to IPNI compared between WCVP and WFO.

Table S2 Genera and family statistics.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.